Thèse de doctorat

Université de Limoges

UNIVERSITÉ DE LIMOGES

ÉCOLE DOCTORALE N° 610

Sciences et Ingénierie des Systèmes, Mathématiques, Informatique

FACULTÉ DES SCIENCES ET TECHNIQUES

Thèse

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE LIMOGES

Discipline : Electronique des hautes fréquences, photonique et systémes

présentée et soutenue par

David Martínez Martínez

Jeudi 20 juin 2019

Méthodologies et Outils de Synthèse pour des Fonctions de Filtrage Chargées par des Impédances Complexes

Thèse dirigée par Stéphane BILA et Laurent BARATCHART

JURY :

Président du jury	
M. Paul ARMAND	Professeur à l'Université de Limoges-XLIM
Rapporteurs	
M. Alejandro ÁLVAREZ MELCÓN	Professeur à l'Université Politécnica de Cartagena
M. Stanislas KUPIN	Professeur à l'Université de Bordeaux-IMB
Examinateurs	
M. Paul ARMAND	Professeur à l'Université de Limoges-XLIM
M. Laurent BARATCHART	Directeur de Recherche à l'INRIA, Sophia Antipolis
M. Stéphane BILA	Directeur de Recherche au CNRS-XLIM
Mme. Martine OLIVI	Chargée de Recherche à l'INRIA, Sophia Antipolis
M. Fabien SEYFERT	Chargé de Recherche à l'INRIA, Sophia Antipolis
Invités	
M. Ludovic CARPENTIER	Ingénieur au CNES, Toulouse
M. Johann LE NEZET	Ingénieur à la DGA, Bruz
M. Damien PACAUD	Ingénieur à Thales Alenia Space, Toulouse
M. François TORRÈS	Chargé de Recherche au CNRS-XLIM

To Vero

Acknowledgements

Before I begin this text, I can not miss the opportunity to thank the countless people who have helped me or supported me throughout a life stage, which concludes with the document here presented.

In the first place, I need to thank Fabien SEYFERT, Laurent BARATCHART and Stéphane BILA for the opportunity to carry out this thesis within the INRIA Institute and the XLIM laboratory. Fabien, thank you for considering that I could be the right candidate for the position available when I still did not have any outstanding qualities for it.

Martine, I also have to give special thanks to you and Fabien, for all the help received during these three years and the time spent explaining to me many and complicated concepts of complex analysis. Especially with relation to the Nevanlinna-Pick interpolation. I loved learning all those new mathematical concepts. Thanks to them, I now have a unique ability that helps me see the world and address each problem more analytically.

In the same way, I also have to thank the collaboration with all members of the old APICS team as well as the renewed team FACTAS at INRIA for all the help and dedication offered.

Sylvain, thank you very much for the patience and the time wasted with the numerous and to a certain extent strange problems, primarily computer related ones, with which I have addressed you throughout my stay at INRIA.

Julliete, thank you for the books and the interesting readings that you have recommended to me during this time. Thanks also for all the good moments, many of them around a coffee, that we have enjoyed all the members of APICS and FACTAS.

To Dmitry, the first person, and the only one, that I found on my first arrival at INRIA at 8 in the morning. Thank you very much for all the help offered and the incredible involvement in the problems of all the members of the team. Thank you also for each of your stories, especially those that occurred during your various trips. The work in INRIA would not have been the same without them.

Many thanks also to Christos and Konstantinos for promoting and carrying out the numerous parties and events that we have all attended on some occasion at INRIA. You guys, along with the rest of the members of the Greek community in INRIA, have an innate talent to organize parties. Besides, you were able to lead to success, without any external help and with much effort and dedication, the doctoral students' seminar,

which nobody believed at the beginning, and which turned out to be an excellent idea to exchange ideas among the works carried out by the different students.

Matthias, thank you very much for the exciting conversations about microwave filters, and all the coaching sessions in the gym teaching us new exercises. Without any doubt, you were the absolute master working-out.

Thank you very much, Sebastien, for reviewing my mathematical calculations even knowing in advance that they were utterly wrong. Discussing with you has always allowed me to deepen my knowledge of French culture and language from a genuine French person willing to argue about everything around him.

Besides, I must especially thank the collaboration with Gibin and Adam throughout this thesis, whose help has made possible the existence of this manuscript.

Gibin, thanks for all the help received, especially in the demonstration of all my strange conjectures, correct or not. Thank you very much for the good times, whether hiking, sightseeing or during conferences.

Adam, since you came to INRIA, you have always had an essential presence in all of us. Thanks to you, continually pushing me to write code well organized, commented and robust, my programming skills have improved incredibly to a level that I could not have imagined a few years ago. I can say without fear of being wrong that your collaboration has been essential in the development of PUMA, the toolbox obtained as a result of this thesis.

Thank you very much, Adam, for all the good times and discussions over the years, hiking, eating, swimming, travelling, eating, or having a beer. Also, I must admit that you are an excellent tour guide in Brussels. Thank you very much for the goodbye farewell, with the help of Vero, organized for my relocation to XLIM. Good luck with your new life in the Netherlands. I hope that our professional paths will cross again in the future.

I must also thank all members of the MACAO team at XLIM. To Stéphane BILA for the help and support received throughout the thesis. To Olivier TANTOT and to Nicolas DELHOTE for allowing me to realise my 3D designs, thanks to the marvels of additive manufacturing.

François, thank you very much for all the lessons on antennas and radiation. These lessons have made me discover the fascinating world of antennas, helping me to carry out the work done and at the same time acquiring new varied skills.

It is also essential to acknowledge and appreciate the work done at XLIM, including during my stay at INRIA by Aurelien and Johann, which have had to face my numerous designs and instructions, not always providing a satisfactory result.

I would also like to thank all the others who have continuously accompanied and

helped me in various ways since my arrival at XLIM: Thibault, Philipe, Etienne, Anthony, Julien, Ahmad, Ali, Andrés, Chaimaa and Oualid.

I can not forget to thank Stéphanie, Marie-Line and Marie-Claude for their constant dedication every day to make the APICS, FACTAS and MACAO project teams work. Thank you very much and my apologies for all the headaches when trying to organise all my crazy missions around the world.

I also want to thank all the members of the jury who have dared to accept the task of reviewing this work: Alejandro ÁLVAREZ MELCÓN, Stanislas KUPIN, Paul ARMAND, Laurent BARATCHART, Stephane Bila, Martine OLIVI and Fabien SEYFERT. Thank you very much for your numerous comments and at the same time, my most sincere apologies for the size of the manuscript.

Finally, and most importantly, I thank my family for all the support received for so many years and in particular to my lovely wife Vero, who has allowed me to finish this thesis alive. I will never thank you enough for all the dedication offered during these years, either listening to my boring conversations or supporting me during the many sleepless nights in which I always had an article, report or presentation to finish before the deadline. You have cared for me and have worked hard in this work as much as I have. That is why this work is more yours than mine. With this manuscript concludes an important stage in the life of we two which has helped us grow both emotionally and intellectually. I wish that the path we are about to start now is even more fruitful and enriching for both of us.

Many thanks also to all the people not mentioned who have contributed or helped in one way or another to the development of this work.

Sincerelly
David Martínez Martínez

Making of the document

This document is based on the latex template available at http://www.unilim.fr/sigmadocx/wp-content/uploads/2012/11/latex-template-linux-utf8.zip.

The latex editor *GNOME Latex* [1], formely known as *LaTeXila* has served in the process of writing and formatting the document.

All 2D plots displayed in this document have been plotted with matlab and converted to *tikz* format with the aid of the *MATLAB2TikZ* toolbox [2]. Figures showing layouts of electric circuits, block diagrams and other various contents have been generated with the LaTeXpackage *circuitikz* [3].

The 3D objects shown in figs. 15.5, 15.17 and 15.18 has been created with the software *Asymptote* [4] by parsing a stereolithography (STL) file to created a *.prc* file. These objects are embedded in the pdf document using the LaTeXpackage *media9* [5].

References

[1] S. Wilmet, "GNOME Latex," 2019. [Online]. Available: https://wiki.gnome.org/Apps/GNOME-LaTeX

[2] E. Geerardyn, "MATLAB2TikZ," 2019. [Online]. Available: https://github.com/matlab2tikz/matlab2tikz

[3] S. Erhardt, R. Giannetti, S. Lindner, and M. Redaelli, "CircuiTikZ," 2019. [Online]. Available: https://ctan.org/pkg/circuitikz

[4] A. Hammerlindl, J. Bowman, T. Prince, and S. Healy, "Asymptote," 2019. [Online]. Available: http://asymptote.sourceforge.net/

[5] A. Grahn, "media9." [Online]. Available: https://www.ctan.org/pkg/media9

Source of figures and tables

All figures and tables included in this document, with the exception of those cited below, have been originally created by the main author.

The following graphics have been made with the collaboration of the cited authors.

Figure 10.23	François Torres
Figure 11.1	François Torres
Figure 11.6	François Torres
Figure 11.8	Aurelien Perigaud
Figure 11.9	Aurelien Perigaud

The following photos have been shot by the authors cited below.

Figure 11.10	François Torres
Figure E.2	Olivier Tantot
Figure E.3	François Torres
Figure E.4	François Torres
Figure E.5	Olivier Tantot
Figure E.6	Olivier Tantot
Figure E.7	Olivier Tantot
Figure E.10	Ali Dia
Figure E.11	Ali Dia
Figure E.12	Oualid Ourya

Abstract

The problem of impedance matching in electronics and particularly in RF engineering consists on minimising the reflection of the power that is to be transmitted, by a generator, to a given load within a frequency band. The matching and filtering requirements in classical communication systems are usually satisfied by using a matching circuit followed by a filter. We propose here to design matching filters that integrate both, matching and filtering requirements, in a single device and thereby increase the overall efficiency and compactness of the system.

In this work, the matching problem is formulated by introducing convex optimisation on the framework established by the matching theory of Fano and Youla. As a result, by means of modern non-linear semi-definite programming techniques, a convex problem, and therefore with guaranteed optimality, is achieved.

Finally, to demonstrate the advantages provided by the developed theory beyond the synthesis of filters with frequency varying loads, we consider two practical applications which are recurrent in the design of communication devices. These applications are, on the one hand, the matching of an array of antennas with the objective of maximizing the radiation efficiency, and on the other hand the synthesis of multiplexers where each of the channel filters is matched to the rest of the device, including the filters corresponding to the other channels.

Part I

Introduction

Chapter 1:

Structure and organization of the manuscript

The work presented here is the result of a long collaboration between two institutions of different nature and different backgrounds. On one side INRIA, the national institute of research in computer science and automatic and on the other hand XLIM, a laboratory belonging to the University of Limoges, the CNRS and the University of Poitiers. These two laboratories provide different competencies to this thesis, which at the same time are complete. INRIA is an institute focused on fundamental research, which in many cases provides some degree of rigorous solutions to external problems, which may come from the private or public sector. In particular, the work presented here has been done within the teams APICS/FACTAS at INRIA and MACAO at XLIM.

The APICS team (which has subsequently been converted into FACTAS) is specialised in functional analysis applied to communications. In particular, among the usual topics are the inverse problems in electromagnetism for the location of electromagnetic sources and the problems of rational approach for the identification and modelling of communication systems.

XLIM presents a more applied background in continuous contact with the most relevant agents in the space communications sector. Most of the products developed in the laboratory are directly transferred to the industry. Among the most recurrent topics are the design of filters for radio-frequency systems using diverse technologies. We could highlight, for example, the microwave filters built from resonant cavities and other devices such as multiplexers, direct application of the filters presented. Besides, the MACAO team of XLIM has a great experience in the prototyping of these components through additive manufacturing processes.

As a result of the collaboration between both institutes, the present thesis contains a fairly extensive theoretical part, but without forgetting the implementation of real devices, which are manufactured in the laboratory XLIM and measured to validate the theory developed.

In addition to the mentioned collaboration, the work presented here is also the result of double founding between two organisations of diverse nature and interests. On the one hand, the French armaments directorate (DGA) and, on the other hand, the national centre for space studies (CNES). This double financing implies a dual thematic because of the difference of interests between the two organisations. For this reason, the thesis has been divided into two parts which, although they share the same background, can be well differentiated.

Below we summarise the content of each chapter in each of the parts. In this way, the reader can get an idea of the subject of each chapter and the type of background necessary. This practice allows a better organisation during the reading of the manuscript, possibly omitting specific chapters that may not be very interesting for a particular type of reader. The first part consists of a common introduction to both parts 2 and 3. This part is composed of 3 chapters including the present one.

1. The current chapter provides a general vision of the framework in which this thesis has been developed as well as a brief summary of each other chapters. We also

introduce the different institution involved in this work.

2. In the chapter 2, we review some of the most important concepts for the correct understanding of the theory developed in subsequent chapters. Particularly some classical theorems in the field of functional analysis that can be useful for a reader with a background in engineering. At the same time, the most important concepts regarding the design of radio-frequency devices are formulated from a mathematical perspective. In this way, we facilitate reading to a reader with a different background.

3. After having introduced the basic concepts necessary for the understanding of the thesis, it is time to take a journey through the history of the matching problem. Chapter 3 is, therefore, a bibliographic chapter where the main contributions to the matching problem that can be found in the literature are reviewed. Besides, we also take the opportunity to lay the foundations of the problem that will occupy us during the first part of the thesis.

In the second part, we find a study of the problem of matching in radio-frequency devices. This part is based on the theories developed in the 40s and 50s mainly by Fano and Youla. The content of this first part has been distributed in the following seven chapters.

4. In chapter 4, we continue with the study of the problem of matching, from where we left it in the previous chapter. Precisely, we return to the problem originally posed by Fano and reformulated by Youla later. This problem has remained unresolved since the 1950s. Thanks to the theory developed in this chapter, combining the original problem with modern techniques of convex optimisation, we achieve a significant advance towards the solution of this problem, guaranteeing optimality in certain particular cases. Additionally, we present as an example the particular case of the matching of an antenna which is modelled by a rational function of grade 1.

5. In chapter 5, we extend the theory presented previously in the framework of a degree 1 antenna. In this way, we obtain a completely generalised formulation of the problem of matching. Thanks to this formulation, the results obtained can be applied to other problems more complex than the one of grade 1. Also, an important property of said formulation is introduced, which will allow the numerical implementation of the problem in question.

6. After having introduced the general problem that we are trying to solve in this work, in chapter 6 we make a small parenthesis to a particular case of the matching problem which is of relevance for the antenna community and has not been considered before. We also use it to present some preliminary results, comparing the provided lower bounds with the matching level obtained by means of a matching filter of fixed degree. With this chapter, we intend to provide the reader with an additional motivation on the usefulness of the problem of matching before moving on to the numerical implementation.

In part III we deal with a topic that is a little different from the rest of the thesis. In this chapter, we provide a numerical implementation of the matching problem. This

implementation is detailed throughout chapters 7 and 8 and section 8.3. Moreover in chapters 9 and 11 we move a little away from the numerical implementation to resume in some way the theory about the matching problem.

It is also important to note that the algorithm presented in part III has been implemented in MATLAB during the course of this thesis, giving birth to the MATLAB toolbox called PUMA which can be found in [6]. This toolbox has the aim of providing the optimal filtering response for the matching network once the load L is fixed. The aforementioned toolbox also serves as a proof of concept of the previous theory and validates the efficacy of the algorithm obtained. Moreover the implemented toolbox allows to compute in an efficient manner the optimal solution to an optimization problem with a considerable number of variables and constraints. Finally several examples obtained through the presented algorithm and some applications of the matching problem are provided.

7. Chapter 7 introduces a reformulation of the matching problem as a non-linear optimisation program which includes matrix inequalities. This numerical formulation corresponds to the field of SDP in optimisation. However, when including non-linear constraints, a program of type NL-SPD is obtained. This program even being a convex problem, is one of the most complicated problems in optimisation that can be solved optimally.

8. Chapter 8 deals with the resolution of the formulated SDP. This chapter provides the details of the numerical implementation of the matching problem as it has been programmed in the core of PUMA toolbox. We introduce some classical techniques in the field of optimization such as the elimination of linear equalities by the substitution method. In this way, the non-linear SDP derived in chapter 7 is simplified. Additionally, as we use an interior-point algorithm, an introduction to the concept of barrier functions is performed.

9. In chapter 9 we discuss some different heuristic approaches for the computation of a sub-optimal matching network of finite degree. These matching networks approach as close as possible the optimal lower bounds obtained as a result of the convex formulation of the matching problem.

10. Chapter 10 contains a compilation of the results obtained during the development of this thesis, except those that have already been presented in chapter 6. In this chapter, we can find some examples related to the matching of more complex antennas destined to applications of different nature. In addition we also provide an interesting discussion comparing the lower bounds issue of the previously formulated optimization problem and the sub-optimal practical results obtained when to approximate those bounds. Both results, the lower bounds and the sub-optimal responses are calculated by means of the PUMA toolbox providing a valuable information on the optimality of the obtained results when comparing with the lower bounds.

11. In chapter 11 we present an additional application of the matching problem. This is one of the applications that has occupied us most of the time during the course of the thesis and the main reason for the conception of the PUMA toolbox. This

application deals with the problem of maximizing the efficiency of an array of antennas. This differs from the classic matching problem previously discussed in the fact that the dissipation in both the matching filters and the antenna is taken into account. In addition, upon working with an antenna array we face the matching of a multi-port device. This problem opens up one of the main and most interesting lines of future work due to its perspective results and potential applications.

After the exhaustive revision of the matching theory carried out in parts II and III, in part V we apply this theory to the design of multiplexers. The motivation for this application stems from the fact that multiplexer design can also be seen as a somewhat peculiar matching problem. However, when trying to adapt this theory, we face some difficulties that prevent us for the direct application of the theory developed for matching filters synthesis. Nevertheless we find the path to overcome those difficulties in one of the scientific contributions to the literature of the problem of matching in the recent years.

12. The third part of this work begins directly with a bibliographic study of the different multiplexer design techniques. In addition, in chapter 12 we include a small state of the art, detailing the most important contributions of the most renowned authors to literature in this field. Finally, we also take the opportunity to highlight the main problems that engineers face in the design of this type of device.

13. In chapter 13 we are dedicated to one of the most recurrent problems in the design of multiplexers. This is the problem of manifold peaks. In the previous parts of the thesis, we have devoted considerable effort in the synthesis of filters. However, in the case of multiplexers, the mastery of filter synthesis is not enough. In this chapter, we propose a practical algorithm for the manifold design that minimise the appearance of such manifold peaks.

14. In chapter 14, we present an original technique for the synthesis of the channel filters once the manifold of the multiplexer is designed. This technique consists of an adaptation for the design of multiplexers of a point-wise matching algorithm available in the literature and reviewed in Chapter 3. Thanks to the reformulated algorithm, the synthesis of the filters is done simultaneously obtaining the circuital model for all of them. In this way, we obtain a much higher computational efficiency compared to the traditional techniques for the optimisation of multiplexers.

15. In chapter 15 we present the results obtained through the previously developed algorithm. These results consist of the design of a manifold coupled triplexer for space applications. Finally, a prototype made by additive manufacturing is presented to serve as a proof of concept to the novel algorithm for multiplexer synthesis. This prototype is constructed in plastic where the inner surface of the structure is metallised to obtain a conductive boundary.

Finally, we conclude the present thesis providing a list of interesting lines of future work which have emerged from the different applications treated.

16. Chapter 16 of this thesis provides an overview of the accomplished work and some concluding remarks drawn from this work. With this chapter, we also take the opportunity to add some considerations and comments about the synthesis and manufacturing of matching filters and multiplexers.

References

[6] D. M. Martinez, F. Seyfert, A. Cooman, and M. Olivi, "Software PUMA-HF: https://project.inria.fr/puma," 2008. [Online]. Available: https://project.inria.fr/puma/

Chapter 2:

Fundamental concepts

In this work, the problem of matching arises for the first time as an optimization problem where the requirements in terms of matching simply represent an additional condition. The said problem is related to the traditional problem of synthesis of transfer functions. In fact, the classical problem of synthesis can be seen as a particularisation of the problem of matching where the matching condition has been relaxed. For this reason, the classic problem of synthesis of filtering functions is reformulated in the present chapter using the same approach and the same terminology as in the problem of matching, which will be introduced in later chapters. In this way, after introducing the matching problem, it will be easier to see it as a generalization of the traditional synthesis problem, established in this chapter.

We begin this chapter, namely the first one dedicated to the topic dealt with in this thesis, with an introduction to the concept of matching in engineering. This introduction includes a list of some classical definitions and theorems, which although they may seem trivial for someone with a certain background in the topic, are necessary for a rigorous statement of the problem. The convenience of such definitions is further demonstrated when dealing with the proofs of the numerous theorems and lemmas stated in next chapter. Hence a quick review of the main definitions provided here is recommended before tackling the exhaustive theory developed in chapter 4 .

Broadband matching is one of the classic problems in circuit theory. This problem arises in communication systems when the power that is intended to be transmitted, by a generator, to a load is reflected. The said reflected power, on the one hand, represents a loss of the useful power provided to the system and on the other hand, it will deteriorate the elements prior to the load, such as amplifiers or the generator which are not usually prepared to receive this reflected power from load. Therefore, by improving the matching between the generator and the load, we are litigating with a double problem that could have very harmful consequences both in terms of power loss and in terms of the reception of this power by devices not prepared for it.

2.1 Transmitted power

Consider the simple circuit in fig. 2.1 consisting of a resistor connected to a generator, the power delivered to the load, or in this case, dissipated in the resistor, is computed as [7, Chapter 3, section 2]

$$P = \frac{1}{2}\Re\left(\overline{I_R}\cdot V_R\right). \tag{2.1}$$

Note that the bar in eq. (2.1) and in the rest of this thesis indicates complex conjugate. Introducing the expressions for V_R and I_R, namely

$$V_R = E_1\frac{R}{R+R_1},$$
$$I_R = \frac{E_1}{R+R_1},$$

we have

$$P = \frac{1}{2}\Re\left(R\right)\left|\frac{E_1}{R+R_1}\right|^2.$$

Equivalently we write

$$P = \frac{|E_1|^2}{2}\frac{\Re\left(R\right)}{\Re\left(R+R_1\right)^2+\Im\left(R+R_1\right)^2}.$$

It can be easily proved that the previous quantity is maximum the assumptions of $\Im\left(R+R_1\right)=0$ and $\Re\left(R+R_1\right)=2\Re\left(R_1\right)$. Therefore the transmitted power is maximum when the resistor equals the conjugate of the internal impedance of the generator, namely $R=\overline{R_1}$, thus

$$P_{max} = \frac{|E_1|^2}{8\Re\left(R_1\right)}.$$

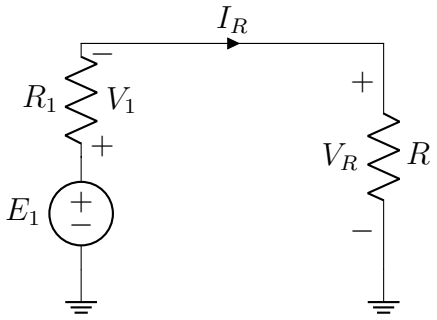


Figure 2.1: Simple transmitting circuit

2.2 Power waves

For a better understanding of the phenomena occurring in fig. 2.1, lets consider a now a transmission line which interconnect both sides of the circuit as shown in fig. 2.2. This transmission line might be consider later on to be of zero length, leading to the same schematic. Take now the transmission line segment of differential length from fig. 2.3a and consider the distributed circuit model shown in fig. 2.3b as in [8, Chapter 2]. It should be noted that a lossless transmission line is assumed and then a pure reactive model is considered, namely with no dissipating elements.

Appliying Kirchhoff's laws we can express

$$\begin{aligned} v(z,t) &= Ldz\cdot\frac{\partial}{\partial t}i(z,t)+v(z+dz,t),\\ i(z,t) &= i(z+dz,t)+Cdz\cdot\frac{\partial}{\partial t}v(z+dz,t). \end{aligned}$$

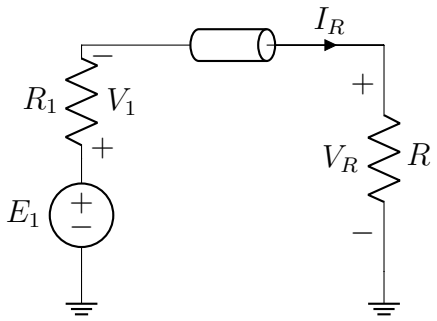
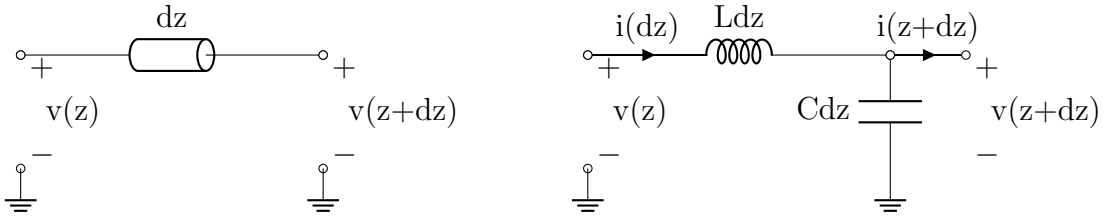


Figure 2.2: Simple transmitting circuit interconnected by a transmission line.



(a) Transmission line differential element (b) Lumped equivalent circuit

Figure 2.3: Differential lenght element of a tramission line and LC equivalent.

Letting now the differential dz tend to zero we obtain the derivatives of $v(z,t)$ and $i(z,t)$

$$\lim_{dz\to 0}\frac{v(z+dz,t)-v(z,t)}{dz}=\frac{\partial}{\partial z}v(z,t),$$
$$\lim_{dz\to 0}\frac{i(z+dz,t)-i(z,t)}{dz}=\frac{\partial}{\partial z}i(z,t).$$

Therefore Kirchhoff's laws lead to

$$L\frac{\partial}{\partial t}i(z,t)+\frac{\partial}{\partial z}v(z,t)=0,$$
$$C\frac{\partial}{\partial t}v(z,t)+\frac{\partial}{\partial z}i(z,t)=0.$$

In order to eliminate one of the unknowns, we derivative both expression with respect to t obtaining

$$L\frac{\partial^2}{\partial t^2}i(z,t)+\frac{\partial^2}{\partial t\partial z}v(z,t)=0,$$
$$C\frac{\partial^2}{\partial t^2}v(z,t)+\frac{\partial^2}{\partial t\partial z}i(z,t)=0.$$

Combining both expression and eliminating the derivative of $i(z,t)$, we obtain the wave equation for the transmission line

$$\frac{\partial^2}{\partial z^2}v(z,t)-LC\frac{\partial^2}{\partial t^2}v(z,t)=0.$$

We obtain a differential homogeneous equation with partial derivatives which admits a particular solution in the form

$$v(z,t)=A(t-t_0(z))+B(t+t_0(z)), \tag{2.2}$$

with

$$t_0 = z\sqrt{LC}.$$

Note therefore that eq. (2.2) is the sum of two functions displaced in time. The time displacement t_0 depends on the position along the transmission line and the constant $1/\sqrt{LC}$ represents the displacement speed. Similarly, if we denote $Z_0 = \sqrt{L/C}$ we obtain the expression for $i(z,t)$, namely

$$i(z,t) = \frac{1}{Z_0}\left(A(t-t_0) - B(t+t_0)\right). \tag{2.3}$$

Expressing now eqs. (2.2) and (2.3) we have

$$\begin{aligned} v(z,w) &= a(\omega)e^{-j\omega z\sqrt{LC}} + b(\omega)e^{j\omega z\sqrt{LC}}, \\ i(z,w) &= \frac{1}{Z_0}\left(a(\omega)e^{-j\omega z\sqrt{LC}} - b(\omega)e^{j\omega z\sqrt{LC}}\right). \end{aligned}$$

We define the propagation constant $\beta(\omega)$ as

$$\beta(\omega) = \omega\sqrt{LC}.$$

This function $\beta(\omega)$ determines the speed at which the phase of the functions $a(\omega), b(\omega)$ varies along the transmission line. We have

$$v(z,w) = a(\omega)e^{-j\beta(\omega)z} + b(\omega)e^{j\beta(\omega)z}, \tag{2.4}$$

$$i(z,w) = \frac{1}{Z_0}\left(a(\omega)e^{-j\beta(\omega)z} - b(\omega)e^{j\beta(\omega)z}\right). \tag{2.5}$$

Each of the expressions eqs. (2.4) and (2.5) can be considered to be composed of a progressive wave $a(\omega)e^{-j\beta(\omega)z}$ which moves forward along the transmission line and a regressive wave $a(\omega)e^{j\beta(\omega)z}$ moving backward in the transmission line.

2.2.1 Reflection coefficient

Under the assumption that the function $a(\omega)$ is a function of minimum phase, we can define the reflection coefficient $s(\omega)$ as the ratio between the regressive and the progressive wave

$$s(z,\omega) = \frac{b(\omega)e^{j\beta(\omega)z}}{a(\omega)e^{-j\beta(\omega)z}} = \frac{b(\omega)}{a(\omega)}e^{2j\beta(\omega)z}.$$

Therefore we can express

$$v(z,w) = a(\omega)\left(e^{-j\beta(\omega)z} + s(\omega)e^{j\beta(\omega)z}\right), \tag{2.6}$$

$$i(z,w) = \frac{a(\omega)}{Z_0}\left(e^{-j\beta(\omega)z} - s(\omega)e^{j\beta(\omega)z}\right). \tag{2.7}$$

2.3 Reflection coefficient

As in section 2.1, we can now compute the power transmitted to the load by meas of eqs. (2.6) and (2.7) as

$$P(z,\omega) = \frac{1}{2}\Re\left(\overline{v(z,\omega)}i(z,\omega)\right).$$

Thus we have

$$P(z,\omega) = \frac{1}{2}\frac{|a(\omega)|^2}{Z_0}\Re\left(1 - 2j\Im\left(s(\omega)e^{2j\beta(\omega)z}\right) - |s(\omega)|^2\right),$$

and therefore taking the real part

$$P(z,\omega) = \frac{1}{2}\frac{|a(\omega)|^2}{Z_0}\left(1 - |s(\omega)|^2\right). \tag{2.8}$$

Remark 2.3.1. *Note that the transmitted power $P(z,\omega)$ does not depends on the position z along the transmission line because of the non dissipation assumption. Therefore the power is conserved through the circuit.*

Remark 2.3.2. *Note as well that maximum available power is given by the progressive wave $a(\omega)$. This maximum power is obtained when $s(\omega) = 0$ as*

$$P_{max} = \frac{1}{2}\frac{|a(\omega)|^2}{Z_0}.$$

In this case, assuming $s(\omega) = 0$, the relation between the current and the voltage in the transmission line is given by the impedance Z_0. Nevertheless at the terminal of the load in fig. 2.2 this relation is given by the value of R. Therefore if the load R in fig. 2.2 is distinct from the impedance Z_0 at a given frequency, we have $s(\omega) \neq 0$. Lets define now $z = 0$ as the position of the load in fig. 2.2 and compute the relation between both quantities, we have

$$R(\omega) = \frac{v(0,\omega)}{i(0,\omega)} = Z_0\frac{a(\omega) + b(\omega)}{a(\omega) - b(\omega)}.$$

Thus

$$R(\omega)\left(a(\omega) - b(\omega)\right) = Z_0\left(a(\omega) + b(\omega)\right).$$

Expressing now $b(\omega)$ in function of $a(\omega)$ we have

$$b(\omega) = \frac{R(\omega) - Z_0}{R(\omega) + Z_0}a(\omega).$$

We obtain the expression of the reflection coefficient $s(\omega)$ induced by the load $R(\omega)$

$$s(0,\omega) = \frac{R(\omega) - Z_0}{R(\omega) + Z_0}.$$

Similarly we can compute the input impedance seen at any point $z < 0$ of the transmission line. The impedance $Z_{in}(z, \omega)$ is therefore obtained as

$$Z_{in}(z,\omega) = \frac{v(z,\omega)}{i(z,\omega)} = Z_0 \frac{a(\omega)e^{-j\beta(\omega)z}\left(1 + s(z,\omega)\right)}{a(\omega)e^{-j\beta(\omega)z}\left(1 - s(z,\omega)\right)} = Z_0 \frac{1 + s(z,\omega)}{1 - s(z,\omega)}.$$

If we set $z = -l$ where l is the length of the transmission line, eventually 0, we can translate the condition of conjugate matching, namely $Z_{in}(z, \omega) = \overline{R_1}$ onto the reflection coefficient $s(z, \omega)$. We have

$$\overline{R_1} = Z_0 \frac{1 + s(-l,\omega)}{1 - s(-l,\omega)}.$$

Therefore the optimal reflection coefficient $s_{opt}(\omega)$ which maximises the transmitted power takes the expression

$$s_{opt}(\omega) = \frac{\overline{R_1} - Z_0}{\overline{R_1} - Z_0}.$$

If we denote by $s_g(\omega)$ the reflection coefficient when looking to the input of the resistor R_1, namely

$$s_g(\omega) = \frac{R_1 - Z_0}{R_1 - Z_0},$$

then we have

$$s_{opt}(\omega) = \overline{s_g(\omega)}. \tag{2.9}$$

Therefore the condition of conjugate impedance translates to conjugate reflection coefficients.

Remark 2.3.3. *It should be noted that a generic impedance Z_0 has been considered in the preceding results. Nevertheless those results still hold for a zero length transmission line, namely $l = 0$. In this case an arbitrary impedance Z_0, commonly denoted as normalising impedance, can be considered.*

Remark 2.3.4. *In this work we consider Z_0 to be a pure real value. Additionally, we also assume that the value of R_1 is a real constant. Therefore we can set $Z_0 = R_1$ obtaining $s_g(\omega) = 0$ for all $\omega \in \mathbb{R}$. In this case eq. (2.9) becomes*

$$s_{opt}(\omega) = 0 \qquad \forall \omega \in \mathbb{R}.$$

2.4 Scattering Parameters

Consider again fig. 2.1, When the power dissipated in the load equals the maximum power P_{max}, we say the load matches the generator. In this case we have

$$P_{max} = \frac{1}{2}\frac{|a|^2}{Z_0} = \frac{1}{2}\frac{|E_1|^2}{4R_1}.$$

Assuming now R_1 real and taking $Z_0 = R_1$ we obtain

$$a = \frac{E_1}{2}.$$

Conversely, if $R \neq \overline{R_1}$, part of the incident power is said to be reflected. We have in this case a reflection coefficient s of the form

$$s = \frac{R - R_1}{R + R_1}.$$

The amount of reflected power is given by eq. (2.8) and denotes the power not dissipated in the load

$$P_{ref} = \frac{1}{2}\frac{|a|^2}{Z_0}s^2 = \frac{1}{2}\frac{|E_1|^2}{4R_1}\left(\frac{R - R_1}{R + R_1}\right)^2.$$

The reflected wave b takes the expression

$$b = \frac{E_1}{2}\frac{R - R_1}{R + R_1}.$$

In this case, we say that the load is unmatched. In this work we are interested in passive, linear and invariant microwave devices over time. These devices can be modelled traditionally by means of the scattering matrix, which relates linearly the outputs and inputs to the system, namely the parameters b and a previously defined. This motivates the formulation of the problem of matching in terms of the scattering matrix of each device. The aforementioned coupling matrix is defined in the following section.

Parameters a and b, denoted before as *power waves* represents the amount of power delivered to the load by the generator and the amount of reflected power. Using the previous definitions of delivered and reflected power, we can consider now a two-port device, where either port 1 or 2 can be excited by the generator. When port 1 is connected to a generator with internal impedance Z_0 and port 2 is closed by a resistor Z_0 (fig. 2.4 with $E_2 = 0$), we denote as a_1^2 the maximum available power and by b_2^2 the power actually delivered to the resistor Z_0. Also, denote b_1^2 the reflected power (not dissipated in the resistor). Scattering parameters associated to port 1 are then defined as the ratio between b_1, b_2 with respect to a_1

$$S_{11} = \left.\frac{b_1}{a_1}\right|_{E_2=0}, \qquad S_{21} = \left.\frac{b_2}{a_1}\right|_{E_2=0}. \tag{2.10}$$

Similarly, if port 2 is connected to the generator with internal impedance Z_0 and port 1 is closed by the resistor Z_0 (fig. 2.4 with $E_1 = 0$) we denote a_2^2 the maximum available power, b_1^2 the power delivered to resistor Z_0 and b_2^2 the reflected power. Then define scattering parameters associated to port 2 as the ration between b_1, b_2 and a_2

$$S_{12} = \left.\frac{b_1}{a_2}\right|_{E_1=0}, \qquad S_{22} = \left.\frac{b_2}{a_2}\right|_{E_1=0}.$$

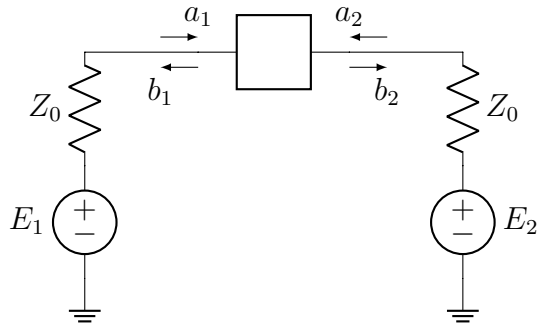


Figure 2.4: Incident and reflected waves

The principle of superposition, allows to compute b_1 and b_2 when both generator are connected by means of the scattering parameters.

$$b_1 = S_{11}a_1 + S_{12}a_2,$$
$$b_2 = S_{21}a_1 + S_{22}a_2.$$

These parameters are of customary use in RF-circuits as they completely characterise the behaviour of a linear device. We can therefore express parameters b_1, b_2 in terms of parameters a_1, a_2

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12}, \\ S_{21} & S_{22} \end{pmatrix} \cdot \begin{pmatrix} a_1, \\ a_2 \end{pmatrix} = [S] \cdot \begin{pmatrix} a_1, \\ a_2 \end{pmatrix}. \tag{2.11}$$

with the left and right column vectors containing the output and input power waves respectively. The matrix S, called scattering matrix is the main object of interest in this work.

2.4.1 Definitions

In this section, some basic notions in engineering as passivity or stability are linked to traditional concepts in complex analysis as the Schur class of functions. Those concepts are later used to tackle a well-known problem in electronics from a functional analysis point of view. From the functional perspective, the matrix S presents some important properties that will be useful for the theory developed in the following sections. However, before discussing such features, we need to provide some definitions.

Definition 2.4.1 (Analyticity domain)**.** *In this work we consider the real variable $\omega \in \mathbb{R}$ where $\mathbb{R}$ represents the real line. The extended real line $\overline{\mathbb{R}}$ is also considered in some cases, which is defined as*

$$\overline{\mathbb{R}} = \mathbb{R} \cup \infty.$$

Consider further the complex variable $\lambda = \omega + j\alpha$. We denote by $\mathbb{C}^+$ ($\mathbb{C}^-$) the open upper (lower) half plane (figs. 2.5c and 2.5e):

$$\mathbb{C}^+ = \{\lambda : \Im(\lambda) > 0\},$$
$$\mathbb{C}^- = \{\lambda : \Im(\lambda) < 0\}.$$

Equivalently $\overline{\mathbb{C}^+}$ *(*$\overline{\mathbb{C}^-}$*) denotes the closed upper (lower) half plane (figs. 2.5d and 2.5f)*

$$\overline{\mathbb{C}^-} = \{\lambda : \Im(\lambda) \geq 0\} \cup \infty,$$
$$\overline{\mathbb{C}^+} = \{\lambda : \Im(\lambda) \leq 0\} \cup \infty.$$

In this work we consider $\mathbb{C}^-$ *as the analyticity domain unless it is specified otherwise.*

An analytic function in $\mathbb{C}^-$ is a function that has convergent series in a neighbourhood of every point $\lambda \in \mathbb{C}^-$ (i.e. it is complex differentiable) [9, chapter 2, section 1]. For the electronic world, an analytic function in a domain is a function with no singularities in that domain. Therefore an analytic function is a stable function.

Definition 2.4.2 (Unit disk)**.** *We denote the open unit disk (fig. 2.5a) by* $\mathbb{D}$

$$\mathbb{D} = \{\lambda : |\lambda| < 1\}.$$

Similarly we use $\overline{\mathbb{D}}$ *to refer to the closed unit disk (fig. 2.5b)*

$$\overline{\mathbb{D}} = \{\lambda : |\lambda| \leq 1\}.$$

Remark 2.4.1. *In contrast to the usual definition in electronics or control theory, where the stable functions are defined as analytic in the right half of the* λ*-plane*

$$\mathbb{\Pi}^+ = \{\lambda \in \mathbb{C} : \Re(\lambda) > 0\}.$$

We consider analyticity in the lower half plane for convenience of notation in many parts of this thesis, we define $\lambda = \omega + j\alpha$ *as the frequency variable and* ω *the frequency axis.*

Remark 2.4.2. *With the chosen domain of analyticity, we also redefine a* stable *(Hurwitz) polynomial as the polynomial having all its roots in the open upper half plane (*$\mathbb{C}^+$*).*

Definition 2.4.3 (Star operation)**.** *We use the notation* $S^\star$ *to denote the transpose conjugate matrix of* S*.*

$$S(\lambda)^\star = \overline{S(\lambda)^T}.$$

Additionally, we also define the star *of a (matrix) function* S *as*

$$S^\star(\lambda) = S(\overline{\lambda})^\star.$$

It should be noted that $S(\omega)^\star$ *is the particularisation of* $S^\star(\lambda)$ *on the frequency axis*

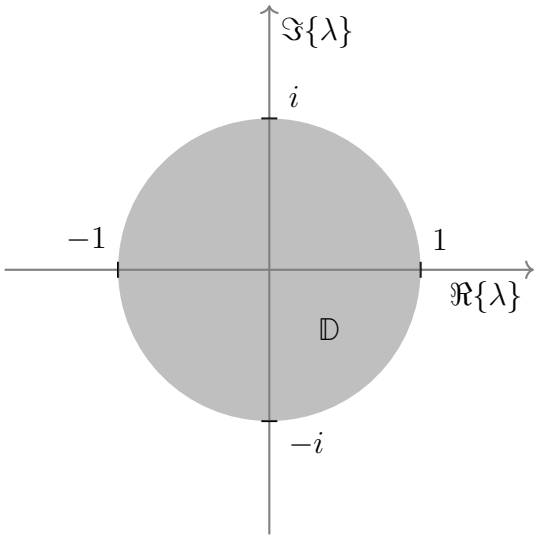
$$S^\star(\omega) = S(\omega)^\star \qquad\qquad \omega \in \mathbb{R}.$$

Therefore the star of the function S *(*$S^\star(\lambda)$*) represents the analytic continuation of* $S(\omega)^\star$ *with* $\omega \in \mathbb{R}$ *to the complex plane.*

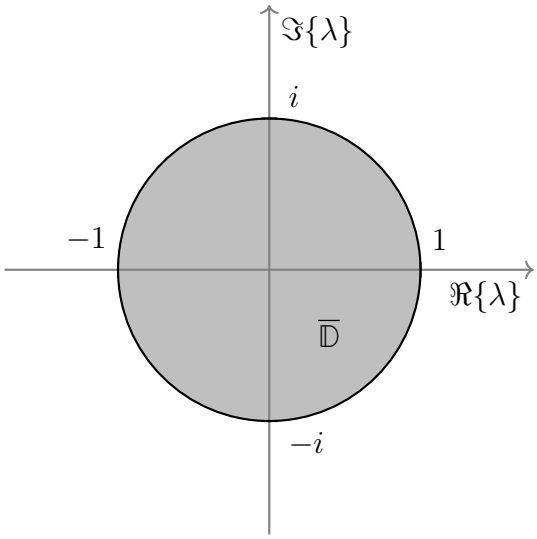
Similarly, the definition of the star operation over polynomials become just the complex conjugation of its roots. The starred polynomial (i.e. the polynomial with conjugate roots) is obtained by taking the complex conjugate of all coefficients.

Remark 2.4.3 (Star of a polynomial)**.** *Given the polynomial* $p(\lambda)$*. Define the starred polynomial* $p^\star(\lambda)$ *as*

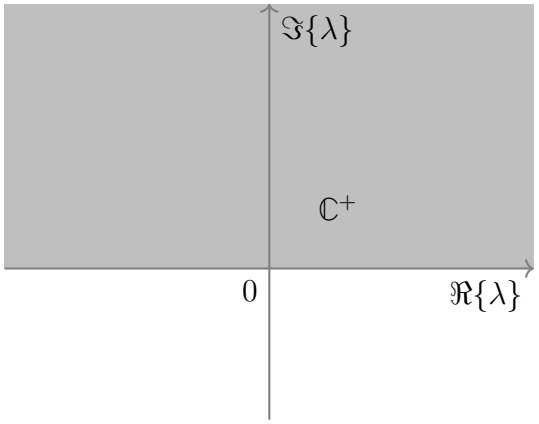
$$p^\star(\lambda) = \overline{p(\overline{\lambda})}$$



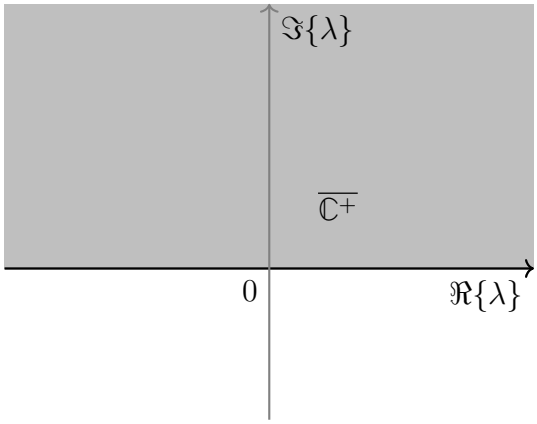
(a) Open unit disk



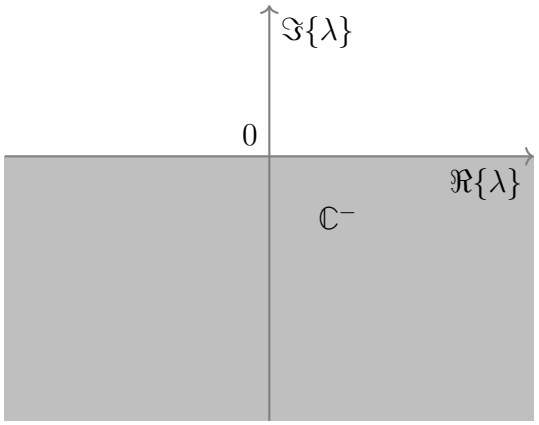
(b) Closed unit disk



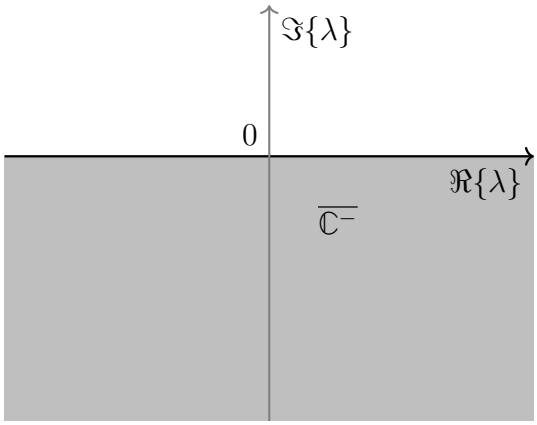
(c) Open upper half plane



(d) Closed upper half plane



(e) Open lower half plane



(f) Closed lower half plane

Figure 2.5: Representation of the different domains defined in this work

As an example we plot in fig. 2.6 the roots of the polynomial p

$$p(\lambda) = \lambda^5 - 1.5\lambda^4 + 2i\lambda^3 + (0.5 - 3i)\lambda^2 - 1.8\lambda + 1.5,$$

along with the conjugate roots corresponding to the polynomial $p^\star$

$$p(\lambda) = \lambda^5 - 1.5\lambda^4 - 2i\lambda^3 + (0.5 + 3i)\lambda^2 - 1.8\lambda + 1.5.$$

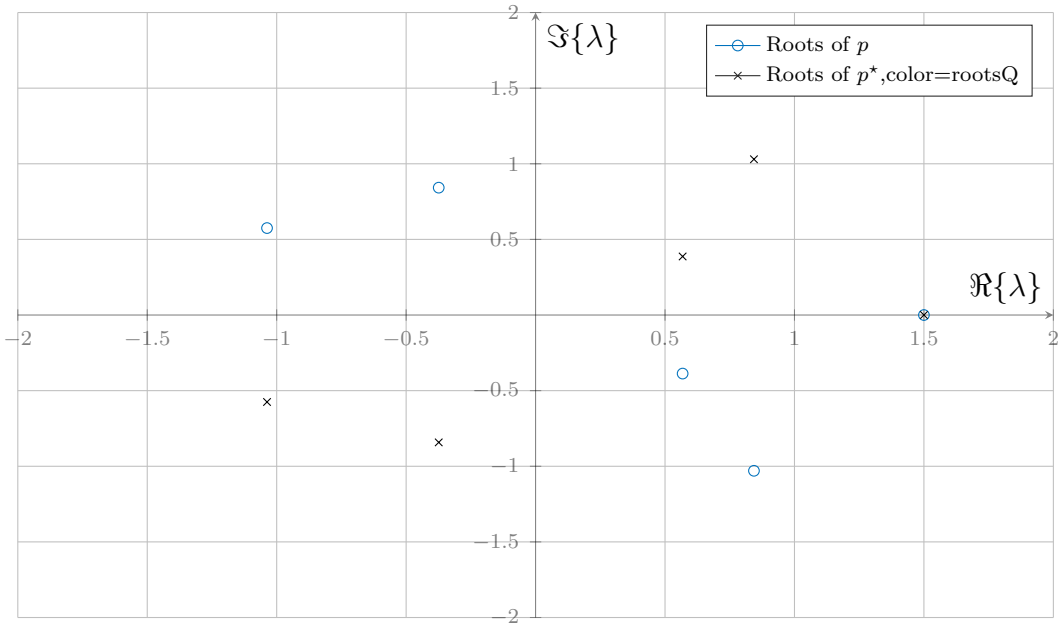


Figure 2.6: Roots of polynomial p and $p^\star$.

Let now introduce the Schur class of functions. Schur functions is a well-known class in functional analysis. It consists of analytic functions whose modulus is bounded by one in the analyticity domain. Traditionally, they are analytic functions from the unit disk to the unit disk. However, to fit the convention used in this work, they are redefined as analytic functions from the open lower half plane to the open unit disk. Schur functions repeatedly appear in this work, and therefore we denote the set of Schur functions by $\mathbb{S}$.

Definition 2.4.4 (Schur functions)**.** *Denote by $\mathbb{S}$ the class of functions analytic in $\mathbb{C}^-$ from the lower half plane to the disc.*

$$\mathbb{S} \equiv \left\{ f : \mathbb{C}^- \longrightarrow \mathbb{D} \right\}.$$

Among the class of Schur function, one particular kind of functions is remarked namely the functions composed by the product of factors in the form $(\lambda - \beta_i)(\lambda - \overline{\beta_i})^{-1}$. Those are functions whose modulus equals one at the boundary of the analyticity domain. We provide then the following additional definition [10, chapter 1, section 2]

Definition 2.4.5 (Blaschke product)**.** *A Blaschke product $b(\lambda)$ of degree N with zeros at the points β_i is the function in the form*

$$b(\lambda) = \prod_{i=1}^{N} \frac{\lambda - \beta_i}{\lambda - \overline{\beta_i}}, \tag{2.12}$$

with $\beta_1, \beta_2, \cdots, \beta_N \in \mathbb{C}^-$.

Note that $|b(\lambda)| = 1$ for all $\lambda \in \mathbb{R}$. In the domain of filter design, this function can be seen as the transmission coefficient of an all-pass filter .

Before continuing with the basic theory on scattering matrices, let us define the sets of Hermitian matrices of size $N \times N$

Definition 2.4.6 (Hermitian matrices)**.**

$$\mathbb{H}^N = \left\{ S \in \mathbb{C}^{N\times N} : \ S_{kl} = \overline{S_{lk}} \ \forall k, l \in [1, N] \right\}.$$

Now we define some matrix inequalities as matrix algebra plays a crucial role in this thesis.

Definition 2.4.7 (Matrix inequalities)**.** *Let us consider the $N \times N$ matrices $A, B \in \mathbb{H}^N$ such that the matrix $A - B$ is positive semi-definite ($A - B \succeq 0$). Then we denote*

$$A \succeq B.$$

Equivalently to represent the fact that $B - A \preceq 0$ we write

$$B \preceq A.$$

Now consider $C, D \in \mathbb{H}^N$ such that $C - D$ is positive definite ($C - D \succ 0$). Then

$$C \succ D.$$

Similarly we have $D - C \prec 0$. Then we denote

$$D \prec C.$$

Finally if $X \succeq Y$, with $X, Y \in \mathbb{H}^N$ we define the notation

$$X \succneqq Y.$$

to represent that the matrix $X - Y$ is singular.

These inequalities are particularly important for chapters 5 and 7.

Next we continue by providing some properties of scattering matrices. Note that, on the frequency axis $\lambda \in \mathbb{R}$, scattering parameters have modulus smaller than 1 for passive systems (if the system does not create energy, the output power cannot be greater than maximum available input power). Additionally, if the system is stable, scattering parameters have no poles in $\mathbb{C}^-$. At this point, it is important to remember the maximum modulus principle for holomorphic functions [9, theorem 12]

Theorem 2.4.1 (Maximum modulus principle)**.** *Given a function f analytic on an open subset $\Omega \subset \mathbb{C}$, there exist a point $\lambda_0 \in \Omega$ such that $|f(\lambda_0)| \geq |f(\lambda)|$ for all λ in a neighbourhood of λ_0 if and only if f is a constant function.*

2.5 Losslessness

We discuss now another important property of Scattering matrices related to losses. In reality, energy is always dissipated inside communication devices or any other system in

the form of heat or lost in any other form. The amount of power lost inside the device is called losses. Throughout this work, we often assume scattering matrices S are lossless, this is a physic assumption saying no energy is lost inside the device characterised by S. Equivalently, we mean the system is lossless. Note this assumption is never correct in practice. Nevertheless, losslessness is still an entirely reasonable approximation for the cases where losses are small providing us with an essential property of scattering matrices.

Note first that, by the maximum modulus principle, and from the fact that scattering parameters S_{ij} have modulus bounded by 1 in $\mathbb{R}$ (the boundary of the analyticity domain), we conclude that scattering parameters have also modulus bounded by 1 inside the analyticity domain $\mathbb{C}^-$.

$$|S_{ij}(\lambda)| < 1 \qquad\qquad \lambda \in \mathbb{C}^-.$$

Additionally, if there exist a point $\lambda_0 \in \mathbb{C}^-$ such that $|S_{ij}(\lambda_0)| = 1$, then S_{ij} is a constant function. Consider now the complex parameters a and b of the variable $\lambda \in \mathbb{C}$. Let review first the case where the device is strictly passive, namely some energy is dissipated inside of it. In this case, using the notation in fig. 2.4 again the sum of the power entering the system must be greater or equal than the power leaving the system

$$|a_1(\omega)|^2 + |a_2(\omega)|^2 - |b_1(\omega)|^2 - |b_2(\omega)|^2 \geq 0 \qquad\qquad \omega \in \mathbb{R}. \tag{2.13}$$

Denote now by B the vector of output waves and by A the vector of input waves

$$A(\omega) = \begin{pmatrix} a_1(\omega) \\ a_2(\omega) \end{pmatrix} \qquad\qquad B(\omega) = \begin{pmatrix} b_1(\omega) \\ b_2(\omega) \end{pmatrix}.$$

Note that $\overline{a_1(\omega)}a_1(\omega) = |a_1(\omega)|^2$. Thus imposing eq. (2.13) we have

$$A(\omega)^\star A(\omega) - B(\omega)^\star B(\omega) \geq 0 \qquad\qquad \omega \in \mathbb{R}.$$

Now introduce the expression of B provided by the scattering matrix

$$B(\omega) = S(\omega) \cdot A(\omega), \qquad\qquad B(\omega)^\star = A(\omega)^\star S(\omega)^\star.$$

Therefore

$$\begin{aligned} A(\omega)^\star A(\omega) - A(\omega)^\star S(\omega)^\star S(\omega) A(\omega) \geq 0 \qquad & \omega \in \mathbb{R}, \\ A(\omega)^\star \left(I - S(\omega)^\star(\omega) S(\omega)\right) A(\omega) \geq 0 \qquad & \omega \in \mathbb{R}. \end{aligned}$$

where I represents the identity matrix of size 2×2, and for any input vector A. Therefore $I - S(\omega)^\star S(\omega)$ is positive semi-definite for $\omega \in \mathbb{R}$ or equivalently

$$S(\omega)^\star S(\omega) \preceq I \qquad\qquad \omega \in \mathbb{R}.$$

When the previous relation holds, we refer to the scattering matrix as a lossless or conservative matrix. If now we assume that $B(\omega)^\star B(\omega) = A(\omega)^\star A(\omega)$ for a non-zero vector $A(\omega)$ and $\omega \in \mathbb{R}$, we obtain

$$I - S(\omega)^\star S(\omega) = 0.$$

This is the unitary property on the real axis. We can state now the following definition

Definition 2.5.1 (Losslessness). *A 2×2 scattering matrix S is lossless if and only if for any $\omega \in \mathbb{R}$*

$$S(\omega)^\star S(\omega) = I,$$

where I represents the 2x2 identity matrix.

We suppose hereinafter that scattering S matrices are lossless. Thus scattering parameters verify at ω real

$$S_{11}(\omega)S_{11}(\omega)^\star + S_{12}(\omega)S_{12}(\omega)^\star = 1, \tag{2.14}$$
$$S_{22}(\omega)S_{22}(\omega)^\star + S_{21}(\omega)S_{21}(\omega)^\star = 1, \tag{2.15}$$
$$S_{11}(\omega)S_{21}(\omega)^\star + S_{12}(\omega)S_{22}(\omega)^\star = 0. \tag{2.16}$$

Corollary 2.5.1 (Absolute value at real frequencies). *Note that, if the matrix is unitary, then we also have*

$$|S_{11}(\omega)| = |S_{22}(\omega)| \qquad \forall \omega \in \mathbb{R}. \tag{2.17}$$

Finally the last definition before introducing the general form of the scattering matrix the notion of reciprocity. A system is said to be reciprocal if the transmission from port k to port l equals the transmission from port l to port k. In other words, a reciprocal matrix is a symmetric matrix.

Definition 2.5.2 (Reciprocity). *A matrix S is reciprocal if and only if*

$$S(\lambda) = S(\lambda)^T \qquad \forall \lambda \in \mathbb{C}.$$

Reciprocity property is often found in passive communication devices. It should be noted that it is not impossible to implement a non-reciprocal passive device; however, the physics requirements behind it make it extremely complicated. This difficulty is the reason why usually scattering matrices are constrained to be reciprocal, even though non-reciprocal systems exist. For the theory developed in this thesis, most devices are assumed to be non-reciprocal, although reciprocal matrices also appear when convenient.

Remark 2.5.1 (Analytic continuation of scattering parameters to the complex plane). *It should be remarked that the theory developed up to this point considering the evaluation of scattering matrices on the real line ($\omega \in \mathbb{R}$) extends to the complex plane by replacing $S(\omega)^\star$ by $S^\star(\lambda)$ with $\lambda \in \mathbb{C}$.*

With the expressions derived in this section, we state the definition of the scattering matrix

Definition 2.5.3 (Scattering matrix). *We call* scattering matrix *a 2×2 matrix of the complex variable λ, analytic in $\mathbb{C}^-$ whose elements are scalar Schur functions.*

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \qquad S_{ij} \in \mathbb{\Sigma}. \tag{2.18}$$

Moreover scattering matrices satisfies the following property

$$S^\star(\lambda)S(\lambda) \succeq I \qquad \forall \lambda \in \mathbb{C}, \tag{2.19}$$

where equality in eq. (2.19) holds everywhere in the complex plane if the matrix S is lossless.

Corollary 2.5.2 (Determinant of unitary matrices)**.** *The unitary property allows us to obtain the determinant of a lossless scattering matrix as a function of the coefficients* S_{11} *and* S_{22} *only*

$$\det(S) = S_{11}S_{22} - S_{12}S_{21}. \tag{2.20}$$

Note that eq. (2.20) *is a stable function if the matrix* S *is stable. From (2.16)*

$$S_{12} = -\frac{S_{11}S_{21}^{\star}}{S_{22}^{\star}}.$$

Thus

$$\det(S) = S_{11}S_{22} + S_{11}\frac{S_{21}^{\star}S_{21}}{S_{22}^{\star}} = S_{11}\left(S_{22} + \frac{S_{21}^{\star}S_{21}}{S_{22}^{\star}}\right) = \frac{S_{11}}{S_{22}^{\star}}\left(S_{22}S_{22}^{\star} + S_{21}S_{21}^{\star}\right).$$

Using now (2.15) we obtain

$$\det(S) = \frac{S_{11}}{S_{22}^{\star}}, \tag{2.21}$$

which is uni-modular on the frequency axis as showed in eq. (2.17)*. Thus* $\det(S)$ *is a Blaschke product in the form given by eq.* (2.12)*.*

$$\det(S) = \epsilon\frac{q^{\star}}{q}, \tag{2.22}$$

with ϵ *an uni-modular constant and* q *a stable polynomial.*

Finally, we introduce the class of rational Schur functions of degree N.

Definition 2.5.4 (Rational Schur Functions)**.** *Denote by* $\mathbb{\Sigma}^N$ *the class of rational Schur functions where both numerator and denominator are polynomials of degree at most* $N \in \mathbb{N}$*. Additionally we denote by* $\mathbb{P}^N$ *the set of polynomials of degree at most* N*.*

$$\mathbb{\Sigma}^N = \left\{ f = \frac{p}{q} : \ p, q \in \mathbb{P}^N; \ f \in \mathbb{\Sigma} \right\},$$

where q *is a stable polynomial, namely* q *has no roots in* $\overline{\mathbb{C}^-}$*.*

2.6 Impedance and admittance matrices

The impedance and admittance parameters express the relations between the input-output voltage v_S, v_L and current i_S, i_L of a two-port network as the one shown in fig. 2.7 (see [7]). We can compare the circuit in fig. 2.7 and in fig. 2.4. Using now eqs. (2.4) and (2.5) and setting the reference $z = 0$ at the input terminal we have $v_S(\omega) = v(0, \omega)$ and $i_S(\omega) = i(0, \omega)$. Therefore

$$\begin{aligned} v_S(\omega) &= a_1(\omega) + b_1(\omega), \\ i_S(\omega) &= \frac{1}{Z_0}\left(a_1(\omega) - b_1(\omega)\right). \end{aligned}$$

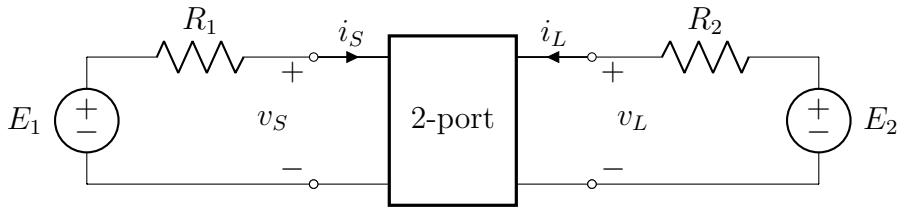


Figure 2.7: Voltage and current definition on a generic 2-port device

Similarly at port 2 we set the reference $z = 0$ at the output terminal to obtain

$$v_L(\omega) = a_2(\omega) + b_2(\omega),$$
$$i_L(\omega) = \frac{1}{Z_0}\left(a_2(\omega) - b_2(\omega)\right).$$

Let us now omit the dependence of the frequency variable ω for economy of notation. We can then write

$$a_1 = \frac{1}{2}\left(v_S + Z_0 i_S\right),$$
$$b_1 = \frac{1}{2}\left(v_S - Z_0 i_S\right),$$
$$a_2 = \frac{1}{2}\left(v_L + Z_0 i_L\right),$$
$$b_2 = \frac{1}{2}\left(v_L - Z_0 i_L\right).$$

Using eq. (2.11) we can relate the voltage and current at the input-output terminals of the network in fig. 2.7 by means of the scattering parameters as

$$\begin{pmatrix} v_S - Z_0 i_S \\ v_L - Z_0 i_L \end{pmatrix} = S \cdot \begin{pmatrix} v_S + Z_0 i_S \\ v_L + Z_0 i_L \end{pmatrix}.$$

Or equivalently, if we define the column vectors $v = [v_S,\ v_L]^T$ and $i = [i_S,\ i_L]^T$ we have

$$v - Z_0 i = S \cdot (v + Z_0 i). \tag{2.23}$$

2.6.1 Impedance matrix

The impedance parameters represents the relation between the voltage at the input-output terminals v_S, v_L and the current values i_S, i_L.

$$\begin{pmatrix} v_S \\ v_L \end{pmatrix} = \begin{pmatrix} Z_{1,1} & Z_{1,2} \\ Z_{2,1} & Z_{2,2} \end{pmatrix} \cdot \begin{pmatrix} i_S \\ i_L \end{pmatrix}.$$

The impedance parameters provide us with simple equivalent circuit of the network represented by the matrix Z which is shown in fig. 2.9. Additionally, in the case of a passive network we have $Z_{12} = Z_{21}$. In this case the network in fig. 2.9 is simplified as in fig. 2.10.

Finally note that from eq. (2.23) we have

$$v - S \cdot v = Z_0 (I + S) i,$$

and expressing the vector v as a function of the vector i

$$v = Z_0 (I - S)^{-1} (I + S) i = Z \cdot i,$$

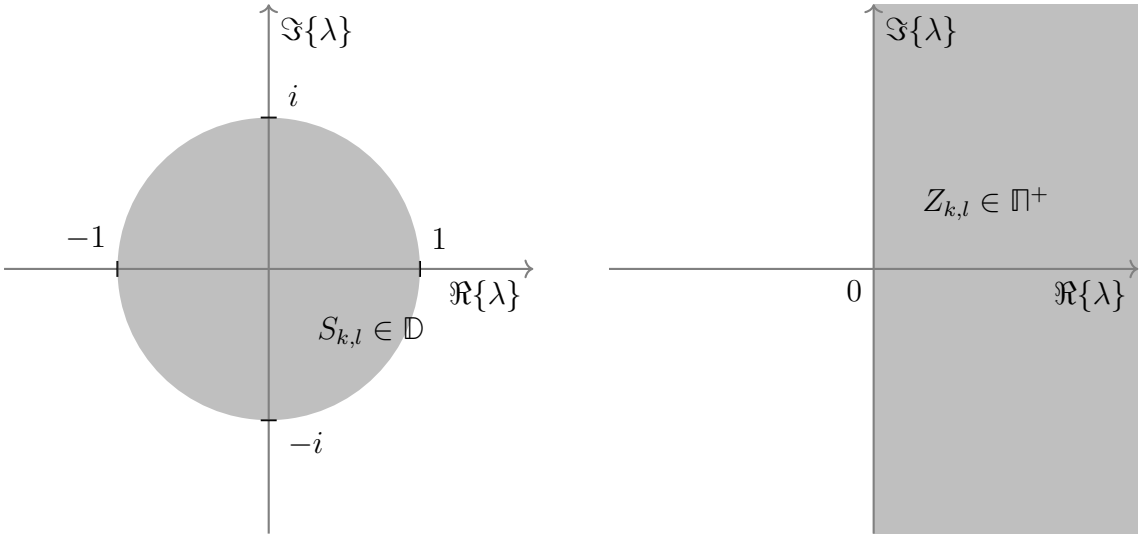
Therefore the impedance parameters are obtained from the scattering parameters as

$$Z = Z_0 (I - S)^{-1} (I + S). \tag{2.24}$$

Equivalently if the Z parameters are know we can compute

$$S = (Z - Z_0 I) (Z + Z_0 I)^{-1}.$$

Remark 2.6.1. *Note that eq.* (2.24) *sends the unit disk to the right half plane* $\mathbb{\Pi}^+$ *as illustrated in fig. 2.8. Therefore given a Schur S matrix, the corresponding Z parameters are positive real.*



(a) $S_{k,l} : \mathbb{C}^- \mapsto \mathbb{D}$ (b) $Z_{k,l} : \mathbb{C}^- \mapsto \mathbb{\Pi}^+$

Figure 2.8: For $\lambda \in \mathbb{C}^-$, the scattering parameters belong to the unit disk and impedance parameters to the right half plane

2.6.2 Admittance matrix

Similarly to the impedance matrix introduced in the previous section, the admittance matrix allows to express the values of i_S, i_L in fig. 2.7 as a function of the voltage at the input and output terminals, namely v_S, v_L. This admittance matrix is traditionally denoted by $Y(\omega)$ in circuit design. Note that the matrix Y is related to the impedance matrix Z by

$$Y(\omega) = Z(\omega)^{-1}.$$

Furthermore, it is also possible to construct a simple circuit (shown in fig. 2.11) equivalent to the network represented by the Y matrix. Finally, as in the case of the Z parameters, we have for a passive network $Y_{12} = Y_{21}$ and the simplified circuit shown in fig. 2.12 is obtained.

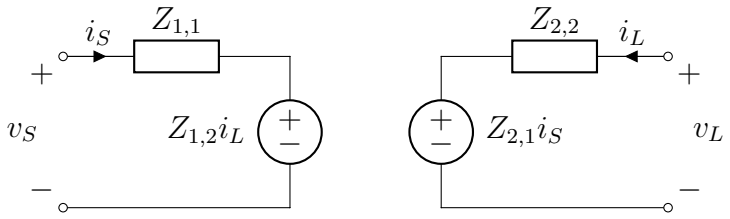


Figure 2.9: Equivalent circuit of the network represented by the Z matrix

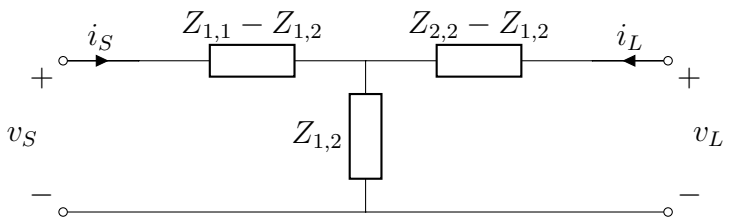


Figure 2.10: Equivalent circuit of a passive network ($Z_{12} = Z_{21}$) represented by the Z matrix .

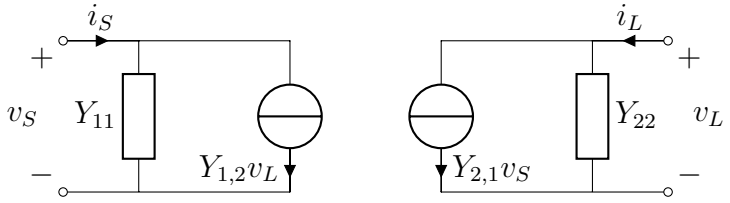


Figure 2.11: Equivalent circuit of the network represented by the Y matrix

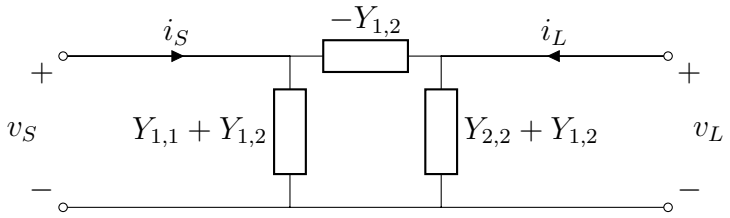


Figure 2.12: Equivalent circuit of a passive network ($Z_{12} = Z_{21}$) represented by the Y matrix .

2.7 Rational model of scattering matrices

So far we have considered the scattering parameters in terms of Schur functions only without any additional restriction. However, in most of this thesis we will consider that the scattering matrices follow a rational model of a certain degree. Furthermore, it is important to note that this approach is legitimate even when we are representing the scattering matrix associated with a physical microwave device since the response of said device can be faithfully approximated by a rational function within a not too large frequency interval.

In this section we provide a more in-depth analysis of linear systems in the field of communications. Linear systems have already been widely studied in the literature and are therefore familiar to many readers with a wide range of different backgrounds. In any case, we provide here some basic concepts related to the state space representation of a linear systems, which will serve as an introduction for other more specific topics discussed later. Nevertheless, for those who are interested in this topic and wish to read more in detail, I must recommend T. Kailath's book [11] on linear systems.

2.7.1 State space representation

Rational scattering matrices represents the impulse response of a linear, time-invariant (LTI) system with 2 input and 2 outputs which can be written in the general state-space form as

$$\frac{dx}{dt} = U \cdot x(t) + V \cdot a(t), \tag{2.25}$$

$$b(t) = C \cdot x(t) + D \cdot a(t), \tag{2.26}$$

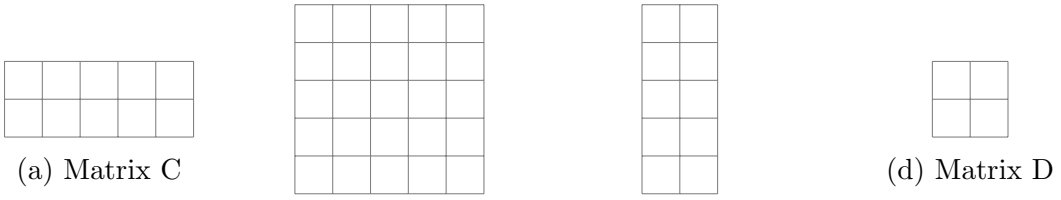
where $a, b \in \mathbb{R}^2$ are the input and output vector respectively and $x(t)$ is denoted the state vector (as we show below, the number of states corresponds to the McMillan degree of the system). Additionally, U, V, C, D are matrices with the following sizes

- C: number of inputs $\times$ number of states
- U: number of states $\times$ number of states
- V: number of states $\times$ number of outputs
- D: number of inputs $\times$ number of outputs

The aspect of these matrices is illustrated in fig. 2.13 for a system with 2 inputs, 2 outputs, and 5 states.

The frequency response of the previous system is trivially obtained by taking the Laplace transform in eqs. (2.25) and (2.26)

$$\lambda X(\lambda) = UX(\lambda) + VA(\lambda),$$
$$B(\lambda) = CX(\lambda) + DA(\lambda),$$



(a) Matrix C (b) Matrix U (c) Matrix V (d) Matrix D

Figure 2.13: Size of the matrices appearing in the state space representation of a system with 5 states, 2 inputs and 2 outputs

where we consider the initial state equal to zero and upper-case variable as the Laplace transform to the corresponding lower-case ones. Then we have

$$X(\lambda) = (\lambda I - U)^{-1} V A(\lambda),$$
$$B(\lambda) = C\,(\lambda I - U)^{-1} V A(\lambda) + D A(\lambda).$$

Introducing the definition of the scattering matrix to express $B(\lambda) = S(\lambda)A(\lambda)$ we have

$$S(\lambda) = D - C\,(U - \lambda I)^{-1} V. \tag{2.27}$$

Remark 2.7.1. *It should be noted that given a rational matrix $S(\lambda)$, the matrices D, V, U appearing in eq.* (2.27) *are not unique.*

2.7.2 Rational form of the transfer function obtained from its state space representation.

The numerator of each element in the matrix $S(\lambda)$ as well as the denominator can be easily obtained from the matrices U, V, C, D. First note that there exists a unique matrix D such that eq. (2.27) holds. This matrix D contains the values at infinity, namely

$$D = \lim_{\lambda \to \infty} S(\lambda).$$

Note further that the eigenvalues of the matrix U represents the values of λ for which $\lambda I - U$ is singular. These eigenvalues correspond to the poles of the matrix $S(\lambda)$.

Finally, we shall compute the zeros of the element i, j of $S(\lambda)$ to fully determine the rational matrix S. Let us consider now an element i, j and denote by c the i-th row of the matrix C, by v the j-th column of V meanwhile d represents the i, j element of the matrix D. We have

$$S_{i,j}(\lambda) = d - c(U - \lambda I)^{-1} v.$$

Consider now the matrix $M(\lambda)$ defined as

$$M(\lambda) = \left[\begin{array}{c|c} U & v \\ \hline c & d \end{array}\right] - \left[\begin{array}{c|c} \lambda I & 0 \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} u(\lambda) & v \\ \hline c & d \end{array}\right]$$

where $u(\lambda) = U - \lambda I$. For compacity denote by z the inverse $z = u(\lambda)^{-1}$. The inverse of the matrix M takes then the block-wise expression

$$M^{-1} = \left[\begin{array}{c|c} z + uzv(d - czv)^{-1}cz & -zv(d - czv)^{-1} \\ \hline -(d - czv)^{-1}cz & (d - czv)^{-1} \end{array}\right].$$

We have already determined the values of λ for which $z = u(\lambda)^{-1}$ is singular. We seek now the values of λ such that $d - czv = 0$, namely $S_{i,j}(\lambda) = 0$. Equivalently we compute the values of λ such that the matrix $M(\lambda)$ is singular

$$\det\left(\left[\begin{array}{c|c} U & v \\ \hline c & d \end{array}\right] - \lambda \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right]\right) = 0. \tag{2.28}$$

Equation (2.28) is a generalised eigenvalue problem where the generalised eigenvalues solution to *eq.* (2.28) correspond to the roots of the function $S_{i,j}(\lambda)$.

Remark 2.7.2. *It can be remarked that the above discussion allows us to write the function $S_{i,j}$ in a rational form*

$$S_{i,j}(\lambda) = \frac{p(\lambda)}{q(\lambda)},$$

where the polynomials p, q are given by

$$p(\lambda) = \det\left(\left[\begin{array}{c|c} U & v \\ \hline c & d \end{array}\right] - \lambda \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array}\right]\right),$$

$$q(\lambda) = \det\left(U - \lambda I\right).$$

We have reached Rosenbrock's *expression which relates the state space representation C, D, U, V of a system with its rational matrix form, which was first proposed in [12].*

2.7.3 McMillan degree

After this introduction to state space representation, let us provide the definition of McMillan degree of a matrix S.

Definition 2.7.1 (McMillan degree)**.** *The McMillan degree of the rational scattering matrix $S(\lambda)$ is defined as the minimum size of the square matrix U in eq.* (2.27) *required to realise $S(\lambda)$.*

2.8 The coupling matrix

Once in possession of a rational model of the optimal matching filter, there are different tools in the literature to assist in the design of physical structures which provide a frequency behaviour as close as possible to the behaviour of the optimal model.

One of the tools which are commonly used microwave engineering for the design of coupled resonators filters is the coupling matrix. The coupling matrix is a ruse that allows us to convert the scattering parameters, determined by means of a rational matrix, to a set of coefficients directly related to the physical dimensions of the structure, which provides the desired response. We provide next a basic introduction to the coupling matrix formalism. Nevertheless note that an exhaustive lecture on this topic can be found, for instance, in [7].

The circuit in fig. 2.14 represents the low-pass prototype of a coupled resonator network. This circuit provides a response of type low-pass, where the elements $M_{k,k}$ denote frequency invariant reactances which are included to consider asynchronous responses.

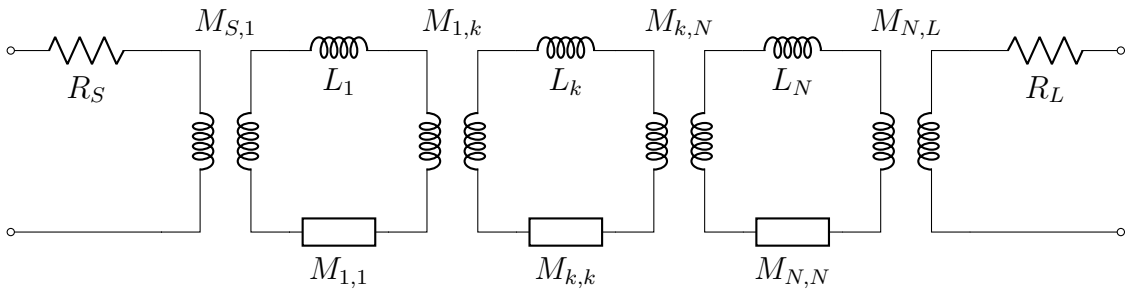


Figure 2.14: Lowpass prototype of a coupled resonators network

Remark 2.8.1. *Note that, with the appropriate choice of the values $M_{i,k}$, we can normalise all inductance and resistor values to one without loss of generality, namely*

$$\begin{aligned} R_S &= 1, \\ L_k &= 1 \qquad\qquad \forall k \in [1, N], \\ R_L &= 1. \end{aligned}$$

Remark 2.8.2. *Note that the circuit in fig. 2.14 allows for the representation of band limited responses in a high frequency interval $\omega_1 \leq \omega \leq \omega$ by means of the frequency transformation*

$$\omega_T = \frac{2\omega - (\omega_2 + \omega_1)}{(\omega_2 - \omega_1)}.$$

This transformation sends the frequency ω_1 to $\omega_T = -1$ and ω_2 to $\omega_T = 1$ while the band $\omega_1 \leq \omega \leq \omega$ is transformed to the interval $[-1, 1]$.

Let us now compute the admittance matrix provided by the circuit in fig. 2.14. In order to apply a Kirchoff analysis, the current i_k is defined clockwise in the k-the resonator

while v_S, v_L together with i_S, i_S represent the voltage and current at the input-output terminals as illustrated in fig. 2.15 where the normalised value 1 is considered for the elements R_S, R_L, L_k with $1 \leq k \leq N$.

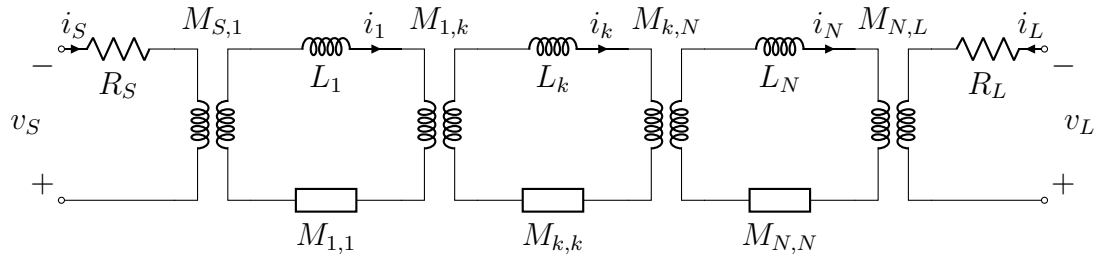


Figure 2.15: Voltage and current definition for the circuit in fig. 2.14

Considering the section k in fig. 2.15 other than the source or load loop, we can write the current law as the sum of currents equal to zero

$$i_S M_{k,S} + i_1 M_{k,1} + \cdots + i_k(\omega L_k + M_{k,k}) + \cdots + i_N M_{N,1} + i_L M_{L,1} = 0.$$

Setting the elements L_k equal to one we can write in matrix form

$$\left(\omega I + \begin{bmatrix} M_{1,1} & \cdots & M_{1,N} \\ \vdots & \ddots & \vdots \\ M_{N,1} & \cdots & M_{N,N} \end{bmatrix}\right) \cdot \begin{bmatrix} i_1 \\ \vdots \\ i_N \end{bmatrix} = - \begin{bmatrix} M_{1,S} & M_{1,L} \\ \vdots & \vdots \\ M_{N,S} & M_{N,L} \end{bmatrix} \cdot \begin{bmatrix} i_S \\ i_L \end{bmatrix}.$$

Therefore we express

$$\begin{bmatrix} i_1 \\ \vdots \\ i_N \end{bmatrix} = - \left(\omega I + \begin{bmatrix} M_{1,1} & \cdots & M_{1,N} \\ \vdots & \ddots & \vdots \\ M_{N,1} & \cdots & M_{N,N} \end{bmatrix}\right)^{-1} \begin{bmatrix} M_{1,S} & M_{1,L} \\ \vdots & \vdots \\ M_{N,S} & M_{N,L} \end{bmatrix} \cdot \begin{bmatrix} i_S \\ i_L \end{bmatrix}. \tag{2.29}$$

Considering the input and source loops we have

$$\begin{aligned} v_S &= i_1 M_{S,1} + i_2 M_{S,2} + \cdots + i_N M_{S,N} + i_S R_S + i_L M_{S,L}, \\ v_L &= i_1 M_{L,1} + i_2 M_{L,2} + \cdots + i_N M_{L,N} + i_S M_{L,S} + i_L R_L. \end{aligned}$$

Again in matrix form we obtain

$$\begin{bmatrix} v_S \\ v_L \end{bmatrix} = \begin{bmatrix} R_S & M_{S,L} \\ M_{L,S} & R_L \end{bmatrix} \cdot \begin{bmatrix} i_S \\ i_L \end{bmatrix} + \begin{bmatrix} M_{S,1} & \cdots & M_{S,N} \\ M_{L,1} & \cdots & M_{L,N} \end{bmatrix} \cdot \begin{bmatrix} i_1 \\ \vdots \\ i_N \end{bmatrix}. \tag{2.30}$$

Let us define now the matrices C, D, U, V

$$C = \begin{bmatrix} M_{S,1} & \cdots & M_{S,N} \\ M_{L,1} & \cdots & M_{L,N} \end{bmatrix} \tag{2.31}$$

$$D = \begin{bmatrix} R_S & M_{S,L} \\ M_{L,S} & R_L \end{bmatrix} \tag{2.32}$$

$$U = -\begin{bmatrix} M_{1,1} & \cdots & M_{1,N} \\ \vdots & \ddots & \vdots \\ M_{N,1} & \cdots & M_{N,N} \end{bmatrix} \tag{2.33}$$

$$V = \begin{bmatrix} M_{1,S} & M_{1,L} \\ \vdots & \vdots \\ M_{N,S} & M_{N,L} \end{bmatrix}. \tag{2.34}$$

Introducing now eq. (2.29) in eq. (2.30) we can express the values of v_S, v_L as a function of the currents i_S, i_L only, by means of the matrices C, D, U, V

$$\begin{bmatrix} v_S \\ v_L \end{bmatrix} = D \cdot \begin{bmatrix} i_S \\ i_L \end{bmatrix} - C\left(\omega I - U\right)^{-1} V \cdot \begin{bmatrix} i_S \\ i_L \end{bmatrix}.$$

Thus the 2×2 impedance matrix Z of the network in fig. 2.14 takes the following rational expression

$$Z(\omega) = D - C(\omega I - U)^{-1} V. \tag{2.35}$$

Remark 2.8.3. *Note we can compare eq.* (2.35) *to eq.* (2.27) *to conclude that the matrices* C, D, U, V *represents a space state realisation of the impedance matrix* Z*.*

We can define finally the coupling matrix by using the matrices C, D, U, V defined before

Definition 2.8.1 (Coupling matrix)**.** *Given a 2-port network with rational* 2×2 *impedance matrix* Z *such that*

$$Z(\omega) = D - C(\omega I - U)^{-1} V, \tag{2.36}$$

where C, D, U, V *are defined as in eqs.* (2.31) *to* (2.34)*.*

The coupling matrix M *associated to the given 2-port device is defined as the matrix composed by the elements in the matrices* C, D, U, V

$$M = \begin{bmatrix} R_S & M_{S,1} & \cdots & M_{S,N} & M_{S,L} \\ M_{1,S} & M_{1,1} & \cdots & M_{1,N} & M_{1,L} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ M_{N,S} & M_{N,1} & \cdots & M_{N,N} & M_{N,L} \\ M_{L,S} & M_{L,1} & \cdots & M_{L,N} & R_L \end{bmatrix}. \tag{2.37}$$

Remark 2.8.4. *Note that the element highlighted in red in eq.* (2.37) *constitute the matrix* C *in eq.* (2.36)*. Similarly the blue corners corresponds to the matrix* D*, the purple column to* V *meanwhile the inner sub-matrix highlighted in green in eq.* (2.37) *equals to* $-U$*.*

Remark 2.8.5. *We have derived an expression of the impedance matrix of the network represented by the coupling matrix* M*, which is shown in fig. 2.14, as a function of the sub-matrices* C, D, U, V*. This expression is given by eq.* (2.35)*. Therefore we conclude that the coupling matrix corresponds simply to a state space representation of the matrix* $Z(\omega)$*. Nevertheless it should be noted that this understanding the the coupling matrix as a state space realisation of the impedance matrix* $Z(\omega)$*, being more general, differs from the classical definition found in [7] which is commonly taught in microwave courses.*

Remark 2.8.6. *Due to the equivalent network in fig. 2.14 where the elements* $M_{k,l}$ *are defined as the mutual coupling between the different sections, it is usually assumed in the design of microwave filters that, for a lossless passive network, all elements* $M_{k,l}$ *are pure imaginary*

$$\begin{aligned}
\Re\left(M_{k,S}\right) &= 0 && \forall k \in [1, N], \\
\Re\left(M_{S,l}\right) &= 0 && \forall l \in [1, N], \\
\Re\left(M_{k,l}\right) &= 0 && \forall k, l \in [1, N], \\
\Re\left(M_{k,L}\right) &= 0 && \forall k \in [1, N], \\
\Re\left(M_{L,l}\right) &= 0 && \forall l \in [1, N].
\end{aligned}$$

Together with the normalisation $R_S = R_L = 1$*, this assumption allows us to obtain the coupling matrix* M *only by specifying its imaginary part, namely* $-jM$*. Nevertheless note that we can multiply the matrices* C, U, V *by a complex basis change matrix* T *which is not singular such that the matrices* C, D, U, V *represents the same systems as the matrices* $\hat{C}, D, \hat{U}, \hat{V}$ *defined by*

$$\begin{aligned}
\hat{A} &= T \cdot A \cdot T^{-1}, \\
\hat{C} &= C \cdot T^{-1}, \\
\hat{V} &= T \cdot V.
\end{aligned}$$

Note that we have

$$D - C(\omega I - U)^{-1} V = D - \hat{C}(\omega I - \hat{U})^{-1} \hat{V}.$$

Therefore we obtain an equivalent representation of the same lossless network where the matrices $\hat{C}, \hat{U}, \hat{V}$ *are complex. Additionally note that we can also consider a complex* D *matrix if we allows for frequency independent phase shift elements. In this case the value of* $\lim_{\omega \to \infty} Z(\omega)$ *can be implemented by a given input-output reactance which is frequency independent.*

Remark 2.8.7. *The impedance matrix* $Z(\lambda)$ *expressed as in eq.* (2.36)*, the corresponding coupling matrix* M *can be immediately obtained. Nevertheless, as pointed out in remark 2.7.1, the matrices* C, U, V*, and therefore the coupling matrix are not unique. As a result we can obtain different coupling topologies in the form of the circuit in fig. 2.14, all of them providing the same impedance matrix* $Z(\lambda)$*. Taking into account the non-unicity*

of the matrix M is crucial for the physical implementation of the network represented in fig. 2.14 as the realisation of each topology might have a different level of complexity. This fact is specially important when some of the elements $M_{k,l}$ vanish in a particular topology, providing a simplified network.

2.9 Belevitch form

The rational scattering matrices are defined here following the Belevitch model used traditionally to parametrise lossless 2-port networks. This parametrisation is customary in the synthesis of electrical network, specially coupled resonator networks (see for instance [13]). Belevitch stated that, any 2×2 matrix S rational and stable that satisfies $S^\star S = I$ can be parametrised a function of 3 polynomials $p, r, q \in \mathbb{P}^N$ with q a stable polynomial of maximum degree N which can be obtained by spectral factorisation of the positive polynomial $qq^\star = pp^\star + rr^\star$. We state next the original Belevitch theorem [14]

Theorem 2.9.1 (Belevitch form)**.** *Any rational 2×2 unitary (lossless) matrix of McMillan degree $N \in \mathbb{N}$ can be parametrised in the Belevitch form as*

$$S = \frac{1}{q}\begin{pmatrix} \epsilon p^\star & -\epsilon r^\star \\ r & p \end{pmatrix}, \tag{2.38}$$

with ϵ a uni-modular constant, and $q, p, r \in \mathbb{P}^N$ with q a stable polynomial of maximum degree N satisfying $qq^\star = pp^\star + rr^\star$.

Proof. Consider the rational matrix S. If S is lossless equality holds in eq. (2.19). Hence we have

$$S^\star = S^{-1}. \tag{2.39}$$

Using now the co-factors matrix to express S^{-1}

$$S^{-1} = \frac{\mathrm{cof}(S)}{\det(S)} = \frac{\begin{pmatrix} S_{22} & -S_{12} \\ -S_{21} & S_{11} \end{pmatrix}}{\det(S)}.$$

Using now eqs. (2.22) and (2.39) we express

$$\epsilon q^\star \cdot S^\star = q \cdot \mathrm{cof}(S). \tag{2.40}$$

Note that, $\mathrm{cof}(S)$ has poles only in $\mathbb{C}^+$ meanwhile the poles of $S^\star$ belong to $\mathbb{C}^-$. Therefore, since equality holds in eq. (2.40), all poles of $S^\star$ are simplified by $q^\star$ and all poles of $\mathrm{cof}(S)$ cancel with the roots of q. Thus S can be written as a polynomial matrix divided by q

$$S = \frac{1}{q}\begin{pmatrix} p_1 & p_2 \\ p_3 & p_4 \end{pmatrix}.$$

Finally using eq. (2.39) we have

$$\frac{1}{q^\star}\begin{pmatrix} p_1^\star & p_3^\star \\ p_2^\star & p_4^\star \end{pmatrix} = \frac{1}{q \cdot \det(S)}\begin{pmatrix} p_4 & -p_2 \\ -p_3 & p_1 \end{pmatrix} = \frac{\overline{\epsilon}}{q^\star}\begin{pmatrix} p_4 & -p_2 \\ -p_3 & p_1 \end{pmatrix}.$$

Thus

$$p_1 = \epsilon p_4^\star,$$
$$-p_2 = \epsilon p_3^\star.$$

Finally note that when computing the determinant as in eq. (2.21), the denominator of S_{11} can not simplify with the denominator of $S_{22}^\star$ while both numerators cancels out. Also note that the eigenvalues of matrix U in eq. (2.27) correspond to the poles of S. □

2.9.1 Transmission zeros and transmission polynomial

Next let us provide the definition of transmission zeros. This notion is widely used in the design of RF devices and references the frequencies at witch a device is able to completely isolate port 2 and 1. This is the reason for the denomination, since the transmission from one port to the other is zero at those frequencies.

Definition 2.9.1 (Transmission zeros)**.** *We define the transmission zeros associated to a matrix function S in the form (2.18) as the zeros in $\overline{\mathbb{C}^-}$ (possibly at ∞) of $S_{12}S_{21}(\lambda)$:*

$$tz\,[S] = \left\{\lambda \in \overline{\mathbb{C}^-} : S_{12}S_{21}(\lambda) = 0\right\}$$

where we consider the classical multiplicity of the transmission zeros in $\mathbb{C}^-$ and half of the multiplicity for the transmission zeros in $\mathbb{R}$.

Remark 2.9.1. *If α_0 is a transmission zero of S and $S_{12}(\alpha_0) = 0$, from property (2.14) we obtain $S_{11}(\alpha_0)S_{11}^\star(\alpha_0) = 1$; conversely if $S_{21}(\alpha_0) = 0$ then (2.15) gives $S_{22}(\alpha_0)S_{22}^\star(\alpha_0) = 1$. Additionally, at a transmission zero $\alpha_0 \in \mathbb{R}$ from eq. (2.17) we have $|S_{11}(\alpha_0)| = |S_{22}(\alpha_0)| = 1$ and $S_{21}(\alpha_0) = S_{12}(\alpha_0) = 0$. Consequently transmission zeros $\alpha_0 \in \mathbb{R}$ are zeros of both S_{12} and S_{21} and therefore have even multiplicity.*

Note that it is possible, even in the case of reciprocal matrices, that a pole-zero cancellation occurs in $S_{12}(\lambda)$ or $S_{21}(\lambda)$ at a point $\lambda \in \mathbb{C}^+$. Nevertheless, with the given definition, the transmission zeros, being in $\overline{\mathbb{C}^-}$, can never simplify with the roots of q, which belong to $\mathbb{C}^+$. Related to the notion of transmission zeros, we find the concept of transmission polynomial.

Definition 2.9.2 (Transmission polynomial)**.** *We denote by transmission polynomial of a scattering matrix S the positive polynomial $R = rr^\star$ where r is the polynomial appearing in the Belevitch form of S (eq. (2.38)). This polynomial R contains among its roots all transmission zeros counting multiplicity, apart for those at infinity.*

As an example, we show in fig. 2.16 the poles and zeroes of the functions S_{21} and S_{22} corresponding to an arbitrary scattering matrix of McMillan degree 5. Particularly we indicate the roots of each polynomial appearing in the Belevitch form of the system. It can first be noted how the denominator polynomial q is stable, namely its roots belong to the open upper half plane. Additionally we can remark that the system is reciprocal as the roots of the transmission polynomial r appear either on the real axis or in complex conjugate pairs.

Remark 2.9.2. *Note that for the purposes of this work, the distribution of the roots of R between the polynomials r and $r^\star$ is not relevant and can therefore be done arbitrary. Furthermore, we consider most of the time, scattering matrices to be non-reciprocal. Thus no additional condition is imposed on the polynomial R or its roots, apart for it being positive. Nevertheless if matrix S is reciprocal, then all roots of R have even multiplicity.*

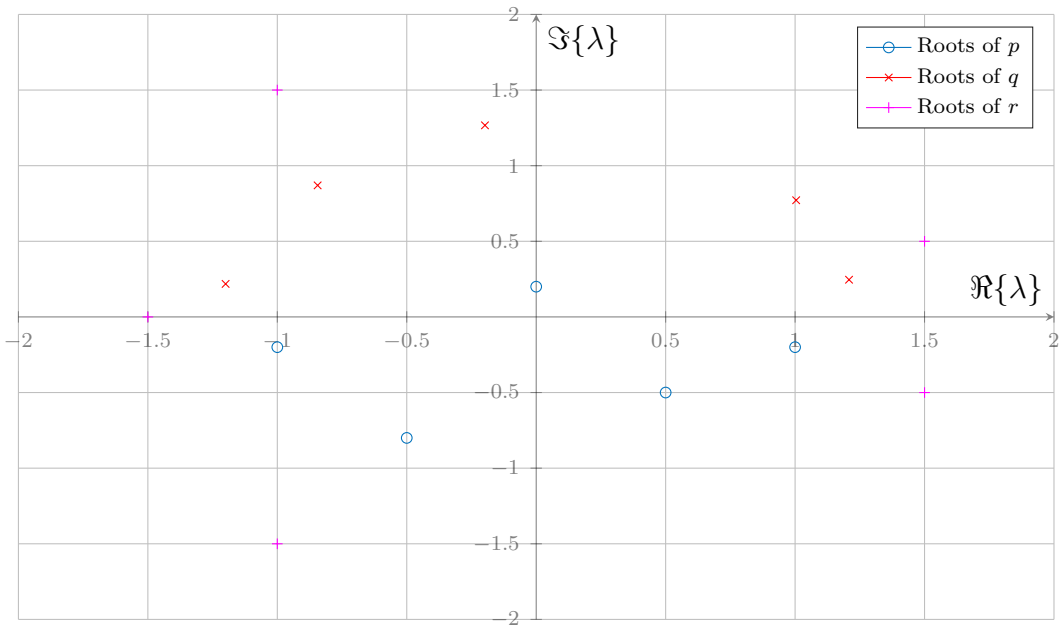


Figure 2.16: Pole-zero diagram with the roots of each polynomial p, r, q from the Belevitch form of an arbitrary reciprocal system.

2.9.2 Belevitch form of an all-pass device

An all-pass device is characterised for having zero reflection coefficients S_{11}, S_{22} while presenting a non-constant uni-modular transmission coefficient, either S_{12}, S_{21} or both. Therefore we have a zero reflection polynomial $p = p^\star = 0$ and an arbitrary transmission polynomial r. Note that the denominator q obtained as the stable factor of $qq^\star = rr^\star$, have roots at all zeros of $rr^\star$ in $\mathbb{C}^+$. As a result, all zeros of r belonging to $\mathbb{C}^+$ simplify out in S_{21} with the corresponding zeros of q while the zeros of $r^\star$ in $\mathbb{C}^+$ simplifies in S_{12}. Additionally, no transmission zeros are introduced on the frequency axis by an all-pass device as if $r(\alpha_i) = 0$ with $\alpha_i \in \mathbb{R}$ then we have $r(\alpha_i) = r^\star(\alpha_i) = q(\alpha_i) = 0$ obtaining a pole-zero cancellation in both S_{21} and S_{12}.

As an example consider the transmission polynomial of degree $B = B_1 + B_2$ with roots at the points $\beta_i, \alpha_i \in \mathbb{C}^-$ such that

$$r = \prod_{i=1}^{B_1}(\lambda - \alpha_i)\prod_{i=1}^{B_2}(\lambda - \overline{\beta_i}).$$

If this polynomial is reciprocal, namely $B_1 = B_2$ and $\alpha_i = \beta_i$ for all $i \in [1, B_1]$, then we obtain a transmission coefficients $S_{12} = S_{21}$ of the form

$$S_{21} = S_{21} = \prod_{i=1}^{B_1}\frac{(\lambda - \alpha_i)}{(\lambda - \overline{\alpha_i})}.$$

However, in the general case of a non-reciprocal all-pass device, the function S_{21} has

roots at the points $\alpha_i \in \mathbb{C}^-$ while S_{12} vanishes at the points $\beta_i \in \mathbb{C}^-$. Thus

$$S_{21} = \prod_{i=1}^{B_1} \frac{(\lambda - \alpha_i)}{(\lambda - \overline{\alpha_i})},$$
$$S_{12} = \prod_{i=1}^{B_2} \frac{(\lambda - \beta_i)}{(\lambda - \overline{\beta_i})}.$$

Remark 2.9.3. *Note that the degree of both parameters S_{21} and S_{12} might not be the same in this case. Indeed if $B_1 \neq B_2$ a different number of simplifications occurs in each of the functions. Nevertheless the McMillan degree (B) of the device does not drop unless a simplification occurs in both S_{12} and S_{21}.*

As an example, we can consider the limiting case where the polynomial r has all roots in $\mathbb{C}^-$, namely

$$r = \prod_{i=1}^{B} (\lambda - \alpha_i).$$

In this case we have

$$S_{11} = 0,$$
$$S_{22} = 0,$$
$$S_{12} = 1,$$
$$S_{21} = \prod_{i=1}^{B} \frac{(\lambda - \alpha_i)}{(\lambda - \overline{\alpha_i})}.$$

Note that transmission zeros are only present in the coefficient S_{21}, namely the transmission from port 1 to port 2 while the transmission from port 2 to port 1 never vanishes. This fact evidences the non-reciprocity of the device. This case is illustrated in fig. 2.17 where we show the roots of the polynomials q and r corresponding to an all-pass device. Note that the roots of r are all in $\mathbb{C}^-$, therefore no simplification occurs in the coefficient S_{21}. Conversely the roots of $r^\star$ coincide with the roots of q producing a complete pole-zero cancellation in the function S_{12}. As a result we obtain a constant parameter $S_{12} = 1$.

2.9.3 Darlington equivalent

The Belevitch form allows to reconstruct the scattering matrix given the polynomials p, q, r and the uni-modular constant ϵ. Additionally, if only polynomials p and q are known, such that $f = \frac{p}{q} \in \mathbb{\Sigma}$, then we can construct transmission polynomial $R = qq^\star - pp^\star$ such that there exist a scattering matrix S

$$S = \frac{1}{q} \begin{pmatrix} p & -\epsilon r^* \\ r & \overline{\epsilon} p^\star \end{pmatrix},$$

with $rr^\star = R$. This matrix satisfies $S_{11} = f$.

Darlington theorem states

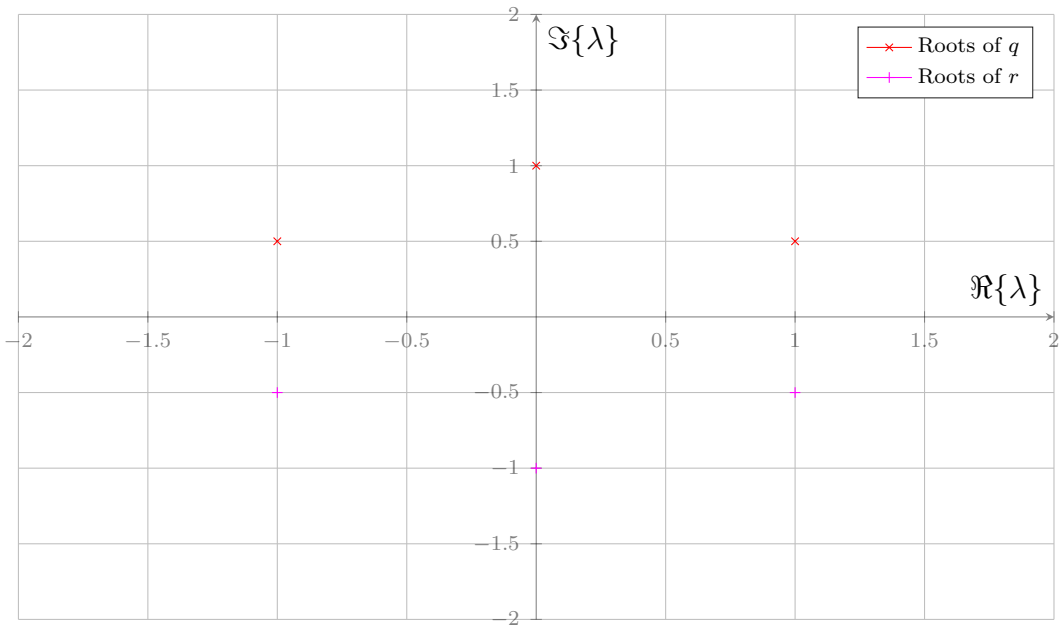


Figure 2.17: Pole-zero diagram of the function S_{21} of an all-pass network.

Theorem 2.9.2 (Darlington equivalent)**.** *Any passive (Schur) reflection S_{11} can be seen as the input reflection of a lossless two-ports network closed at port two by an arbitrary impedance.*

To illustrate the concept of Darlington equivalent, we represent in fig. 2.18 an antenna with input reflection coefficient S_{11} along with a Darlington equivalent of the latter. This equivalent consist of a lossless two-port device which shows the same input reflection S_{11}.

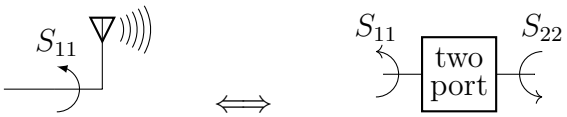


Figure 2.18: Darlington equivalent illustration

Remark 2.9.4. *Note that the Darlington equivalent is not unique. Particularly, it is possible to multiply polynomials p, r by a product of factors $(\lambda - \alpha)$ with $\alpha \in \mathbb{C}^-$ and the polynomial q by the factors $(\lambda - \overline{\alpha})$ without modifying the element S_{11}. This is done at the expenses of increasing the degree of the functions S_{22} and S_{12}, thereby increasing the McMillan degree of the matrix S as well.*

For the better understanding of this concept we can use the diagram provided in fig. 2.19. In this case an all-pass block of an arbitrary degree D has been added at the right port of the Darlington equivalent shown in fig. 2.18. This all-pass block has the effect of multiplying the reflection coefficient S_{22} by an uni-modular factor meanwhile the reflection S_{11} is not modified. We have

$$S_{22}^a = S_{22} \prod_{i=1}^{D} \frac{\lambda - \alpha_i}{\lambda - \overline{\alpha_i}}.$$

Thus a set of reflection zeros are introduced at the points α_i for all $i \in [1, D]$ along with the corresponding poles at the conjugate positions. Note that when looking from port 1, the additional poles and zeros cancel out, therefore keeping the degree of the function S_{11}. Nevertheless the McMillan degree of the Darlington equivalent is increased as the parameter S_{22} increases.

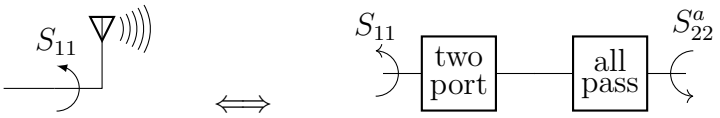


Figure 2.19: Darlington equivalent of not-minimal McMillan degree

Remark 2.9.5. *Note further that if polynomials p, r are multiplied by such factor $(\lambda - \alpha)$ with $\alpha \in \mathbb{C}^+$, a transmission zero is added at $\lambda = \alpha$, i.e. the polynomial $R = rr^\star$ vanish at $\overline{\alpha} \in \mathbb{C}^-$. Therefore, if the matrix S realising the function f as its element $(1,1)$ is desired to be reciprocal, we can distribute the roots of $qq^\star - pp^\star$ between r and $r^\star$, adding some extra roots in $\mathbb{C}^-$ if necessary, such that all roots of polynomial $R = rr^\star$ have even multiplicity.*

Example 2.9.1 (Darlington equivalent)**.** *Consider the example of an antenna (one-port device) where only the reflection coefficient at port one S_{11} can be measured. The antenna is a passive device, therefore there exist a lossless two-port with the same input reflection S_{11} characterised by a 2×2 scattering matrix S as illustrated in fig. 2.18. Note that the function S_{11} can be modelled by a rational function of finite arbitrary degree N*

$$S_{11}(\lambda) = \frac{p(\lambda)}{q(\lambda)} \qquad\qquad p, q \in \mathbb{P}^N,$$

with q a stable polynomial. The reflection coefficient S_{11} can then be written as the $(1,1)$ element of a matrix S

$$S = \begin{pmatrix} \dfrac{p}{q} & S_{12} \\ S_{21} & \overline{\epsilon}\dfrac{p^\star}{q} \end{pmatrix},$$

with functions S_{12} and S_{21} satisfying

$$S_{12}(\lambda)S_{21}(\lambda) = \frac{-\epsilon R(\lambda)}{q(\lambda)^2},$$

and $R = qq^\star - pp^\star$. Note the uni-modular constant ϵ can be arbitrarily chosen.

2.10 Rational form of the impedance matrix Z

The impedance parameters Z of any stable and lossless device can be calculated from the polynomials scattering matrix S as follows

$$Z = Z_0 \frac{I + S}{I - S}$$

Introducing now the Belevitch form the scattering matrix S from eq. (2.38) we have

$$I + S = \begin{pmatrix} q + \epsilon p^\star & -\epsilon r^\star \\ r & q + p \end{pmatrix} \frac{1}{q}, \tag{2.41}$$

$$(I - S)^{-1} = \begin{pmatrix} q - p & -\epsilon r^\star \\ r & q - \epsilon p^\star \end{pmatrix} \frac{1}{q + \epsilon q^\star - p - \epsilon p^\star}. \tag{2.42}$$

From eqs. (2.41) and (2.42) we reach

$$Z = \begin{pmatrix} q - \epsilon q^\star - (p - \epsilon p^\star) & -2\epsilon r^\star \\ 2r & q - \epsilon q^\star + p - \epsilon p^\star \end{pmatrix} \frac{1}{q + \epsilon q^\star - p - \epsilon p^\star}. \tag{2.43}$$

2.11 Coupling matrix derivation from the Belevitch form

We have obtained in the above section a rational form for the 2×2 impedance matrix $Z(\omega)$ which is computed from the polynomials p, q, r taking part in the Belevitch model of the Scattering matrix $S(\omega)$. In order to obtain the corresponding coupling matrix, the matrix $Z(\omega)$ should be expressed by a State Space realisation. We denote by ζ_i the poles of the matrix $Z(\omega)$ with $1 \leq i \leq N$ and where N is the McMillan degree of the system with scattering parameters $S(\omega)$. Furthermore we define the 2×2 matrix θ_i as $\theta_i = \text{res}\,(Z(\omega), \zeta_i)$, namely the residue of the matrix $Z(\omega)$ at the pole ζ_i.

Now we express the matrix $Z(\omega)$ in the form

$$Z(\omega) = D + C(\omega I - U)^{-1} V.$$

Matrix D As reviewed in section 2.8 the matrix D is obtained directly as

$$D = \lim_{\omega \to \infty} Z(\omega).$$

Matrix U Moreover the eigenvalues of the matrix U corresponds to the poles ζ_i for all $i \in [1, N]$. It should be noted that we have at this moment an infinite amount of possibilities for this matrix U, nevertheless the simpler one is the diagonal matrix

$$U = \begin{pmatrix} \zeta_1 & & & \\ & \zeta_2 & & \\ & & \ddots & \\ & & & \zeta_N \end{pmatrix}.$$

Matrices C**,** V With the previous choice for the matrix U, the matrices C and V only depend now of the residues of $Z(\omega)$ at each point ζ_i, in particular we have

$$C \cdot V = \theta_i \qquad \forall i \in [1, N]. \tag{2.44}$$

Now we choose the matrices C and V, which again are not unique, satisfying eq. (2.44). Note that in the case of a reciprocal network, we can additionally impose $C = V^T$.

Remark 2.11.1. *In section 2.7.2 we have derived the expression of the rational form corresponding a any arbitrary state space (C, D, U, V) representation of a system. Note that any system represented by a state space realisation can be written in a rational form. Nevertheless, not every rational matrix admits a state space representation. Indeed there might not exist matrices C, V such that eq. (2.44) holds at every point $i \in [1, N]$. Nevertheless a rational matrix which is proper with residues of rank 1 admits a state space representation. It is the case of every matrix in the form given by eq. (2.43).*

2.11.1 Coupling matrix

Finally, we denote

$$C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix},$$
$$D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix},$$
$$V = \begin{pmatrix} V_1 & V_2 \end{pmatrix},$$

where C_1, C_2 are row vectors, V_1, V_2 are column vectors and $D_{11}, D_{12}, D_{21}, D_{22}$ the scalar elements of the matrix D. Then we construct the coupling matrix M as

$$M = \begin{pmatrix} D_{11} & C_1 & D_{12} \\ V_1 & -U & V_2 \\ D_{21} & C_2 & D_{22} \end{pmatrix}.$$

2.12 Rational Schur functions associated to a prescribed transmission polynomial

The Belevitch parametrisation is customary in classical filter design and allows us to have additional control on the location of the points of zero reflection or zero transmission by properly placing the roots of polynomials p and q respectively. Indeed, when the classical synthesis of a filtering function with a resistive load is performed, the transmission zeros are often set to given positions either inside the complex plane as a complex conjugate pair or on the frequency axis. This practice has the purpose of increasing the out-of-band selectivity (if transmission zeros are placed on the frequency axis) or to have some control on the group delay in the passband (by imposing complex conjugate pairs of zeros). In this work, we inherit the practice of imposing the location of the transmission zeros.

Therefore, it is also interesting to define the set of rational Schur functions $\mathbb{\Sigma}^N$ obtained as the $(2, 2)$ element of a 2×2 rational matrix following the Belevitch parametrisation where the transmission polynomial $R = rr^\star \in \mathbb{P}_+^{2N}$ is prescribed

Definition 2.12.1 (Rational Schur function)**.** *We define the concept of rational Schur functions of degree N associated to the transmission polynomial $R \in \mathbb{P}_+^{2N}$ as the set of input (or output) reflection coefficients provided by a 2-port system that is passive and stable having the polynomial R as transmission polynomial.*

$$\mathbb{\Sigma}_R^N = \left\{ f = \frac{p}{q} : \ p, q \in \mathbb{P}^N \mid qq^\star - pp^\star = R \right\}, \tag{2.45}$$

where polynomial p, q are not required to be co-prime.

Remark 2.12.1. *Remark that, even though the transmission polynomial R of a matrix S has all transmission zeros of S as roots, some transmission zero might not be obtained only from the scalar functions S_{11} or S_{22} since a pole-zero cancellation might occurs in either S_{11} or S_{22} (not both). Therefore, when defining the rational scalar functions associated to a transmission polynomial R ($\mathbb{\Sigma}_R^N$) we allow for some of the transmission zeros not to be present as roots of the positive polynomial $qq^\star - pp^\star$. If this happens, we can multiply both p and q by a factor $(\lambda - \overline{\alpha})$ with $\alpha \in \mathbb{C}^-$, introducing a transmission zero at $\lambda = \alpha$.*

Moreover, if the transmission polynomial R is fixed, then it is possible to parametrise the set $\mathbb{\Sigma}_R^N$ only by the numerator polynomial p. Thus, given $p \in \mathbb{P}^N$ such that $pp^\star(\lambda) + R(\lambda) > 0$ for all $\lambda \in \mathbb{R}$, expression $qq^\star = pp^\star + R$ determines the polynomial $q \in \mathbb{P}^N$. This is possible because $q(\lambda) \neq 0$ for all $\lambda \in \mathbb{C}^-$ (q has all roots in $\mathbb{C}^-$ and $q^\star$ in $\mathbb{C}^+$). However, note that polynomial q is only determined up to an uni-modular constant $\epsilon = e^{j\theta}$ (if $\hat{q} = \epsilon q$, then $qq^\star = \hat{q}\hat{q}^\star$). Nevertheless this constant ϵ can be absorbed by p allowing to normalise q in a unique way. In this work we assume that polynomial q is normalised at a reference frequency $\lambda = \lambda_0$ such that $\arg q(\lambda_0) = \phi \in [-\pi, \pi]$.

2.13 Classical synthesis of transfer functions with a resistive load

After introducing the parametrisation chosen in this work for the function $f \in \mathbb{\Sigma}_R^N$, and in order to show the benefits of such parametrisation, we use it in the classical statement of the filter synthesis problem where a resistive load is considered. The classical synthesis problem consisting in finding, given the frequency band $\mathbb{I}$, the reflection coefficient S_{11} with the minimum return loss level l

Problem 2.13.1 (Classical synthesis problem)**.**

$$l_{opt} = \min_{S_{11} \in \mathbb{\Sigma}_R^N} \max_{\omega \in \mathbb{I}} |S_{11}(\omega)|,$$

with a specified rejection γ on the stop band $\mathbb{J}$

$$|S_{11}(\omega)| \leq \gamma \qquad \forall \omega \in \mathbb{J}.$$

The Belevitch parametrisation allows for the uniform constraints on the modulus of S_{11} to be cast to uniform constraints on the filtering function $pp^\star/R$ with $R = rr^\star$

$$|S_{11}|^2 = \frac{pp^\star}{qq^\star} = \frac{pp^\star}{pp^\star + R} = \left(1 + \frac{R}{pp^\star}\right)^{-1}.$$

The optimal solution to this classical problem is proved to be the *quasi-elliptic* functions [15]. These functions are obtained by solving the following problem for a fixed R over the coefficients of p:

Problem 2.13.2 (Classical synthesis problem.)**.** *Find* $\min_{p \in \mathbb{P}^N}(L)$, *subject to:*

$$\begin{aligned} \frac{pp^\star}{R}(\omega) &\leq L & \omega &\in \mathbb{I}, \\ \frac{pp^\star}{R}(\omega) &\geq \Gamma & \omega &\in \mathbb{J}, \end{aligned} \tag{2.46}$$

where the value of L, obtained as

$$L = \max_{\omega \in \mathbb{I}} \frac{p(\omega)p^{\star}(\omega)}{R(\omega)},$$

represents the reflection level in the passband and (2.46) the rejection requirement on the stop-band. The constant Γ can be computed from the specified rejection constraint:

$$\Gamma = \frac{1}{\gamma^{-1} - 1},$$

where the value γ in the denominator was defined in problem 2.13.1.

Note this problem is a minimisation of a linear criterium under a set of linear constraints only, which can be solved by means of well known methods like linear programming.

2.13.1 Optimal multi-band transfer functions

The optimal solution to the preceding problem is well known in the filter synthesis community in the case in which the band $\mathbb{I}$ is composed of a single interval. Additionally the solution in this case is unique and the polynomial p that provides this solution is the Tchebyshev polynomial weighted with the polynomial R.

In the generic case in which the passband $\mathbb{I}$ is composed of the union of an arbitrary number of real compact intervals, the result of the previous problem is not so well known. However, we must emphasize that the solution of the problem with a single interval is guaranteed optimal since it is the solution to a Zolotarev problem which is characterized in terms of the number of oscillations present in the pass-band. In the same way the solution of the general problem considering a generic band $\mathbb{I}$ is solution to another Zolotarev problem and can be characterized in the same way.

Nevertheless it was not until the recent years when the general version of problem 2.13.2 was formally stated and the optimal solution provided. In [16] problem 2.13.2 is formulated in a generalised form where the optimisation set consists on both polynomials $(p, r) \in \mathbb{P}^{N_1} \times \mathbb{P}^{N_2}$ with the assumption $pp^{\star} = p^2$ and $rr^{\star} = r^2$ as

Problem 2.13.3 (Optimal multi-band filter synthesis)**.**

$$\begin{aligned} \textit{Find :} \quad & \min_{(p,r)}(L) && (p, r) \in \mathbb{P}^{N_1} \times \mathbb{P}^{N_2}, \\ \textit{Subject to :} \quad & \frac{p^2(\omega)}{r^2(\omega)} \leq L && \omega \in \mathbb{I}, \\ & \frac{p^2(\omega)}{r^2(\omega)} \geq \Gamma && \omega \in \mathbb{J}. \end{aligned}$$

This problem is proven to be quasi-convex as for a fixed value of L, the set of polynomials $(p, r) \in \mathbb{P}^{N_1} \times \mathbb{P}^{N_2}$ such that

$$p^2(\omega) \leq Lr^2(\omega),$$

is a convex set. The solution is then obtained by solving a sequence of convex problems

$$\begin{aligned} \text{Find :} \quad & h = \min_{(p,r)} p^2(\omega) - L^{(i)} r^2(\omega) && (p,r) \in \mathbb{P}^{N_1} \times \mathbb{P}^{N_2}, \\ \text{Subject to :} \quad & p^2(\omega) \geq \Gamma r^2(\omega) && \omega \in \mathbb{J}, \end{aligned}$$

where the value of $L^{(i)}$ is fixed and reduced in each iteration until no positive value of h is found.

2.13.2 Non-reciprocal multi-band transfer functions

It should be noted in the previous formulation of the synthesis problem introduced in [16], it is assumed both polynomial to be star-symmetric, namely $p^\star = p$ and $r^\star = r$, thereby obtaining $pp^\star = p^2$ and $rr^\star = r^2$. This assumption is not done without loss of generality as if it is not imposed a better solution can be obtained. For instance, the previous problem without the constrain $p^\star = p$ would in general provide a better criterium. Furthermore if the second constraint $r^\star = r$ is also relaxed, namely a general non-reciprocal response is sought, the best solution for the synthesis problem is attained.

Nevertheless, this formulation can also be applied to the case where both conditions $r^\star = r$ and $p^\star = p$ are relaxed and all properties provided for such problem still hold. The problem is then formulated in an equivalent form by considering, instead of the polynomials (p, r), the couple of positive polynomials $(P, R) \in \mathbb{P}_+^{2N_1} \times \mathbb{P}_+^{2N_2}$ with $P = pp^\star$ and $R = rr^\star$. We obtain the following problem

Problem 2.13.4 (Generalised multi-band filter synthesis)**.**

$$\begin{aligned} \textit{Find :} \quad & \min_{(P,R)} (L) && (P,R) \in \mathbb{P}_+^{2N_1} \times \mathbb{P}_+^{2N_2}, \\ \textit{Subject to :} \quad & P(\omega) \leq LR(\omega) && \forall \omega \in \mathbb{I}, \\ & P(\omega) \geq \Gamma R(\omega) && \forall \omega \in \mathbb{J}. \end{aligned}$$

As before, the optimal solution to problem 2.13.4 is computing by solving a following sequence $\mathcal{P}_i$ of convex problems

Problem 2.13.5 ($\mathcal{P}_i$)**.**

$$\begin{aligned} \textit{Find :} \quad & h = \min_{(P,R)} P(\omega) - L^{(i)} R(\omega) && (P,R) \in \mathbb{P}_+^{2N_1} \times \mathbb{P}_+^{2N_2}, \\ \textit{Subject to :} \quad & P(\omega) \geq \Gamma R(\omega) && \omega \in \mathbb{J}, \end{aligned}$$

where the value of $L^{(i)}$ is reduced on each iteration until no $h \geq 0$ can be found.

This problem is of special importance in this thesis since it constitutes a particular case of the optimization problem studied in later chapters. This formulation is obtained by considering the simplest case of the matching problem, namely the case where the load is a constant impedance. Therefore, the properties of the optimization problem presented in [16] apply directly to our problem in the mentioned case.

References

[7] R. J. Cameron, C. M. Kudsia, and R. R. Mansour, *Microwave Filters for Communication Systems*. Wiley, 2007.

[8] D. M. Pozar, *Microwave Engineering, 4th Edition*. John Wiley & Sons, 2012.

[9] L. V. Ahlfors, *Complex analysis*, 3rd ed. McGraw-Hill Education, 1966. [Online]. Available: https://books.google.fr/books?id=RfYK28TcZEwC

[10] J. B. Garnett, *Bounded Analytic Functions*, ser. Pure and Applied Mathematics. Elsevier Science, 1981. [Online]. Available: https://books.google.fr/books?id=DVLO9gJ66{_}YC

[11] T. Kailath, *Linear systems*. Prentice-Hall, 1980.

[12] H. Rosenbrock, "Transformation of linear constant system equations," *Proceedings of the Institution of Electrical Engineers*, 1967.

[13] F. Seyfert and S. Bila, "General synthesis techniques for coupled resonator networks," *IEEE Microwave Magazine*, 2007.

[14] V. Belevitch, *Classical network theory*, ser. Holden-Day series in information systems. Holden-Day, 1968. [Online]. Available: https://books.google.fr/books?id=q-JSAAAAMAAJ

[15] R. J. Cameron, R. Mansour, and C. M. Kudsia, *Microwave Filters for Communication Systems: Fundamentals, Design and Applications*. Wiley, 2007.

[16] V. Lunot, F. Seyfert, S. Bila, and A. Nasser, "Certified Computation of Optimal Multiband Filtering Functions," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 1, pp. 105–112, 2008. [Online]. Available: https://hal.inria.fr/hal-00663542

Chapter 3:

General matching problem and state of art

In the design of communication devices, the matching problem has as objective the minimisation of the reflected power when the load impedance does not match the conjugate of the internal impedance of the generator. For that, a matching network is introduced between the load and the generator. This matching network is used to convert the load impedance into a different one as close as possible to the conjugate of the generator impedance (fig. 3.1).

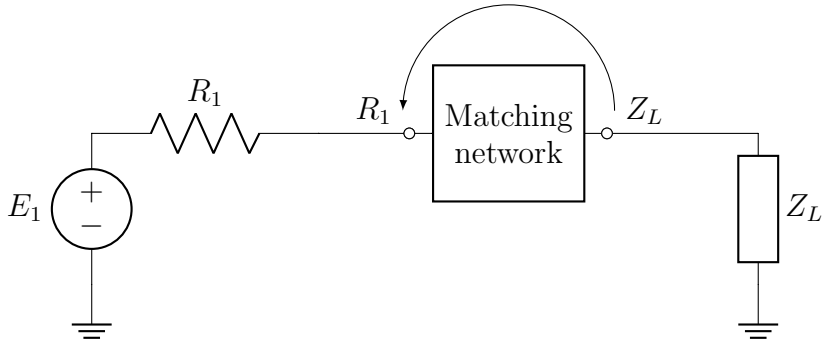


Figure 3.1: Matching circuit

In classical matching problems, the objective is to design a passive network showing the conjugate of the generator impedance at port one when the load impedance closes port two. If the matching network can show precisely the impedance $\overline{R_1}$ at its input, we speak about a perfect match. However, when the load is frequency variant ($Z_L = Z_L(\lambda)$), employing classical matching techniques, ideal matching is usually only attained at one frequency point λ_1. For narrow-band applications, obtaining a perfect match at one frequency is typically acceptable since the reflected power is still low around the frequency of perfect matching. However, it might not be good enough for a broad frequency band or when the design specifications in term of maximum reflected power are stringent. The goal then becomes to obtain uniform matching within the interval of interest with the smallest possible reflected power within the whole band and, possibly, with no perfect match at any frequency.

3.1 Classical matching problem

The matching problem is stated in this work in a different but still equivalent form. To begin with, we normalise all impedances in the circuit using as reference an arbitrary impedance Z_0. We can assume $Z_0 = 50\Omega$ as it is customary in classical matching problems. This technique is standard in the field of circuits design to obtain the reflection coefficient seen when the load is connected to the reference impedance. In practice, it is done just by applying the Möbius transform that maps the class of positive real functions (PR) onto $\mathbb{\Sigma}$.

$$\begin{aligned} f :& PR \longrightarrow \mathbb{\Sigma} \\ & Z_L \longrightarrow f(Z_L) = S_{11}(Z_L), \end{aligned}$$

where

$$f(Z_L) = \frac{Z_L - R_1}{Z_L + R_1}. \tag{3.1}$$

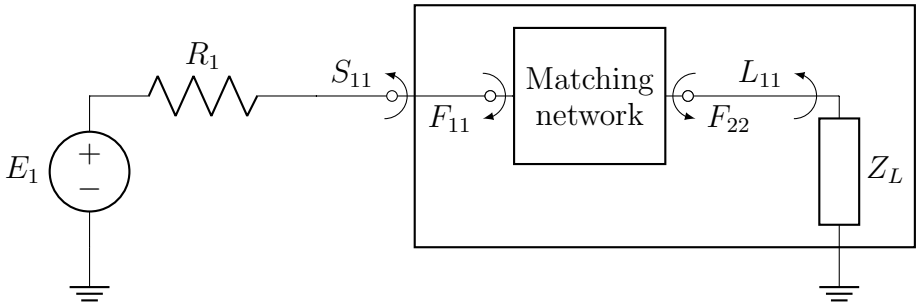


Figure 3.2: Matching circuit after normalising to R_1

This transformation provides us with an alternative representation of the circuit in fig. 3.1 where each component is parametrised by its scattering parameters referenced to the internal impedance of the generator. This representation is shown in fig. 3.2 where the input and output reflection parameters of the matching network are denoted by F_{11} and F_{22} respectively while the reflection coefficient of the load is denoted L_{11}. Additionally, we denote S_{11} the reflection coefficient at the input of the global system composed of the matching network connected to the load. With the new formalism, the objective becomes the minimisation of the reflection coefficient S_{11}, this time, in a frequency band instead of a single point.

This reflection coefficient S_{11} can be computed from the scattering matrix of the matching network F as the chaining of the 2×2 matrix F with the scalar reflection L_{11}. This operation will be referred repeatedly, and therefore it deserves some attention before moving on

Definition 3.1.1 (Scalar chaining)**.** *We denote by scalar chaining the operation (represented by the symbol $\circ$) of closing one of the ports of the two port device F by an impedance (possibly frequency-dependent) Z_L with reflection coefficient L_{11} (normalised to the reference impedance R_1). The result is a one-port device S with the input reflection (normalised to the same reference) S_{11}*

$$S_{11} = F \circ L_{11}.$$

The function S_{11} at each frequency, is computed as

$$S_{11} = F \circ L_{11} = F_{11} + \frac{F_{21}F_{12}L_{11}}{1 - F_{22}L_{11}}. \tag{3.2}$$

Similarly, we define the chaining of two-ports devices F and L as the interconnection of port 2 of F with port 1 of L providing the global scattering matrix S

Definition 3.1.2 (Chaining operation)**.** *The global scattering matrix resulting of chaining two matrices F and L (at each frequency) is computed as:*

$$S = F \circ L = \begin{pmatrix} F_{11} + \dfrac{F_{12}L_{11}F_{21}}{1 - F_{22}L_{11}} & \dfrac{F_{12}L_{12}}{1 - F_{22}L_{11}} \\ \dfrac{L_{21}F_{21}}{1 - F_{22}L_{11}} & L_{22} + \dfrac{L_{21}F_{22}L_{12}}{1 - F_{22}L_{11}} \end{pmatrix}. \tag{3.3}$$

In this chapter we are mostly interested in eq. (3.2), since the absolute value of the function S_{11} is the quantity to be minimised within a given interval if a good matching between the matching network and load is desired. With the purpose of studying such expression, we provide the concept of pseudo-hyperbolic metric which happens to be of great relevance.

Definition 3.1.3 (Pseudo-hyperbolic distance)**.** *The Pseudo-hyperbolic distance is a distance in $\mathbb{D}$. The pseudo-hyperbolic distance between two points $a, b \in \mathbb{D}$ is defined as*

$$\delta(a,b) = \left| \frac{a-b}{1-\overline{a}b} \right|,$$

where $\overline{a}$ denotes again the complex conjugate of the value a. As a distance, it verifies the properties

1. *Non-negativity: $\delta(a,b) \geq 0$; additionally $\delta(a,b) = 0$ if and only if $a = b$.*
2. *Symmetry: $\delta(a,b) = \delta(b,a)$.*
3. *Triangle inequality: $\delta(a,c) \leq \delta(a,b) + \delta(b,c)$*

One more property we should remark is the fact that the pseudo-hyperbolic disc $\mathrm{D}_H(c_0, r)$ of centre c_0 and radius r defined as [17, section 1]

$$\mathrm{D}_H(c_0, r) = \{\lambda : \ \delta(\lambda, c_0) < r\}$$

is also a disc $\mathrm{D}_E(C_0, R)$ in euclidean geometry, namely

$$\mathrm{D}_H(c_0, r) = \mathrm{D}_E(C_0, R) = \{\lambda : \ |\lambda - C_0| < R\}, \tag{3.4}$$

with a different centre C_0 and radius R where

$$C_0 = \frac{1-r^2}{1-r^2|c_0|^2} c_0,$$
$$R = \frac{1-|c_0|^2}{1-r^2|c_0|^2} r.$$

Note that for any point $c_0 \in \mathbb{D}$, the disc $\mathrm{D}_H(c_0, 1)$ is the unit disk

$$\mathrm{D}_H(c_0, 1) = \mathrm{D}_E(0, 1) = \mathbb{D} \qquad \forall c_0 \in \mathbb{D}.$$

Therefore there are not two points $a, b \in \mathbb{D}$ such that $\delta(a,b) \geq 1$. We obtain then

$$\delta(a,b) < 1 \qquad \forall a, b \in \mathbb{D}.$$

Before giving the formal statement of the matching problem, it is interesting to note that some properties of scattering matrices can be used to simplify eq. (3.2). First, at each frequency $\lambda \in \mathbb{C}$ we write

$$\begin{aligned} F \circ L_{11} &= F_{11} + \frac{F_{21}F_{12}L_{11}}{1 - F_{22}L_{11}} \\ &= \frac{F_{11} - F_{11}F_{22}L_{11} + F_{21}F_{12}L_{11}}{1 - F_{22}L_{11}} \\ &= \frac{F_{11} - \det(F)L_{11}}{1 - F_{22}L_{11}}. \end{aligned}$$

Assuming the matrix F is unitary we apply eq. (2.21)

$$F \circ L_{11} = \frac{F_{11} - \dfrac{F_{11}}{F_{22}^\star} L_{11}}{1 - F_{22} L_{11}} = \frac{F_{11}}{F_{22}^\star} \frac{F_{22}^\star - L_{11}}{1 - F_{22} L_{11}}. \tag{3.5}$$

Similarly the chaining operation $F_{22} \circ L$ can be written as

$$F_{22} \circ L = \frac{L_{22}}{L_{11}^\star} \frac{L_{11}^\star - F_{22}}{1 - L_{11} F_{22}}. \tag{3.6}$$

Finally we can use eq. (2.17) to compute $|F(\omega) \circ L_{11}(\omega)|$ in the real axis

$$|F(\omega) \circ L_{11}(\omega)| = \left| \frac{\overline{F_{22}(\omega)} - L_{11}(\omega)}{1 - F_{22}(\omega) L_{11}(\omega)} \right| \qquad \forall \omega \in \mathbb{R}. \tag{3.7}$$

Previous expression is the pseudo-hyperbolic distance between $F_{22}^\star$ and L_{11}.

$$|S_{11}(\omega)| = |F(\omega) \circ L_{11}(\omega)| = \delta\left(\overline{F_{22}(\omega)}, L_{11}(\omega)\right) \qquad \forall \omega \in \mathbb{R}.$$

Hence the modulus of the input reflection obtaining by chaining a two-port scattering matrix F with the scalar reflection coefficient L_{11} is expressed as the pseudo-hyperbolic distance between two scalar functions, namely the conjugate of the reflection coefficient of the matching network ($\overline{F_{11}}$) and the output reflection coefficient of the load (L_{22}).

3.2 General broadband matching problem

Now we are ready to formulate the matching problem as the minimisation of the maximum pseudo-hyperbolic distance $\delta\left(\overline{F_{11}}, L_{22}\right)$, or equivalently, the minimisation of the pseudo-hyperbolic distance $\delta\left(\overline{L_{11}}, F_{22}\right)$ within the set compact set $\mathbb{I} \subset \mathbb{R}$. Note that, as it is done for the classical synthesis of filtering functions in problem 2.13.1, the filter F to be synthesise is parametrised by the scalar reflection coefficient F_{22}. The scalar function F_{22} allows to recover afterwards a 2×2 scattering matrix having F_{22} as output reflection by means of the Darlington equivalent.

Problem 3.2.1 (General broadband matching problem)**.** *Let $\mathbb{I}$, denoted hereinafter as the passband, represents a finite union of compact frequency intervals and $L_{11} \in \mathbb{\Sigma}$ the reflection coefficient of the load. The problem of matching the load with reflection L_{11} within the passband $\mathbb{I}$ is stated as*

$$\textit{Find: } \min_{F_{22} \in \mathcal{E}} \max_{\omega \in \mathbb{I}} \delta\left(\overline{L_{11}(\omega)}, F_{22}(\omega)\right) \qquad \mathcal{E} \subset \mathbb{\Sigma}.$$

This is a problem of finding the best approximant of the function $\overline{L_{11}}$ in a limited band over a given subset $\mathcal{E}$ of the Schur functions. The best approximant is the minimiser of the error $\Psi(F_{22})$ representing the maximum value of the pseudo-hyperbolic distance.

$$\Psi(F_{22}) = \max_{\omega \in \mathbb{I}} \delta\left(\overline{L_{11}(\omega)}, F_{22}(\omega)\right).$$

It is important to note here that $\overline{L_{11}(\lambda)}$ is the restriction to the real axis of the function $L_{11}^{\star}(\lambda)$ which is not a Schur function, and most importantly it is an anti-analytic function. Therefore it can not be perfectly approached by $F_{22} \in \mathcal{E}$.

$$\min_{F_{22}\in\mathcal{E}} \Psi(F_{22}) = \Psi_{opt} > 0.$$

This fact differs already from what happens with a resistive load where perfect matching can be obtain in an arbitrary large interval.

Remark 3.2.1. *Remark that we have not specified yet what is the class $\mathcal{E} \in \mathbb{\Sigma}$ in which the function F_{22} is sought for. In the literature, previous authors have chosen the set $\mathcal{E}$ differently either to cope with realisability constraints of the synthesised matching network F or to guarantee the optimality of the obtained solution to problem 3.2.1. Indeed, the complexity of problem 3.2.1 strongly depends on the choice of the subset $\mathcal{E}$. Furthermore, the optimal solution to problem 3.2.1 also differs depending on the nature of the set $\mathcal{E}$.*

Next, instead of providing a chronological review of the different contributions to the problem, we believe its more important to follow a pragmatic order, allowing the reader to better understand the evolution of the matching problem in terms of the different subsets $\mathcal{E}$ of the Schur functions used to parametrise the function F_{22}.

In first place (although last in the chronology) we review the approach to problem 3.2.1 taken by Helton in the eighties where he considered the set $\mathcal{E}$ as the set of Schur function itself $\mathbb{\Sigma} \subset \mathcal{E}$. This choice allows him to state the problem in a convex form, ensuring the optimal solution is reached. As a result, Helton solution represents a remarkable contribution to the literature of problem 3.2.1 as he computes the best solution F_{22}^{opt} among all Schur functions $F_{22} \in \mathbb{\Sigma}$ providing thus a hard bound on the smallest attainable error $\Psi(F_{22})$

$$\min_{F_{22}\in\mathcal{E}} \Psi(F_{22}) = \Psi_{opt} \geq \Psi\left(F_{22}^{opt}\right).$$

3.3 Helton's solution to the matching problem

Let us consider the fundamental question of finding, for an arbitrary reflection coefficient $L_{11} \in \mathbb{\Sigma}$ what is the best matching level that can be obtained on an interval $\mathbb{I}$. In other words, the minimum value attainable in problem 3.2.1. This question has been already answered by Helton who solved the general matching problem in [18] as the minimisation of the pseudo-hyperbolic distance between the reflection F_{22} and $\overline{L_{11}}$

$$\Psi(F_{22}) = \max_{\omega\in\mathbb{I}} \delta\left(F_{22}(\omega), \overline{L_{11}(\omega)}\right).$$

without any additional constraints on F_{22}. It is only supposed that F_{22} belong to $\mathbb{\Sigma}$ namely an infinite dimensional class of functions. The functional space $\mathbb{\Sigma}$ is a convex space since any convex combination between two functions $f_1, f_2 \in \mathbb{\Sigma}$ is still a Schur function. For instance, we have f_1, f_2 such that $|f_1(\lambda)| < 1$ and $|f_2(\lambda)| < 1$ for all $\lambda \in \mathbb{C}^-$. Then the function

$$f_3(\lambda) = \kappa f_1(\lambda) + (1-\kappa) f_2(\lambda) \qquad 0 \leq \kappa \leq 1$$

satisfies $|f_3(\lambda)| < 1$ for all $\lambda \in \mathbb{C}^-$ and therefore $f_3 \in \mathbb{\Sigma}$. The convexity of $\mathbb{\Sigma}$ represents the main reason of the success of Helton approach.

3.3.1 The problem

In order to compute the optimal solution, the problem is stated as finding a function $F_{22} \in \mathbb{\Sigma}$ belonging to the Pseudo-hyperbolic disc with centre $\overline{L_{11}}$ and radius Ψ

$$\delta\left(F_{22}(\omega), \overline{L_{11}(\omega)}\right) \leq \Psi(\omega) \qquad \forall \omega \in \mathbb{I}, \tag{3.8}$$

with $\Psi(\omega)$ a fix tolerance K within the passband and 1 outside

$$\Psi(\omega) = \begin{cases} K & \omega \in \mathbb{I} \\ 1 & \omega \notin \mathbb{I} \end{cases}. \tag{3.9}$$

Note that if $F_{22}(\omega) \in \mathbb{D}$ for all $\omega \in \mathbb{R}$, since the Pseudo-hyperbolic distance is smaller than 1, then eq. (3.8) imposes a restriction only in the interval $\mathbb{I}$. Conversely, $\delta\left(F_{22}(\omega), \overline{L_{11}(\omega)}\right) > 1$ with $\omega \in \mathbb{R}$ implies that F_{22} is not a Schur function, even though it might be stable.

Thus, if there exist $F_{22} \in \mathbb{\Sigma}$ such that eq. (3.8) is verified, then matching tolerance $\Psi(\omega)$ is admissible. The idea is then to iterate on the tolerance $\Psi(\omega)$ by reducing the value of K until the pseudo-hyperbolic disk $D_H(\overline{L_{11}}, \Psi)$ contains only the single function $F_{22}^{opt} \in \mathbb{\Sigma}$.

3.3.1.1 The Hardy space H^∞

In this work, most theory is developed around the class of Schur functions $\mathbb{\Sigma}$, however a bigger class in functional analysis is the class of functions H^∞. H^∞ represents the space of analytic functions bounded in its analyticity domain, in our case, $\mathbb{C}^-$. Therefore we define

Definition 3.3.1 (Hardy space H^∞)**.**

$$H^\infty \equiv \left\{ f \in \mathcal{H}(\mathbb{C}^-) \mid \sup(|f(\lambda)|) < \infty,\ \lambda \in \mathbb{C}^- \right\},$$

where $\mathcal{H}$ denotes holomorphic functions.

Note H^∞ contains all analytic functions bounded by a finite value $\psi \leq \infty$. In particular, taking $\psi = 1$, we obtain the set of Schur function $\mathbb{\Sigma}$, which is also included in H^∞. In circuit theory, the class H^∞ is often used to denote reflection or transmission coefficients of stable devices when passivity is not a constraint. In other words, it can also represent the reflection or transmission coefficient of active microwave systems.

Nevertheless it is important to remark that the extension of the function $\overline{L_{11}(\omega)}$ to the complex plane is not analytic and hence it does not belong to H^∞. This function $\overline{L_{11}(\omega)}$ belong instead to the broader class L^∞. Therefore, we shall introduce now the concept of $p-norm$, which is useful for the definition of the Lebesgue spaces, among which we find the L^∞ space.

Definition 3.3.2 (p-norm)**.** *Consider again a function $f(\omega)$ with $\omega \in \mathbb{R}$ and an integer p with $1 \leq p < \infty$. The p norm of the function $f(\omega)$ is defined as*

$$\|f\|_p = \left(\int_\omega |f(\tau)|^p d\tau \right)^{\frac{1}{p}}. \tag{3.10}$$

Remark 3.3.1. *Note that we can also define the ∞-norm as the limit of eq.* (3.10) *when $p \to \infty$. We have*

$$\|f\|_\infty = \sup f.$$

We obtain the supremum of the function f. For this reason the ∞-norm is usually known as sup*-norm.*

Definition 3.3.3 (Lebesgue spaces L^p)**.** *Consider again a function $f(\omega)$ with $\omega \in \mathbb{R}$. The space L^p with $1 \leq p < \infty$ consists in all functions f with finite p-norm, namely*

$$\|f\|_p < \infty.$$

Remark 3.3.2. *We can also consider the special case of $p = \infty$. The space L^∞ contains the functions $f(\omega)$ which are bounded for $\omega \in \mathbb{R}$, namely*

$$\sup f < \infty.$$

3.3.2 The approach

Helton's approach is based on the fact that pseudo-hyperbolic disks are also disc in euclidean geometry as showed in eq. (3.4), therefore the problem can be solved by verifying whether the euclidean disk $D_E\left(C_0(\omega), R(\omega)\right)$ with

$$C_0(\omega) = \frac{1-\Psi(\omega)^2}{1-\Psi(\omega)^2|\overline{L_{11}(\omega)}|^2}\Psi(\omega), \qquad R(\omega) = \frac{1-|\overline{L_{11}(\omega)}|^2}{1-\Psi(\omega)^2|\overline{L_{11}(\omega)}|^2}\Psi(\omega),$$

is non-empty, or equivalently, the disk of unit radius $D_E\left(C_0(\omega)R_0(\omega)^{-1}, 1\right)$ after dividing by the minimum phase factor of the function $R(\omega)$, denoted here by $R_0(\omega)$.

Helton approach to problem 3.2.1 consist on relaxing the set $\mathbb{\Sigma}$, obtaining the minimisation problem over the functions $f \in H^\infty$ of the maximum of $|f(\omega) - C_0(\omega)R_0(\omega)^{-1}|$. This formulation of the matching problem by minimising the euclidean distance from a function $f(\omega) \in H^\infty$ to a function belonging to L^∞ is a classical Nehari problem

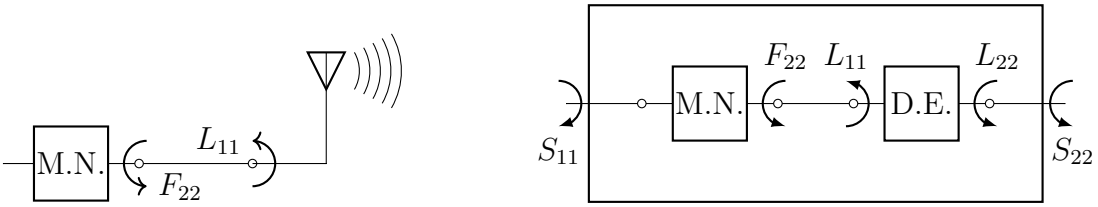
Problem 3.3.1 (Nehari)**.**

$$\textit{Find: } \min_{f\in H^\infty} \max_{\omega\in\mathbb{R}} |f(\omega) - C_0(\omega)R_0(\omega)^{-1}|.$$

Problem 3.3.1 is an approximation problem with a function $f \in H^\infty$ to a function $C_0R_0^{-1} \in L^\infty$ and the optimal solution is given by Nehari's theorem by means of Nehari's Hankel operator. Also, if the obtained minimiser f_{opt} satisfies

$$|f_{opt}(\omega) - C_0(\omega)R_0(\omega)^{-1}| \leq 1 \qquad \forall\omega \in \mathbb{R},$$

then the given tolerance $\Psi(\omega)$ is admissible. In such case the tolerance $\Psi(\omega)$ is reachable with a function $F_{22} \in \mathbb{\Sigma}$.



(a) Matching network + one-port load. (b) Global system (matching circuit + load Darlington equivalent).

Figure 3.3: Different matching approaches: classical matching and global system synthesis

3.3.3 The result

Helton's solution provides a hard lower bound

$$\Psi_{opt} = \min_{F_{22}\in\mathbb{\Sigma}} \Psi(F_{22}),$$

to the best matching level attainable with any arbitrary load of reflection coefficient L_{11}. However, in spite of being a remarkable result, Helton's work did not have a big impact in the electronic community. This is partly because of the complexity of his approach, based on Nehari's Hankel operator theory but also due to the fact that the F_{22} required to attain the best matching Ψ_{opt} is of infinite degree. In electronics a function F_{22} of infinite degree can be translated to either a matching network having an infinite number of lumped components or a coupled microwave structure having an infinite number of resonators. In either case, it is not suitable for practical applications.

3.4 Baratchart-Seyfert-Olivi: point-wise matching

We review next the work presented in [19] as this approach shares the same conceptual line as Helton's problem, this is minimising the pseudo-hyperbolic distance $\delta(F_{22}(\omega), \overline{L_{11}(w)})$ within the band of interest $\omega \in \mathbb{I}$. The formulation chosen by Baratchart et al. to establish the problem of matching differs from that of Helton in two fundamental aspects.

The first aspect is the choice of the set $\mathcal{E}$. In the work of Helton the set $\mathcal{E}$ was selected as the complete set of Schur functions, namely $\mathbb{\Sigma}$. This choice represents a huge relaxation of the problem but allowed to guarantee the optimality of the obtained solution. In the present case, the authors have decided to restrict the set $\mathcal{E}$ to the rational functions of degree N associated with a transmission polynomial $R \in \mathbb{P}_+^{2N}$ fixed in advance. This set corresponds to the set $\mathbb{\Sigma}_R^N$ introduced in eq. (2.45).

The choice $\mathcal{E} = \mathbb{\Sigma}_R^N$ provides a formulation fully equivalent to the problem of synthesis of classical transfer functions introduced in problem 2.13.1 where the parametrisation $F_{22} \in \mathbb{\Sigma}_R^N$ was used. With this parametrisation, the existence of a scattering matrix F of McMillan degree equal to N having the function F_{22} as element $(2,2)$ is guaranteed. The control obtained on both, the maximum McMillan degree and the transmission polynomial R facilitates the implementation of the physical 2-port device with scattering matrix F. Furthermore, the possibility to impose the polynomial R provides an in-

creased versatility as the technology and structure of the device can be decided in advance.

In comparison with the formulation of Helton, the set of rational functions of finite degree introduces a tremendous restriction with respect to the set of Schur functions without degree restriction. As a result of such restriction the convexity of the optimisation set is lost and so is the optimality of the obtained solution.

3.4.1 An algorithm for point-wise matching

The second aspect in which this formulation differs from that of Helton is the type of set of the real axis on which it is intended to minimize the reflection of the system, expressed as the pseudo-hyperbolic distance $\delta(F_{22}(\omega), \overline{L_{11}(w)})$. While in the work of Helton the minimization is done uniformly over the whole interval $\mathbb{I}$, here it has been decided to impose the points of perfect matching, namely a set $\mathbb{X}$ with a maximum of $N+1$ frequency points $[x_1, x_2, \cdots, x_{N+1}]$.

Note as it was pointed out in section 3.2, if the function $L_{11}(\lambda)$ is rational, we can not obtain $F_{22}(\lambda) = L_{11}^{\star}(\lambda)$ for all $\lambda \in \mathbb{I}$ as $\overline{L_{11}(\lambda)}$ corresponds to the evaluation on the real axis of $L_{11}^{\star}(\lambda)$, which is an anti-analytic function. However, we can still find a function $F_{22} \in \mathbb{\Sigma}_R^N$ such that $F_{22}(x_i) = \overline{L_{11}(x_i)}$ for a finite set of points $\mathbb{X}$. At each point x_i we have

$$\delta(F_{22}(x_i), \overline{L_{11}(x_i)}) = 0 \qquad \forall i \in [1, N+1].$$

These points λ_i can a priory be placed either on the real line or inside the analyticity domain $\mathbb{C}^-$, in which case the evaluation of the function $L_{11}^{\star}(\lambda_i)$ should be considered. However to obtain a solution that approach as close a possible the solution to problem 3.2.1 we consider in the present summary that the interval $\mathbb{I}$ is discretised distributing the matching points x_i within the passband $\mathbb{X} \subset \mathbb{I}$. Therefore

$$\delta(F_{22}(\omega), \overline{L_{11}(\omega)}) = 0 \qquad \forall \omega \in \mathbb{X}.$$

Having the transmission polynomial of F_{22} prescribed, what is easily done for the functions in $\mathbb{\Sigma}_R^N$, allows to parametrise the function $F_{22} \in \mathbb{\Sigma}_R^N$ in terms of the polynomial $p \in \mathbb{P}^N$ only. Additionally from eq. (3.7) the condition of perfect matching at the points x_i becomes an interpolation problem where the function F_{22} interpolates the value of $\overline{L_{11}}$ at each point $x_i \in \mathbb{X}$.

$$[F_{22}(p)](x_i) = \overline{L_{11}(x_i)} \qquad 1 \leq i \leq N+1.$$

3.4.2 The result: perfect matching points with a matching network of fixed degree.

The contribution provided in [19] represents the first step toward the solution of the general matching problem when a matching filter of McMillan degree N is considered. The main result is stated in the following theorem.

Theorem 3.4.1 (Baratchart-Seyfert-Olivi: pointwise matching)**.** *Let $L_{11} \in \mathbb{\Sigma}$ be the reflection of the load at port one and fix a transmission polynomial $R \in \mathbb{P}_+^{2N}$. Given*

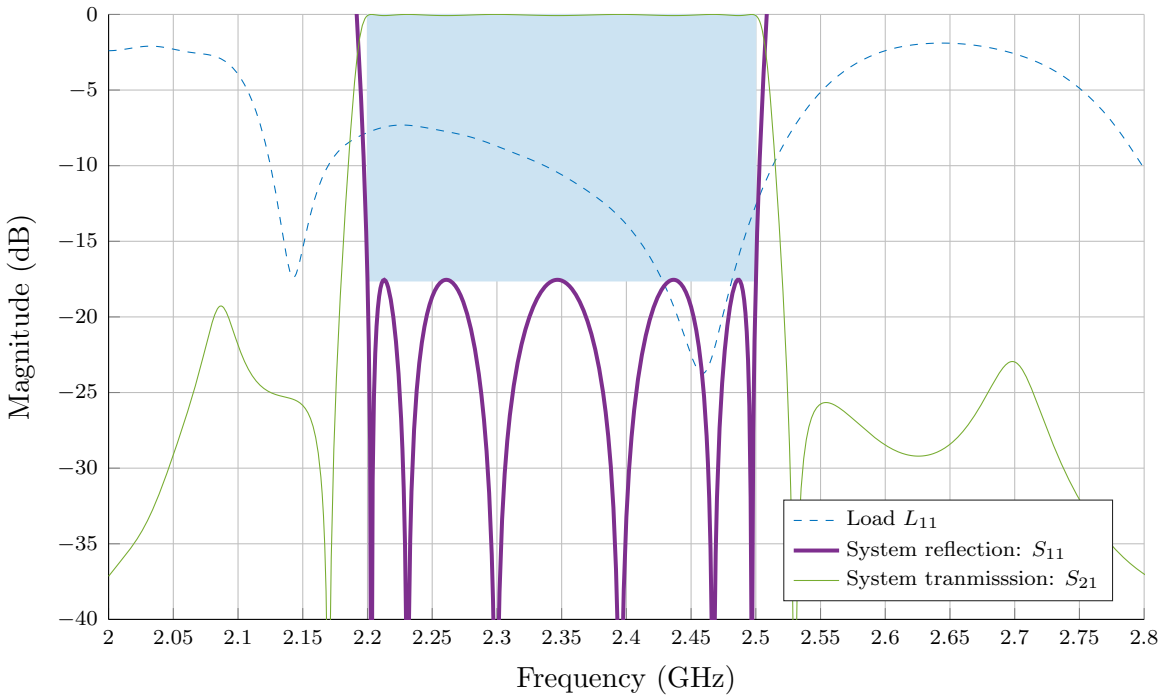


Figure 3.4: Load reflection and global response using the point-wise matching strategy and a matching network of degree 5. Shaded area indicates the interval $\mathbb{I}$.

$N+1$ distinct points on the real axis $x_1, x_2, \cdots, x_{N+1} \in \mathbb{R}$, there exist a unique function $F_{22} \in \mathbb{\Sigma}_R^N$ such that $F_{22}(x_i) = \overline{L_{11}(x_i)}$ for all $i \in [1, N+1]$.

Example 3.4.1. *An example of the kind of results obtained with this point-wise matching algorithm can be seen in fig. 3.4. In this example the passband $\mathbb{I}$ consists on a single frequency interval with range from 2.2 GHz to 2.5 GHz. Within this interval six perfect matching points have been distributed at approximately the following frequencies*

$$\mathbb{X} \approx \{\ 2.2,\quad 2.23,\quad 2.3,\quad 2.4,\quad 2.47,\quad 2.5\ \}\ GHz.$$

Remark 3.4.1. *Note as it is stated by theorem 3.4.1, the result appearing in fig. 3.4 and featuring six matching points inside the passband can be attained with a matching filter of degree $N = 5$. Furthermore two transmission zeros are imposed at the frequencies of 2.17 GHz and 2.53 GHz. We obtain the polynomial $R \in \mathbb{P}_+^{2N}$ as*

$$R = (\lambda - 2.53)^2(\lambda - 2.17)^2.$$

Therefore we have $F_{22} \in \mathbb{\Sigma}_R^N$.

Remark 3.4.2. *Note that nothing has been said yet with respect to the optimal placement of the points $x_i \in \mathbb{X}$. Indeed the set of points $\mathbb{X}$ can not be chosen arbitrarily in order to obtain an equioscillating response such that the one shown in fig. 3.4. The reason comes from the fact that the position of such points influences the reflection level ψ obtained as*

$$\psi = \max_{\omega \in \mathbb{I}} |S_{11}(\omega)|.$$

In order to determine the points $x_i \in \mathbb{X}$, we can define different heuristic approaches which optimises the location of such points with the goal of minimising the criterium ψ. In our case we state the following optimisation problem

Problem 3.4.1 (Placement of the perfect matching points)**.**

$$\textit{Find:} \qquad \psi_{opt} = \min_{\mathbb{X}\in\mathbb{R}^{N+1}} \max_{\omega\in\mathbb{I}} |S_{11}(\omega)|.$$

Remark 3.4.3. *Note further in fig. 3.4 that all maxima of the function $|S_{11}(\omega)|$ has the same value. Nevertheless it has not been proved yet that the optimal location of the points x_i satisfy this property for every possible load.*

3.5 Fano-Youla's contribution to the broadband matching problem

Problem 3.2.1 of minimising the pseudo-hyperbolic distance between the output reflection of the matching network and the conjugate of the load reflection coefficient is illustrated in fig. 3.3a. With this formulation the optimisation is done on the function F_{22} while only the input reflection L_{11} of the load is required. Nevertheless, since the load is a passive device, we can construct a Darlington equivalent (see section 2.9.3) which provides us with a 2-ports extension of the reflection coefficient L_{11}. If the load in fig. 3.3a is now replaced by its Darlington equivalent, the two-ports global system showed in fig. 3.3b is obtained. The Darlington equivalent in fig. 3.3b shows at port 1 the same reflection L_{11} as the load, therefore the global system reflection S_{11} given by eq. (3.5) is not modified. Additionally, by eq. (2.17), we have

$$|S_{22}(\omega)| = |S_{11}(\omega)| = \delta\left(\overline{F_{22}(\omega)}, L_{11}(\omega)\right) \qquad \forall\omega \in \mathbb{R}.$$

Moreover, the obtained reflection of the global system is also a Schur function.

Lemma 3.5.1. *Given the unitary scattering matrix F and the scalar function $L_{11} \in \mathbb{\Sigma}$, the function obtained as the chaining $S_{11} = F \circ L_{11}$ is a Schur function.*

Proof. Consider first the case when $|L_{11}(\omega)| < 1$ for all $\omega \in \mathbb{R}$. In this case the denominator of eq. (3.5) can not vanish inside the analyticity domain. Additionally, if a transmission zero occurs at a point $\alpha \in \mathbb{R}$ such that $F_{22}(\alpha)\overline{L_{11}(\alpha)} = 1$, by lemma A.3.1 a pole-zero simplification happens cancelling the singularity at $\lambda = \alpha$. Moreover the zeros of $L_{11}^{\star}$ in eq. (3.5) cancels with those of L_{22}. Thus $S_{11} = F \circ L_{11}$ is analytic in $\mathbb{C}^{-}$. Finally we check $|S_{11}(\omega)| < 1$ for all $\omega \in \mathbb{R}$. We have

$$|S_{11}(\omega)| = |F(\omega) \circ L_{11}(\omega)| = \delta\left(\overline{F_{22}(\omega)}, L_{11}(\omega)\right) < 1 \qquad \forall\omega \in \mathbb{R}.$$

□.

Therefore problem 3.2.1 can be stated over the scalar Schur function $S_{11} \in \mathbb{\Sigma}$ under the constraint that there exist a device F that is passive and stable such that

$$|F(\lambda) \circ L_{11}(\lambda)| = |S_{11}(\lambda)| \qquad \forall\lambda \in \mathbb{C}. \tag{3.11}$$

Remark 3.5.1. *Note that from eq.* (2.17) *we have* $|S_{11}(\lambda)| = |S_{22}(\lambda)|$ *for all* $\lambda \in \mathbb{R}$. *Therefore the problem can be stated equivalently over the function* S_{22}.

The main contribution of Fano-Youla's matching theory is the characterisation of the global system S such that eq. (3.11) holds for a passive stable matching network F.

3.5.1 Fano-Youla's global system approach

In [20] the matching problem is stated as the synthesis of the two-ports global system issue of the association of the matching network together with the Darlington equivalent of the load (fig. 3.3b). Using the Darlington equivalent, the input reflection of the load can be seen as the input reflection of a lossless 2-port device since by introducing the Darlington equivalent, the global system is considered to be lossless. Note that even if the reflection coefficient L_{11} is dissipative (not lossless), when the Darlington equivalent is computed, it is assumed that all the power that is not reflected is transmitted, thus obtaining a lossless device.

Fano's approach, was built around the characterisation of the global systems containing the Darlington equivalent of the load (fig. 3.3b). This characterisation is done at the transmission zeros α_i of the Darlington equivalent of the load. At these transmission zeros, looking from the right of the global system in fig. 3.3b it is not possible to see the matching network, because the load completely isolates port 1 and 2. Therefore the behaviour of the reflection coefficient S_{22} depends only on the load at these frequencies. In particular at $\lambda = \alpha_i$ we have $S_{22}(\alpha_i) = L_{22}(\alpha_i)$. Additionally, conditions on the derivatives of S_{22} are obtained depending on the nature and multiplicity of the transmission zero at α_i. Fano introduced a set of integral equations involving the function $\log\left|\frac{1}{S_{22}}\right|$ providing the conditions to be satisfied by the global reflection S_{22} at each transmission zero α_i. Some years later, in [21], Fano integral restrictions were reformulated as a complex interpolation problem, and sufficiency proofs were given. In Fano's theory, complex interpolation conditions are imposed at the transmission zeros of the load. It constitutes the necessary and sufficient conditions on the global system to contain the Darlington equivalent of the load. These interpolation conditions are the crux of this work, and therefore they receive special attention in section 3.5.2.

Next, we formulate Fano-Youla's characterisation of the global system. However, before addressing such formulation, we provide in appendix A some important basic concepts related to the chaining operation of two-ports devices.

3.5.2 Fano-Youla characterisation

The remarkable contribution of Fano-Youla to the solution of the matching problem is the characterisation of the set of functions S_{22} that can be obtained as the association of a passive matching network with the Darlington equivalent of the load, which is given. In other words, the conditions satisfied by the scattering parameters of the global system S that depends only on the load (necessary conditions). Furthermore, it was also proved that those conditions are indeed sufficient to guarantee the existence of a network F

satisfying eq. (3.11).

Let L be the 2×2 scattering matrix of the Darlington equivalent of the load and consider the matching network parametrised by the scalar reflection coefficient F_{22} which is denoted by f for the economy of notation. Similarly, let the global system be parametrised by the scalar function $S_{22} \in \mathbb{\Sigma}$ representing the reflection coefficient at port two as we can see in fig. 3.5.

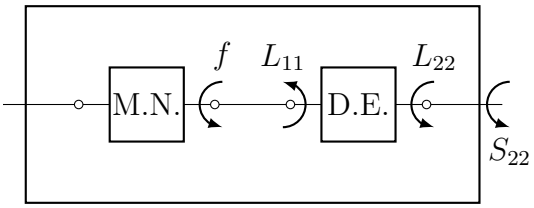


Figure 3.5: Global reflection S_{22} obtained by closing port one of the load with f

Remark 3.5.2. *For simplicity, we consider first the case of a load for which no transmission zero happens at the boundary of the analyticity domain ($\overline{\mathbb{R}}$). In consequence, the reflection coefficients L_{11}, L_{22} do not take uni-modular values on the frequency axis ($\mathbb{R}$).*

$$|L_{11}(\lambda)| < 1 \qquad |L_{22}(\lambda)| < 1 \qquad \forall \lambda \in \mathbb{R}.$$

The necessary conditions can be easily seen from eq. (3.2) of the chaining operation when the port 1 of L is closed by the reflection $f \in \mathbb{\Sigma}$

$$S_{22} = f \circ L = L_{22} + \frac{L_{12}L_{21}f}{1 - L_{11}f}. \tag{3.12}$$

Let $\alpha_1, \alpha_2, \cdots \alpha_M \in \mathbb{C}^-$ be the transmission zeros of L. Note we assume the points α_i have simple multiplicity. Then we have $L_{21}(\alpha_i)L_{12}(\alpha_i) = 0,\ \forall i = [1, M]$. Since all transmission zeros belong to $\mathbb{C}^-$ the reflection of the load is strictly less than 1 $|L_{11}(\alpha_i)| < 1$. Therefore the denominator $1 - L_{11}f$ does not vanish at the points α_i. Computing now the value of $S_{22}(\alpha_i)$

$$S_{22}(\alpha_i) = L_{22}(\alpha_i) + \frac{L_{12}(\alpha_i)L_{21}(\alpha_i)f(\alpha_i)}{1 - L_{11}(\alpha_i)f(\alpha_i)} = L_{22}(\alpha_i),$$

we obtain a set of interpolation conditions on the reflection coefficient L_{22} that does not depends on the matching network.

$$S_{22}(\alpha_i) = L_{22}(\alpha_i) \qquad \forall i \in [1, M].$$

These conditions were first written by Youla who proved that they are indeed necessary and sufficient to guarantee that a function $S_{22} \in \mathbb{\Sigma}$ can be expressed in the form of eq. (3.12). In this case we say the matrix L is de-chainable from the function S_{22}.

Definition 3.5.1 (De-chaining)**.** *A lossless two port, L is said to be de-chainable from a function, $S_{22} \in \mathbb{\Sigma}$ if and only if there exists $f \in \mathbb{\Sigma}$, such that $S_{22} = f \circ L$. Expression for f is obtained by inverting eq.* (3.12)*,*

$$f = \frac{L_{22} - S_{22}}{\det(L) - S_{22}L_{11}}. \tag{3.13}$$

Let us provide now Fano-Youla's theorem [21] assuming only simple transmission zeros $\alpha_i \in \mathbb{C}^-$

Theorem 3.5.1 (De-embedding conditions)**.** *Consider $S_{22} \in \mathbb{\Sigma}$ and let L be the 2×2 lossless scattering matrix of the load with M simple transmission zeros $\alpha_1, \alpha_2, \ldots, \alpha_M \in \mathbb{C}^-$. The matrix L is de-chainable from S_{22} if and only if at each transmission zero of α_i, the following condition is satisfied*

$$S_{22}(\alpha_i) = L_{22}(\alpha_i) \qquad\qquad 1 \leq k \leq M. \tag{3.14}$$

Instead of giving now Youla's original proof of theorem 3.5.1, we provide an alternative one based on functional analysis. Note the proof of necessity has already been done with the previous theory. Before continuing with the proof of sufficiency, I recall Rouche's theorem [22, corollary to theorem 18], on which the second part of the proof is based.

Theorem 3.5.2 (Rouche's theorem)**.** *Let f and g be two complex-valued functions and holomorphic inside a closed domain $\mathbb{X}$ with simple contour $\partial\mathbb{X}$. If $|f(\lambda)| < |g(\lambda)|$ for all $\lambda \in \partial\mathbb{X}$ then $h = f + g$ has the same number of zeros in $\mathbb{X}$ as f (counting multiplicity).*

Sufficiency proof of theorem 3.5.1. Now assume that eq. (3.14) is satisfied at each transmission zero of L. Then consider the function f obtained from eq. (3.13)

$$f = \frac{L_{22} - S_{22}}{\det(L) - S_{22}L_{11}}.$$

We need to show that f is a Schur function, namely $f(\lambda)$ is analytic in $\mathbb{C}^-$ and $|f(\omega)| \leq 1$ for all $\omega \in \mathbb{R}$. We prove first $|f(\omega)| \leq 1$. Using eq. (2.21) we have

$$\begin{aligned} f &= \frac{L_{22} - S_{22}}{\det(L) - S_{22}L_{11}} \\ &= \frac{L_{22} - S_{22}}{\frac{L_{11}}{L_{22}^\star} - S_{22}L_{11}} \\ &= \frac{L_{22}^\star}{L_{11}} \frac{L_{22} - S_{22}}{1 - S_{22}L_{22}^\star} \\ &= \frac{L_{22}^\star}{L_{11}} \delta\left(L_{22}, S_{22}\right). \end{aligned}$$

We have $\left|\frac{L_{22}^\star(\omega)}{L_{11}(\omega)}\right| = 1$ for all $\omega \in \mathbb{R}$. Additionally $\delta\left(L_{22}(\omega), S_{22}(\omega)\right)$ is the pseudo-hyperbolic distance between L_{22} and S_{22} which is bounded by one. Thus

$$|f(\omega)| = \left|\frac{L_{22}^\star(\omega)}{L_{11}(\omega)}\right| \delta\left(L_{22}(\omega), S_{22}(\omega)\right) \leq 1 \qquad\qquad \forall \omega \in \mathbb{R}.$$

Now we prove that f is analytic in $\mathbb{C}^-$. Consider again the expression

$$f(\lambda) = \frac{L_{22}^\star(\lambda)}{L_{11}(\lambda)} \frac{L_{22}(\lambda) - S_{22}(\lambda)}{1 - S_{22}(\lambda)L_{22}^\star(\lambda)}. \tag{3.15}$$

At each transmission zero α_i the numerator of eq. (3.15) vanishes since

$$L_{22}(\alpha_i) - S_{22}(\alpha_i) = 0 \qquad\qquad 1 \leq k \leq M.$$

From remark 2.9.1 we have

$$L_{22}^{\star}(\alpha_i)S_{22}(\alpha_i) = L_{22}^{\star}(\alpha_i)L_{22}(\alpha_i) = 1.$$

Therefore both numerator and denominator vanishes. As a result, M zeros in the denominator of eq. (3.15) gets cancelled out with the M zeros of the numerator. Finally, we apply Rouche's theorem to show that the denominator $(\det(L) - S_{22}L_{11})$ has no other zeros in $\overline{\mathbb{C}^-}$. From unitary property of L we have $|\det(L(\lambda))| = 1$ for all $\lambda \in \mathbb{R}$. Note that $S_{22} \in \mathbb{\Sigma}$, therefore $|S_{22}(\lambda)| \leq 1$ for all $\lambda \in \mathbb{R}$ and given that L has no transmission zero $\alpha_i \in \mathbb{R}$ we obtain for all $\lambda \in \mathbb{R}$, $|L_{22}(\lambda)| < 1$. At the boundary of the analyticity domain we obtain

$$|\det(L(\omega))| > |S_{22}(\omega)L_{22}(\omega)| \qquad \forall \omega \in \mathbb{R}.$$

This implies, using Rouche's Theorem, $\det(L) - S_{22}L_{22}$ has the same number of zeros in $\overline{\mathbb{C}^-}$ as $\det(L)$, which is M. Since we already showed that those M zeros cancel with M zeros in the numerator, we prove that f is analytic in $\overline{\mathbb{C}^-}$. Therefore, we have a function $f \in \mathbb{\Sigma}$, such that $S_{22} = f \circ L$ and hence proving the de-chainability of L from S_{22}. □

We have now proven Youla's interpolation conditions which are necessary and sufficient to ensure that the Darlington equivalent of the load can be de-embedded from the right of the global system. Nevertheless, we shall also provide before moving on the original de-embedded conditions stated by Fano some years earlier and proven to be necessary. These conditions are stated with an integral formulation and introduced for the first time the interpolation problem at the transmission zeros of the load.

3.6 Fano's integral formulation

Youla's interpolation conditions are stated in [20] in an integral form involving the function

$$-\log|S_{22}(\lambda)|.$$

Let us state now Fano's interpolation conditions. Consider the case of a load L whose Darlington equivalent is of McMillan degree M and has an arbitrary number of transmission zeros $\alpha_1, \alpha_2, \cdots, \alpha_m$ with $m \leq M$ and where μ_i denotes the multiplicity of the transmission zero at α_i. Also denote by z_b the zeros of the function S_{22} inside the analyticity domain $\mathbb{C}^-$.

Fano considered different cases where the transmission zeros α_i may be located on the frequency axis $\alpha_i \in \mathbb{R}$ (possibly at $\omega = 0$ or $\omega = \infty$, as a complex pair $\alpha_i, -\alpha_i$ with $\alpha_i \in j\mathbb{R}$ or a quadruplet $\alpha_i, -\alpha_i, \overline{\alpha_i}, -\overline{\alpha_i}$ with $\alpha_i \in \mathbb{C}^-$. Each of those cases were treated separately and necessary conditions were given in every case. We provide next the necessary condition in its integral form which are satisfied at a transmission zero $\alpha_i \in \mathbb{R}$ as it represents a clear example of the work done by Fano. Nevertheless, it is advisable to check Fano's complete work in [20] as it constitutes one of the most remarkable contribution to the synthesis of matching filters.

For each transmission zero $\alpha_i \in \mathbb{R}$ we have for all $k \in [0, 2\mu_i - 2]$

$$\frac{1}{\pi}\int_{\mathbb{R}} \frac{\log|S_{22}(\tau)|}{(\tau-\alpha_i)^{k+1}}dx + \frac{1}{k}\sum_b \left(\frac{1}{(\overline{z_b}-\alpha_i)^k} - \frac{1}{(z_b-\alpha_i)^k}\right) = K_{i,k},$$

while for $k = 2\mu_i - 1$ the following inequality is obtained

$$\frac{1}{\pi}\int_{\mathbb{R}} \frac{\log|S_{22}(\tau)|}{(\tau-\alpha_i)^{2\mu_i}}dx + \frac{1}{k}\sum_b \left(\frac{1}{(\overline{z_b}-\alpha_i)^{2\mu_i-1}} - \frac{1}{(z_b-\alpha_i)^{2\mu_i-1}}\right) \leq K_{i,2\mu_i-1},$$

where

$$K_{i,k} = \Im\left(\frac{\partial^k}{\partial\lambda^k}\log L_{22}(\lambda)\right)_{\lambda=\alpha_i}.$$

In the case where a transmission zero of the load occurs at ∞, the previous integrals are reformulated with the change of variable $\lambda \longrightarrow \lambda^{-1}$. If μ_∞ represents the multiplicity of the transmission zero at ∞, then we have

$$\begin{aligned}
&-\frac{1}{\pi}\int_{\mathbb{R}} x^{k-1}\log|S_{22}(x)|\,dx + \frac{1}{k}\sum_b \left(\overline{z_b}^k - {z_b}^k\right) = K_{\infty,k} \qquad \forall k \in [1, 2\mu_\infty - 2],\\
&-\frac{1}{\pi}\int_{\mathbb{R}} x^{k-1}\log|S_{22}(x)|\,dx + \frac{1}{k}\sum_b \left(\overline{z_b}^k - {z_b}^k\right) \leq K_{\infty,k} \qquad k = 2\mu_\infty - 1,
\end{aligned}$$

where

$$K_{\infty,k} = \Im\left(\frac{\partial^k}{\partial\lambda^k}\log L_{22}\left(\frac{1}{\lambda}\right)\right)_{\lambda=0}.$$

3.6.1 Load of degree 1 with no finite transmission zeros.

These integral conditions were used to provide bounds on the smaller realisable $|S_{22}|$ in the case where the reflection of the load is expressed as a rational function of degree 1 without finite transmission zeros. For instance, the integral restriction for a load of degree 1 with no finite transmission zeros is

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\log\left|\frac{1}{S_{22}(\lambda)}\right|d\lambda \leq \Im\left(\frac{\partial}{\partial\lambda}\log L_{22}\left(\frac{1}{\lambda}\right)\right)_{\lambda=0}. \tag{3.16}$$

Additionally, in the case of a load of degree 1, the optimal function S_{22} was proven to be of minimum phase. Equation (3.16) can be seen as a limitation on the maximum area covered by the function $\log|S_{22}(\lambda)^{-1}|$.

In order to minimise $|S_{22}(\lambda)|$ over an interval $\mathbb{I} = [\omega_1, \omega_2]$ he considered a function S_{22}^{opt} whose modulus is constant within the interval $\mathbb{I}$ and zero outside

$$|S_{22}^{opt}(\lambda)| = \begin{cases} K_{opt} & \lambda \in \mathbb{I} \\ 0 & \lambda \notin \mathbb{I} \end{cases}. \tag{3.17}$$

Intuitively, eq. (3.17) can be seen as the function using all "the available area" inside the interval $[\omega_1, \omega_2]$. Denoting

$$h = \Im\left(\frac{\partial}{\partial\lambda}\log L_{22}\left(\frac{1}{\lambda}\right)\right)_{\lambda=0}. \tag{3.18}$$

We obtain from eq. (3.16) the constraint

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\log\left|\frac{1}{S_{22}^{opt}}\right|d\lambda = h,$$

therefore we have

$$\frac{1}{\pi}\int_{\omega_1}^{\omega_2}\log\left|\frac{1}{S_{22}^{opt}}\right|d\lambda = h,$$

and after solving the integral

$$\frac{1}{\pi}(\omega_2-\omega_1)\log\left|\frac{1}{K_{opt}}\right| = h.$$

The lower bound on the minimum matching tolerance, K_{opt} takes the expression

$$K_{opt} \geq e^{-\frac{\pi h}{(\omega_2-\omega_1)}}. \tag{3.19}$$

Furthermore, apart from the previous bounds, Fano's integral conditions were used to yield a practical realisation of a global system S approaching those bounds. This was possible in the case where the system has no finite transmission zeros and the load is of degree 1. It was done by considering a global reflection S_{22} of Tchebyshev type. We can see an example of the Tchebyshev type response introduced by Fano in fig. 3.6. This function approximates the optimal tolerance indicated in eq. (3.17) with a reflection level that oscillates between two levels λ_1 and λ_2 within the passband and grows toward 1 out of the band.

With this kind of response it is straightforward to compute the values of λ_1 and λ_2 allowing for the de-embeddeding of the load. Additionally it can be shown that the minimum value of the reflection level λ_2 which still allows us to de-embed the load exists and is unique, indeed a formal proof for this statement is provided in appendix D. With this oscillating responses, good results in terms of matching are achieved in this simple case. However, note that this type of responses is known to be non-optimal concerning matching performances unless the load is a constant impedance.

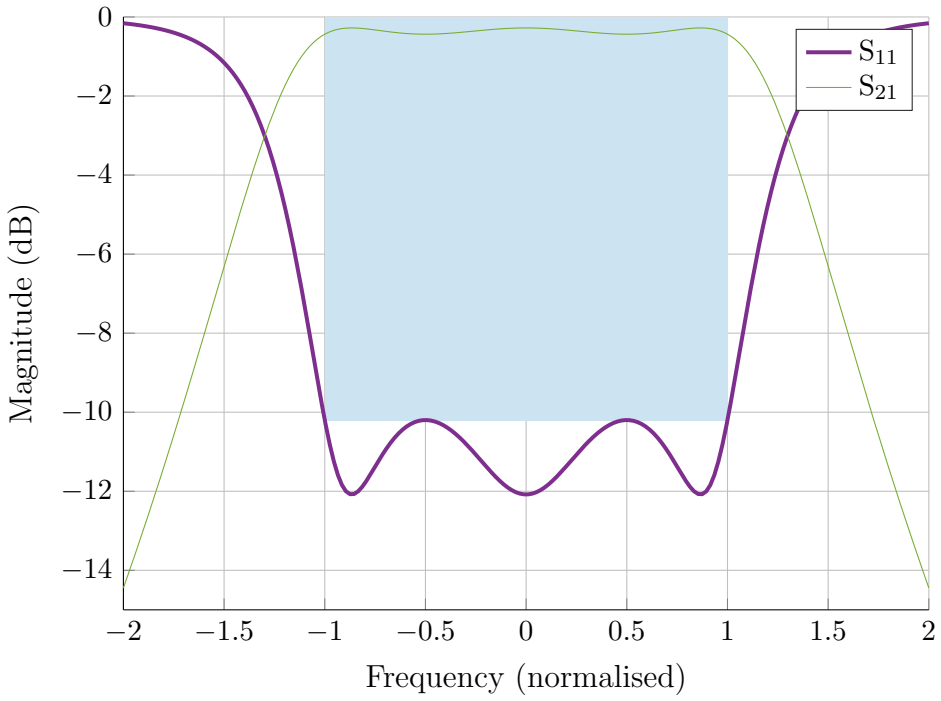


Figure 3.6: Equioscillating response.

3.7 Bode's result for an RC-load

We also review in this section the result first introduced by Bode in [23, chapter 16, section 3] as he realised trying to match a RC-load, that perfect matching $\Psi_{opt} = 0$ cannot be attained. This result can be easily derived now by particularising Fano-Youla's interpolation conditions to the case of a RC-load. In particular, for a RC-load (fig. 3.7) we have

$$Z_L = \left(j\lambda C + R^{-1}\right)^{-1}.$$

Applying eq. (3.1) we obtain the reflection coefficient L_{11}

$$L_{11}(\lambda) = \frac{Z_L - R_1}{Z_L + R_1} = \frac{(j\lambda C + R^{-1})^{-1} - R_1}{(j\lambda C + R^{-1})^{-1} + R_1} = \frac{1 - j\lambda CR_1 - R_1/R}{1 + j\lambda CR_1 + R_1/R}. \tag{3.20}$$

Bode's RC load is included in the class of load of degree 1 with a single transmission zero as $\lambda \to \infty$. It can be seen in eq. (3.20) that L_{11} is expressed as a rational function of degree 1 and $|L_{11}(\lambda)|$ tends to 1 as $\lambda \to \infty$. Therefore the integral constrain given by *eq.* (3.16) is a necessary condition. Computing now the derivative of $\log(L_{22}(\lambda^{-1}))$

$$\begin{aligned}\frac{d}{d\lambda}\log(L_{22}(\lambda^{-1})) &= \frac{d}{d\lambda}\log\left(\frac{\lambda(1 - R_1/R) - jCR_1}{\lambda(1 + R_1/R) + jCR_1}\right) \\ &= \frac{1 - R_1/R}{\lambda(1 - R_1/R) - jCR_1} - \frac{1 + R_1/R}{\lambda(1 + R_1/R) + jCR_1}.\end{aligned}$$

Considering now R_1 as the reference impedance and evaluating at $\lambda = 0$ we have

$$\frac{d}{d\lambda}\log(L_{22}(\lambda^{-1}))\bigg|_{\lambda=0} = \frac{2j}{R_1 C}.$$

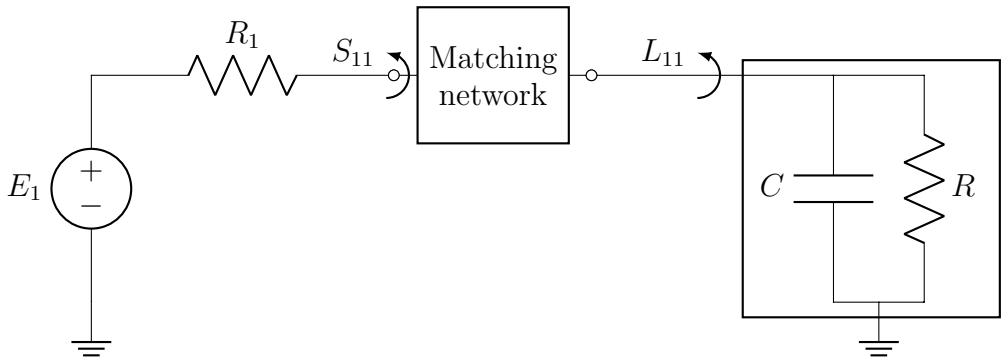


Figure 3.7: Matching circuit for an RC load

Defining now h as in eq. (3.18)

$$h = \Im\left(\frac{d}{d\lambda}\log(L_{22}(\lambda^{-1}))\right)_{\lambda=0} = \frac{2}{R_1 C},$$

and introducing it in eq. (3.19) we obtain

$$K_{opt} = e^{-\frac{2\pi}{(\omega_2-\omega_1)R_1 C}}.$$

Or equivalently, in a logarithmic scale

$$K_{opt}(\text{dB}) = -\frac{2\pi \log_{10} e}{(\omega_2-\omega_1)R_1 C}.$$

The previous result corresponds to the fundamental bound with respect to matching tolerances found by Bode in [23, chapter 16, section 3]. With this result, Bode was able to provide a hard lower bound on Ψ_{opt} for problem 3.2.1 with a load consisting on a shunt RC association. Additionally, as long as the load presents a transmission zero at infinity (i.e. the transmission vanishes when the frequency tends to infinity) the bound provided by Bode still applies since in this case the load can be seen as a generic network ended at port 2 in a shunt capacitor

$$\Psi(F_{22}) = \max_{\lambda\in\mathbb{I}} \delta\left(\overline{L_{11}}, F_{22}\right) \geq e^{-\frac{2\pi}{(\omega_2-\omega_1)RC}}.$$

3.8 Carlin's real frequency technique

Theorem 3.5.1 provides us with the restrictions on the function S_{22} when the load L is fixed. However, despite its undeniable elegance, this theory, like Helton's theory, did not lead to great practical applications in the field of electronics, mainly due to its complexity and the relative rigidity of its practical implementations induced. Finally, Youla's focus was, therefore, progressively forgotten and replaced by the optimization based on Carlin's real frequency technique introduced in [24].

Carlin motivated the use of this technique of real frequency by the fact that the theory previously developed by Fano and Youla needs and assumes a load specified from a rational model. On the contrary, the Carlin real frequency technique does not need the aforementioned rational model of the load. However, this method, as we will see below, also ends up performing a rational approximation.

3.8.1 The problem

Carlin's frequency technique is actually quite equivalent to the problem in pseudo-hyperbolic geometry formulated by Helton. However, instead of considering this problem in an analytical way as indicated in Helton's contribution, Carlin uses a more rudimentary procedure, without the need to use tools as advanced as Nehari's theory.

Carlin made the assumption that the optimal filter F is of minimal phase and then parametrise the output impedance of the matching filter Z_F by means of its real and imaginary parts. The problem consists therefore in maximizing the transmission of the global system S_{21} over the set of minimum phase impedances Z_F and within a given passband $\mathbb{I}$. Additionally, to ensure stability of the matching filter, the imaginary part of the impedance Z_F is computed from its real part thanks to the minimum phase property.

3.8.2 The approach

The simplicity of the approach here comes from the fact that this parametrisation of the matching filter is not done at every frequency within the interval $\mathbb{I}$ but at a finite set of control points $\omega_1, \omega_2, \cdots, \omega_n \in \mathbb{R}$ while an interpolation based on straight line segments is used to obtain the value of the matching filter between two control points.

Note, however, that the matching filter whose output impedance is a function defined by straight line segments represents a function of infinite degree. Indeed, if the number of control points used to parametrize the impedance of the filter tends to infinity, the response of the global system acquires the optimal form shown in eq. (3.9), probably with a higher reflection value than the optimum level provided by the Helton's method due to the minimum phase assumption.

Once the impedance Z_F defined by line segments has been obtained, the last step is to perform a rational approximation of the said impedance Z_F by means of a rational function of the desired degree for the matching filter.

3.8.3 The result

At this point we can argue whether it is better to perform this rational approximation, before solving the matching problem to obtain a rational model of the load or later on to approximate the optimum matching filter of infinite degree. In this context we can find two important arguments in favour of the rational approximation of the load at the beginning

1. In many cases, the load shows a response close to a rational function within a not too large band. For example, an antenna composed of a single resonance will provide a response similar to a rational function of degree 1. In the worst case the reflection of the load is a function of infinite degree while the matching filter obtained by the real frequency technique has an infinite degree.

2. If the rational approximation is made at the beginning of the process to obtain the said rational model of the load, a higher degree can be used to perform the approximation in case this approach is not good enough with a lower degree. On the other hand, if the approximation of the optimum matching filter is made once the optimization process has been completed, the degree used in the rational approach is determined by the desired degree for this filter, and it can not be increased in the case where the rational approximation process does not end well.

3.9 Concluding remarks

In this chapter we have made a quick review of the main contributions to the theory of matching problem. However, due to the long history of this problem since the first formulation made by Bode, it is possible to find many other contributions of interest. As an honourable mention we have to highlight, for example, Carlin's numerous publications on the problem of double matching introduced in [25]. In this problem the matching network, represented by a 2×2 scattering matrix , is connected between two devices, each of them presenting a complex impedance variable in frequency. An example of this problem is the use of a passive matching network to match the impedance shown by a generator (which is considered complex in this case) to the impedance of the load, variable in frequency.

References

[17] J. B. Garnett, *Bounded Analytic Functions*, ser. Pure and Applied Mathematics. Elsevier Science, 1981. [Online]. Available: https://books.google.fr/books?id=DVLO9gJ66{_}YC

[18] J. W. Helton, "Broadbanding:Gain equalization directly from data," *Circuits and Systems, IEEE Transactions on*, vol. 28, no. 12, pp. 1125–1137, dec 1981.

[19] L. Baratchart, M. Olivi, and F. Seyfert, "Boundary Nevanlinna-Pick interpolation with prescribed peak points. Application to impedance matching," *SIAM Journal on Mathematical Analysis*, 2017.

[20] R. M. Fano, *Theoretical Limitations on the Broadband Matching of Arbitrary Impedances*, ser. Technical report: Research Laboratory of Electronics. MIT Res. Lab. of Electronics, 1947.

[21] Youla and D. C. Youla, *A new theory of broadband matching.* Defense Technical Information Center, 1964.

[22] L. V. Ahlfors, *Complex analysis*, 3rd ed. McGraw-Hill Education, 1966. [Online]. Available: https://books.google.fr/books?id=RfYK28TcZEwC

[23] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, ser. Bell Telephone Laboratories series. Van Nostrand, 1945.

[24] H. J. Carlin and P. P. Civalleri, *Wideband circuit design*, ser. Electronic engineering systems series. Boca Raton, Fla. CRC Press, 1998.

[25] H. J. Carlin and B. S. Yarman, "The Double Matching Problem: Analytic and Real Frequency Solutions," *IEEE Transactions on Circuits and Systems*, 1983.

Part II

A convex approach to the problem of uniform broadband matching

Chapter 4:

General broadband matching problem

In this thesis, we pursue the goal of providing more accurate lower bounds than the fundamental limitations obtained for the set of functions H^∞. In particular, bounds that are sharper when matching networks of finite degree are considered, particularly networks with a rational scattering matrix of McMillan degree bounded by $N < \infty$. The problem of finding the best matching network of such type is a very hard one that remains unsolved. Indeed not much is know about that problem apart from the already introduced fundamental bounds.

This chapter concerns a new formulation of the matching problem where the degree of the function f is fixed rendering it suitable for practical applications. In practice, matching networks of low degree are preferred. On the one hand to meet the specifications in term of complexity (i.e maximum size of the devices) and cost, and on the other hand, due to the fact that matching networks of increased degree intrinsically involves (considering not ideal components), an increment of the power dissipation inside the system, reducing the overall performance of the device. In general, networks used in electronics are of degree one to five, and never more than 10.

Our route to the solution of such a problem consists on relaxing the previous set of rational functions of bounded McMillan degree allowing a higher degree, in particular $N + M$ with M finite as well. This relaxation leads to a convex formulation of the matching problem when a matching network of finite degree is considered. As a result, we are provided with hard bounds for the matching tolerance that is attainable with a matching network of degree N. Furthermore, in some cases where those bounds happen to be sharp, we provided the optimal matching network of degree N. Conversely, in the scenarios where the provided bounds are not sharp, we are still computing more accurate bounds as the ones obtained with the class functions H^∞.

4.1 Optimisation problem in Fano-Youla framework

The classical and well-known filter synthesis problem makes use of the Belevitch parametrisation to formulate a convex problem and thereby obtain a guaranteed solution. Nevertheless, for an arbitrary frequency-varying load, the problem is much less studied, and no optimal solution is known to exist.

Following Youla's and Fano's approach, the problem is stated similarly for the global system that is formed by the filter connected to the load, instead of focusing on the matching filter. With this formulation, the objective is to minimise, within the band of interest, the input reflection of the system composed of the matching network chained at port 2 with a two-ports loss-less reciprocal load. The schematic of the system appears in fig. 4.1 where we denote by F the scattering matrix of the matching network, and by L the scattering matrix of the load. Finally we denote $S = F \circ L$ the scattering matrix of the global system obtained as the cascade of the scattering matrix of the matching network with the 2×2 scattering matrix of the load as in fig. 4.1.

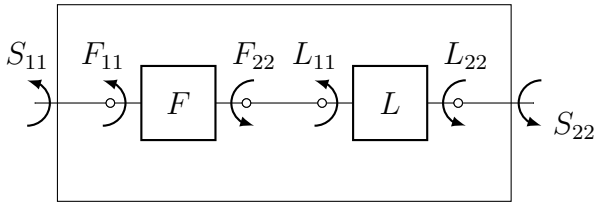


Figure 4.1: Schematic of global system obtained as the cascade of the matching filter with the load.

4.1.1 The load

In this chapter we consider a lossless 2-port load with 2×2 scattering matrix L in the Belevitch form (eq. (2.38)) having McMillan degree $M \in \mathbb{N}$. Note that if the load under study has only one port, the Darlington equivalent can be used to obtain a 2-port extension of the load. This is the case, for instance, of an antenna where only the input reflection coefficient L_{11} is known, namely

$$L_{11}(\lambda) = \frac{p_L(\lambda)}{q_L(\lambda)},$$

with $p_L, q_L \in \mathbb{P}^M$ and q_L a stable polynomial (roots only in $\mathbb{C}^+$). In this case we construct the 2-ports extension of L_{11} as

$$L(\lambda) = \frac{1}{q_L(\lambda)} \begin{pmatrix} p_L(\lambda) & -r_L^\star(\lambda) \\ r_L(\lambda) & p_L^\star(\lambda) \end{pmatrix}.$$

Note the Darlington equivalent is not unique as the roots of the transmission polynomial, denoted by $R_L \in \mathbb{P}_+^{2M}$, with

$$R_L = r_L^\star \cdot r_L,$$

can be arbitrarily distributed between r_L and $r_L^\star$. Nevertheless only the polynomial R_L is relevant in this work. The polynomial R_L, which is non negative on $\mathbb{R}$ can be factored in a unique manner $R_L = r_L r_L^*$ (up to uni-modular constant) if one imposes that r_L has only roots in $\mathbb{C}^-$. The multiplicity of each transmission zero in the such chosen polynomial r_L, we call the multiplicity of the transmission zero.

Additionally note that the uni-modular constant appearing in section 2.9.3 has already been normalised to 1. Unless specified otherwise (this will be the case in chapter 6), we consider here as in chapter 3 a matrix L where no transmission zero happens on the boundary of the analyticity domain, namely $\overline{\mathbb{R}}$. Furthermore we assume every transmission zero of L in $\mathbb{C}^-$ has simple multiplicity, so that the load has exactly M transmission zeros in $\mathbb{C}^-$. We denote then $\alpha_1, \alpha_2, \cdots \alpha_M \in \mathbb{C}^-$ the transmission zeros of the matrix L. Therefore we have

$$L_{12}(\alpha_i)L_{21}(\alpha_i) = 0 \qquad \forall i \in [1, M].$$

4.1.2 The matching network

In eq. (3.8), Helton formulated the matching problem on the reflection of the matching filter F_{22} belonging to $\mathbb{\Sigma}$, namely an infinity dimensional class of functions. In this chapter,

we take $F_{22} \in \mathcal{E}$ with $\mathcal{E}$ a subset of $\mathbb{\Sigma}$ containing only rational functions of bounded degree. In particular, we consider a lossless matching network with scattering matrix F parametrised in the Belevitch form having McMillan degree K

$$F(\lambda) = \frac{1}{q_F(\lambda)} \begin{pmatrix} p_F(\lambda) & -r_F^\star(\lambda) \\ r_F(\lambda) & p_F^\star(\lambda) \end{pmatrix},$$

with $p_F, q_F, r_F \in \mathbb{P}^K$ and q_F a stable polynomial that satisfies $q_F q_F^\star = p_F p_F^\star + r_F r_F^\star$. We denote further $R_F = r_F^\star r_F$ the transmission polynomial of the matching network. We have $R_F \in \mathbb{P}_+^{2K}$. Furthermore, as it is customary in classical filter synthesis we assume transmission zeros of the matching network to be fixed at prescribed positions in the complex plane (possibly at infinity). Hence the polynomial R_F is fixed. Note that in problem 2.13.2, convexity is obtained when the transmission polynomial of the network is prescribed. We denote $\nu_1, \nu_2, \cdots \nu_k \in \overline{\mathbb{C}^-}$ with $k \leq K$ the finite transmission zeros of F. Thus

$$F_{12}(\nu_i)F_{21}(\nu_i) = 0 \qquad \forall i \in [1, K].$$

4.1.3 Necessary conditions on the global system

Now, with the previous assumptions made on the load L and the matching filter F, we obtain a set of properties that are necessarily satisfied by the global system whose scattering matrix we call S.

1. The matrix S is obtained as the chaining of a rational matrices F of McMillan degree K and the scattering matrix of the load L of McMillan degree M, both of them in the Belevitch form. Hence the scattering matrix of the global system is also rational and has McMillan degree $N = K + L$.

2. The transmission zeros of both F and L are prescribed. Therefore the transmission zeros of the global system (the zeros in $\overline{\mathbb{C}^-}$ of $S_{12} \cdot S_{21}$) are also fixed to the given positions in the complex plane. Moreover, the transmission polynomial $R \in \mathbb{P}_+^{2N}$ of the global system is defined by

$$R = R_F \cdot R_L,$$

   with $R_F = \prod_{i=1}^{k} (\lambda - \nu_i)(\lambda - \overline{\nu_i})$ the transmission polynomial of the matching network and $R_L = \prod_{i=1}^{M} (\lambda - \alpha_i)(\lambda - \overline{\alpha_i})$ the transmission polynomial of the load. Thus R has roots at the transmission zeros α_i of the load as well as any other possible transmission zero ν_i fixed in advance.

3. The last necessary condition to ensure that the global system can be expressed in the form $F \circ L$ is given by Youla's interpolation conditions. Indeed, if the global system S is obtained by chaining a load L at port 2 of the matching network, at the transmission zeros of the load, the system verifies

$$S_{22}(\alpha_i) = L_{22}(\alpha_i) \qquad \forall i \in [1, M].$$

As it has already been discussed in chapter 3, note these conditions are actually necessary and sufficient to ensure that the system can be obtained as $S = F \circ L$ with F a rational matrix of McMillan degree K with transmission zeros ν_i as indicated before. Particularly Youla's interpolation conditions ensure there exists a scattering matrix F passive and stable such that $F \circ L = S$ meanwhile the bound on the McMillan degree of the system together with the fact that the transmission polynomial R of the system is fixed, guarantees the matrix F has McMillan degree $K = N - L$ and prescribed transmission zeros at the points ν_i.

Our goal is then to minimise the maximum value of the function $|S_{11}(\omega)|$ within the passband $\omega \in \mathbb{I}$. It should be noted that the global system, being lossless, it can also be written using the Belevitch form as

$$S(\lambda) = \frac{1}{q(\lambda)} \begin{pmatrix} p^\star(\lambda) & -r^\star(\lambda) \\ r(\lambda) & p(\lambda) \end{pmatrix},$$

with $p, q, r \in \mathbb{P}^N$. Additionally note the transmission polynomial $R = rr^\star$ is fixed. Thus q can be obtained as the stable polynomial satisfying $qq^\star = pp^\star + R$.

From the losslessness property we have

$$|S_{11}(\omega)|^2 = |S_{22}(\omega)|^2 = \frac{p(\omega)p^\star(\omega)}{q(\omega)q^\star(\omega)} \qquad \forall \omega \in \mathbb{R}.$$

We can now reformulate the necessary conditions on the 2-port global system S over the scalar function S_{22} such as

1. The function S_{22} is expressed as the ratio of the polynomials p, q of degree at most N.

2. The transmission polynomial R of the global system is fixed. Thus

$$qq^\star - pp^\star = R.$$

.

3. Youla's interpolation conditions bearing on $S_{22} = p/q$ are verified.

4.1.4 Class of feasible reflection coefficients

Next, let us define the class of functions S_{22} that are feasible for the given load L. In other words, once the load is fixed, the class of functions S_{22} satisfying the previous necessary conditions. We first define the class of functions $\mathbb{F}$ satisfying a set of M interpolation conditions.

Definition 4.1.1 (The load)**.** *Let consider a lossless two-port load with 2×2 scattering matrix L. We assume that the load presents only simple transmission zeros $\alpha_1, \alpha_2 \cdots \alpha_M \in \mathbb{C}^-$. Note that a one-port load can also be considered, in this case the 2×2 matrix L and the transmission zeros α_i, $1 \leq i \leq M$ refers to the Darlington equivalent of the load instead.*

Definition 4.1.2 (Feasible functions)**.** *We define the class of feasible functions $\mathbb{F}$ for the load L as the set of Schur functions satisfying Fano-Youla's interpolation conditions.*

$$\mathbb{F} = \{S_{22} \in \mathbb{\Sigma} : \ S_{22}(\alpha_i) = L_{22}(\alpha_i), \ \forall i \in [1, M]\}.$$

Remark 4.1.1. *For ease of notation, we are not explicitly writing the dependency of the set $\mathbb{F}$ with respect of the load L. Nevertheless each time that feasible functions $\mathbb{F}$ appears in this work, it always refers to a load with scattering matrix L and transmission zeros α_i with $1 \leq i \leq M$. Therefore, in order to make clear the kind of load we are referring to, the load is introduced at the beginning of the section or chapter where it is relevant.*

Rational Schur functions of degree bounded by N, namely of the class $\mathbb{\Sigma}^N$ are of high importance in this chapter. Particularly the functions satisfying the interpolation conditions $S_{22}(\alpha_i) = L_{22}(\alpha_i)$ for all $i \in [1, M]$ among the class $\mathbb{\Sigma}^N$, Therefore we define the following class

Definition 4.1.3 (Rational feasible functions)**.** *Let again L be a scattering matrix of a load of degree M. We denote the class of rational feasible Schur functions as*

$$\mathbb{F}^N = \mathbb{F} \cap \mathbb{\Sigma}^N = \left\{ S_{22} = \frac{p}{q} : \ p, q \in \mathbb{P}^N; \ S_{22} \in \mathbb{\Sigma}; \ S_{22} \in \mathbb{F} \right\},$$

where q is a stable polynomial, namely q has no roots in $\overline{\mathbb{C}^-}$.

Additionally, the reflection of the global system is a rational Schur function of degree at most N with the transmission polynomial R. This is the class of functions $\mathbb{\Sigma}_R^N$ defined in eq. (2.45). Let us now denote $\mathbb{F}_R^N$ the set of functions $S_{22} \in \mathbb{F} \cap \mathbb{\Sigma}_R^N$.

Definition 4.1.4 (Rational feasible functions with transmission polynomial R)**.** *Let L be as before the 2×2 rational scattering matrix of a load L of degree M with transmission polynomial R_L. Consider again the polynomial $R \in \mathbb{P}_+^{2N}$ defined as $R = R_F R_L$ where R_F represents the transmission polynomial of the matching network F. The set of rational functions of degree N with $N \geq M$ feasible for the load L is defined as*

$$\mathbb{F}_R^N = \left\{ S_{22} = \frac{p}{q} : \ p, q \in \mathbb{P}^N \mid qq^\star - pp^\star = R; \ S_{22} \in \mathbb{F} \right\},$$

where polynomial p, q are not required to be co-prime.

4.1.5 Statement of the problem

We state now our version of the matching problem as an optimisation problem while considering the class of rational functions satisfying Fano-Youla interpolations conditions, namely the class $\mathbb{F}_R^N$. This problem represents a particular version of problem 3.2.1 where the scattering matrix of the matching network is restricted to have a rational form with bounded McMillan degree.

Problem 4.1.1 (Matching problem with bounded degree)**.**

$$\begin{aligned} &\textit{Find:} && \psi_{opt} = \min_{S_{22} \in \mathbb{F}_R^N} \max_{\omega \in \mathbb{I}} |S_{22}(\omega)|^2, && \\ &\textit{Subject to:} && |S_{22}(\omega)| \geq \gamma && \forall \omega \in \mathbb{J}. \end{aligned}$$

Note that in problem 4.1.1 we use Fano-Youla's theory by imposing that the global reflection S_{22} belongs to the set of feasible functions $\mathbb{F}_R^N$. Next we state an optimisation problem on that framework by asking for the best function S_{22} in terms of matching level while satisfying Fano-Youla's conditions. This optimisation was never considered together with Fano-Youla's interpolation conditions, what is the main reason behind the rigidity of the original matching theory developed in the sixties.

This problem is equivalent to the classical filter synthesis problem (problem 2.13.1) with the function S_{22} belonging to $\mathbb{F}_R^N$ instead of $\mathbb{\Sigma}_R^N$. This fact introduces an additional constraint to ensure the de-chaining of the load. Nevertheless note that problem 2.13.1 can be seen as a matching problem with a resistive load and since resistive loads are of degree 0 (they have no transmission zeros) the set $\mathbb{F}_R^N$ coincides with $\mathbb{\Sigma}_R^N$ in that case.

Now we argue towards a formulation of problem 4.1.1 in terms of a polynomial $P = pp^\star \in \mathbb{P}_+^{2N}$ only, such that

$$|S_{22}(\omega)|^2 = \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R},$$

with R the transmission polynomial of the global system. Constraints on the modulus of S_{22} can easily be cast to constraints on the polynomial P as it is done in problem 2.13.2. However the Fano-Youla's interpolation conditions $S_{22}(\alpha_i) = L_{22}(\alpha_i)$ are stated on the function S_{22} and not on its modulus squared. This interpolation conditions over S_{22} renders problem 4.1.1 not convex, as the spectral factorisation of the positive polynomial $P + R$ is required to compute the function S_{22}.

Because of the non-convexity of problem 4.1.1, no guarantee can be provided about the optimality of the solution. To overcome such an issue, we introduce a relaxation of problem 4.1.1 where only the minimum phase factor of the function S_{22} is considered. First, decompose the function S_{22} as the product of a function $u(\lambda)$ of minimum phase and an uni-modular function $b(\lambda)$:

$$S_{22}(\lambda) = u(\lambda)b(\lambda)$$

This decomposition is trivial in the case of rational functions and it is known as *inner-outer* factorisation [26, Chapter 8, p.132]. The function $b(\lambda)$ is a Blaschke product with zeros at the points in the lower half plane where S_{22} vanishes.

$$b(\lambda) = \frac{\prod_{i=1}^{M}(\lambda - \xi_i)}{\prod_{i=1}^{M}(\lambda - \overline{\xi_i})} \qquad \xi_i \in \mathbb{C}^- \qquad M \leq N.$$

Note that $|b(\omega)| = 1$ for all $\omega \in \mathbb{R}$, therefore

$$|u(\omega)|^2 = |S_{22}(\omega)|^2 = \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R}.$$

Hence constraints on the function $|S_{22}(\omega)|$ with $\omega \in \mathbb{R}$ also hold on the function $|u(\omega)|$. Note that the function $b(\lambda)$ is analytic by definition as it has poles at the points $\overline{\xi}_i \in \mathbb{C}^+$.

Furthermore $b(\lambda)$ satisfies the following interpolation conditions at the transmission zeros of the load

$$b(\alpha_i) = \frac{L_{22}(\alpha_i)}{u(\alpha_i)} \qquad \forall \alpha_i \in [1, M]. \tag{4.1}$$

We obtain therefore the following two conditions

$$|u(\omega)|^2 = \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R}, \tag{4.2}$$

$$\exists f \in \mathbb{\Sigma} \mid f(\alpha_i) = \frac{L_{22}(\alpha_i)}{u(\alpha_i)} \qquad \forall i \in [1, M]. \tag{4.3}$$

Equations (4.2) and (4.3) are necessarily satisfied by the minimum phase factor of every function $S_{22} \in \mathbb{F}_R^N$. Note that in eq. (4.3) we are not seeking a Blaschke product but a function $f \in \mathbb{\Sigma}$. This constitutes the second part of the introduced relaxation as the function f belong to a bigger class of functions, namely the class of Schur functions. The proposed relaxation consist therefore in solving problem 4.1.1 over the set of minimum phase functions $u(\lambda)$ satisfying eqs. (4.2) and (4.3). The relaxation comes from the fact that the connection between the function $b(\lambda)$ and $u(\lambda)$ that ensures the product $u \cdot b$ remains of degree N is lost.

4.2 A convex relaxation to the matching problem

In this section, we present the relaxation of problem 4.1.1 which provides us with hard lower bounds Ψ on the reflection level attainable with a rational global system of finite degree N, or equivalently, with a matching network of degree $K = N - M$. Particularly we have

$$\Psi \leq \min_{S_{22} \in \mathbb{F}_R^N} \max_{\omega \in \mathbb{I}} \frac{P(\omega)}{R(\omega)}.$$

To begin with, we provide some concepts around Schur interpolation, particularly considering the interpolation conditions developed in eq. (4.1). These concepts are closely related to the theory provided in appendix B. Nevertheless, to avoid the necessity of having to go through the appendix B at this moment, we provide here the required result, extracted from theorem B.1.3 and directly applied it to the interpolation conditions in eq. (4.1) at the transmission zeros of the load $\alpha_1, \alpha_2, \cdots, \alpha_M$. We define then the set of Schur interpolant functions.

Definition 4.2.1 (Schur interpolant)**.** *Given the matrix L with transmission zeros α_i with $1 \leq i \leq M$. Let $u \in \mathbb{\Sigma}$ be a minimum phase function. Denote*

$$\mathbb{E}(u) \equiv \left\{ f \in \mathbb{\Sigma} \mid f(\alpha_i) = \frac{L_{22}(\alpha_i)}{u(\alpha_i)},\ \forall i \in [1, M] \right\}.$$

Note if there exists a function $f \in \mathbb{\Sigma}$ such that $f(\alpha_i) = \gamma_i$ with $1 \leq i \leq M$, then Nevanlinna's theory [27, Chapter IV, section 6] states that there exists as well a

Blaschke product of degree at most M that satisfies the same interpolation conditions (see appendix B). Hence if the set $\mathbb{E}(u)$ is not empty, it contains a Blaschke product of degree at most M. This motivates the relaxation made in eq. (4.3) where only the existence of a Schur function $f \in \mathbb{E}(u)$ is required. Particularly if $\mathbb{E}(u)$ is a singleton, then by theorem B.1.3 $\mathbb{E}(u)$ contains only a Blaschke product f of degree $m < M$. On the contrary, if $\mathbb{E}(u)$ is not a singleton, then it contains a strictly contractive function $f \in \mathbb{\Sigma}^M$ as it is stated next.

Theorem 4.2.1 (Existence of strictly contractive interpolant)**.** *Given the matrix L of McMillan degree M with transmission zeros α_i and let $u \in \mathbb{\Sigma}$ be a minimum phase function. Suppose $\mathbb{E}(u)$ is not a singleton. Then $\mathbb{E}(u)$ contains a function $f \in \mathbb{\Sigma}^M$ verifying eq. (4.5) such that $|f(\omega)| < 1$ for all $\omega \in \mathbb{R}$ and $\lim_{\omega\to\infty} |f(\omega)| < 1$.*

Proof. Consider $\mathbb{E}(u)$ which is not a singleton. Then the set of interpolant $\mathbb{E}(u)$ can be parametrised by eq. (B.7). Take $f(\lambda) = 0$ in eq. (B.7). Then by remark B.1.4 we have a function $f \in \mathbb{E}(u)$ such that

$$f(\lambda) = \frac{A(\lambda)}{C(\lambda)} \qquad\qquad A, C \in \mathbb{P}^M,$$

where $|f(\omega)| < 1$ for all $\omega \in \mathbb{R}$ and $\lim_{\omega\to\infty} |f(\omega)| < 1$. □

Lemma 4.2.1 suggests once the minimum phase function u is given, there exist a connection between the set of functions $f \in \mathbb{E}(u)$ and the functions $\rho \in \mathbb{F}$ such that $|\rho(\omega)| \leq |u(\omega)|$ for all $\omega \in \mathbb{R}$. This relation is expressed by the following lemma.

Lemma 4.2.1. *Consider the 2×2 scattering matrix L with simple transmission zeros $\alpha_1, \alpha_2 \cdots \alpha_M \in \mathbb{C}^-$ as introduced before. The following statements are equivalent:*

1. *The set $\mathbb{E}(u)$ is not empty.*
2. *There exist a function $\rho \in \mathbb{F}$ verifying $|\rho(\omega)| \leq |u(\omega)|$ for all $\omega \in \mathbb{R}$.*

We shall prove before moving on that both statements are equivalent.

Proof. Consider the function $u(\lambda)$ of minimum phase. Assume first that $\mathbb{E}(u)$ is not empty. Then there exists a rational function $f \in \mathbb{E}(u)$ which is of degree at most M. This function satisfies eq. (4.1) at each transmission zero α_i. Take now the function $\rho(\lambda) = u(\lambda) \cdot f(\lambda)$. This function belong to $\mathbb{F}^{N+N}$ as it has a degree bounded by the sum of the maximum degree of $u(\lambda)$ and $f(\lambda)$ and it verifies the interpolation conditions $\rho(\alpha_i) = u(\alpha_i)f(\alpha_i) = L_{22}(\alpha_i)$ for all $i \in [1, M]$. Additionally we have $|f(\omega)| \leq 1$ for all $\omega \in \mathbb{R}$, hence

$$|\rho(\omega)| \leq |u(\omega)| \qquad\qquad \forall\omega \in \mathbb{R}. \tag{4.4}$$

Conversely, suppose $u(\lambda)$ is a minimum phase function and eq. (4.4) is verified with $\rho \in \mathbb{F}$, i.e. $\rho(\lambda)$ verifies $\rho(\alpha_i) = \gamma_i$ for all $i \in [1, M]$. Given the minimum phase property of $u(\lambda)$, the function $f(\lambda)$ constructed as

$$f(\lambda) = \frac{\rho(\lambda)}{u(\lambda)}$$

is analytic on $\mathbb{C}^-$. Additionally from eq. (4.4) we have

$$|f(\omega)| \leq 1 \qquad \forall \omega \in \mathbb{R}$$

hence $f \in \mathbb{\Sigma}$. Finally notice f satisfies the interpolation conditions

$$f(\alpha_i) = \frac{L_{22}(\alpha_i)}{u(\alpha_i)},$$

for all $i \in [1, M]$ proving the equivalence. □

We formulate now a relaxed version of problem 4.1.1 allowing to express constraint on the function S_{22} in a convex form. We start by providing the definitions of the relaxed set of functions described before and a convex set of positive polynomials.

4.2.1 Admissible functions

Let us define a relaxation of the set $\mathbb{F}_R^N$ containing the functions $u(\lambda)$ of minimum phase satisfying eqs. (4.2) and (4.3). Note that eq. (4.2) translates into the fact that $u \in \mathbb{\Sigma}_R^N$.

Definition 4.2.2 (Admissible minimum phase functions)**.** *Given the load L with transmission zeros $\alpha_1, \alpha_2 \cdots \alpha_M \in \mathbb{C}^-$, we denote as admissible the set of functions $u \in \mathbb{\Sigma}_R^N$ of minimum phase such that there exists a Schur function $f(\lambda)$ satisfying*

$$f(\alpha_i) = \frac{L_{22}(\alpha_i)}{u(\alpha_i)} \qquad \forall i \in [1, M], \tag{4.5}$$

or equivalently, there exist a function $\rho \in \mathbb{F}$ verifying

$$|\rho(\omega)| \leq |u(\omega)| \qquad \forall \omega \in \mathbb{R}.$$

We parametrise the function $u(\lambda)$ in terms of the positive polynomial $P \in \mathbb{P}_+^{2N}$ as in eq. (4.2) since minimum phase property of the function $u(\lambda)$ allows for it to be recovered in a unique form from its modulus squared up to an uni-modular constant. Notice that the function $u(\lambda)$ does not vanish inside $\mathbb{C}^-$. Thus by imposing the normalisation $\Im(u(-j)) = 0$ the phase ambiguity when determining $u(\lambda)$ is eliminated. Let us now provide the formal definition of $u(\lambda)$.

Definition 4.2.3 (Minimum phase factor u_P)**.** *Given the polynomials $P, R \in \mathbb{P}_+^{2N}$, define $u_P(\lambda)$ as the minimum phase function satisfying*

$$\begin{aligned} |u_P(\omega)|^2 &= \frac{P(\omega)}{P(\omega) + R(\omega)} && \forall \omega \in \mathbb{R} \setminus \mathbb{X}, \\ \Im(u_P(\lambda)) &= 0 && \lambda = -j, \end{aligned}$$

where $\mathbb{X} \subset \mathbb{R}$ is a set containing at most $2N$ points on the real line where the polynomial R might vanish.

Remark 4.2.1. *The function $u_P(\omega)$ is well defined in $\mathbb{C}^-$ and on the real line apart from the points $\omega_0 \in \mathbb{X}$. Nevertheless note that if it happens that P and R vanish at a point $\omega_0 \in \mathbb{X}$ then both polynomials have a zero of even multiplicity at the point ω_0 due to the positivity property. Denote the multiplicity of the zero of P at ω_0 by $2n_0$. In this case we can divide P and R by the positive polynomial $(\omega - \omega_0)^{2n_0}$ removing the singularity of u_P at the point ω_0.*

Remark 4.2.2. *Remark that the dependence of $u_P(\omega)$ with respect to R is not indicated as the polynomial R is fixed containing the transmission zeros of the load together with the transmission zeros prescribed for the matching filter $R = R_F R_L$.*

4.3 Admissible polynomials

We now define the set of polynomials $P \in \mathbb{P}_+^{2N}$ such that the associated minimum phase function u_P is an admissible function.

Definition 4.3.1 (Admissible polynomials)**.** *Denote $\mathbb{A}_R^N$ the set of polynomials $P \in \mathbb{P}_+^{2N}$ such that the function $u_P(\lambda)$ is admissible.*

Lemma 4.3.1 (Bounds on the degree of the feasible function)**.** *Let $P \in \mathbb{A}_R^N$. Then there exists a function $\rho \in \mathbb{F}^{M+N}$, that is rational with degree at most $M+N$, satisfying*

$$|\rho(\omega)|^2 \leq \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R}.$$

Proof. Follows from lemma 4.2.1. If the set $\mathbb{E}(u_P)$ is not empty, it contains a rational function $f \in \mathbb{\Sigma}$ of degree at most M. Then take the function $\rho \in \mathbb{F}^{M+N}$ as $\rho = f \cdot u_P$. It should be noted that u_P is rational of degree N. Therefore the product $\rho = f \cdot u_P$ is of degree at most $M+N$. □

We state the following lemma

Lemma 4.3.2 (Concavity of $|u_P(\omega)|^2$)**.** *Denote by $\mathbb{X} \subset \mathbb{R}$ the set of at most $2N$ points where $R(\omega) = 0$ (assuming the polynomial R is not identically 0). The function $U_\omega : P \longrightarrow |u_P(\omega)|^2$ with $\omega \in \mathbb{R} \setminus \mathbb{X}$ and $P \in \mathbb{P}^{2N}$ is strictly concave with respect to the coefficients of polynomial P.*

$$|U_\omega(\kappa P_1 + (1-\kappa)P_2))|^2 > \kappa |U_\omega(P_1)|^2 + (1-\kappa)|U_\omega(P_2)|^2.$$

Proof. Consider the function $g_a : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ with

$$g_a(x) = \frac{x}{x+a} \qquad a > 0.$$

This function is concave with respect to x as its second derivative is negative.

$$\mathrm{D}_x^2 g_a(x) = \frac{-2a}{(x+a)^3}$$

Note that we have strict concavity $\mathrm{D}_x^2 g_a(x) < 0$ for all $x \geq 0$ assuming $a \neq 0$. Therefore the composition $U_\omega(P) = g_a(P(\omega))$ with $a = R(\omega)$ is concave with respect to P since the evaluation $P \to P(x)$ is an affine map. Additionally $P \to |u_P(\omega)|^2$ is strictly concave for all $\omega \in \mathbb{R} \setminus \mathbb{X}$. □

Lemma 4.3.2 is of vital importance as it contributes to prove many valuable theorems in this work. For instance, we use next this lemma to prove an important theorem which will provide us with the convexity of the set of admissible polynomials.

Theorem 4.3.1 (Convex combinations in $\mathbb{A}_R^N$)**.** *Let $P_1, P_2 \in \mathbb{A}_R^N$ be distinct polynomials and $\rho_1, \rho_2 \in \mathbb{F}$ the Schur functions satisfying*

$$|\rho_1(\omega)|^2 \leq \frac{P_1(\omega)}{P_1(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R} \setminus \mathbb{X}, \tag{4.6}$$

$$|\rho_2(\omega)|^2 \leq \frac{P_2(\omega)}{P_2(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R} \setminus \mathbb{X}. \tag{4.7}$$

Let $P_3 = \kappa P_1 + (1-\kappa)P_2$ with $0 \leq \kappa \leq 1$. The function $\rho_3 = \kappa\rho_1 + (1-\kappa)\rho_2$ is feasible, namely $\rho_3 \in \mathbb{F}$, and it satisfies

$$|\rho_3(\omega)|^2 \leq |u_{P_3}(\omega)|^2 \qquad \forall \omega \in \mathbb{R} \setminus \mathbb{X},$$

where equality may hold only on set $\mathbb{L} \subset \mathbb{R} \setminus \mathbb{X}$ of at most $2N$ points. $\mathbb{A}_R^N$ is therefore a convex set.

Before providing the proof, we need to remark the fact that the modulus square function is convex.

Lemma 4.3.3. *The function $f : \mathbb{C} \longrightarrow \mathbb{R}_+$ that associates to each complex value its modulus square $f(\lambda) = |\lambda|^2$ is convex.*

Proof. The proof is immediate if we remark that the modulus function $m : \mathbb{C} \longrightarrow \mathbb{R}_+$ with $m(\lambda) = |\lambda|$, as a norm in $\mathbb{C}$ satisfies

$$|\kappa\lambda_1 + (1-\kappa)\lambda_2| \leq \kappa|\lambda_1| + (1-\kappa)|\lambda_2| \qquad \forall \lambda_1, \lambda_2 \in \mathbb{C}$$

is a convex function. Additionally note the function $g : w \in \mathbb{R}_+ \longrightarrow w^2$ is convex and non decreasing, which proves the convexity of $\lambda \in C \longrightarrow g(m(\lambda))$. □

Lemma 4.3.3 can now be used to prove theorem 4.3.1.

Proof of theorem 4.3.1. Note P_1, P_2 are polynomials of degree at most $2N$, therefore $P_1(\omega) = P_2(\omega)$ with $P_1 \neq P_2$ can only holds in at most $2N$ points. Then we define

$$\mathbb{L} \equiv \{\omega \in \mathbb{R} \mid P_1(\omega) = P_2(\omega)\},$$

where the set $\mathbb{L}$ can only contain $2N$ points at most. Take a point $\omega_0 \in \mathbb{R} \setminus (\mathbb{X} \cup \mathbb{L})$. From the strict concavity of the function $U_{\omega_0} : P \longrightarrow |u_P(\omega_0)|^2$ with respect to P as stated in lemma 4.3.2 we have

$$|U_{\omega_0}(P_3)|^2 > \kappa|U_{\omega_0}(P_1)|^2 + (1-\kappa)|U_{\omega_0}(P_2)|^2. \tag{4.8}$$

Now take $\rho_3 = \kappa\rho_1 + (1-\kappa)\rho_2$. The functions ρ_1, ρ_2 are feasible, namely $\rho_1(\alpha_i) = L_{22}(\alpha_i)$ and $\rho_2(\alpha_i) = L_{22}(\alpha_i)$ for all $i \in [1, M]$. Then we have that ρ_3 also satisfies the interpolation conditions $\rho_3(\alpha_i) = L_{22}(\alpha_i)$. Hence $\rho_3 \in \mathbb{F}$. From the convexity of the modulus square we have

$$|\rho_3(\omega_0)|^2 \leq \kappa|\rho_1(\omega_0)|^2 + (1-\kappa)|\rho_2(\omega_0)|^2. \tag{4.9}$$

Given the statement of the theorem, namely from eqs. (4.6) and (4.7) we have

$$|\rho_1(\omega_0)|^2 \leq |u_{P_1}(\omega_0)|^2, \tag{4.10}$$
$$|\rho_2(\omega_0)|^2 \leq |u_{P_2}(\omega_0)|^2. \tag{4.11}$$

Introducing now eqs. (4.10) and (4.11) in eq. (4.9) we obtain

$$|\rho_3(\omega_0)|^2 \leq \kappa|U_{\omega_0}(P_1)|^2 + (1-\kappa)|U_{\omega_0}(P_2)|^2. \tag{4.12}$$

Finally combining eqs. (4.8) and (4.12) we reach the inequality

$$|\rho_3(\omega_0)|^2 <|u_{P_3}(\omega_0)|^2 \qquad . \tag{4.13}$$

Note that eq. (4.13) holds for all $\omega_0 \in \mathbb{R} \setminus (\mathbb{X} \cup \mathbb{L})$. Additionally note if $\omega_0 \in \mathbb{L}$, then we have equality in eq. (4.8), and therefore

$$|\rho_3(\omega)|^2 \leq|u_{P_3}(\omega)|^2 \qquad \forall \omega \in \mathbb{L}.$$

This concludes the proof of theorem 4.3.1 □

Corollary 4.3.1 (Convexity)**.** *The set of polynomials $\mathbb{A}_R^N$ is convex.*

Next we show that the set $\mathbb{A}_R^N$ is closed.

Theorem 4.3.2 (Closure of admissible set)**.** *$\mathbb{A}_R^N$ is closed.*

Before providing the proof of theorem 4.3.2, note that the set of polynomials $P \in \mathbb{P}^N$ identifies with $\mathbb{C}^N$, using the coefficients associated to a given basis as coordinates. Also note that all norms defined on the space $\mathbb{R}^N$ with N finite are equivalent. Therefore when speaking about convergence of a sequence of polynomials, we consider convergence in terms of any of the equivalent norms of $\mathbb{C}^N$. Furthermore, we provide the following lemma.

Lemma 4.3.4. *Consider a sequence $\{f_n\}_1^\infty \in \mathbb{F}^N$. There exists a sub-sequence $\left\{f_{\phi(n)}\right\}_1^\infty$ that converges uniformly on every compact $\mathbb{K} \subset \mathbb{C}^-$ and point-wise on $\mathbb{R} \setminus \mathbb{L}$ where $\mathbb{L} \in \mathbb{R}$ is a set of at most N points to a function $p/q \in \mathbb{F}^N$.*

Proof. Let $\{f_n\}_1^\infty \in \mathbb{F}^N$. The function f_n can be written as the ratio of two polynomials $p_n, q_n \in \mathbb{P}^N$

$$f_n = \frac{p_n}{q_n},$$

with q_n a stable polynomial. Now denote by $a_n^{(k)}$ with $k \in [0, N]$ the coefficients of the polynomial q_n with respect to the canonical basis $[\lambda^n, \cdots 1]$. The polynomial q_n takes the expression

$$q_n = \sum_{k=0}^{N} a_n^{(k)} \lambda^k.$$

Now as $q_n \neq 0$, we can divide numerator and denominator by the l_2 norm $\|q_n\|_2$ (any other norm on polynomials would do here)

$$\|q_n\|_2 = \left(\sum_{k=0}^{N} |a_n^{(k)}|^2\right)^{\frac{1}{2}},$$

and obtain

$$f_n = \frac{p}{\|q_n\|_2} \left(\frac{q_n}{\|q_n\|_2} \right)^{-1}.$$

Note the sequence of polynomials $\{A_n\}_1^\infty$ with

$$A_n = \frac{q_n}{\|q_n\|_2},$$

is bounded in norm as every element of $\{A_n\}_1^\infty$ has norm 1. Therefore there exists a convergent subsequence $\left\{A_{\phi(n)}\right\}_1^\infty$ which converges to a polynomial q. Considering again the l_2 norm we have

$$\lim_{n\to\infty} \|q - A_n\|_2 = 0.$$

By an argument using *Rouche's* theorem it is easy to show that the polynomial q is stable in the broad sense, that is has all its roots in $\mathbb{C}^-$. The sequence of functions $\left\{{A_n}^{-1}\right\}_1^\infty$ therefore converges uniformly on every compact $\mathbb{K} \in \mathbb{C}^-$ and on the real line $\mathbb{R}$ apart from the points where the polynomial q might have zeros. This is a set $\mathbb{L} \subset \mathbb{R}$ of at most N points.

$$\mathbb{L} = \{\omega \in \mathbb{R} \mid q(\omega) = 0\}.$$

Now note that the modulus of the functions f_n is bounded $|f_n(\lambda)| \leq 1$ for all $\lambda \in \overline{\mathbb{C}^-}$. Hence

$$|p_n(\lambda)| \leq |q_n(\lambda)| \qquad \forall \lambda \in \overline{\mathbb{C}^-}.$$

Dividing again by $\|q_n\|$

$$\frac{|p_n(\lambda)|}{\|q_n\|_2} \leq \frac{|q_n(\lambda)|}{\|q_n\|_2} \qquad \forall \lambda \in \overline{\mathbb{C}^-}. \tag{4.14}$$

Consider now the sup-norm on polynomials

$$\|p\|_\infty = \max_{w\in[-1,1]} |p(w)|.$$

The convergence of the sequence of polynomials $A_{\phi(n)}$ and eq. (4.14) is to the consequence that the sequence $B_{\phi(n)} = p_{\phi(n)}/\|q_{\phi(n)}\|_2$ is bounded for the sup-norm, and hence for any norm by the equivalence of norms in finite dimensions. Thus we can extract a convergent sub-sequence $\left\{B_{\varphi(\phi(n))}\right\}_1^\infty$ that converges to a polynomial p, and we have:

- $f_{\varphi(\phi(n))}$ converges uniformly to p/q on every compact of $\mathbb{C}^-$
- $f_{\varphi(\phi(n))}$ converges point-wise to p/q on $\mathbb{R} \setminus \mathbb{L}$
- eventually,

$$\forall \lambda \in \mathbb{R} \setminus \mathbb{L}, \ \left| \frac{p}{q}(\lambda) \right| \leq 1.$$

Pole-zero simplifications in the fraction p/q, at the real zeros of q, that by previous inequality are necessarily zeros of p (with same or higher multiplicity) yield the expected result. □

We do now the proof of theorem 4.3.2.

Proof. Consider now any convergent sequence $\{P_n\}_1^\infty$ in $\mathbb{A}_R^N$. Note that any convergent sequence $\{P_n\}_1^\infty \in \mathbb{P}_+^{2N}$ converges to a polynomial $P \in \mathbb{P}_+^{2N}$ since $\mathbb{P}_+^{2N}$ is closed. Since $\{P_n\}_1^\infty \in \mathbb{A}_R^N$, by lemma 4.3.1 there exists a sequence of rational function $\{\rho_n\}_1^\infty \in \mathbb{F}^{M+N}$ satisfying

$$|\rho_n(\omega)|^2 \leq \frac{P_n(\omega)}{P_n(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R} \setminus \mathbb{X}.$$

The degree of the Schur functions ρ_n is bounded by $M+N$, therefore by lemma 4.3.4 we can extract a subsequence $\left\{\rho_{\phi(n)}\right\}_1^\infty$ that converges point-wise to a function $\rho \in \mathbb{F}^{M+N}$ point-wise on $\mathbb{R}$ apart from a set $\mathbb{L}$ of at most $M+N$. By continuity (at the points of $\mathbb{L}$) we have,

$$|\rho(\omega)|^2 \leq \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R} \setminus \mathbb{X}.$$

Therefore $P \in \mathbb{A}_R^N$ and the set $\mathbb{A}_R^N$ is closed. □

Let us now provide a characterisation of the polynomials P belonging to the boundary of $\mathbb{A}_R^N$. First note that any polynomial $P \in \mathbb{P}_+^{2N}$ that vanishes at a point $\omega_0 \in \mathbb{R}$ belongs to the boundary of $\mathbb{A}_R^N$ since it belongs to the boundary of the set of positive polynomials.

Definition 4.3.2 (Boundary of positivity)**.** *Denote by $\partial_0\mathbb{A}_R^N$ the set of polynomials $P \in \mathbb{A}_R^N$ vanishing at some point ω_0 on the real line:*

$$\partial_0\mathbb{A}_R^N \equiv \left\{P \in \mathbb{A}_R^N \mid \exists \omega_0 \in \mathbb{R} : P(\omega_0) = 0\right\}.$$

Then take a polynomial $P \in \mathbb{P}_+^{2N}$ such that $P(\omega) \neq 0$ for all $\omega \in \mathbb{R}$. We next provide the following lemma

Lemma 4.3.5 (Interior of $\mathbb{A}_R^N$)**.** *The polynomial $P \in \mathbb{A}_R^N \setminus \partial_0\mathbb{A}_R^N$ belongs to the interior of the set $\mathbb{A}_R^N$, denoted here by $\mathring{\mathbb{A}}_R^N$ if and only if the set of interpolant functions $\mathbb{E}(u_P)$ is not a singleton.*

Proof of sufficiency. Consider a polynomial $P \in \mathbb{A}_R^N$. Suppose there exist at least two functions in $\mathbb{E}(u_P)$. Then by theorem 4.2.1 we have a function $b \in \mathbb{E}(u_P)$ such that $|b(\omega)| < 1$ for all $\omega \in \mathbb{R}$ and $\lim_{\omega\to\infty} |b(\omega)| < 1$. Hence the function $\rho \in \mathbb{F}$ constructed as $\rho = b \cdot u_P$ verifies

$$|\rho(\omega)|^2 < \epsilon < \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R}, \tag{4.15}$$

with $\epsilon > 0$. Then considering definition 4.2.2, there exist an open set $\mathbb{\Omega}$ around polynomial P such that eq. (4.15) still holds. Thus u_P with $P \in \mathbb{\Omega}$ is admissible. Therefore P is not on the boundary of the set $\mathbb{A}_R^N$. □

Proof of necessity. Let $P \in \mathbb{A}_R^N$ and assume set of interpolating functions $\mathbb{E}(u_P)$ is a singleton, then by theorem B.1.3 $\mathbb{E}(u_P)$ contains only a Blaschke product b of degree $m < M$. The Blaschke product B is the unique function in $\mathbb{\Sigma}$ such that

$$b(\alpha_i) = \frac{L_{22}(\alpha_i)}{u_P(\alpha_i)} \qquad \forall i \in [1, M].$$

In this case the function $\rho = b \cdot u_P$ is the unique function $\rho \in \mathbb{F}$ satisfying $|\rho(\omega)| \leq |u_P(\omega)|$ for all $\omega \in \mathbb{R}$. Additionally we have

$$|\rho(\omega)|^2 = \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R}.$$

Now take the sequence $\{P_n\}_1^\infty$ with $P_n = \left(1 + \frac{-1}{n}\right) P$ converging to the polynomial P. Suppose that $P \in \overset{\circ}{\mathbb{A}}_R^N$ then for n big enough $P_n \in \mathbb{A}_R^N$ and there exists $\rho_n \in F$ such that $\forall \omega \in \mathbb{R}\ |\rho_n(\omega)|^2 \leq P_n(\omega)/(P_n(\omega) + R(\omega))$. But

$$\forall \omega \in \mathbb{R},\ P_n(\omega)/(P_n(\omega) + R(\omega)) < P(\omega)/(P(\omega) + R(\omega))$$

which indicates $|\rho_n(\omega)|^2 < |\rho(\omega)|^2 = P(\omega)/(P(\omega) + R(\omega))$: a contradiction. □

Lemma 4.3.5 also provide us with the characterisation of the polynomials $P \in \mathbb{P}_+^{2N}$ such that $P(\omega) > 0$ for all $\omega \in \mathbb{R}$ belonging to the boundary of $\mathbb{A}_R^N$. We state it as a corollary of lemma 4.3.5.

Corollary 4.3.2 (Boundary of admissibility)**.** *Polynomial $P \in \mathbb{A}_R^N \setminus \partial_0 \mathbb{A}_R^N$ belong to the boundary of $\mathbb{A}_R^N$ if and only if $\mathbb{E}(u_P)$ is a singleton.*

Corollary 4.3.2 is defining a second type of boundary of $\mathbb{A}_R^N$, different from $\partial_0 \mathbb{A}_R^N$. The definition of this boundary is of great utility for the for the relaxed matching problem stated in next section as it contributes to the characterisation of the optimal solution. We provide therefore a formal definition.

Definition 4.3.3 (Boundary of admissibility)**.** *We denote by $\partial \mathbb{A}_R^N$ the set of polynomials P such that $\mathbb{E}(u_P)$ is a singleton.*

Note if we consider now the boundary of the set $\mathbb{A}_R^N$ we can distinguish between two boundaries of different nature namely the boundary of the set of positive polynomials $\partial_0 \mathbb{A}_R^N$ and the boundary of admissibility $\partial \mathbb{A}_R^N$. The set $\mathbb{A}_R^N$ is then obtained as the union

$$\mathbb{A}_R^N = \overset{\circ}{\mathbb{A}}_R^N \cup \partial_0 \mathbb{A}_R^N \cup \partial \mathbb{A}_R^N$$

4.4 Statement of the problem

We are now in position to formulate the relaxed version of problem 4.1.1 where the optimisation is performed over the minimum phase factor u_P of the function S_{22} under the restriction that u_P is admissible for the given load. Note that u_P is parametrised by the positive polynomial $P \in \mathbb{A}_R^N$, therefore the problem is stated in terms of the polynomial P only as it is usually done in the classic synthesis of filter functions (problem 2.13.2)

Problem 4.4.1 (Relaxed matching problem)**.**

$$\begin{aligned} &\textit{Find:} && \min_{P\in\mathbb{A}_R^N} \max_{\omega\in\mathbb{I}} \frac{P(\omega)}{R(\omega)}, \\ &\textit{Subject to:} && \frac{P(\omega)}{R(\omega)} \geq \Gamma && \omega \in \mathbb{J}. \end{aligned} \tag{4.16}$$

The belonging to the set $\mathbb{A}_R^N$ synthesis all the conditions a minimum phase factor $S_{22} \in \mathbb{F}_R^N$ verifies. As already mentioned the sets $\mathbb{I}$ and $\mathbb{J}$ are union of compact intervals of the real line. Whereas J can be empty, we will suppose here that $\mathbb{I}$ is not discrete, that is contains a continuum in order to avoid trivial situations. We review next the properties of problem 4.4.1.

4.4.1 Properties

Theorem 4.4.1 (Feasability of problem 4.4.1)**.** *There exists a polynomial $P \in \mathbb{A}_R^N$ satisfying eq.* (4.16)*.*

Proof. We are looking for a polynomial $P \in \mathbb{P}_+^{2N}$ satisfying eq. (4.16) such that there exist a function $\rho \in \mathbb{\Sigma}$ verifying $\rho(\alpha_i) = L_{22}(\alpha_i)$ for all $i \in [1, M]$ and

$$|\rho(\omega)|^2 \leq \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \overline{R},$$

where

$$L_{22}(\lambda) = \frac{p_L(\lambda)}{q_L(\lambda)} \qquad p_L, q_L \in \mathbb{P}^M,$$

and the transmission polynomial $R_L = q_L q_L^\star - p_L p_L^\star$ of the load is a divisor of R. Take $\rho = L_{22}$. We have

$$|\rho(\omega)|^2 = \frac{p_L(\omega)p_L^\star(\omega)}{q_L(\omega)q_L^\star(\omega)} \qquad \forall \omega \in \mathbb{R}.$$

Let us look now for a polynomial $P \in \mathbb{P}_+^{2N}$ such that

$$\frac{p_L(\omega)p_L^\star(\omega)}{q_L(\omega)q_L^\star(\omega)} \leq \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \overline{R}.$$

Solving for P we reach

$$\begin{aligned} P(\omega)\left(q_L(\omega)q_L^\star(\omega) - p_L(\omega)p_L^\star(\omega)\right) &\geq R(\omega)p_L(\omega)p_L^\star(\omega) && \forall \omega \in \overline{\mathbb{R}}, \\ P(\omega)R_L(\omega) &\geq R(\omega)p_L(\omega)p_L^\star(\omega) && \forall \omega \in \overline{\mathbb{R}}. \end{aligned}$$

Let denote $\hat{R} = \frac{R}{R_L}$ and $P_L = p_L p_L^\star$ where $\hat{R} \in \mathbb{P}_+^{2N-2M}$ and $P_L \in \mathbb{P}_+^{2M}$. Thus

$$P(\omega) \geq \hat{R}(\omega)P_L(\omega) \qquad \forall \omega \in \overline{\mathbb{R}}. \tag{4.17}$$

Note that $\hat{R} \cdot P_L$ belong to $\mathbb{P}_+^{2N}$. Therefore the polynomial $\hat{R} \cdot P_L$ belong to $\mathbb{A}_R^N$. Finally, to ensure that eq. (4.16) is satisfied we take the polynomial P as

$$P = \hat{R} \cdot P_L + \Gamma \cdot R$$

Note $P(\omega) \geq \Gamma \cdot R(\omega)$ for all $\omega \in \mathbb{R}$, therefore P verifies eq. (4.16). Additionally eq. (4.17) holds, hence $P \in \mathbb{A}_R^N$. □

Lemma 4.4.1 (Existence of P^{opt})**.** *There exists a polynomial $P^{opt} \in \mathbb{A}_R^N$ solution of problem 4.4.1.*

Proof. We proved in theorem 4.4.1 that the set of admissible P satisfying eq. (4.16) is not empty. Then consider the image of the set $\mathbb{A}_R^N$ by the application

$$\begin{aligned} \Psi : \mathbb{A}_R^N &\longrightarrow \mathbb{R}_+ \\ P &\longrightarrow \Psi(P) = \max_{\omega \in \mathbb{I}} \frac{P(\omega)}{R(\omega)}. \end{aligned}$$

Denote Ψ the image set of $\Psi(P)$ with $P \in \mathbb{A}_R^N$ and eq. (4.16) satisfied.

$$\Psi \equiv \left\{ \Psi(P) \mid P \in \mathbb{A}_R^N : P(\omega) \leq \Gamma R(\omega) \ \forall \omega \in \mathbb{J} \right\}.$$

The set Ψ is a set of real numbers which is bounded below by 0 and therefore it has an infimum Ψ_{opt} . Thus for all $\Psi \in \Psi$ we have $\Psi \geq \Psi_{opt}$ with $\Psi_{opt} \geq 0$. Now take a minimising sequence of numbers $\{\Psi_n\}_1^\infty \in \Psi$ converging to Ψ_{opt} such that

$$\lim_{n \to \infty} \Psi_n = \Psi_{opt}.$$

Take then a sequence of polynomials $\{P_n\}_1^\infty \in \mathbb{A}_R^N$ such that

$$\begin{aligned} \Psi(P_n) &= \Psi_n \\ P_n(\omega) &\leq \Gamma R(\omega) \qquad\qquad \forall \omega \in \mathbb{J}. \end{aligned}$$

As $\mathbb{I}$ contains a continuum, the sequence $\{P_n\}_1^\infty$ which is bounded on I is bounded for any norm on polynomials of $\mathbb{P}^{2N}$. From the sequence $\{P_n\}_1^\infty$ we can therefore extract a convergent subsequence $\left\{P_{\phi(n)}\right\}_1^\infty$ converging to a polynomial P_{opt}. By theorem 4.3.2 the set $\mathbb{A}_R^N$ is closed, which implies $P_{opt} \in \mathbb{A}_R^N$. The continuity of ψ, as a weighted sup norm, implies

$$\lim_{n \to \infty} \Psi(P_{\phi(n)}) = \Psi_{opt}.$$

Thus we have a polynomial P_{opt} which attain the optimum criterium Ψ_{opt}. This concludes the proof. □

Next we provide a necessary condition for the optimality of the solution to problem 4.4.1 using the definition of the set $\mathbb{E}(u_P)$

Theorem 4.4.2 (Optimality)**.** *If $P_{opt} \in \mathbb{A}_R^N$ is the solution to problem 4.4.1, and eq.* (4.16) *is not binding, then $P_{opt} \in \partial\mathbb{A}_R^N$*

Proof. Let P_{opt} be the optimal solution to problem 4.4.1. Now assume $P_{opt} \notin \partial\mathbb{A}_R^N$, namely the set $\mathbb{E}(u_P)$ is not a singleton, then by theorem 4.2.1 there exist a Schur function $b(\lambda)$ satisfying

$$b(\alpha_i) = \frac{L_{22}(\alpha_i)}{u_P(\alpha_i)} \qquad\qquad \forall i \in [1, M],$$

such that

$$\begin{aligned} |b(\omega)| &< 1 && \forall \omega \in \mathbb{R}, \\ \lim_{\omega \to \infty} |b(\omega)| &< 1. \end{aligned}$$

Now construct the function $\rho \in \mathbb{F}$ as $\rho = u_{P_{opt}} \cdot b$. We have

$$|\rho(\omega)|^2 < \mu < \frac{P_{opt}(\omega)}{P_{opt}(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R},$$

where $\mu > 0$. Therefore, since eq. (4.16) is not saturated we can multiply P_{opt} by a constant κ with $0 < \kappa < 1$ such that $\kappa P_{opt} \in \mathbb{A}_R^N$. Additionally we have

$$\frac{\kappa P_{opt}(\omega)}{R(\omega)} < \frac{P_{opt}(\omega)}{R(\omega)} \qquad \forall \omega \in \mathbb{R} \setminus \{w \in \mathbb{R},\ P_{opt}(w) = 0\}$$

which shows that κP_{opt} improves the criterium in problem 4.4.1. A contradiction. □

Corollary 4.4.1. *If $P_{opt} \in \mathbb{A}_R^N$ is the solution to problem 4.4.1, and eq.* (4.16) *is not binding, the unique function $b \in \mathbb{\Sigma}$ satisfying*

$$b(\alpha_i) = \frac{L_{22}(\alpha_i)}{u_{P_{opt}}(\alpha_i)} \qquad \forall i \in [1, M]$$

is a Blaschke product of degree at most $M - 1$.

Corollary 4.4.2. *In the preceding case, if we assume further a load of degree 1, the obtained Blaschke product b is of degree $M - 1 = 0$. Therefore this is a case in which the introduced relaxation exact, namely the polynomial P_{opt} is also solution to problem 4.1.1.*

We have proven in lemma 4.4.1 the existence of a sequence $\{P_n\}_1^\infty$ converging to a polynomial P_{opt} such that $\Psi(P_{opt}) = \psi_{opt}$. Nevertheless it is not clear for the moment if the polynomial P_{opt} is the only polynomial $P \in \mathbb{A}_R^N$ such that $\Psi(P) = \psi_{opt}$. To answer that question, we shall prove next the unicity of the solution to problem 4.4.1

Theorem 4.4.3 (Unicity)**.** *The optimal polynomial P_{opt} solution to problem 4.4.1 is unique.*

Proof. To study the unicity of the solution to problem 4.4.1, two different scenarios must be considered.

1. Suppose there exists a polynomial P_{opt} solution to problem 4.4.1 which does not belong to $\partial\mathbb{A}_R^N$. In this case problem 4.4.1 is only constrained by eq. (4.16) and therefore P_{opt} is the solution to the classical filter convex synthesis problem on $\mathbb{P}_{2N}^+$ which is known to have a unique solution, characterised by an equioscillation property. Hence we conclude P_{opt} is unique.

2. Suppose now $P_1, P_2 \in \partial\mathbb{A}_R^N$ are distinct optimal solutions to problem 4.4.1, for which eq. (4.16) is saturated or not, and that provide the optimal criterium ψ_{opt}

$$\begin{aligned} P_1(\omega) &\leq \psi_{opt} R(\omega) && \forall \omega \in \mathbb{I}, \\ P_2(\omega) &\leq \psi_{opt} R(\omega) && \forall \omega \in \mathbb{I}. \end{aligned}$$

Note that polynomials P_1 and P_2 also satisfy eq. (4.16), therefore

$$\begin{aligned} P_1(\omega) &\leq \Gamma R(\omega) && \forall \omega \in \mathbb{J}, \\ P_2(\omega) &\leq \Gamma R(\omega) && \forall \omega \in \mathbb{J}. \end{aligned}$$

Then take $P_3 = \kappa P_1 + (1-\kappa)P_2$ with $0 \leq \kappa \leq 1$. Polynomial P_3 obtained as a combination of P_1 and P_2 also satisfies

$$\begin{aligned} P_3(\omega) &\leq \psi_{opt} R(\omega) && \forall \omega \in \mathbb{I}, \\ P_3(\omega) &\leq \Gamma R(\omega) && \forall \omega \in \mathbb{J}. \end{aligned}$$

Furthermore since P_1, P_2 are admissible, from theorem 4.3.1 and corollary 4.3.1 we have a function $\rho_3 \in \mathbb{F}$ with $\rho_3 = \kappa\rho_1 + (1-\kappa)\rho_2$ such that

$$|\rho_3(\omega)| < |u_{P_3}(\omega)| \qquad \forall \omega \in \mathbb{R} \setminus \mathbb{L}, \tag{4.18}$$

where $\mathbb{L} \subset \mathbb{R}$ is a finite set of points. Hence P_3 belongs to $\mathbb{A}_R^N$ but not to $\partial\mathbb{A}_R^N$ because of the strict inequality in eq. (4.18) and we are back to the first case: a contradiction. □

4.5 Characterisation of the optimal solution

Up to this point, we have stated the matching problem using a relaxed convex formulation and provided some basic properties of the solution to this relaxed problem. Nevertheless, we have not provided any effective characterisation of the optimal solution to problem 4.4.1. In this section, we describe some property verified by the optimal polynomial P_{opt} in terms of number of extremal points. First we need to provide a particular definition for the multiplicity of a real root ξ of a polynomial $p \in \mathbb{P}^N$.

Definition 4.5.1 (Multiplicity of polynomial roots)**.** *Consider a polynomial $p \in \mathbb{P}^N$ with roots at the points $x_i \in \mathbb{R}$ for all $i \in [1, N]$. Let m_i be the standard multiplicity of the root x_i of p. We define $\mu(x_i)$ as the smallest even integer such that $m_i \leq \mu(x_i)$.*

We can now state the main theorem.

Theorem 4.5.1 (Number of optimal extrema points)**.** *Take the polynomial $P_{opt} \in \mathbb{P}_+^{2N}$ which provides the optimal criterium Ψ_{opt} in problem 4.4.1 and denote by $x_i \in \mathbb{I}$ with $i \in [1, n \leq N]$ all the roots of the polynomial $P_{opt} - \Psi_{opt}R$ within the interval $\mathbb{I}$. We have*

$$\frac{1}{2}\sum_{i=1}^{n} \mu(x_i) \geq N + 1.$$

Before providing the proof, we shall remember following basic property

Theorem 4.5.2 (Positive polynomials with prescribed roots)**.** *Given $x_i \in \mathbb{R}$ a finite number of points, and $m_i \in \mathbb{N}$ a set of associated multiplicities. Let $M = \sum_i 2m_i$. There exists a polynomial $\Phi \in \mathbb{P}_+^{2N}$ with roots at the points x_i of multiplicity $2m_i$ if and only if $N \geq M$.*

Proof of theorem 4.5.1. To prove theorem 4.5.1, we assume

$$\frac{1}{2}\sum_{i=1}^{n}\mu(x_i) \leq N,$$

and then we construct an admissible polynomial $P \neq P_{opt}$ such that $P(\omega) \leq \psi_{opt}R(\omega)$ for all $\omega \in \mathbb{I}$. This contradicts the unicity of P_{opt} stated in theorem 4.4.3. We have

$$\sum_{i=1}^{n}\mu(x_i) \leq 2N.$$

From theorem 4.5.2 there exists a polynomial $\Phi \in \mathbb{P}_{+}^{2N}$ with roots of multiplicity $\mu(x_i)$ at each points x_i. Then consider the polynomial

$$P(\omega) = P_{opt}(\omega) + \epsilon\Phi(\omega),$$

with ϵ is a positive constant. We show next that there exist $\epsilon > 0$ such that

$$\begin{aligned} P(\omega) - \Psi_{opt}R(\omega) &\leq 0 \qquad && \forall\omega \in \mathbb{I}, \\ P(\omega) - \Gamma R(\omega) &\geq 0 && \forall\omega \in \mathbb{J}. \end{aligned} \tag{4.19}$$

Note that eq. (4.19) is satisfied for any $\epsilon > 0$ since $P(\omega) \geq P_{opt}(\omega)$ for all $\omega \in \mathbb{R}$. Furthermore since $P_{opt} \in \mathbb{A}_R^N$ then we have $\hat{P} \in \mathbb{A}_R^N$ as well. Define

$$S(\omega) = P_{opt}(\omega) - \Psi_{opt}R(\omega),$$

which is by construction negative or zero on I. The Taylor expansions of S and Φ at x_i, a zero of order m_i of S, writes as:

$$S(x_i + h) = \frac{\mathrm{D}^{m_i}S(x_i)}{(m_i)!}h^{m_i} + o(h^{m_i+1}),$$

$$\Phi(x_i) = \begin{cases} \dfrac{\mathrm{D}^{m_i}\Phi(x_i)}{(m_i)!}h^{m_i} + o(h^{m_i+1}) & \text{if } m_i \text{ is even} \\ \dfrac{\mathrm{D}^{m_i+1}\Phi(x_i)}{(m_i+1)!}h^{m_i+1} + o(h^{m_i+2}) & \text{if } m_i \text{ is odd} \end{cases}$$

We claim that we can find an open neighbourhood $\mathbb{\Omega}_i$ of x_i and $\epsilon > 0$ sufficiently small such that $S + \epsilon\Phi$ has same sign as S on $\mathbb{\Omega}_i$. If m_i is odd, $\Phi(x_i + h)$ is an $o(h^{m+1})$, so that any neighbourhood $\mathbb{\Omega}_i$ small enough will do. For m_i even, the first term in the Taylor expansion of $S + \phi$ around x_i is

$$\frac{\mathrm{D}^{m_i}S(x_i) + \epsilon\mathrm{D}^{m_i}\Phi(x_i)}{m_i!}h^{m_i}$$

so that taking

$$\epsilon \leq \frac{1}{2}\left|\frac{\mathrm{D}^{m_i}S(x_i)}{\mathrm{D}^{m_i}\Phi(x_i)}\right|$$

and $\mathbb{\Omega}_i$ small enough will do. Doing this repeatedly for all x_i and taking the minimum of all found values for ϵ, we have:

$$\forall w \in \bigcup_i \mathbb{\Omega}_i \cap I, \;\; S(w) + \epsilon\Phi(w) \leq 0.$$

Now on the compact set $I \setminus \bigcup_i \mathbb{\Omega}_i$, which contains no zeros of S, the maximum of S is a strictly negative value $a < 0$. Upon one final lowering of $\epsilon > 0$ we have therefore,

$$\forall \omega \in \mathbb{I}, \;\; S(\omega) + \epsilon\Phi(\omega) = P_{opt}(\omega) - \Psi_{opt}R(\omega) + \epsilon\Phi(\omega) \leq 0.$$

Hence the polynomial $P = P_{opt} + \epsilon\Phi$ satisfies $P(\omega) \leq \Psi_{opt}R(\omega)$ for all $\omega \in \mathbb{I}$ what contradicts theorem 4.4.3. □

Theorem 4.5.3 (Number of points where the optimal criterium is attained)**.** *If all points x_i with $i \in [1, n]$ have "simple" multiplicity $m_i \leq 2$ then the optimal criterium Ψ_{opt} is attained by the function $P_{opt}(\omega)/R(\omega)$ at least $N + 1$ times within the interval $\omega \in \mathbb{I}$.*

Proof. Let $m_i \leq 2$ for all $i \in [1, n]$, considering definition 4.5.1 we have

$$\frac{1}{2}\sum_{i=1}^{2} \mu(x_i) \leq \frac{1}{2}\sum_{i=1}^{n} 2 = n$$

and from theorem 4.5.1 we conclude $n \geq N + 1$. Hence the the polynomial $P_{opt} - \Psi_{opt}R$ has at least $N + 1$ roots in the interval $\mathbb{I}$. □

4.6 Characterisation of $\mathbb{A}_R^N$ for a load of degree 1

Note that parametrisation of the set $\mathbb{A}_R^N$ is still very abstract since it requires to establish the existence, or not, of a Schur function satisfying a set of interpolation conditions. We have used for the moment Nevanlinna parametrisation of the set of such Schur interpolant (provided in theorem B.1.3) for this purpose. Nevertheless it still does not allow for a simple numerical implementation of problem 4.4.1. Nevertheless there is one case where this characterisation become quite simple. It is the case of a load of degree 1 with one single transmission zero $\alpha_1 \in \mathbb{C}$. In this case we have $P \in \mathbb{A}_R^N$ if and only if there exist a Schur function $f(\lambda)$ satisfying the single interpolation condition

$$f(\alpha) = \frac{L_{22}(\alpha)}{u_P(\alpha)}.$$

For a load of degree 1, we have the following lemma

Lemma 4.6.1. *There exist a Schur function $f(\lambda)$ of degree at most one satisfying $f(\alpha) = \gamma$, with $\alpha \in \mathbb{C}^-$ and $\gamma \in \mathbb{D}$ if and only if $|\gamma| \leq 1$.*

Proof. Follows from eq. (B.2). □

From the previous lemma we obtain the following characterisation

Theorem 4.6.1 (Characterisation of $\mathbb{A}_R^N$)**.** *The positive polynomial P belong to the set $\mathbb{A}_R^N$ if and only if the function u_P satisfies*

$$|u_P(\alpha)| \geq |L_{22}(\alpha)|. \tag{4.20}$$

By definition, the evaluation of a the modulus of a minimum phase function inside its analyticity domain can be computed by means of the Poisson integral of its log modulus functions. See for instance [28] where the Poisson integral is studied using the Poisson kernel of the unit disk. Remember now that the Poisson kernel of the lower plane is:

$$P_y(x) = \frac{1}{\pi}\frac{-y}{x^2+y^2}.$$

The Poisson integral for the lower half plane is computed as

$$\begin{aligned}
\log(u_p(\alpha = x+iy)) &= \frac{1}{\pi}\int_{\mathbb{R}} \log|u_P(\tau)| P_y(x-\tau)d\tau \\
&= \frac{1}{\pi}\int_{\mathbb{R}} \frac{-y\log|u_P(\tau)|}{|x-\tau|^2+y^2}d\tau \\
\log(u_p(\alpha = x+iy)) &= \frac{1}{\pi}\int_{\mathbb{R}} \log\left(\sqrt{\frac{P(\tau)}{P(\tau)+R(\tau)}}\right)\frac{\Im(\overline{\alpha})}{|\alpha-\tau|^2}d\tau
\end{aligned}$$

which yields for $\alpha \in \mathbb{C}^-$

$$= \frac{\Im(\overline{\alpha})}{2\pi}\int_{\mathbb{R}} \frac{\log\left(\frac{P(\tau)}{P(\tau)+R(\tau)}\right)}{|\alpha-\tau|^2}d\tau.$$

Finally we take logarithm in eq. (4.20) to obtain

$$\log|u_P(\alpha)| = \frac{\Im(\overline{\alpha})}{2\pi}\int_{\mathbb{R}} \frac{\log\left(\frac{P(\tau)}{P(\tau)+R(\tau)}\right)}{|\alpha-\tau|^2}d\tau \geq \log|L_{22}(\alpha)|,$$

or equivalently

$$\int_{\mathbb{R}} \frac{\log\left(1+\frac{R(\tau)}{P(\tau)}\right)}{|\alpha-\tau|^2}d\tau \leq K,$$

with

$$K = \frac{2\pi\log|L_{22}(\alpha)|}{\Im(\alpha)},$$

where both quantities are real positive.

Additionally, we can characterise the subset $\partial\mathbb{A}_R^N \subset \mathbb{A}_R^N$ as the set of polynomials $P \in \mathbb{P}_+^{2N}$ where eq. (4.20) is saturated.

Theorem 4.6.2 (Characterisation of $\partial\mathbb{A}_R^N$)**.** *The positive polynomial P belong to $\partial\mathbb{A}_R^N$ if and only if*

$$|u_P(\alpha)| = |L_{22}(\alpha)|.$$

4.6.1 Statement of the problem

We are now prepared to state problem 4.4.1 in the case where the load is of degree 1 by using the provided parametrisation of the set $\mathbb{A}_R^N$

Problem 4.6.1 (Relaxed matching problem of degree 1)**.**

$$\begin{aligned} &\textit{Find:} && L_{opt} = \min_{P\in\mathbb{P}_R^{2N}} \max_{\lambda\in\mathbb{I}} \frac{P(\lambda)}{R(\lambda)}, \\ &\textit{Such that:} && \frac{P(\lambda)}{R(\lambda)} \geq \Gamma && \forall\lambda\in\mathbb{J}, \qquad (4.21) \\ & && \int_{\mathbb{R}} \frac{\log\left(1+\frac{R(\tau)}{P(\tau)}\right)}{|\alpha-\tau|^2} d\tau \leq K && K>0, \qquad (4.22) \end{aligned}$$

where

$$K = \frac{2\pi \log|L_{22}(\alpha)|}{\Im(\alpha)}.$$

If $J = \emptyset$ (no selectivity constraints are considered) the relaxed matching problem (problem 4.6.1) is in fact exact.

Problem 4.6.1 is a convex optimization program with linear cost function and a single non linear scalar convex constraint. The latter can therefore be solved efficiently with a guarantee of optimality: to the best of our knowledge this result is new and constitutes one of the rare situation, but already an interesting one (loads of degree one are quite common), where problem 4.1.1 can be solved optimally for any degree N.

4.7 Extraction of the matching filter

At this point we believe it is convenient to include a small overview of the algorithm introduced in this work for the computation of matching filters as it differs from the classical design procedure. Throughout this chapter, we have developed an optimisation problem focused on the minimisation of the magnitude of the global system reflection $|S_{22}(\omega)|$ within a given frequency interval. Equivalently, the synthesis problem over the function S_{22} also provides the optimal system when the input reflection coefficient S_{11} is considered since for lossless devices we have $|S_{11}(\omega)| = |S_{22}(\omega)|$ for all $\omega \in \mathbb{R}$. This is, indeed, the ultimate goal of the matching problem as we are interested on minimising the input reflection of the system. Nevertheless, in order to obtain the optimal reflection in practice, it is still necessary to compute the matching filter which provides such reflection S_{11}.

In the case of degree one, the matching filter providing the optimal global reflection can be computed directly from the Belevitch model of the global system after the de-embedding of the load. This is possible if the relaxation made in problem 4.4.1 is exact, namely the solution obtained by solving problem 4.4.1 is also the solution to problem 4.1.1.

Therefore when eq. (4.21) is not saturated in problem 4.6.1 the optimal polynomial $P_{opt} \in \mathbb{A}_R^N$, once it is factorised in its inner and outer factors such that $P_{opt} = p_{opt}p_{opt}^\star$ where p_{opt} has all roots in $\mathbb{C}^+$ and $p_{opt}^\star$ in $\overline{\mathbb{C}^-}$, the polynomial p_{opt} satisfies

$$S_{22} = \frac{p_{opt}}{q_{opt}} \in \mathbb{F}_R^N,$$

where q_{opt} is the stable polynomial satisfying $q_{opt}q_{opt}^\star = P_{opt} + R$. Since $S_{22} \in \mathbb{F}_R^N$, there exist a filter which chained at the input of the antenna provides the global reflection S_{22}. The output reflection F_{22} of this filter can be computed by eq. (3.13) as

$$F_{22} = \frac{L_{22} - S_{22}}{\det(L) - S_{22}L_{11}}.$$

The reflection coefficient expressed in the rational form $F_{22} = p_F/q_F$ with $p_F, q_F \in \mathbb{P}^N$ provides us with the Belevitch model of the scattering matrix F of the matching filter

$$F = \frac{1}{q_F}\begin{pmatrix} p_F^\star & -r_F^\star \\ r_F & p_F \end{pmatrix},$$

where the polynomial $r_F \in \mathbb{P}^N$ satisfies $r_F r_F^\star = R_F$ and $R_F \cdot R_L = R$.

Remark 4.7.1. *Note that the function F_{22} does not depend on the distribution of the roots of R_F between r_F and $r_F^\star$. Therefore this distribution can be done arbitrarily. Note further that in the where all roots of the polynomial R_F has even multiplicity, we can obtain a reciprocal matching filters by assigning the same roots to r_F and $r_F^\star$ such that $r_F = r_F^\star$.*

4.7.1 Overview of the proposed algorithm

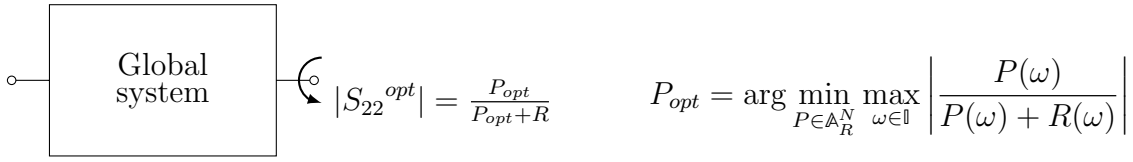
We provide finally a conceptual overview of the proposed algorithm in the case of a load of degree one, and assuming eq. (4.21) is not saturated in problem 4.6.1.

1. Minimisation of the global system reflection (fig. 4.2a). The filter synthesis approach begins with the synthesis of the global system by minimising the output reflection $|S_{22}(\omega)|$ within the passband $\mathbb{I}$. Note that being the global system lossless we have $|S_{22}(\omega)| = |S_{11}(\omega)|$ for all $\omega \in \mathbb{R}$.

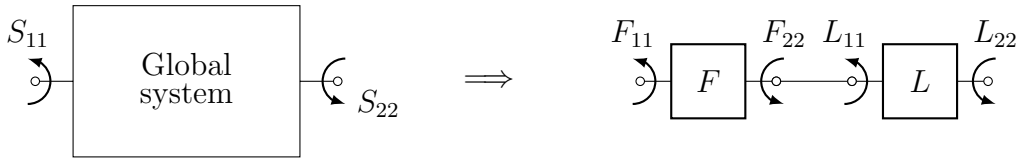
2. De-embedding of the load model (fig. 4.2b). Extraction of the matching filter after the de-embedding of the Belevitch model of the load or its Darlington equivalent in the case where only the input reflection of the load is known.

3. Realisation of the matching filter (fig. 4.2c). Use of the rational model for the physical implementation of the matching filter as it is also done with the classical filter synthesis. Note that this time, as we minimised the global reflection, the output impedance of the filter is as close as possible to the conjugate of the input impedance of the load.

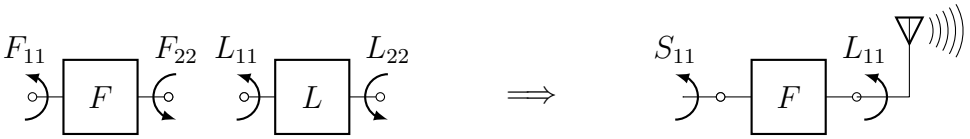
4. Cascade of the implemented filter with the load (fig. 4.2c). Once the filter has been built, we can replace the Darlington equivalent by the actual load. Since the Darling equivalent shown the same input impedance as the load, when the matching filter is plugged to the input port of the load, the global reflection S_{11} synthesised at the beginning is recovered.



(a) Step 1: synthesis of the global system.



(b) Step 2: De-embedding of the load.



(c) Step 3&4: realisation of the matching filter and cascade to the load.

4.8 Concluding remarks

This result obtained in this chapter represents an extension of the classical filter synthesis problem, where the optimality is guaranteed, to the case of rational loads of degree one. Note that the classical problem can be seen as a matching problem with a resistive load. Indeed when a constant load is considered, eq. (4.22) is not required as no transmission zeros are present, removing therefore such constraint from the problem.

Additionally note that, in the case of a not constant load, if constraint eq. (4.21) is removed (i.e. there are no selectivity requirements), problem 2.13.1 is unbounded meanwhile problem 4.1.1 is not, showing how the load imposes a limitation in terms of matching regardless of the selectivity requirements.

It should also be noted that in the case of a load of degree one, we have derived a practical characterisation of the set of admissible polynomials P, which is given by eq. (4.22). We will now consider the case of load of general degree and work towards a constructive description of the set $\mathbb{A}_R^N$ in chapter 5.

References

[26] K. Hoffman, *Banach Spaces of Analytic Functions*, new editio ed. PRENTICE-HALL, INC., 1962.

[27] J. B. Garnett, *Bounded Analytic Functions*, ser. Pure and Applied Mathematics. Elsevier Science, 1981. [Online]. Available: https://books.google.fr/books?id=DVLO9gJ66{_}YC

[28] W. Rudin, *Real and complex analysis*, ser. Mathematics series. McGraw-Hill, 1987.

Chapter 5:

Practical characterisation of the admissible set: a perspective of admissibility as a classical Nevanlinna-Pick interpolation problem

We have introduced in chapter 4, a relaxed version of problem 4.1.1 by using the notion of admissible polynomials. This relaxed problem happens to be of special interest as we demonstrated in corollary 4.3.1 the convexity of this admissible set. Furthermore in section 4.6 the theory developed through chapter 4 is applied to the most simple kind of load that is considered in this work, namely those whose reflection coefficient is modelled as a rational function of degree one, which provides us with several remarkable results. These results are reached by introducing a practical characterisation of the set of admissible polynomials for the particular case of a load of degree one. Such characterisation allows for problem 4.4.1 to be solved numerically as a standard convex optimisation problem.

Next, to generalise the results obtained in chapter 6, we shall first generalise the characterisation of the set of admissible polynomials for a load of degree $M \in \mathbb{N}$. This characterisation is achieved in a very straightforward manner utilising the Nevanlinna-Pick interpolation theorem. However, we also perform a small parenthesis in this chapter to introduce some more advanced theoretical concepts, very well known in the field of functional analysis, which is used for an in-depth study of a valuable property satisfied by the given characterisation. The reason behind the importance of this additional property of the admissible set (which is, in fact, stronger that convexity) will become clear in chapter 7.

Nevertheless, note that in chapter 7 we come back to the matching problem which we left in chapter 4 but already having the generalised characterisation of the admissible set. Furthermore, a different style is adopted in chapter 7, which is more computationally oriented. Therefore an impatient reader might skip the present chapter and go directly to chapter 7 as quick as possible, perhaps returning to chapter 5 afterwards. To study such characterisation of the admissible polynomials, we include a more bibliographic chapter here, where some classic concepts related to the Nevanlinna-Pick interpolation problem are extracted from the literature and formulated in the framework established in this thesis, so it can be directly applied to our problem.

Additionally, we provide some required basic notions on Hilbert spaces and more particularly the Hardy space H^2. Those basic concepts are briefly revised to give the reader an overall intuition on where the story is going and immediately applied in a slightly more complex form to the problem we are dealing with in this chapter. Nevertheless, a quick revision of some of the books treating Hilbert spaces and Hardy spaces is suggested, for instance, [29, Chapter 4] for a basic introduction to Hilbert spaces, and [29, Chapter 17] to acquire some notions on Hardy spaces. Similarly, a reader who is interested in the field of functional analysis should not miss the opportunity to review the reference book [30] for a more advanced lecture.

5.1 Nevanlinna-Pick interpolation

Schur interpolation is recurrent in this thesis to the point that it could be considered as the second main topic after matching problem. Note further that the general matching problem, namely problem 4.1.1 is indeed a Schur interpolation problem as the set of

feasible functions $\mathbb{F}$ consists of the Schur functions satisfying an interpolation problem.

We start this section with an introduction to Nevanlinna-Pick interpolation and the Pick theorem. After having introduced the problem in its standard form, the motivation for this chapter will become clear. The Nevanlinna-Pick interpolation problem has already been introduced in theorem B.1.3 where the Schur recursion and the original characterisation of the interpolant solutions by Nevanlinna are presented. Let us recall the problem we are referring to

Definition 5.1.1 (Nevanlinna-Pick interpolation)**.** *Given γ_1, γ_2 ... $\gamma_M \in \mathbb{D}$ and the set of points α_1, α_2 ... $\alpha_M \in \mathbb{C}^-$. The Nevanlinna-Pick interpolation problem consists on determining the functions $f \in \mathbb{\Sigma}$ such that*

$$f(\alpha_i) = \gamma_i \qquad \forall i \in [1, M]. \tag{5.1}$$

Let us now provide the Pick theorem, a powerful tool for the characterisation of the set of interpolant functions satisfying eq. (5.1).

Theorem 5.1.1 (Pick Theorem)**.** *There exist a Schur function $f : \mathbb{C}^- \to \mathbb{D}$ satisfying eq.* (5.1) *if and only if the* Pick matrix Δ *defined as*

$$\Delta = \frac{1}{j}\begin{bmatrix} \frac{1-\gamma_1\overline{\gamma_1}}{\alpha_1-\overline{\alpha_1}} & \frac{1-\gamma_1\overline{\gamma_2}}{\alpha_1-\overline{\alpha_2}} & \cdots & \frac{1-\gamma_1\overline{\gamma_M}}{\alpha_1-\overline{\alpha_M}} \\ \frac{1-\gamma_2\overline{\gamma_1}}{\alpha_2-\overline{\alpha_1}} & \frac{1-\gamma_2\overline{\gamma_2}}{\alpha_2-\overline{\alpha_2}} & \cdots & \frac{1-\gamma_2\overline{\gamma_M}}{\alpha_2-\overline{\alpha_M}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1-\gamma_M\overline{\gamma_1}}{\alpha_M-\overline{\alpha_1}} & \frac{1-\gamma_M\overline{\gamma_2}}{\alpha_M-\overline{\alpha_2}} & \cdots & \frac{1-\gamma_M\overline{\gamma_M}}{\alpha_M-\overline{\alpha_M}} \end{bmatrix} \tag{5.2}$$

is positive semidefinite. Furthermore, f is unique if and only if Δ is singular. In this case f is a Blaschke *product of degree equal to the rank of the matrix in eq.* (5.2).

Now after all the theory developed in chapter 4 around the realisability of the global system reflection with a prescribed load and its application to the case of a load of degree 1, we are ready to apply such theory to the general case. First let us illustrate how the Pick theorem can be used for that purpose.

Note that in the same way as in chapter 4, we consider a load with a 2×2 scattering matrix L of McMillan degree M and simple transmission zeros $\alpha_1, \alpha_2, \cdots, \alpha_M \in \mathbb{C}^-$. The matrix L present a rational form and its parametrised by means of the polynomials $p_L, q_L, r_L \in \mathbb{P}^M$ as in eq. (2.38)

$$L = \frac{1}{q_L}\begin{pmatrix} p_L^\star & -r_L^\star \\ r_L & p_L \end{pmatrix}.$$

We denote by $R_L = r_L r_L^\star$ the transmission polynomial of L and q_L the stable polynomial such that $q_L q_L^\star = p_L p_L^\star + R_L$. From definition 4.2.2, a minimum phase function $u \in \mathbb{\Sigma}$ is

admissible if and only if the set of interpolant functions $\mathbb{E}(u)$ (see definition 4.2.1) which contains the functions f satisfying

$$f(\alpha_i) = \frac{L_{22}(\alpha_i)}{u(\alpha_i)} \qquad \forall i \in [1, M] \tag{5.3}$$

is not empty. Note in eq. (5.3) all points α_i lies in the interior of $\mathbb{C}^-$ and the values $L_{22}(\alpha_i)$ belong to $\mathbb{D}$. Additionally, since the function u is of minimum phase, the values of $L_{22}(\alpha_i)u(\alpha_i)$ are bounded. Hence $\mathbb{E}(u)$ is the set of solutions of a *Nevanlinna-Pick interpolation* problem.

Therefore by means of eq. (5.2) a more tangible characterisation of admissibility is provided in terms of the positivity of the matrix Δ. This new characterisation is equivalent to definition 4.2.2, Particularly we obtain that the minimum phase function u is admissible if and only if the matrix in eq. (5.2) is positive semidefinite $\Delta \succeq 0$ where the interpolation values γ_i for all $i \in [1, M]$ are computed as $\gamma_i = L_{22}(\alpha_i)u(\alpha_i)^{-1}$.

5.2 Introduction to Hardy spaces of vector valued functions

The purpose of this section is to provide a general notion on Hardy spaces providing some properties that applies to the characterisation of the set of admissible polynomials. Nevertheless it might be pertinent to remember the definition of a Hilbert space. A Hilbert space is a Banach space, namely it is normed and complete, provided with an inner product. For instance, the euclidean space with N dimensions $\mathbb{R}^N$ provided with the usual inner product $\langle a, b\rangle = a^T b$ with $a = (a_1, a_2, \cdots, a_N)^T$, $b = (b_1, b_2, \cdots, b_N)^T$ and $a, b \in \mathbb{R}^N$ is a Hilbert space.

Let us now introduce the Hilbert space L^2 which contains the functions $f : \mathbb{R} \longrightarrow \mathbb{C}$ that are square integrable

$$\int_{\mathbb{R}} |f(\lambda)|^2 d\lambda < \infty.$$

The space L^2 is endorsed with the standard inner product $\langle f, g\rangle$ defined as

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(\omega)\overline{g(\omega)}d\omega \qquad f, g \in L^2. \tag{5.4}$$

This is actually, the only Hilbert space among the L^P spaces.

In section 3.3.1.1 we already introduced the Hardy space H^∞. Now we provide a more general view of Hardy spaces, namely H^p spaces, where the space H^2 outstands over the rest by showing some properties of particular interest for our purposes. We give next the general definition of H^p

Definition 5.2.1 (Hardy spaces)**.** *Denote by H^p, with $1 \leq p \leq \infty$ the following class of functions*

$$H^p = \left\{ f \in \mathcal{H}(\mathbb{C}^-) \mid \sup_{\sigma<0} \left(\int_{-\infty}^{\infty} |f(\omega + j\sigma)|^p d\omega \right)^{\frac{1}{p}} < \infty \right\},$$

where $\|f\|_p$*, defined as*

$$\|f\|_p = \sup_{\sigma<0} \left(\int_{-\infty}^{\infty} |f(\omega + j\sigma)|^p d\omega \right)^{\frac{1}{p}} \qquad 1 \leq p < \infty,$$

can be proved to be a norm. Particularly, it corresponds to the norm associated to the space H^P*. Additionally note that if* $f \in H^P$*, then the non-tangential limit of* $f(\omega + j\sigma)$ *when* $\sigma \to 0$ *exist and allows to define the function* f *on the real axis as*

$$f(\omega)|_{\omega\in\mathbb{R}} = \lim_{\substack{\sigma\to 0 \\ \sigma<0}} f(\omega + j\sigma).$$

For the case of H^∞ *the norm* $\|f\|_\infty$ *is defined separately as in section 3.3.1.1*

$$\|f\|_\infty = \sup_{\omega\in\mathbb{R}} |f(\omega)|.$$

In this chapter, among the previously introduced Hardy spaces, we are mostly interested in the space H^2. The space H^2 is a Hilbert space equipped with the scalar product in eq. (5.4) and the norm

$$\|f\|_2 = \left(\int_{-\infty}^{\infty} |f(\omega)|^2 d\omega \right)^{1/2}.$$

Consider further the space of $1 \times k$ vector valued functions where each element belong to H^2. We denote such space $H^2_{1\times k}$. Equivalently we can define the inner product of two functions $f, g \in \mathbb{H}^2_{1\times k}$ as

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(\omega) g(\omega)^\star d\omega \qquad f, g \in H^2_{1\times k}.$$

Additionally, we extend now some of the concepts introduced in section 2.1 to the case of matrix valued functions. First we extend the set of Schur functions to allow for the inclusion of matrix functions

Definition 5.2.2 (Contractive matrix)**.** *We denote by* $\mathbb{\Sigma}_{k\times l}$ *the set of analytic and contractive matrix functions.*

$$\mathbb{\Sigma}_{k\times l} = \left\{ S \in \mathcal{H}(\mathbb{C}^-) \mid S : \mathbb{C} \longrightarrow \mathbb{C}^{k\times l}; S(\lambda)^\star S(\lambda) \preceq I_k \ \forall \lambda \in \mathbb{C}^- \right\},$$

where $\mathbb{C}^{k\times l}$ *represents the complex* $k \times l$ *matrices.*

Note also that when referring to a column vector, we omit the column index and use the simplified notation $\mathbb{C}^k$ referring to a complex valued column vector or $\mathbb{\Sigma}_k$ to represents the analytic vector valued functions S such that $S(\lambda)^\star S(\lambda) \leq 1$ for all $\lambda \in \mathbb{C}^-$.

Furthermore we also have the equivalent to the maximum modulus theorem for matrix functions. Particularly, if $S \in \mathcal{H}(\mathbb{C}^-)$, $S : \mathbb{C} \longrightarrow \mathbb{C}^{k\times l}$ and $S(\omega)^\star S(\omega) \preceq I$ for all $\omega \in \mathbb{R}$ therefore $S \in \mathbb{\Sigma}_{k\times l}$. We also develop next the Cauchy's formula for functions in the space $H^2_{1\times k}$ and a formula to evaluate the projection of a function belonging to the space L^2_k onto $H^2_{1\times k}$. We follow now by remembering Cauchy's integral formula and we develop the remaining theory from there, in that way no major mathematical knowledge is required.

Theorem 5.2.1 (Cauchy's integral theorem)**.** *Let $f \in \mathcal{H}(\mathbb{K})$ where $\mathbb{K} \subset \mathbb{C}$ is an open subset of $\mathbb{C}$ such that $f : \mathbb{K} \longrightarrow \mathbb{C}$. Then if $\gamma \subset \mathbb{K}$ is a simple closed path going clockwise around a point $\lambda_0 \in \mathbb{K}$, we have*

$$f(\lambda_0) = \frac{1}{2\pi j} \int_\gamma \frac{f(\lambda)}{\lambda_0 - \lambda} d\lambda. \tag{5.5}$$

With the choice of the analyticity domain used in this work namely $\mathbb{C}^-$, if $f \in H^2$ and $\lambda_0 \in \mathbb{C}^-$ then we can express Cauchy's integral by means of the inner product defined above

$$f(\lambda_0) = \frac{1}{2\pi j} \int_\gamma \frac{f(\lambda)}{\lambda_0 - \lambda} d\lambda = \left\langle f, \frac{1}{2\pi j \left(\lambda - \overline{\lambda_0}\right)} \right\rangle.$$

It should be noted the function $K_{\lambda_0}(\lambda)$

$$K_{\lambda_0}(\lambda) = \frac{1}{2\pi j \left(\lambda - \overline{\lambda_0}\right)}$$

belongs to H^2. The expression $f(\lambda) = \langle f, K_{\lambda_0} \rangle$ is the Cauchy's formula for H^2 which allows to evaluate any function $f \in H^2$ at a point $\lambda_0 \in \mathbb{C}^-$ as the scalar product of f with a function $K_{\lambda_0} \in H^2$.

5.2.1 Reproducing kernel Hilbert space

Next let us provide the notion of reproducing kernel Hilbert space is required. A Hilbert space H is a reproducing kernel Hilbert space if and only if the evaluation of any function $f \in H$ at a point λ inside the analyticity domain, in our case $\mathbb{C}^-$ is obtained as the scalar product of the function f with a function $K_\lambda \in H$.

$$f(\lambda) = \langle f, K_\lambda \rangle \qquad \forall \lambda \in \mathbb{C}^- \qquad \forall f \in H.$$

This function K_λ is called reproducing kernel of H, giving the name to this kind of Hilbert spaces. Hence we obtain the function K_λ as the reproducing kernel of H^2.

Equivalently if $f \in \mathbb{H}^2_{1\times k}$ (a row vector) we obtain a similar expression by applying eq. (5.5) to the evaluation of the function $f(\lambda_0)\xi$ with $\xi \in \mathbb{C}^k$ (a column vector) we have

$$f(\lambda_0)\xi = \frac{1}{2\pi j} \int_\gamma \frac{f(\lambda)\xi}{\lambda_0 - \lambda} d\lambda = \left\langle f, \frac{\xi^\star}{2\pi j \left(\lambda - \overline{\lambda_0}\right)} \right\rangle.$$

Therefore

$$f(\lambda_0)\xi = \langle f, \xi^\star K_{\lambda_0} \rangle. \tag{5.6}$$

Similarly eq. (5.6) is known as the Cauchy's formula for the space $H^2_{1\times k}$.

Corollary 5.2.1. *Note that lemma 5.2.1 can be used to compute*

$$\langle S, \xi^\star K_{\lambda_0} \rangle = S(\lambda_0)\xi \qquad \forall \xi \in \mathbb{C}^k \qquad \forall S \in H^2_{1\times k}. \tag{5.7}$$

In particular note that applying eq. (5.7) to compute the norm of $\|f\|^2$ with

$$f = \sum_{i=1}^{M} \xi_i^\star K_{\alpha_i} \qquad \alpha_i \in \mathbb{C}^- \qquad \xi_i \in \mathbb{C}^k$$

yields

$$\|f\|^2 = \left\langle \sum_{j=1}^{M} \xi_j^\star K_{\alpha_j}, \sum_{i=1}^{M} \xi_i^\star K_{\alpha_i} \right\rangle = \sum_{i=1}^{M}\sum_{j=1}^{M} \xi_j^\star K_{\alpha_j}(\alpha_i)\xi_i \geq 0,$$

where the positivity of the norm implies the right side is positive.

Next note that a function $f \in H^2(\mathbb{C}^-)$ where the parenthesis indicates the choice of analyticity domain, if it is evaluated in the real line we obtain a function belonging to L^2, namely for $-\infty < \omega < \infty$ we have $f(\omega) \in L^2(\mathbb{R})$. Furthermore the same property holds if $f \in H^2(\mathbb{C}^+)$. Indeed it can be shown that the space $L^2(\mathbb{R})$ can be decomposed in two orthogonal spaces $H^2(\mathbb{C}^-)$ and $H^2(\mathbb{C}^+)$. Denote then by $p : L_k^2 \longrightarrow H_{1\times k}^2$ the projection function from L_k^2 to $H_{1\times k}^2$. Lemma 5.2.1 states that the projection onto $H_{1\times k}^2$ of the function $\xi^\star S^\star K_{\lambda_0}$ is obtained as $\xi^\star S(\lambda_0)^\star K_{\lambda_0}$.

Lemma 5.2.1. *Let $S \in \mathbb{\Sigma}_{l\times k}$. Therefore*

$$p\left(\frac{\xi^\star S^\star(\lambda)}{2\pi j(\lambda - \overline{\lambda_0})}\right) = \frac{\xi^\star S(\lambda_0)^\star}{2\pi j(\lambda - \overline{\lambda_0})} = \xi^\star S(\lambda_0)^\star K_{\lambda_0}.$$

Proof. The function $S^*(\lambda)/(\lambda - \overline{\lambda_0})$ is analytic in $\mathbb{C}^+$ up to the point $\overline{\lambda_0}$ where it might have a pole of degree one. In the decomposition

$$\frac{S^*(\lambda)}{\lambda - \overline{\lambda_0}} = \left(\frac{S^*(\lambda)}{(\lambda - \overline{\lambda_0})} - \frac{S^*(\overline{\lambda_0})}{\lambda - \overline{\lambda_0}}\right) + \frac{S^*(\overline{\lambda_0})}{\lambda - \overline{\lambda_0}},$$

the first term is a function in $\mathbb{H}^2$ while the second one is in $\overline{\mathbb{H}}^2$. Recalling that such a decomposition is unique completes the proof. □

This lemma is particularly useful in the proof of the Nevanlinna-Pick interpolation theorem as it is shown in the following section.

5.3 Vectorial formulation of the Nevanlinna-Pick interpolation problem: generalised Pick matrix

We provide now an extended version of theorem 5.1.1 to the case where the interpolating function is a matrix valued function and the interpolation conditions are composed by a set of M directions and M vectors. This problem shares many properties with the standard one already mentioned, the reason why we believe it should be reviewed in this chapter. Nevertheless only a first glance of this problem and theory is provided, for more advanced information, as it is done for the literature on Hilbert spaces, we refer to [31, Chapters 1,2]. Let us provide now one of the vectorial formulations of the Nevanlinna-Pick interpolation theorem.

Theorem 5.3.1 (Left interpolation problem)**.** *Consider the left interpolation*

$$S(\alpha_i)\xi_i = v_i \qquad \forall i \in [1, M], \tag{5.8}$$

with $\alpha_1, \alpha_2, \cdots, \alpha_M$ *complex values in* $\mathbb{C}^-$ *the set of complex vectors* $v_1, v_2, \cdots, v_M \in \mathbb{C}^k$ *and with the set of directions* $\xi_1, \xi_2, \cdots, \xi_M \in \mathbb{C}^l$*. There exist a function* $S \in \mathbb{S}_{k\times l}$ *satisfying eq.* (5.8)*, if and only if* $\Delta \succeq 0$ *where*

$$\Delta = \frac{1}{j}\begin{bmatrix} \frac{\xi_1^\star\xi_1 - v_1^\star v_1}{\alpha_1 - \overline{\alpha_1}} & \frac{\xi_2^\star\xi_1 - v_2^\star v_1}{\alpha_1 - \overline{\alpha_2}} & \cdots & \frac{\xi_M^\star\xi_1 - v_M^\star v_1}{\alpha_1 - \overline{\alpha_M}} \\ \frac{\xi_1^\star\xi_2 - v_1^\star v_2}{\alpha_2 - \overline{\alpha_1}} & \frac{\xi_2^\star\xi_2 - v_2^\star v_2}{\alpha_2 - \overline{\alpha_2}} & \cdots & \frac{\xi_M^\star\xi_2 - v_M^\star v_2}{\alpha_2 - \overline{\alpha_M}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\xi_1^\star\xi_M - v_1^\star v_M}{\alpha_M - \overline{\alpha_1}} & \frac{\xi_2^\star\xi_M - v_2^\star v_M}{\alpha_M - \overline{\alpha_2}} & \cdots & \frac{\xi_M^\star\xi_M - v_M^\star v_M}{\alpha_M - \overline{\alpha_M}} \end{bmatrix}. \tag{5.9}$$

Proof of necessity. We only prove the necessity part here. Suppose there exists $S \in \mathbb{S}_{k\times l}$ verifying eq. (5.8). Define the Toeplitz operator,

$$\mathcal{T} : \begin{array}{ccc} \mathbb{H}^2_{1\times l} & \mapsto & \mathbb{H}^2_{k\times 1} \\ h & \mapsto & p(gS^*). \end{array}$$

The operator $\mathcal{T}$ is contractive, as $\|gS^\star\|_2 \leq \|g\|_2$ and the orthogonal projection on $\mathbb{H}^2_{1\times k}$ also is. The Hermitian form $\mathcal{D}$ on $\mathbb{H}^2_{l\times 1} \times \mathbb{H}^2_{1\times l}$ defined by:

$$(h, g) \in (\mathbb{H}^2_{l\times 1})^2 \quad \mathcal{D}(h, g) = \langle h, g\rangle - \langle \mathcal{T}(h), \mathcal{T}(g)\rangle$$

is therefore positive semi-definite.

Let $\mathcal{V}$ be the finite dimensional vector subspace of dimension M of $\mathbb{H}^2_{l\times 1}$ spanned by the vectorial functions $\xi_i^*.\sqrt{2\pi}K_{\alpha_i}(\lambda)$ that is ,

$$\mathcal{V} = span\{i = 1 \ldots M,\ \xi_i^*.\sqrt{2\pi}K_{\alpha_i}(\lambda)\}.$$

We claim that the matrix Δ, is the Hermitian matrix representing $\mathcal{D}$ restricted to V^2. To see this we compute $\mathcal{D}_{i,j}$, the element with index (i, j) in the matrix representing the restriction of $\mathcal{D}$ in the canonical basis of V. We have,

$$\begin{aligned} \mathcal{D}_{i,j} &= 2\pi\mathcal{D}(\xi_j K_{\alpha_j}(\lambda), \xi_i K_{\alpha_i}(\lambda)) \\ &= 2\pi\left(\langle \xi_j^* K_{\alpha_j}, \xi_i^* K_{\alpha_i}\rangle + \langle P(\xi_j^* K_{\alpha_j} S^*(\lambda)), P(\xi_i^* K_{\alpha_i}) S^*(\lambda)\rangle\right) \\ &= \frac{1}{j}\frac{\xi_j^*\xi_i}{\alpha_i - \alpha_j *} + 2\pi\langle \xi_j^* S^*(\alpha_j)K(\alpha_j), \xi_i^* S^*(\alpha_i)K(\alpha_i)\rangle \\ &= \frac{1}{j}\frac{\xi_j^*\xi_i}{\alpha_i - \alpha_j *} + 2\pi\langle v_j^* K(\alpha_j), v_i^* K(\alpha_i)\rangle \\ &= \frac{1}{j}\frac{\xi_j^*\xi_i - v_j^* v_i}{\alpha_i - \alpha_j *} \\ &= \Delta_{i,j}. \end{aligned}$$

□

We are now disposed to argue towards the parametrisation of the set $\mathbb{A}_R^N$ with the aid of theorem 5.3.1.

5.4 Parametrisation of $\mathbb{A}_R^N$

We deal in this section with functions that associates a $M \times M$ matrix to a polynomial P. Particularly we have the function $\Delta(P)$ that associates positive definite matrices to positive polynomials. This pleads in flavours of defining now the set of positive definite Hermitian matrices

Definition 5.4.1 (Positive semi-definite Hermitian matrices)**.**

$$\mathbb{H}_+^N = \left\{ S \in \mathbb{H}^N \mid S \succeq 0 \right\}.$$

Furthermore we redefine the Pick matrix in eq. (5.2) as a matrix function

$$\Delta : \mathbb{P}_+^{2N} \longrightarrow \mathbb{H}_+^M.$$

We can now parametrise the set of admissible polynomials P by means of the Pick matrix $\Delta(P)$ associated to the interpolation problem in eq. (5.3). let us apply theorem 5.1.1 to determine the existence of a function $b \in \mathbb{\Sigma}$ satisfying eq. (5.3). Thus by taking in eq. (5.2) the interpolation values γ_i defined as

$$\gamma_i = \frac{L_{22}(\alpha_i)}{u_P(\alpha_i)} \qquad\qquad \forall i \in [1, M],$$

we obtain the matrix

$$\Delta(P) = \begin{bmatrix} \dfrac{1 - \dfrac{L_{22}(\alpha_1)}{u_P(\alpha_1)}\dfrac{\overline{L_{22}(\alpha_1)}}{\overline{u_P(\alpha_1)}}}{\alpha_1 - \overline{\alpha_1}} & \dfrac{1 - \dfrac{L_{22}(\alpha_1)}{u_P(\alpha_1)}\dfrac{\overline{L_{22}(\alpha_2)}}{\overline{u_P(\alpha_2)}}}{\alpha_1 - \overline{\alpha_2}} & \cdots & \dfrac{1 - \dfrac{L_{22}(\alpha_1)}{u_P(\alpha_1)}\dfrac{\overline{L_{22}(\alpha_M)}}{\overline{u_P(\alpha_M)}}}{\alpha_1 - \overline{\alpha_M}} \\ \dfrac{1 - \dfrac{L_{22}(\alpha_2)}{u_P(\alpha_2)}\dfrac{\overline{L_{22}(\alpha_1)}}{\overline{u_P(\alpha_1)}}}{\alpha_2 - \overline{\alpha_1}} & \dfrac{1 - \dfrac{L_{22}(\alpha_2)}{u_P(\alpha_2)}\dfrac{\overline{L_{22}(\alpha_2)}}{\overline{u_P(\alpha_2)}}}{\alpha_2 - \overline{\alpha_2}} & \cdots & \dfrac{1 - \dfrac{L_{22}(\alpha_2)}{u_P(\alpha_2)}\dfrac{\overline{L_{22}(\alpha_M)}}{\overline{u_P(\alpha_M)}}}{\alpha_2 - \overline{\alpha_M}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{1 - \dfrac{L_{22}(\alpha_M)}{u_P(\alpha_M)}\dfrac{\overline{L_{22}(\alpha_1)}}{\overline{u_P(\alpha_1)}}}{\alpha_M - \overline{\alpha_1}} & \dfrac{1 - \dfrac{L_{22}(\alpha_M)}{u_P(\alpha_M)}\dfrac{\overline{L_{22}(\alpha_2)}}{\overline{u_P(\alpha_2)}}}{\alpha_M - \overline{\alpha_2}} & \cdots & \dfrac{1 - \dfrac{L_{22}(\alpha_M)}{u_P(\alpha_M)}\dfrac{\overline{L_{22}(\alpha_M)}}{\overline{u_P(\alpha_M)}}}{\alpha_M - \overline{\alpha_M}} \end{bmatrix}.$$

Note that, as an element of the space $\mathbb{H}^M$, we can apply any change of basis to the matrix Δ. Consider then the transition matrix T

$$T = \begin{bmatrix} u_P(\alpha_1) & 0 & \cdots & 0 \\ 0 & u_P(\alpha_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_P(\alpha_M) \end{bmatrix}.$$

The matrix T is invertible as the function u_P, being of minimum phase, can not vanish at the points α_i. After applying the change of basis, the matrix $T^\star\Delta(P)T$ is positive if and only if the original matrix $\Delta(P)$ is positive. Therefore we have the transformed matrix

$$T^\star\Delta(P)T = \mathbf{U}(P) - \mathbf{\Pi},$$

where

$$\mathbf{U}(P) = \frac{1}{j}\begin{bmatrix} \frac{u_P(\alpha_1)\overline{u_P(\alpha_1)}}{\alpha_1-\overline{\alpha_1}} & \frac{u_P(\alpha_1)\overline{u_P(\alpha_2)}}{\alpha_1-\overline{\alpha_2}} & \cdots & \frac{u_P(\alpha_1)\overline{u_P(\alpha_M)}}{\alpha_1-\overline{\alpha_M}} \\ \frac{u_P(\alpha_2)\overline{u_P(\alpha_1)}}{\alpha_2-\overline{\alpha_1}} & \frac{u_P(\alpha_2)\overline{u_P(\alpha_2)}}{\alpha_2-\overline{\alpha_2}} & \cdots & \frac{u_P(\alpha_2)\overline{u_P(\alpha_M)}}{\alpha_2-\overline{\alpha_M}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{u_P(\alpha_M)\overline{u_P(\alpha_1)}}{\alpha_M-\overline{\alpha_1}} & \frac{u_P(\alpha_M)\overline{u_P(\alpha_2)}}{\alpha_M-\overline{\alpha_2}} & \cdots & \frac{u_P(\alpha_M)\overline{u_P(\alpha_M)}}{\alpha_M-\overline{\alpha_M}} \end{bmatrix}, \tag{5.10}$$

and

$$\mathbf{\Pi} = \frac{1}{j}\begin{bmatrix} \frac{L_{22}(\alpha_1)\overline{L_{22}(\alpha_1)}}{\alpha_1-\overline{\alpha_1}} & \frac{L_{22}(\alpha_1)\overline{L_{22}(\alpha_2)}}{\alpha_1-\overline{\alpha_2}} & \cdots & \frac{L_{22}(\alpha_1)\overline{L_{22}(\alpha_M)}}{\alpha_1-\overline{\alpha_M}} \\ \frac{L_{22}(\alpha_2)\overline{L_{22}(\alpha_1)}}{\alpha_2-\overline{\alpha_1}} & \frac{L_{22}(\alpha_2)\overline{L_{22}(\alpha_2)}}{\alpha_2-\overline{\alpha_2}} & \cdots & \frac{L_{22}(\alpha_2)\overline{L_{22}(\alpha_M)}}{\alpha_2-\overline{\alpha_M}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{L_{22}(\alpha_M)\overline{L_{22}(\alpha_1)}}{\alpha_M-\overline{\alpha_1}} & \frac{L_{22}(\alpha_M)\overline{L_{22}(\alpha_2)}}{\alpha_M-\overline{\alpha_2}} & \cdots & \frac{L_{22}(\alpha_M)\overline{L_{22}(\alpha_M)}}{\alpha_M-\overline{\alpha_M}} \end{bmatrix}. \tag{5.11}$$

Note that only the matrix $\mathbf{U}(P)$ depends on the polynomial P while the matrix $\mathbf{\Pi}$ depends only on the values of the reflection coefficient of the load at port 2 at the transmission zeros α_i. This decomposition of the Pick matrix in $\mathbf{U}(P)$ and $\mathbf{\Pi}$ provides us with an elegant characterisation of the set admissible polynomials

Theorem 5.4.1 (Admissibility)**.** *The polynomial $P \in \mathbb{P}_+^{2N}$ belong to the set of admissible polynomials $\mathbb{A}_R^N$ if and only if*

$$\mathbf{U}(P) \succeq \mathbf{\Pi}. \tag{5.12}$$

This parametrisation is equivalent to the parametrisation of $\mathbb{A}_R^N$ obtained in chapter 6 as we compare a function of P with a fix matrix $\mathbf{\Pi}$ which is imposed by the load. Indeed note the similarity between theorem 5.4.1 and theorem 4.6.1. Particularly, taking $M = 1$ we have

$$\frac{u_P(\alpha_1)\overline{u_P(\alpha_1)}}{-2\Im\alpha_1} \succeq \frac{L_{22}(\alpha_1)\overline{L_{22}(\alpha_1)}}{-2\Im\alpha_1},$$

therefore

$$|u_P(\alpha_1)|^2 \geq |L_{22}(\alpha_1)|^2,$$

what correspond to the result obtained in eq. (4.20).

We can now express some of the results obtained in chapter 4 in terms of the matrix $\mathbf{U}(P)$. By corollary 4.3.1, the set of polynomials $P \in \mathbb{P}_+^{2N}$ such that $\mathbf{U}(P) \succeq \mathbf{\Pi}$ is a convex set. Additionally, the subspace $\partial\mathbb{A}_R^N \subset \mathbb{A}_R^N$ can be characterised as the positive polynomials P such that $\mathbf{U}(P) \succeqq \mathbf{\Pi}$, namely the matrix $\mathbf{U}(P) - \mathbf{\Pi}$ is positive semi-definite and singular.

Definition 5.4.2 (Boundary of admissibility)**.** *The polynomial $P \in \mathbb{A}_R^N$ belong to $\partial\mathbb{A}_R^N$ if and only if*

$$\mathbf{U}(P) \succeqq \mathbf{\Pi}.$$

Note this definition constitutes a generalisation of theorem 4.6.2. Similarly, we are now in measure to provide a more concise formulation of theorem 4.4.2 by means of the previous definition and the matrices $\mathbf{U}(P)$ and $\mathbf{\Pi}$ which characterise the optimal solution to problem 4.4.1 with a matrix inequality

Theorem 5.4.2 (Optimality)**.** *If $P_{opt} \in \mathbb{A}_R^N$ is the optimal solution to problem 4.4.1, and eq.* (4.16) *is not binding, then $\mathbf{U}(P) \succeqq \mathbf{\Pi}$.*

Also note the degree of the function $b \in \mathbb{\Sigma}$ solution to the interpolation problem in eq. (5.3) is linked to the rank of the Pick matrix in eq. (5.2) in the sense that if $\mathbf{U}(P) \succeq \mathbf{\Pi}$ then there exist a function $b \in \mathbb{\Sigma}$ of degree at most the rank of the matrix $\mathbf{U}(P) - \mathbf{\Pi}$ verifying eq. (5.3).

Similarly, if we consider the function $\rho = b \cdot u_P$, where b is an interpolant to *eq.* (5.2) of minimum degree, this function satisfies the interpolation conditions $\rho(\alpha_i) = L_{22}(\alpha_i)$ with $1 \leq i \leq M$ and also $\deg(\rho) = \deg(u_P) + \operatorname{rank}(\mathbf{U}(P) - \mathbf{\Pi})$. This allows us to state a more precise version of lemma 4.3.1

Lemma 5.4.1. *Let $P \in \mathbb{A}_R^N$ and $\mathit{\Lambda} = \operatorname{rank}(\mathbf{U}(P) - \mathbf{\Pi})$. There exists a function $\rho \in \mathbb{F}^{N+\mathit{\Lambda}}$ such that*

$$|\rho(\omega)|^2 \leq \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R}.$$

We introduce next a vectorial version of the Nevanlinna-Pick interpolation problem as it is required to prove a crucial property of the matrix $\mathbf{U}(P)$.

5.4.1 Concavity of the matrix function $\mathbf{U}(P)$

This section is devoted to study a remarkable property of the matrix function $\Delta(P)$ which is of vital importance for the numeric solution of problem 4.4.1 in chapter 7. We are referring to the concavity property. Note the relaxation version of the matching problem formulated in problem 4.4.1 is convex thanks to the convexity of the admissible set of polynomials $\mathbb{A}_R^N$. Moreover the convexity of such set is demonstrated in corollary 4.3.1 as

a consequence of the concavity of the function $U_P(\lambda)$ at a point $\lambda \in \mathbb{C}$ with respect to polynomial P (shown in lemma 4.3.2).

In this chapter we characterise the set $\mathbb{A}_R^N$ such that $P \in \mathbb{A}_R^N$ if and only if $\mathbf{U}(P) \succeq \mathbf{J\!I}$. This characterisation is particularly interesting because of the properties of the matrix function $\mathbf{U}(P)$. In particular, we use now again lemma 4.3.2 to prove the concavity of $\mathbf{U}(P)$ with respect to P, namely the main theorem of this chapter

Theorem 5.4.3 (Concavity of $\mathbf{U}(P)$)**.** *The matrix function $\mathbf{U}(P)$ defined in eq. (5.10) is concave in the matrix sense. In other words it satisfies*

$$\mathbf{U}(\kappa P_1 + (1-\kappa)P_2) \succeq \kappa \mathbf{U}(P_1) + (1-\kappa)\mathbf{U}(P_2) \qquad \forall P_1, P_2 \in \mathbb{P}_+^{2N}.$$

Proof. We shall prove now that for every $P_1, P_2 \in \mathbb{P}_+^{2N}$ and $\kappa \in [0,1]$ we have

$$\mathbf{U}(P_3) \succeq \kappa \mathbf{U}(P_1) + (1-\kappa)\mathbf{U}(P_2), \tag{5.13}$$

where $P_3 = \kappa P_1 + (1-\kappa)P_2$. Equivalently we prove the positivity of the matrix Λ defined as

$$\Lambda = \mathbf{U}(P_3) - \kappa \mathbf{U}(P_1) - (1-\kappa)\mathbf{U}(P_2).$$

Once again, we apply a particular change of basis to the matrix Λ. This time we use the transition matrix

$$T_3 = \begin{bmatrix} \frac{1}{u_{P_3}(\alpha_1)} & 0 & \cdots & 0 \\ 0 & \frac{1}{u_{P_3}(\alpha_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{u_{P_3}(\alpha_M)} \end{bmatrix}.$$

Since the function u_P is of minimum phase we have $u_{P_3}(\alpha_i) \neq 0$ for all $i \in [1, M]$, therefore T_3 is invertible. Also note that when computing the matrix $T_3^\star \Lambda T_3$ the element (i,j) of the matrix Λ is divided by $u_{P_3}(\alpha_i)\overline{u_{P_3}(\alpha_j)}$. Thus we have for all $i, j \in [1, M]$

$$[T_3^\star \Lambda T_3]_{i,j} = \frac{1}{j(\alpha_i - \overline{\alpha_j})} \left(1 - \kappa \frac{u_{P_1}(\alpha_i)\overline{u_{P_1}(\alpha_j)}}{u_{P_3}(\alpha_i)\overline{u_{P_3}(\alpha_j)}} - (1-\kappa)\frac{u_{P_2}(\alpha_i)\overline{u_{P_2}(\alpha_j)}}{u_{P_3}(\alpha_i)\overline{u_{P_3}(\alpha_j)}} \right). \tag{5.14}$$

Consider the column function F

$$F(\lambda) = \begin{bmatrix} \sqrt{\kappa}\frac{u_{P_1}(\lambda)}{u_{P_3}(\lambda)} & \sqrt{1-\kappa}\frac{u_{P_2}(\lambda)}{u_{P_3}(\lambda)} \end{bmatrix}^T.$$

First we show that $F \in \mathbb{\Sigma}_2$. We have

$$\begin{aligned} F^\star(\omega)F(\omega) &= \frac{\kappa|u_{P_1}(\omega)|^2 + (1-\kappa)|u_{P_2}(\omega)|^2}{|u_{P_3}(\omega)|^2} && \forall \omega \in \mathbb{R} \\ &= \frac{\kappa\left(\frac{P_1(\omega)}{P_1(\omega)+R(\omega)}\right) + (1-\kappa)\left(\frac{P_2(\omega)}{P_2(\omega)+R(\omega)}\right)}{\left(\frac{P_3(\omega)}{P_3(\omega)+R(\omega)}\right)} && \forall \omega \in \mathbb{R} \\ &= \frac{\kappa U_\omega(P_1) + (1-\kappa)U_\omega(P_2)}{U_\omega(P_3)} && \forall \omega \in \mathbb{R}. \end{aligned}$$

Consider now the function $U_\omega : P \longrightarrow |u_P(\omega)|^2$. This function $U_\omega(P)$ was introduced in the proof of lemma 4.3.2. Moreover by lemma 4.3.2 we have

$$U_\omega(P_3) \geq \kappa U_\omega(P_1) + (1-\kappa)U_\omega(P_2).$$

Hence $F(\omega)^\star F(\omega) \leq 0$ for all $\omega \in \mathbb{R}$, concluding by the maximum modulus theorem that $F \in \mathbb{\Sigma}_2$. Furthermore the function F satisfies the right interpolation condition

$$F(\alpha_i)\xi_i = v_i \qquad\qquad \forall i \in [1, M],$$

with $\xi_i = 1$ for all $i \in [1, M]$ and

$$v_i = \begin{bmatrix} \sqrt{\kappa}\dfrac{u_{P_1}(\alpha_i)}{u_{P_3}(\alpha_i)} & \sqrt{1-\kappa}\dfrac{u_{P_2}(\alpha_i)}{u_{P_3}(\alpha_i)} \end{bmatrix}^T.$$

Therefore by eq. (5.9) we have $\Delta \succeq 0$ where

$$\Delta = \frac{1}{j}\begin{bmatrix} \dfrac{\xi_1^\star\xi_1 - v_1^\star v_1}{\alpha_1 - \overline{\alpha_1}} & \dfrac{\xi_2^\star\xi_1 - v_2^\star v_1}{\alpha_1 - \overline{\alpha_2}} & \cdots & \dfrac{\xi_M^\star\xi_1 - v_M^\star v_1}{\alpha_1 - \overline{\alpha_M}} \\ \dfrac{\xi_1^\star\xi_2 - v_1^\star v_2}{\alpha_2 - \overline{\alpha_1}} & \dfrac{\xi_2^\star\xi_2 - v_2^\star v_2}{\alpha_2 - \overline{\alpha_2}} & \cdots & \dfrac{\xi_M^\star\xi_2 - v_M^\star v_2}{\alpha_2 - \overline{\alpha_M}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\xi_1^\star\xi_M - v_1^\star v_M}{\alpha_M - \overline{\alpha_1}} & \dfrac{\xi_2^\star\xi_M - v_2^\star v_M}{\alpha_M - \overline{\alpha_2}} & \cdots & \dfrac{\xi_M^\star\xi_M - v_M^\star v_M}{\alpha_M - \overline{\alpha_M}} \end{bmatrix}.$$

Setting $\xi_i = 1$ and introducing the expression for v_i for all i we obtain

$$[\Delta]_{i,j} = \frac{1}{j(\alpha_i - \overline{\alpha_j})}\left(1 - \kappa\frac{u_{P_1}(\alpha_i)\overline{u_{P_1}(\alpha_j)}}{u_{P_3}(\alpha_i)\overline{u_{P_3}(\alpha_j)}} - (1-\kappa)\frac{u_{P_2}(\alpha_i)\overline{u_{P_2}(\alpha_j)}}{u_{P_3}(\alpha_i)\overline{u_{P_3}(\alpha_j)}}\right),$$

which coincides with eq. (5.14). Thus $\Delta = T_3^\star \Lambda T_3$ implying $T_3^\star \Lambda T_3 \succeq 0$. Hence eq. (5.13) follows. □

5.5 Remarks about the positivity of the Pick matrix

Before this chapter concludes, we must point out that only the proof of necessity of theorem 5.3.1 is provided here as it is essential for the general parametrisation of the set $\mathbb{A}_R^N$. Nevertheless the positivity of the matrix in eq. (5.9) is indeed necessary and sufficient to guarantee that there exist a function $S \in \mathbb{\Sigma}_{k\times l}$ satisfying eq. (5.8).

Furthermore, the equivalent right interpolation problem can be considered if ξ_i and v_i are row vectors. An exhaustive revision of matricial interpolation is available in [32] or [33], for instance, where the reader will find the remarkable theory behind the Nevanlinna-Pick interpolation problem without any lack of detail.

5.6 Remarkable contributions

To sum up, in this chapter we have introduce a characterisation of the set of admissible polynomials in terms of the matrix inequality provided by eq. (5.12). Additionally we have stated and proved the concavity property of $\mathbf{U}(P)$ in the matrix sense which is given in theorem 5.4.3. This property allows in chapters 7 to 9 for the implementation of eq. (5.12) by means of barrier functions and therefore, for the numerical resolution of problem 4.4.1.

References

[29] W. Rudin, *Real and complex analysis*, ser. Mathematics series. McGraw-Hill, 1987.

[30] J. B. Garnett, *Bounded Analytic Functions*, ser. Pure and Applied Mathematics. Elsevier Science, 1981. [Online]. Available: https://books.google.fr/books?id=DVLO9gJ66{_}YC

[31] H. Dym, *J Contractive Matrix Functions, Reproducing Kernel Hilbert Spaces and Interpolation*, ser. Conference board of the mathematical sciences: Regional conference series in mathematics. American Mathematical Soc., 1989, no. 71. [Online]. Available: https://books.google.fr/books?id=dpNF5OvPqqAC

[32] J. A. Ball, I. Gohberg, and L. Rodman, *Interpolation of rational matrix functions*, ser. Operator theory, advances and applications. Birkhäuser Verlag, 1990. [Online]. Available: https://books.google.fr/books?id=iR{_}vAAAAMAAJ

[33] L. V. Ahlfors, *Complex analysis*, 3rd ed. McGraw-Hill Education, 1966. [Online]. Available: https://books.google.fr/books?id=RfYK28TcZEwC

Chapter 6:

Solution to the problem of matching a rational load of degree one with a transmission zero on the boundary

We have provided in chapter 4 a convex relaxation of the general matching problem commonly studied in electrical engineering. To formulate the mentioned problem, we have considered an arbitrary rational load of degree M. However, in the development of the theory presented in chapter 4 we have considered a load L with only a simple transmission zero $\alpha \in \mathbb{C}^-$. The reason comes from the fact that when a transmission zero of the load approaches the boundary of the analyticity domain, previous formulas degenerates to an indetermination. Nevertheless, a limiting case of problem 4.6.1 where $\alpha \to \overline{\mathbb{R}}$ can be considered.

In this chapter, we consider the case of a load of degree one having a transmission zero on the boundary of the analyticity domain, namely the real line, and possibly at infinity. The result obtained in this case is particularly enlightening due to the simplicity of the associated expressions and also because the problem of matching with a load of degree 1 has some additional properties, which provide sufficient motivation for said case of degree 1 to be studied separately. In this case the optimal bound is exactly attained with a matching filter of fixed McMillan degree. Furthermore, this particular case is of greater importance due to the fact that the loads that are commonly faced in electronics have a transmission zero at infinite or sufficiently high frequencies.

Finally, some applications of the theory already developed are provided where the function to be matched is represented by the reflection coefficient of an antenna in a limited frequency band. In those examples, real data coming from different antennas is considered obtaining sharp bounds on the best attainable matching level.

6.1 The load

In this chapter we consider, exceptionally, a load that differs from the one used in the rest of the thesis. This load is represented by the scattering matrix $\tilde{L}$

$$\tilde{L}(\lambda) = \frac{1}{\tilde{q}_L(\lambda)} \begin{pmatrix} \tilde{p}_L^\star(\lambda) & -\tilde{r}_L^\star(\lambda) \\ \tilde{r}_L(\lambda) & \tilde{p}_L(\lambda) \end{pmatrix},$$

with $\tilde{q}_L$ the stable polynomial that satisfies $\tilde{q}_L\tilde{q}_L^\star = \tilde{p}_L\tilde{p}_L^\star + \tilde{r}_L\tilde{r}_L^\star$. We assume the load $\tilde{L}$ has McMillan degree 1, namely $\tilde{p}_L, \tilde{r}_L, \tilde{q}_L \in \mathbb{P}^1$. Additionally the single root of $\tilde{r}_L$ is on the real line. We define

$$\tilde{r}_L(\lambda) = (\alpha\lambda - 1) \qquad\qquad \alpha \in \mathbb{R}.$$

Note the polynomial $\tilde{r}_L$ has one single zero at the frequency $\lambda = \frac{1}{\alpha}$. Additionally, the transmission polynomial $\tilde{R}_L$ of the matrix $\tilde{L}$ is computed as

$$\tilde{R}_L(\lambda) = \tilde{r}_L(\lambda)\tilde{r}_L^\star(\lambda) = \tilde{r}_L(\lambda)^2 = (\alpha\lambda - 1)^2 .$$

With this definition we include as well the case of a transmission zero at infinity. Indeed note that if $\alpha \to 0$ then the roots of polynomial $\tilde{R}_L(\lambda)$ tends to infinity, obtaining at the limit the positive polynomial $\tilde{R}_L(\lambda) = 1$. Eventually, we consider the change of variable

$\lambda = \tau^{-1}$. This change of variable sends infinite to the origin and the origin to infinity.

$$L(\tau) = \frac{1}{\tau\tilde{q}_L(\tau^{-1})}\begin{pmatrix} \tau\tilde{p}_L^\star(\tau^{-1}) & -\tau\tilde{r}_L^\star(\tau^{-1}) \\ \tau\tilde{r}_L(\tau^{-1}) & \tau\tilde{p}_L(\tau^{-1}) \end{pmatrix} = \frac{1}{q_L(\tau)}\begin{pmatrix} p_L^\star(\tau) & -r_L^\star(\tau) \\ r_L(\tau) & p_L(\tau) \end{pmatrix}.$$

Note transmission zeros of the matrix $L(\tau)$ happens at a finite frequency $\alpha \in \mathbb{R}$ as the transmission polynomials becomes

$$R_L(\tau) = \tau^2\left(\frac{\alpha}{\tau} - 1\right)^2 = (\tau - \alpha)^2 \qquad \alpha \in \mathbb{R}.$$

6.2 The feasible set

The first requirement here is a characterisation of the set of feasible functions in the case of a load L with a transmission zero on the real line. In appendix A we have provided a generalised version of Fano-Youla's characterisation of the global reflection when the load present a transmission zero $\alpha \in \overline{\mathbb{R}}$. Let us now restate here theorem 6.2.1

Theorem 6.2.1 (Generalised de-embedding conditions)**.** *Consider $S_{22} \in \mathbb{\Sigma}$ and let L be the 2×2 lossless scattering matrix with simple transmission zeros $\alpha_1, \alpha_2, \ldots, \alpha_{Mr} \in \mathbb{R}$ and $\alpha_{Mr+1}, \ldots, \alpha_M \in \mathbb{C}^-$. The matrix L is de-chainable from S_{22} if and only if at each transmission zero α_i, the following condition is satisfied*

$$\begin{aligned} S_{22}(\alpha_i) &= L_{22}(\alpha_i) && i \in [1, M], \\ \mathrm{ang}S_{22}(\alpha_i) &\leq \mathrm{ang}L_{22}(\alpha_i) && i \in [1, M_r]. \end{aligned}$$

This characterization allows for the load to present any number of transmission zeros on the boundary of the analyticity domain, namely $\overline{\mathbb{R}}$. However an additional condition is imposed at each transmission zero $\alpha_i \in \overline{\mathbb{R}}$. This condition bears on the angular derivatives of the system reflection $\mathrm{ang}S_{22}(\alpha_i)$ which are defined as

$$\mathrm{ang}S_{22}(\alpha_i) = j\frac{d}{d\lambda}\log S_{22}(\lambda) \qquad \lambda = \alpha_i. \tag{6.1}$$

Remark 6.2.1. *Note that eq.* (6.1) *is not well defined when $\alpha_i \to \infty$. As transmission zeros at infinity are of interest in this section, to overcome this issue we apply here the change of variable $\lambda \to \tau^{-1}$ introduced before, allowing to handle a possible transmission zero at $\lambda = \infty$. We assume then that transmission zeros cannot happen at $\lambda = 0$.*

We apply now theorem 6.2.1 to the load $L(\tau)$ considered in this section. This load has one single transmission zero at $\alpha \in \mathbb{R}$. Hence we obtain the following necessary conditions over the reflection S_{22} of the global system

$$S_{22}(\alpha) = L_{22}(\alpha) \tag{6.2}$$

$$j\frac{d}{d\tau}\log S_{22}(\tau) \leq j\frac{d}{d\tau}\log L_{22}(\tau) \qquad \tau = \alpha. \tag{6.3}$$

Let us now provide a modified version of the feasible set for the load of degree 1 $L(\tau)$.

Definition 6.2.1 (Feasible functions for a load with a boundary tranmission zero)**.** *We denote by* $\mathbb{G}$ *the set of functions* $S_{22} \in \mathbb{\Sigma}$ *satisfying eqs.* (6.2) *and* (6.3)*.*

$$\mathbb{G} = \left\{ S_{22} \in \mathbb{\Sigma} \mid S_{22}(\alpha) = K_0; \left[j\frac{d}{d\tau} \log S_{22}(\tau) \right]_{\tau=\alpha} \leq K_1 \right\},$$

where K_0 *and* K_1 *are defined as follows:*

$$\begin{aligned} K_0 &= L_{22}(\alpha) \\ K_1 &= \left[j\frac{d}{d\tau} \log L_{22}(\tau) \right]_{\tau=\alpha}. \end{aligned}$$

6.3 The matching problem

With this new set of feasible functions, the general problem of matching still without degree restriction becomes

Problem 6.3.1 (General matching problem of degree 1 with transmission zeros on the boundary.)**.**

$$\begin{aligned} \textit{Find:} \qquad & l_{opt} = \min_{S_{22} \in \mathbb{\Sigma}} \max_{\tau \in \mathbb{I}} |S_{22}(\tau)|, \\ \textit{Such that:} \qquad & |S_{22}(\lambda)| \geq \gamma \qquad \forall \lambda \in \mathbb{J}, \qquad (6.4) \\ & S_{22}(\alpha) = K_0, \qquad (6.5) \\ & \left[j\frac{d}{d\tau} \log S_{22}(\tau) \right]_{\tau=\alpha} \leq K_1. \qquad (6.6) \end{aligned}$$

It should be noted that the function S_{22} can be multiplied by any uni-modular value ϵ without modifying neither the criterium of problem 6.3.1 nor the constrains eq. (6.4) or eq. (6.6) since

$$\frac{d}{d\lambda} \log\left(\epsilon S_{22}(\tau)\right) = \frac{d}{d\lambda} \left[\log S_{22}(\tau) + \log(\epsilon)\right] = \frac{d}{d\lambda} \log S_{22}(\tau).$$

Therefore eq. (6.5) can be replaced by the fact that α is also a transmission zero of S_{22}, namely $|S_{22}(\alpha)| = 1$. Let us now redefine as well the set of rational Schur function $\mathbb{\Sigma}_R^N$ that are feasible for the load L. Note the restriction of S_{22} to the set $\mathbb{\Sigma}_R^N$ allows us to easily impose the transmission zeros of the system S by the choice of the positive polynomial $R \in \mathbb{P}_+^{2N}$. Then we can now drop eq. (6.5) with the assumption that $R(\alpha) = 0$ obtaining the following definition for the set of rational functions $\mathbb{G}_R^N$.

Definition 6.3.1 (Feasible rational functions)**.** *We denote by* $\mathbb{G}_R^N$ *the set of rational functions* $S_{22} \in \mathbb{\Sigma}_R^N$ *satisfying eq.* (6.3)

$$\mathbb{G}_R^N = \left\{ S_{22} \in \mathbb{\Sigma}_R^N, |S_{22}(\alpha)| = 1 : \left[j\frac{d}{d\tau} \log S_{22}(\tau) \right]_{\tau=\alpha} \leq \left[j\frac{d}{d\tau} \log L_{22}(\tau) \right]_{\tau=\alpha} \right\},$$

where $R \in \mathbb{P}_+^{2N}$ *satisfies* $R(\alpha) = 0$*.*

Next, we introduce a special version of problem 4.6.1 where the transmission zero α is on the real axis. The difference from the original formulation of problem 4.6.1 comes from the fact that the redefined feasible set $\mathbb{G}_R^N$ in definition 6.2.1 is used. Similarly to the previous case, we denote by $\mathbb{I}$ the set in which the reflection level is minimized (passband) and by $\mathbb{J}$ the set where rejection constraints are imposed (stop-band).

Problem 6.3.2 (Matching problem with transmission zeros on the boundary.)**.**

$$\begin{aligned} &\textit{Find:} && l_{opt} = \min_{S_{22}\in\mathbb{G}_R^N} \max_{\tau\in\mathbb{I}} |S_{22}(\tau)|, \\ &\textit{Such that:} && |S_{22}| \geq \gamma && \forall \lambda \in \mathbb{J}. \end{aligned} \tag{6.7}$$

The most important difference of the problem of matching with a zero of transmission in the real axis with respect to the original problem where all the zeros of transmission are inside the domain of analyticity is found in the fact that the relaxation of the set of feasible functions is not necessary. Indeed, as shown below, the set of feasible functions $\mathbb{G}_R^N$ allows to obtain a formulation of the previous problem in terms of a positive polynomial P only.

6.4 Characterisation of $\mathbb{G}_R^N$

In this section, as it has been done for the classical synthesis problem and for the matching problem with transmission zeros inside the analyticity domain, we express condition on the angular derivative of S_{22} in terms of the modulus square of the function S_{22} along the real line. First we express from eq. (A.2)

$$\operatorname{ang} S_{22}(\alpha_i) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\log|S_{22}^O(\tau)|^2}{(\tau-\alpha_i)^2} d\tau + 2\sum_{n=1}^{N} \Im\left(\frac{1}{\beta_n - \alpha_i}\right).$$

We can now state the characterisation of the functions $S_{22} \in \mathbb{G}_R^N$ as a function of the modulus square $|S_{22}(\tau)|^2$ for $\tau \in \mathbb{R}$ and the position of the hypothetical zeros of S_{22} inside the analyticity domain.

Definition 6.4.1 (Characterisation of $\mathbb{G}_R^N$)**.** *Given the load $L(\tau)$ introduced in this section with a transmission zero $\alpha \in \mathbb{R}$, and let $R \in \mathbb{P}_+^{2N}$ such that $R(\alpha) = 0$. We have $S_{22} \in \mathbb{G}_R^N$ if and only if $S_{22}(\omega)$ is well defined at the point α and additionally*

$$\frac{1}{2\pi}\int_{\mathbb{R}} \frac{\log\left(|S_{22}(\tau)|^{-2}\right)}{(x-\alpha)^2} dx + 2\sum_k \frac{\Im\overline{\tau_k}}{|\alpha-\tau_k|^2} \leq \left[j\frac{d}{dx}\log L_{22}(\tau)\right]_{\tau=\alpha},$$

where the values τ_k correspond to the zeros of $S_{22}(\tau)$ inside the analyticity domain $\mathbb{C}^-$.

This characterization of the set of feasible functions $\mathbb{G}_R^N$ provides us with the necessary tool to derive the property we were looking for at the beginning of this section, which motivates the fact of considering this particular case of the matching problem separately.

6.5 Minimum phase property

We provide next the theorem stating that the optimal function S_{22} solution to problem 6.3.2 is of minimal phase. It is important to remember that in previous sections

when the transmission zeros of the load were inside the analytic domain we have obtained similar results with respect to the minimum phase property of the function S_{22} under the assumption that the condition of Selectivity was not active. However, it is impossible to determine a priori whether the aforementioned selectivity condition will be active or not without having to solve the problem.

In this case, the result obtained does not depend on the selectivity condition. The function S_{22} solution to problem 6.3.2 will be of minimum phase regardless of whether the selectivity condition is saturated or not, obtaining a true characterization of the optimal solution.

Lemma 6.5.1 (Minimum phase of p^{opt})**.** *Denote by $S_{22}^{opt} \in \mathbb{G}_R^N$ the optimal solution to problem 6.3.2. This solution is of minimum phase, therefore it can be written in the form*

$$S_{22}^{opt} = \frac{p_{opt}}{q(p_{opt})},$$

where the polynomial p has no roots in $\mathbb{C}^-$ and $q(p_{opt})$ is the stable polynomial satisfying $qq^\star = p_{opt}p_{opt}^\star + R$.

Proof. To prove this statement assume $p(\tau_k) = 0$ for some $\tau_k \in \mathbb{C}^-$. Then we have

$$2\sum_k \frac{\Im\overline{\tau_k}}{|\tau - \tau_k|^2} = C > 0.$$

Now multiply p by the inverse Blaschke $b^{-1}(\tau)$ such that the roots of p in $\mathbb{C}^-$ are flipped to $\mathbb{C}^+$. Note that $b^\star = b^{-1}$, therefore the function $pp^\star$ is not modified

$$\hat{p}\hat{p}^\star = (p \cdot b^{-1})(p \cdot b^{-1})^\star = (p \cdot b^{-1})(p^\star \cdot b) = pp^\star.$$

Therefore we have $|S_{22}^{opt}(\tau)| = |\hat{S}_{22}(\tau)|$ for all $\tau \in \mathbb{R}$ where

$$\hat{S}_{22} = \frac{\hat{p}}{q(\hat{p})}.$$

We consider now three possible scenarios, note that when the solution S_{22}^{opt} is obtained, at least one of the constraints in problem 6.3.2 must be binding, either eq. (6.7) or eq. (6.6)

1. Only eq. (6.6) is binding. In this case the function $\hat{S}_{22}$ provides the same criterium as S_{22}^{opt} and none of the constraints are saturated

$$\begin{aligned} \hat{S}_{22} &> \gamma && \forall \tau \in \mathbb{J}, \\ \left[j\frac{d}{d\tau} \log S_{22}(\tau) \right]_{\tau=\alpha} &\leq K_1 - C. \end{aligned}$$

   Thus we can multiply $\hat{S}_{22}$ by a positive constant improving the criterium in problem 6.3.2.

2. Only eq. (6.7) is binding. In this case the solution is only constrained by the selectivity requirements. Therefore the solution is the well-known quasi-elliptic response and S_{22}^{opt} has all roots in $\mathbb{R}$.

3. Both eqs. (6.6) and (6.7) are binding. This is a particular case since $\hat{S}_{22}$ still saturates eq. (6.7). However if the value of the criterium provided by $\hat{S}_{22}$, where only eq. (6.7) is active, can not be improved, then we are in the previous case where the solution is the quasi-elliptic response. On the other hand, if the criterium can be improved, we have a better solution that contradicts the assumption of the optimality of S_{22}^{opt}. □

Lemma 6.5.1 allows to restrict the class of polynomial on which the solution of problem 6.3.2 is sought, to the minimum phase functions $S_{22} \in \mathbb{G}_R^N$, were a limiting case of S_{22}^{opt} having only real roots may occur if the quasi-elliptic solution is obtained.

6.6 An exact convex relaxation

Problem 6.3.2 is not convex since the spectral factorization of the positive polynomial $qq^\star$ is needed. Nevertheless, in a similar form as it is done for the transmission zeros inside the analyticity domain in the previous section by using the *Poisson integral*, this time we can use *Hilbert Transform* to express eq. (6.6) as a function of the modulus $|S_{22}|^2$

$$|S_{22}(\tau)|^2 = \frac{p(\tau)p^\star(\tau)}{p(\tau)p^\star(\tau) + R(\tau)} \qquad \forall \tau \in \mathbb{R}$$

Furthermore, due to the minimum phase property of S_{22}, we can apply the change of variable $P(\tau) = p(\tau)p^\star(\tau)$ where $P \in \mathbb{P}_+^{2N}$. Therefore problem 6.3.2 can then be restated as a function of the positive polynomial P as:

Problem 6.6.1 (Convex matching problem with boundary tranmission zeros.)**.**

$$\begin{aligned} \textit{Find:} \qquad & L_{opt} = \min_{P \in \mathbb{P}_+^{2N}} \max_{\tau \in \mathbb{I}} \frac{P(\tau)}{R(\tau)}, \\ \textit{Subject to:} \qquad & P(\tau) \geq \Gamma \cdot R(\tau) \qquad \tau \in \mathbb{J}, \\ & f(P) \leq K_1. \end{aligned} \tag{6.8}$$

where the function $f(P)$ is now defined as

$$f(P) = \int_{\mathbb{R}} \log\left(1 + \frac{R(x)}{P(x)}\right)(x-\alpha)^{-2}\, dx, \tag{6.9}$$

and K_1 is computed as

$$K_1 = 2\pi j \left[\frac{d}{d\tau} \log L_{22}(\tau)\right]_{\tau=\alpha}.$$

Remark 6.6.1. *Note that in problem 6.6.1 the condition $P(\alpha) \neq 0$ is also required to ensure the feasibility of the obtained reflection coefficient S_{22}. Nevertheless it is important to remark that if the polynomial P in problem 6.6.1 tends to a polynomial P_S such that $P_S(\alpha) = 0$ then the integral in eq.* (6.9) *tends to infinity which implies that such polynomial P is not admissible for a finite value of K_1. Therefore we can relax this additional constraint.*

We obtain an alternative formulation of problem 4.6.1 allowing for the $\alpha \in \mathbb{R}$. Furthermore, theorems 4.3.2, 4.4.1 and 4.5.1, lemma 4.4.1, and corollary 4.3.1 still holds.

6.6.1 Different kinds of solutions

Note that, in problem 6.6.1, two different kinds of solutions can be distinguished depending on which constraints are active:

- *Quasi-elliptic* response: if eq. (6.8) is not active, the solution P_{opt} solves a classical filter synthesis problem. If p_{opt} is defined as $p_{opt}p^*_{opt} = P_{opt}$ then p_{opt} is the extended-Tchebychev polynomial of the interval $\mathbb{I}$.
- *Minimum-area responses*: if eq. (6.8) is active, the solution realizes the best possible matching level under the specified selectivity requirement.

6.7 Transmission zeros at infinity

As a particular case, note that if the transmission zero α occurs at the origin, the previous formulation is still valid. Introducing $\alpha = 0$ in eq. (6.9) we obtain

$$f(P) = \frac{1}{2\pi}\int_{\mathbb{R}} \log\left(|S_{22}(\tau)|^2\right)\tau^{-2}d\tau.$$

If we now undo the change of variable performed at the beginning of the section we have $\tau = \lambda^{-1}$ and $d\tau = -\lambda^{-2}d\lambda$. Therefore eq. (6.9) becomes

$$f(P) = \frac{1}{2\pi}\int_{\mathbb{R}} \log\left(|\tilde{S}_{22}(\lambda)|^2\right)d\lambda,$$

representing the surface covered by the function $\log|\tilde{S}_{22}(\lambda)|^2$ where $\tilde{S}_{22}(\lambda) = S_{22}(\lambda^{-1})$. We obtain then a particular version of the classical synthesis problem where the maximum area under the magnitude of the reflection coefficient of the global system (in a logarithmic scale) is constrained.

.

This allows to easily derive Fano bounds introduced in [34] for a load of degree 1. If we consider, for instance, an interval $\mathbb{I} = [-1, 1]$, the optimal function $\tilde{S}_{22}$ is a function that *employees* all the available area on the interval $\mathbb{I}$ with a constant reflection level on that interval and zero everywhere else

$$|\tilde{S}_{22}(\tau)|^2 = \begin{cases} L_{opt} & \tau \in \mathbb{I} \\ 0 & \tau \notin \mathbb{I} \end{cases}.$$

Computing now the value L_{opt} such that eq. (6.9) is maximum

$$\begin{aligned} f(P) &= -\frac{1}{2\pi}\int_{\mathbb{R}} \log\left(|\tilde{S}_{22}(\lambda)|^2\right)d\lambda = K_1 \\ &= -\frac{1}{2\pi}\int_{-1}^{1} \log\left(L_{opt}\right)d\lambda = K_1 \\ &= -\frac{1}{\pi}\log\left(L_{opt}\right) = K_1. \end{aligned}$$

Thus

$$L_{opt} = e^{-\pi K_1}.$$

We obtain the bound known in the literature of the matching problem and attainable with a function S_{22} of infinite degree.

6.8 Sharpness of the provided bounds

We obtain in this chapter the optimal solution to problem 4.1.1, namely the provided bounds are sharp, in two different cases.

1. The case of a load of degree 1 with a transmission zero on the frequency axis (possible at infinity) with any possible constraint on the rejection with the stop-band $\mathbb{J}$ imposed by eq. (4.16).

2. The case of a load of degree 1 with a transmission zero inside the complex plane and no selectivity requirement.

Note that in practice, these results are suitable for many antennas since they often feature one single resonance, particularly small antennas.

6.9 Examples

We present now some practical results concerning the broadband matching of some antennas whose reflection coefficient is modelled as a rational function of degree odd. Furthermore, we illustrate the procedure to follow from the measurement of the load reflection to the physical design of the 3D filter structure which implements the optimal response in terms of matching. In the example provided in this chapter, the practical design of the filter is done via the classical coupling matrix approach [35] using the full wave simulation software *Ansoft Electronic Desktop*. The target coupling matrix (M_T) is obtained from the scattering matrix F as discussed in section 2.11. The algorithm to follow in the case of a load of degree odd remains

1. Computation of the Darlington equivalent of the load

   (a) Perform a rational approximation of the reflection coefficient L_{11} of the load within the passband of interest. In other words, find the polynomials p_L, q_L of degree one with q_L stable such that

   $$L_{11}(\omega) \approx \frac{p_L^\star(\omega)}{q_L(\omega)} \qquad \forall \omega \in \mathbb{I}.$$

   (b) Obtain the Darlington equivalent L as the two port extension of the function p_L/q_L, namely

   $$L(\omega) = \begin{pmatrix} p_L^\star(\omega) & -r^\star(\omega) \\ r(\omega) & p_L(\omega) \end{pmatrix}, \tag{6.10}$$

   such that $r_L r_L^\star = q_L q_L^\star - p_L p_L^\star$.

   (c) The single root of the positive polynomial $R_L = r_L r_L^\star$ in the lower half plane ($\mathbb{C}^-$) provides us with the point α_1 where the interpolation condition on the global system is imposed.

   $$S_{22}(\alpha_1) = L_{22}(\alpha_1) \qquad \alpha_1 \in \mathbb{C}^+ \qquad R_L(\alpha_1) = 0. \tag{6.11}$$

2. Design of the global system

   (a) Pick a degree N for the global system and fix the transmission polynomial of the matching filter $R_F \in \mathbb{P}_+^{2N-2}$. Once the polynomial R_F is fixed we have $R = R_L \cdot R_F$. Finally problem 6.6.1, possibly with some rejection constraints $P(\tau) \geq \Gamma R(\tau)$ if additional filtering constraints are required within a given stop-band $\mathbb{J}$.

   (b) Factorise the polynomial $P = pp^\star$ such that all roots in the upper half plane ($\mathbb{C}^+$) are assigned to p. In other words, p is a stable (Hurwitz) polynomial.

   (c) Ensure the interpolation condition in eq. (6.11). To satisfy this condition the function S_{22} is constructed as

   $$S_{22}(\omega) = \frac{\epsilon p(\omega)}{q(\omega)}$$

   with $|\epsilon| = 1$ and q the stable polynomial which satisfies $qq^\star + pp^\star = R$. Note that we also have $R(\alpha_1) = 0$, therefore since $\alpha_1 \in \overline{\mathbb{R}}$ we have $|S_{22}(\alpha_1)| = |L_{22}(\alpha_1)| = 1$. Thus the constant ϵ is computed such that eq. (6.11) is satisfied as

   $$\epsilon = \frac{q(\alpha_1)}{p(\alpha_1)} L_{22}(\alpha_1)$$

   (d) 2-port extension of the global system. Note that, similarly to what was done for the load, only the S_{22} coefficient of the global system has been synthesized. Nevertheless we can again use the Darlington equivalent to extend the reflection coefficient S_{22} to the 2×2 matrix S

   $$S = \frac{1}{q} \begin{pmatrix} p^\star & -\epsilon r^\star \\ r & \epsilon p \end{pmatrix}. \tag{6.12}$$

3. Computation of the matching filter

   (a) De-embedding of the load. Equation (3.13) allows us to obtain the rational model of the matching filter. Introducing eqs. (6.10) and (6.12) in eq. (3.13) we have

   $$F_{22} = \frac{\frac{p_L}{q_L} - \epsilon \frac{p}{q}}{\epsilon \frac{p_L p_L^\star + R_L}{q_L q_L} - \epsilon \frac{p p_L^\star}{q q_L}}.$$

   Using now the relation $q_L q_L^\star = p_L p_L^\star + r_L r_L^\star$ we obtain

   $$F_{22} = \frac{\overline{\epsilon} p_L q - p q_L}{q_L^\star q - p_L^\star p}.$$

   Therefore we have

   $$p_F = \overline{\epsilon} p_L q - p q_L,$$
   $$q_F = q_L^\star q - p_L^\star p.$$

   Finally we also perform the 2-port extension of the function F_{22} as

   $$F = \frac{1}{q_F} \begin{pmatrix} p_F^\star & -r_F^\star \\ r_F & p_F \end{pmatrix}.$$

(b) Extraction of the coupling matrix corresponding to the coupled resonator model of the matching filter following the procedure proposed in section 2.11;

(c) Design of the 3D structure of the coupled resonators matching filters with the aid of an electromagnetic simulation software.

6.10 Single band matching and example of matching filter design.

We solve first the matching problem considering a microstrip patch antenna for a *GNSS* (Global Navigation Satellite System) receiver. The specifications are the coverage of the band L_1 (from 1.55GHz to 1.6GHz). The reflection of the antenna to be matched along with the passband interval are shown in fig. 6.1.

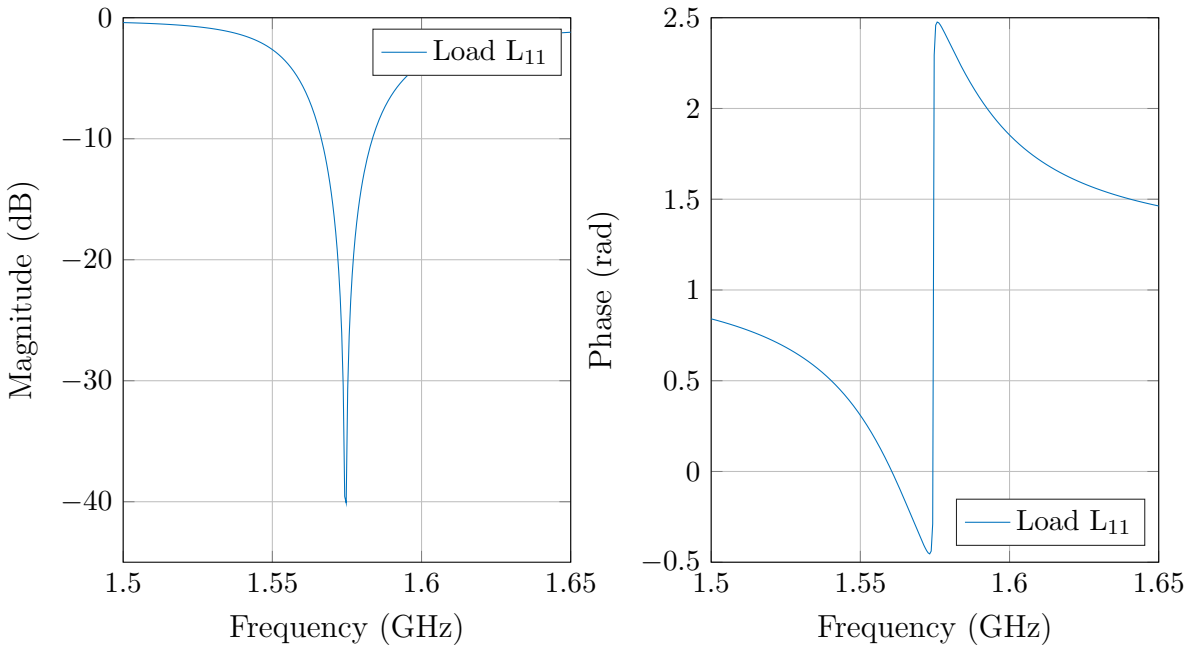


Figure 6.1: Load of degree 1

We fix a transmission polynomial for the global filter with no finite transmission zeros $R_F = 1$ and then solve optimaly the matching problem for a matching network of McMillan degree K from 1 to 13. We can see in fig. 6.2 the level of optimal matching, which coincides with the lower bound since the load is of degree 1, depending on the degree of the matching filter K. We can see how this limit tends to a value around -14 dB when the value of K tends to infinity.

As an example we have selected the value $K = 3$ to implement the filter that provides the optimal solution in terms of matching. First we show in fig. 6.3 the optimal reflection of the global system S_{22} with $K = 3$. We obtain an optimal reflection level $\psi_{opt} = -11.47dB$. In this example it is interesting to note, on the one hand, that the global system is of degree $N = 4$, namely the degree of the matching filter $K = 3$ plus the antenna degree $M = 1$; On the other hand, we can check

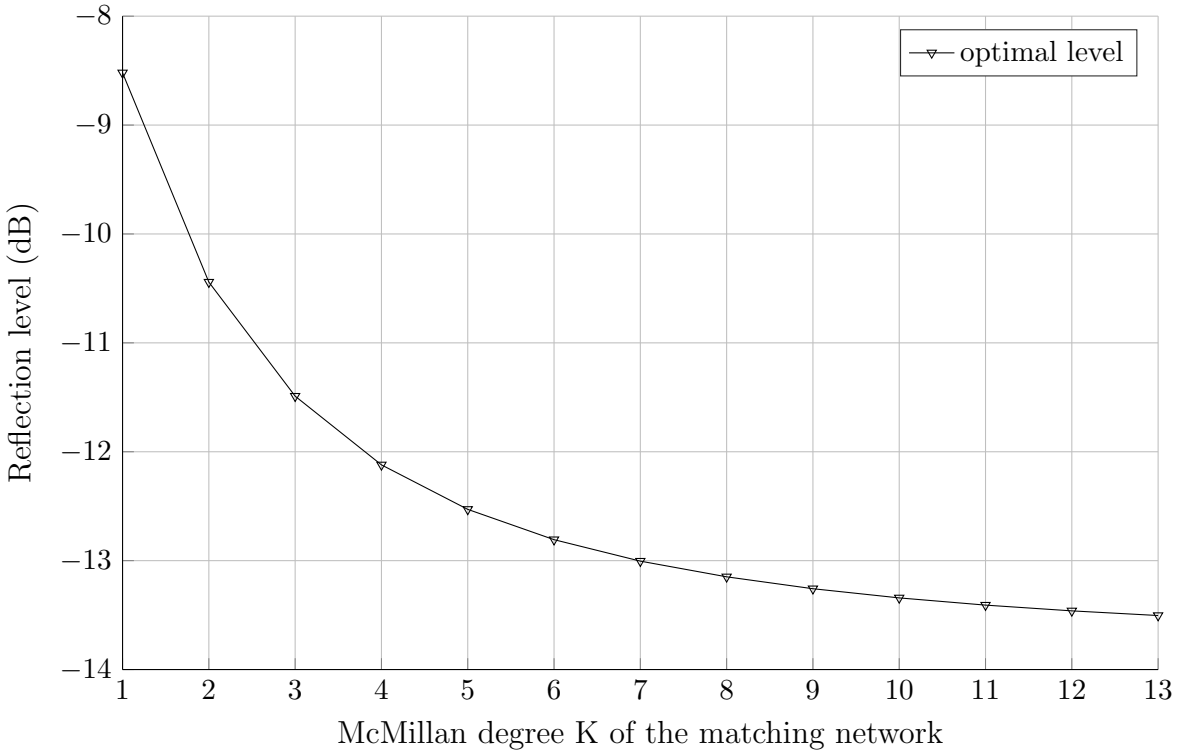


Figure 6.2: Lower bound for the best matching level in band GPS $L1$

the optimality condition obtained in the previous chapter in terms of the number of times that the optimum matching level is reached within the band, in particular $N+1 = 5$.

We provide, with an illustrating purpose, the location of the roots and poles of the relevant functions in the complex plane. We indicate in fig. 6.4a the poles and zeros of the function S_{22} which is plotted in fig. 6.3. Remember that the degree of the global system is obtained as $N = K + 1 = 4$. It should be noted how all zeros of both p and q belong to $\mathbb{C}^+$, namely both are Hurwitz polynomials. This fact indicates that the function S_{22} is indeed of minimum phase as stated by lemma 6.5.1.

After de-embedding the Darlington equivalent of the load from the port 2 of the global system, the rational function $F_{22} = \frac{p_F}{q_F}$ is obtained. Note that we have also indicated in fig. 6.4b the poles and zeros of the function F_{22}. Note that after the de-embedding of the load, a drop in degree is produced, leading to the function F_{22} of degree $K = 3$. This rational function F_{22} allow us to compute the *Belevitch* model of the matching filter. We also plot in fig. 6.3 the scattering parameters of the matching filter, namely the functions F_{22} and F_{21} that provide the aforementioned global reflection.

6.10.1 Effect of the filter transmission zeros

It can be remarked that the algorithm proposed in this thesis does not tells anything about the optimal choice of the filter transmission polynomial R_F. Indeed, in this work we assume that the function F_{22} belong to the set $\mathbb{\Sigma}_{R_F}^K$, what implies that the transmission

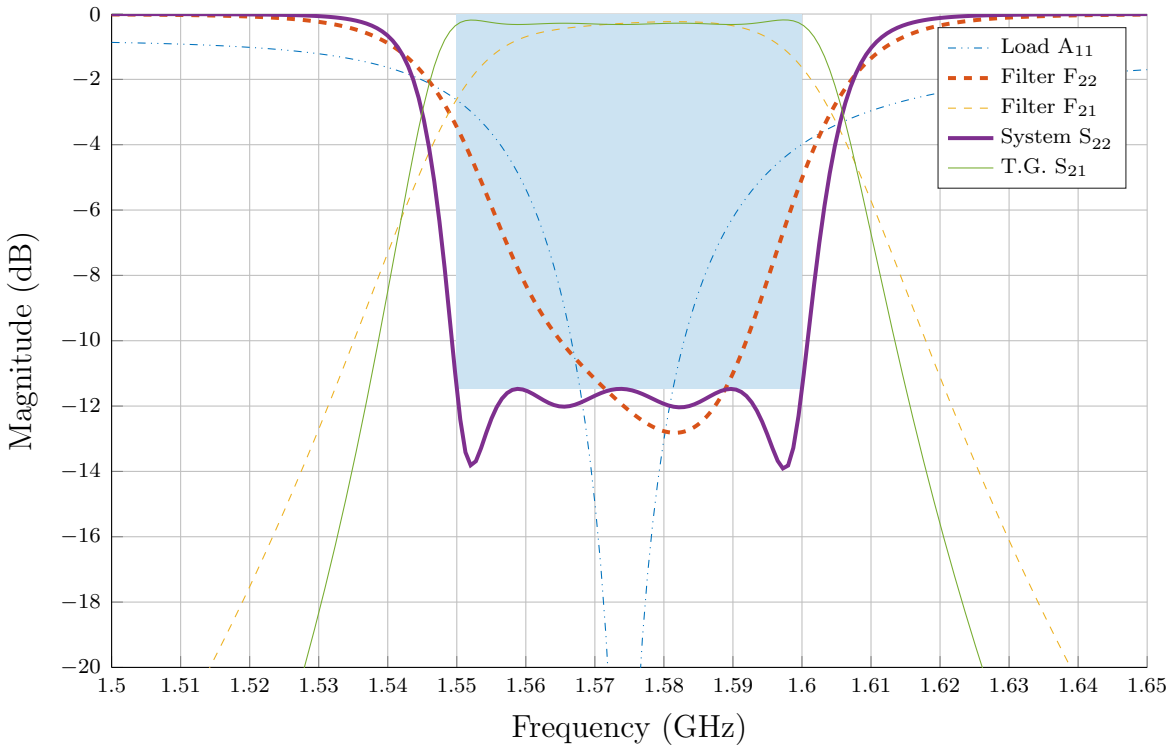
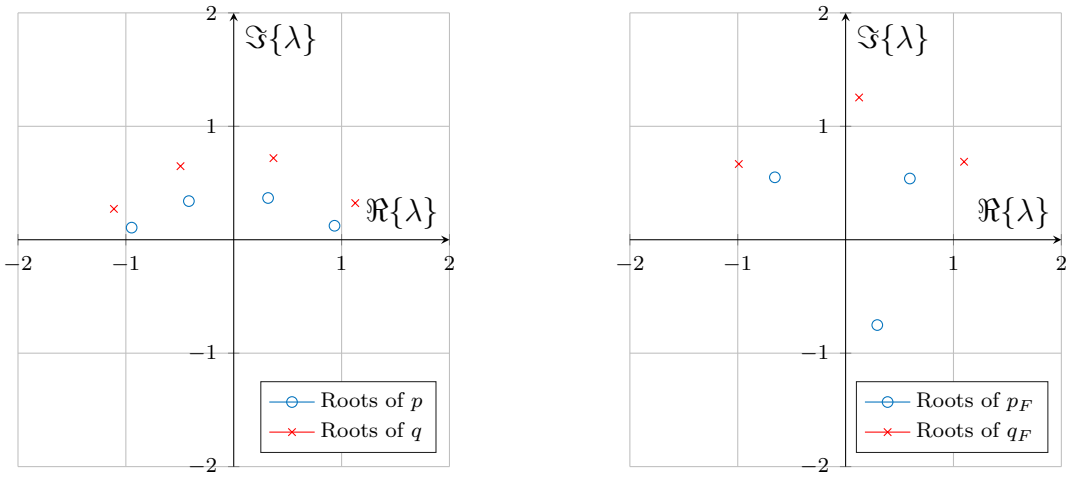


Figure 6.3: Matching result with a filter of degree $K = 3$ in band GPS $L1$



(a) Global system's $S_{22} = \epsilon\frac{p}{q}$

(b) Matching filter's $F_{22} = \frac{p_F}{q_F}$

Figure 6.4: Location of poles and zeros in the complex plane.

polynomial R_F is fixed and not part of the optimisation. Therefore the task of deciding which is the right choice for the said transmission polynomial R_F falls on the user.

Remark 6.10.1. *Note, in fact, that only the roots of the polynomial $R_F = r_F r_F^\star$ need to be fixed by the user. Note that, since the leading terms of the polynomial p are q are computed in the optimisation procedure, adding a multiplicative constant to the polynomial R_F has no effect as the algorithm would provide the polynomials p, q multiplied by the corresponding factor. Therefore the multiplicative constant added to the polynomial R_F is simplified.*

Fixing the transmission zeros of the filter is not new in filter design since in most of the classic design techniques which are commonly used in this field, it is the designer who is in charge of selecting the appropriate position for the filter transmission zeros. In the previous example, we have choose to place all transmission zeros of the filter at infinity, therefore we have fixed $R_F = 1$. Nevertheless we should also investigate the disadvantages or benefits obtained in terms of matching when those transmission zeros are set at different positions.

1. Position A. We consider first the case of one single transmission zero positioned next to the passband at the frequency of $1.54GHz$, namely $r_F^A = \lambda - 1.54 \cdot 10^9$. With this choice for the polynomial r_F we obtain the result shown in fig. 6.5 where the transmission and reflection of both the global system and the matching filter are provided. It should be noted that in this case the optimal reflection level is improved reaching $\psi_{opt}^A = -11.76dB$.

2. Position B. Next we investigate the effect of shifting this transmission zero toward the interior of the complex plane. This time we take $r_F^B = \lambda - (1.54 + 0.001j) \cdot 10^9$. Note that this polynomial r_F implies a non reciprocal matching filter as $r_F \neq r_F^\star$ which provides us with the response in fig. 6.6. The optimal reflection level in this case is $\psi_{opt}^B = -11.63dB$.

3. Position C. Finally we provide a third example illustrated in fig. 6.6 where two transmission zeros have been symmetrically located at both edges of the passband at the frequencies of $1.54GHz$ and $1.61GHz$. Therefore we have $r_F^C = (1 - 1.54 \cdot 10^9)(1 - 1.61 \cdot 10^9)$. In this case we obtain the best reflection level among the three cases considered, namely $\psi_{opt}^C = -12.15dB$.

We summarise in table 6.1 the optimal matching level ψ_{opt} obtained with each of the choices made in this section for the position of the filter transmission zeros. It should be noted that the original result obtained with $r_F = 1$, namely $\psi_{opt} = -11.47dB$ is only slightly improved with the presence of transmission zeros at finite frequencies.

It can be noted that get the most benefit in terms of matching when a transmission zero is placed at each side of the passband. This case coincides with the global response providing the strongest rejection at both edges of the passband due to the presence of such transmission zeros. We are therefore encouraged to explore further this configuration

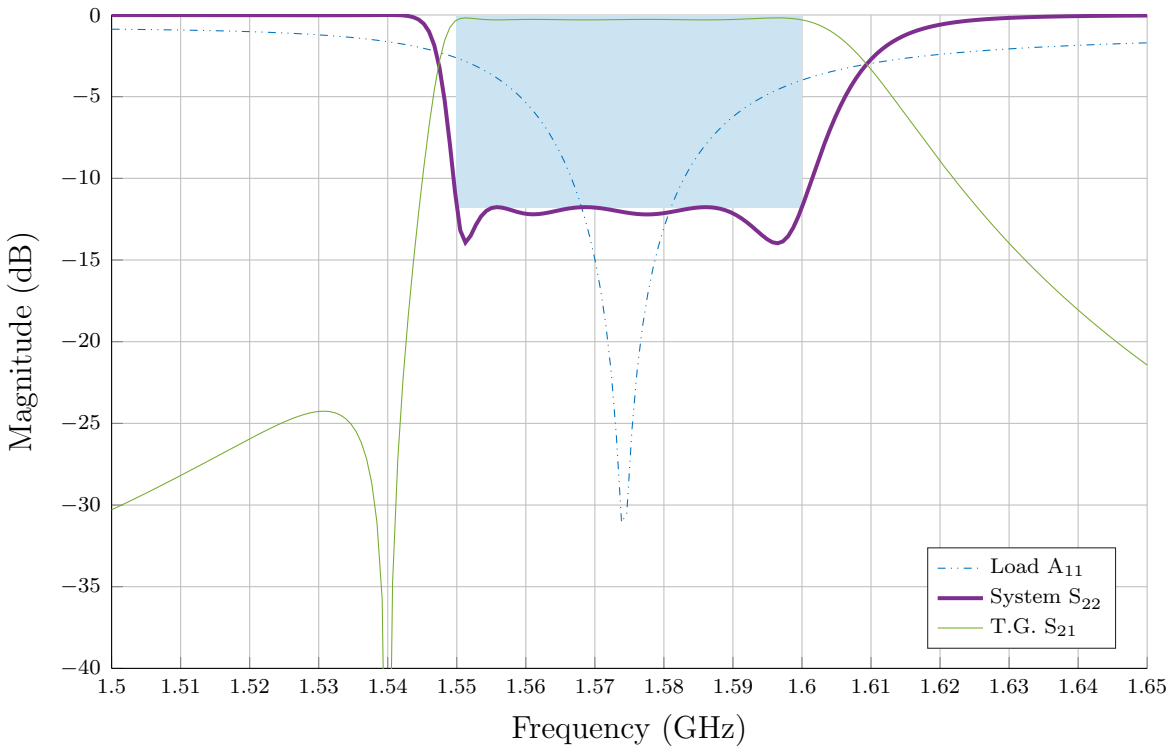


Figure 6.5: Matching result in case A by placing a transmission zero next to the passband obtaining a reflection level $\psi_{opt}^{A} = -11.76dB$.

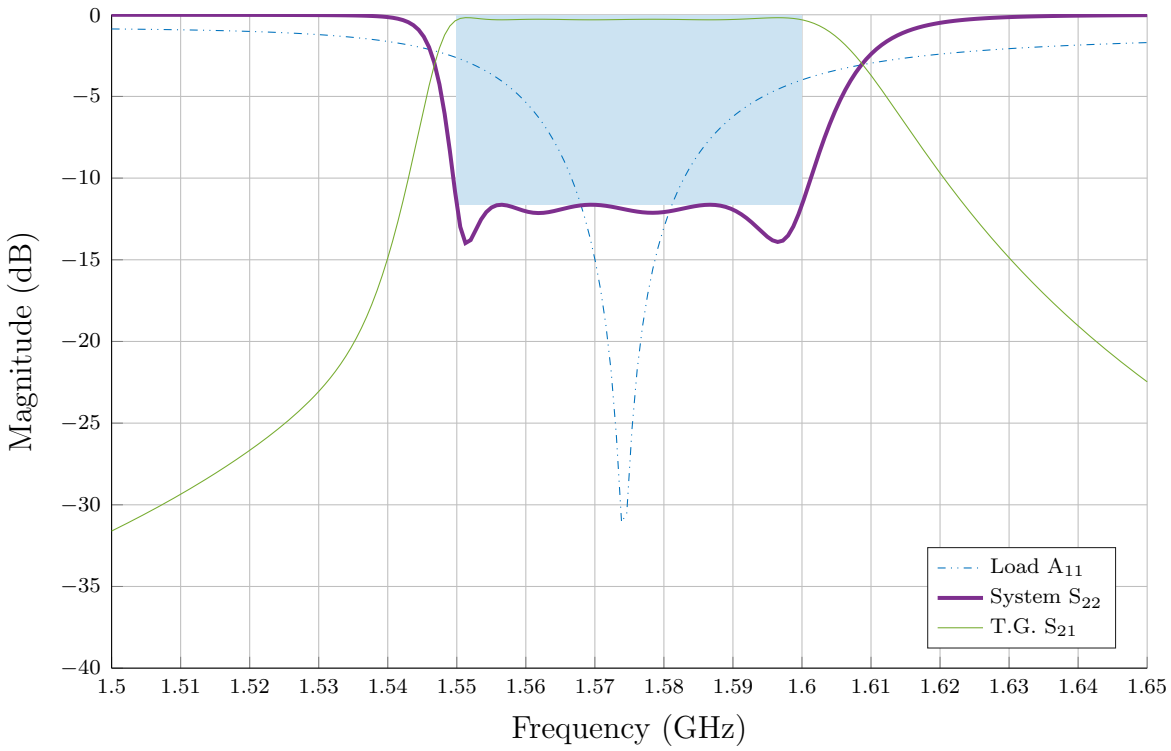


Figure 6.6: Matching result in case A by placing a transmission zero in the complex plane reaching a reflection level $\psi_{opt}^{B} = -11.63dB$.

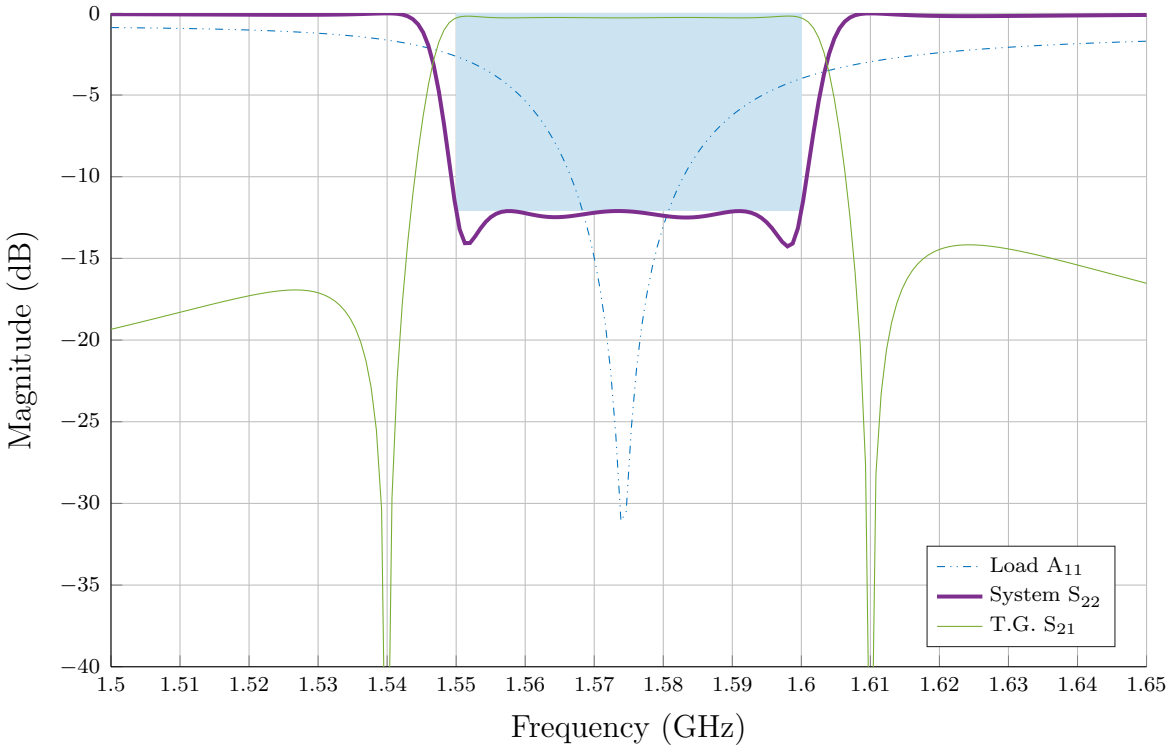


Figure 6.7: Matching result in case A by placing a pair of transmission zeros, one at each side of the passband. The optimal reflection level is $\psi^B_{opt} = -12.15dB$.

Case	Transmission zeros position	Optimal matching level
Ref.	No finite t.z.	$\psi_{opt} = -11.47dB$
A	$\sigma_1 = 1.54 \cdot 10^9$	$\psi^A_{opt} = -11.76dB$
B	$\sigma_1 = (1.54 + 0.001j) \cdot 10^9$	$\psi^B_{opt} = -11.63dB$
C	$\sigma_1 = 1.54 \cdot 10^9;\ \sigma_2 = 1.61 \cdot 10^9$	$\psi^C_{opt} = -12.15dB$

Table 6.1: Matching result obtained in each of the cases of study provided above

with a matching filter having two finite transmission zeros. We choose two transmission zeros σ_1, σ_2 symmetrically located around the passband centre $f_C = 1.575GHz$ such that

$$\sigma_1 = f_C - \delta,$$
$$\sigma_2 = f_C + \delta.$$

Let now compute, for different values of δ, the optimal reflection level $\psi_{opt}(\delta)$. In fig. 6.8 we show the transducer gain, namely the parameter S_{21} of the global system for each value of δ. We can see the displacement along the frequency axis of the transmission zeros along with the different out-of-band rejection.

Furthermore we provide in fig. 6.9 the optimal global system reflection S_{22} which is computed by the algorithm proposed in this work. We verify that the displacement along the axis of the transmission zeros has indeed an influence on the optimal matching. The chosen values of δ along with the correspondent value of $\psi_{opt}(\delta)$ are listed in table 6.2 and plotted in fig. 6.10. We can see in fig. 6.10 that the minimum reflection level is obtained around $\delta = 32.5\ MHz$ while as long as the transmission zeros are shifted toward infinity, this optimal level approaches the value $\psi_{opt} = -11.47\ dB$ obtained in the reference case with no finite transmission zeros. We obtain therefore a minimum value around $-12.21\ dB$ when two transmission zeros are placed at the frequencies

$$\sigma_1 = 1.575GHz - 32.5MHz = 1.5425GHz,$$
$$\sigma_2 = 1.575GHz + 32.5MHz = 1.6075GHz.$$

Remark 6.10.2. *It must also be noted that, although the optimal reflection level does depends on the position of the transmission zeros, the variation of this level in fig. 6.10 occurs within a range of less than* $1\ dB$*. This result helps to justify fixing the transmission zeros of the matching filter in the matching problem under the reasoning that the obtained optimal reflection level would not vary much if the polynomial* r_F *is modified.*

Nevertheless, the solution to the matching problem where the transmission zeros of the matching filter are optimally positioned is still unknown. Note that this solution corresponds to problem 3.2.1 when the function F_{22} *belong to the class of rational Schur function of degree* K *(*$\mathbb{\Sigma}^K$*). This problem is stated as*

$$\textit{Find:} \qquad \min_{F_{22}\in\mathbb{\Sigma}^K} \max_{\omega\in\mathbb{I}} |F_{22} \circ L_{11}(\omega)|.$$

6.10.2 SIW filter

We have performed in section 6.10 a study of the theoretical limitations regarding the matching of the load with the reflection shown in fig. 6.1 within the GPS band L_1. Before concluding the study of the previous load, we shall offer the reader an example of matching filter synthesis with the aid of the information provided the proposed algorithm.

We consider again the solution to the matching problem where the matching filter is set to have no finite transmission zeros, namely $R_F = 1$. This solution has already been illustrated in fig. 6.3 where the optimal response for the global system is plotted along

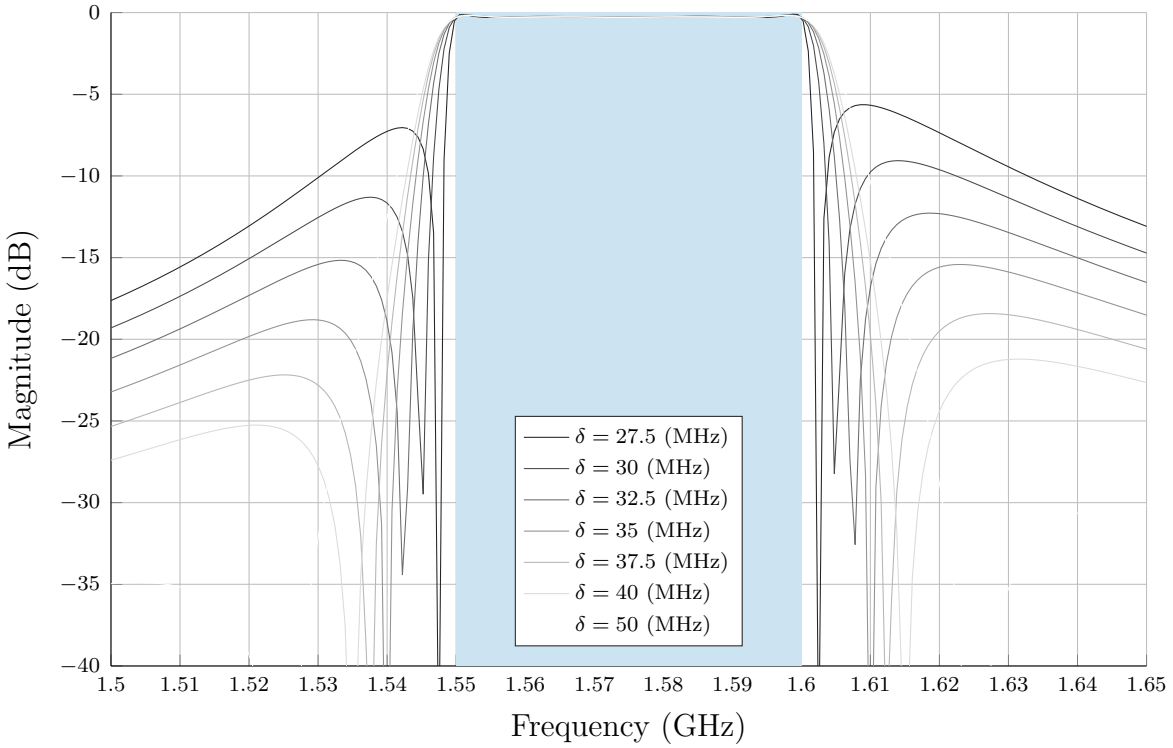


Figure 6.8: Global transducer gain (S_{21}). Different position of the transmission zeros along the frequency axis according to the value of δ.

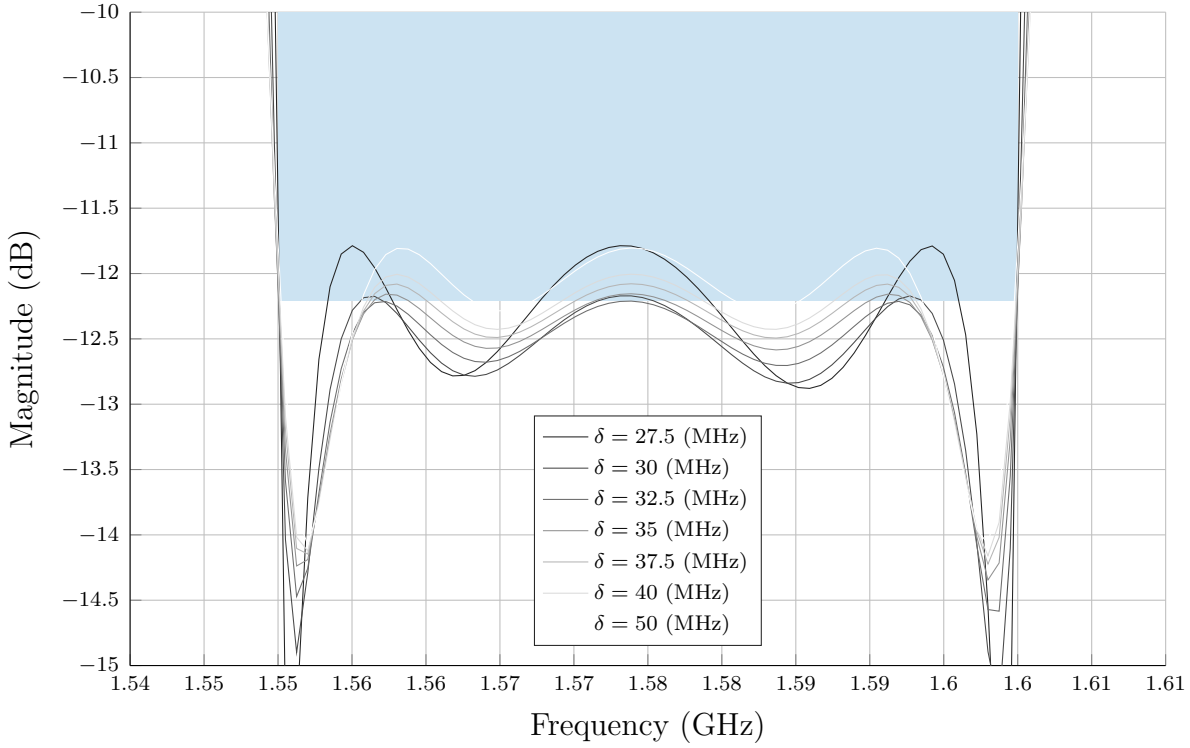


Figure 6.9: Optimal global reflection (S_{22}) attainable with each value of δ.

δ (MHz)	$\psi_{opt}(\delta)$ (dB)
27.5	-11.7836
30	-12.1663
32.5	-12.2071
35	-12.1513
37.5	-12.0747
40	-12.0015
50	-11.8006
60	-11.6967
70	-11.6371
80	-11.5999
90	-11.5749
100	-11.5574

Table 6.2: Values of $\psi_{opt}(\delta)$

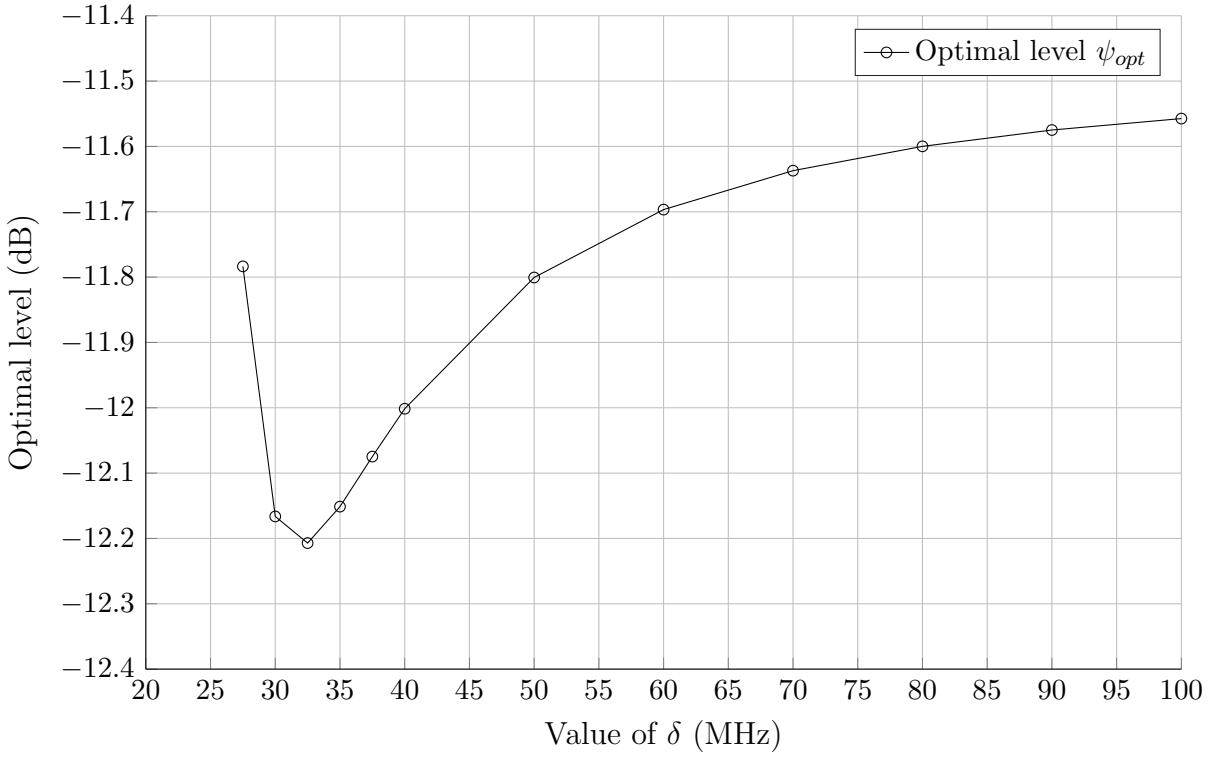


Figure 6.10: Optimal reflection level as a function of the displacement of the transmission zeros (δ).

with the scattering parameters of the matching filter. In this example, since we have $R_F = 1$, the matching filter has no additional transmission zeros and can be implemented by a coupled resonators network with an in-line topology.

The term *in-line topology* refers to a coupled resonator network as depicted in fig. 2.14 where only the coupling elements $M_{S,1}$, $M_{i,i+1}$ and $M_{K,L}$ with $i \in [1, K-1]$ are non-zero. Note that in this case the matching filter is of degree $K = 3$, therefore the filter response can be implemented with a network of 3 resonators, each of them coupled to the previous and the following one while the first and last are coupled to the input and output ports respectively as represented in fig. 6.11.

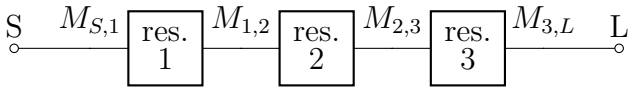


Figure 6.11: Structure of the in-line coupled resonators filter of order 3. Boxes represent microwave resonators while the connecting lines denotes inter-resonators couplings.

This filter is realized in *SIW* (Substrate Integrated Waveguide) planar technology fed with *CPWG* (Co-Planar WaveGuide) input and output lines as illustrated in fig. 6.12 by using the substrate *Rogers RT/duroid 6010LM*. The *SIW* structure consists on a dielectric substrate with a bottom and top copper layers where resonant cavities are formed with the aid of metallic post walls interconnecting the upper and lower metal surfaces. The inter resonator couplings are implemented by properly sized openings in the resonator walls which connect the adjacent cavities. We can also notice the input/output feeding lines which provides a tunable input/output coupling by varying the penetration inside the resonator.

The electromagnetic (EM) response provided by the structure in fig. 6.12 is computed with the aid of the EM-simulation software *Ansys HFSS* [36]. In order to fit the EM response to the desired transfer function, the width and length of the resonators, as

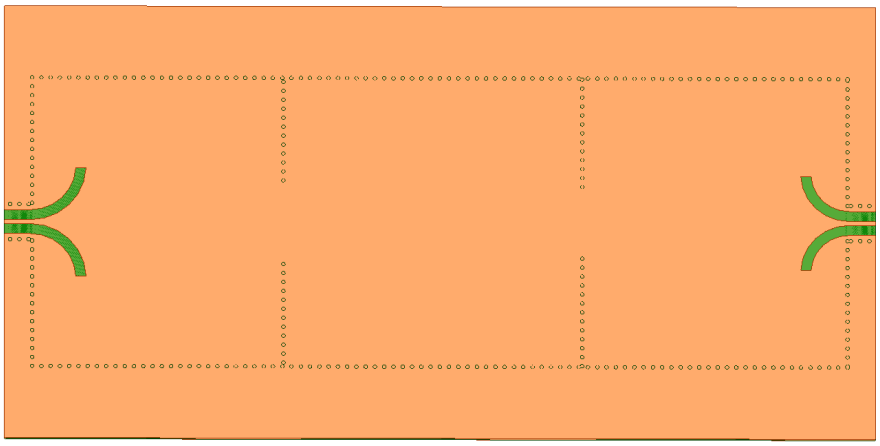


Figure 6.12: Designed SIW filter.

well as the size of the inter-resonator windows and input-output feeding lines are adjusted.

In this example we obtain the target coupling matrix M_T after performing the steps from 1 to 3(b) in section 6.9

$$M_T = \begin{bmatrix} 0 & 1.195 & 0 & 0 & 0 \\ 1.195 & 0 & 1.018 & 0 & 0 \\ 0 & 1.018 & -0.007 & 0.7 & 0 \\ 0 & 0 & 0.7 & -0.404 & 1.009 \\ 0 & 0 & 0 & 1.009 & 0 \end{bmatrix}.$$

However before performing the design of the 3D structure, it should be noted that this kind of filtering functions differs from the classical *Tchebyshev* responses in the sense that they do not present all reflection zeros distributed on the frequency axis but inside the complex plane. For this reason and in order to achieve a good agreement between the circuital response and the EM response, the design has been assisted with the circuit - extraction software *PRESTO-HF* which can be consulted in [37]. This software is used to perform the rational approximation of the EM-simulated response F^{EM} provided by the structure in fig. 6.12 such that

$$F^{EM} \approx \begin{pmatrix} \varepsilon p^{\star}_{EM} & -\varepsilon r^{\star}_{EM} \\ r_{EM} & p_{EM} \end{pmatrix}.$$

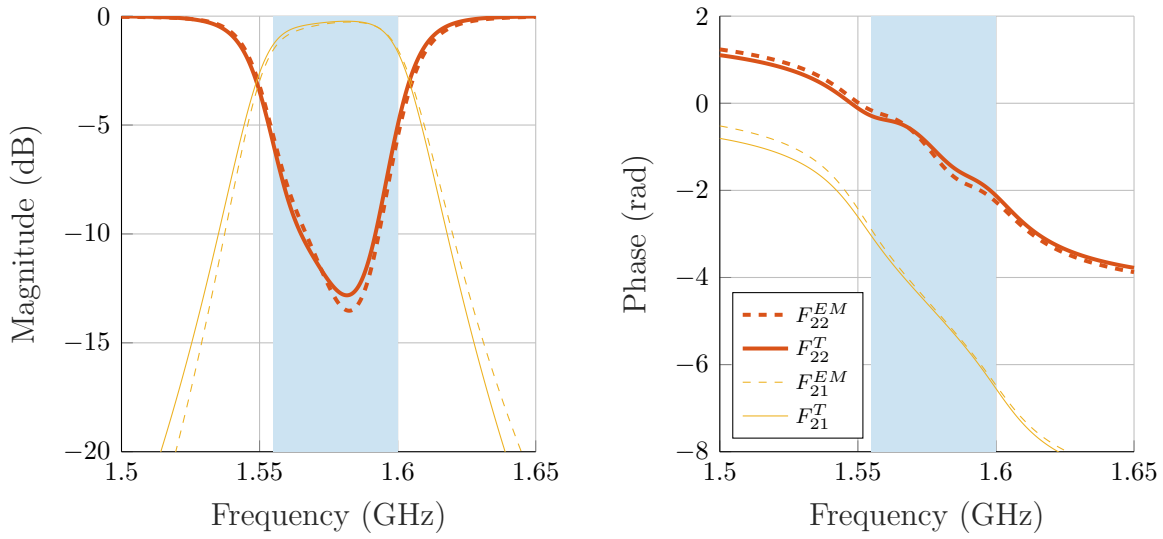
This rational matrix allows us to compute a coupling matrix M_{EM} representing the simulated response.

Remark 6.10.3. *As discussed in remark 2.8.7, it is important to remember that the matrix M_{EM} associated to a given rational response is not unique. This not uniqueness of the coupling matrix is an issue which has already been rigorously studied in the literature of filter synthesis, for instance in [35]. In our case, we seek a matrix M_{EM} which is equivalent to the target coupling matrix M_{EM}, or with the nomenclature commonly used in the field of filter design, we choose the coupling matrix M_{EM} with the same topology as the matrix M_T.*

In order to optimise the 3D structure of the filter shown in fig. 6.12, we use a heuristic procedure that compares the target coupling matrix M_T to the matrix M_{EM} extracted from the EM response adjusting the physical dimensions in consequence. The goal is to minimise the Frobenius norm $\|E\|$ where the error matrix E is computed as $E = M_T - M_{EM}$. Once the optimisation procedure concludes, we obtain the error

$$E = \begin{bmatrix} 0 & -3.2 & 0 & 0.5 & 0 \\ -3.2 & -0.4 & 0.3 & -0.3 & 0.5 \\ 0 & 0.3 & -1.1 & 0.8 & 0 \\ 0.5 & -0.3 & 0.8 & -0.6 & -2.1 \\ 0 & 0.5 & 0 & -2.1 & 0 \end{bmatrix} \cdot 10^{-2}.$$

It is important to remark that, in contrast with the traditional filter synthesis, synthesising the right phase of F_{22} is a crucial point here in order to obtain a filter that is matched at port 2 to the antenna. Nevertheless as the coupling matrix corresponding



(a) Magnitude (b) Phase

Figure 6.13: EM simulation of the matching filter and target response.

to the filter response has already been adjusted to fit the target matrix M_T, the phase of the EM simulated F_{22}^{EM} would only differ from the target F_{22}^{T} by a constant factor. This factor, namely a constant phase shift, is approximated within the passband by a transmission line. Therefore, a transmission line of $10.5mm$ has been required to adjust the phase of $S22$.

Figure 6.13 shows the comparison between the S-parameters of the filter, obtained in one case from the EM simulation (F_{22}^{EM} and F_{21}^{EM}) and in the other case from the circuital analogue (F_{22}^{T} and F_{21}^{T}). Moreover, the line S_{22}^{EM} in fig. 6.14 represents the input reflection obtained when the designed SIW filter is connected to the antenna. Note that an excellent match between the reflection parameters S_{22}^{EM} and S_{22}^{T} is obtained, validating the employed synthesis technique for matching filters.

Remark 6.10.4. *It should be remarked that the present example constitutes an academic exercise which is meant to confirm the benefits of using a matching filter with this kind of loads regarding the input reflection coefficient of the system. For this reason, although the simulated SIW filter introduces a certain level of losses in the system, dissipation losses have not been considered during the optimisation of the SIW structure, obtaining the lossless response shown in fig. 6.13. Similarly the result in fig. 6.14 is obtained when the matching filter is assumed to be lossless.*

In practice a certain level of power dissipation occurs inside the filter structure shown in fig. 6.12 which in the microwave field corresponds to a finite quality factor Q_0. Particularly, for the filter shown in this section we obtain an unloaded quality factor per resonator of about $Q_0 \approx 200$. If we now consider real materials in the matching filter, which means that a given amount of power is dissipated due to heat dissipation inside the matching filter, we obtain the response shown in fig. 6.15. We can now observe that the global response has been degraded, obtaining even a lower reflection than in the lossless

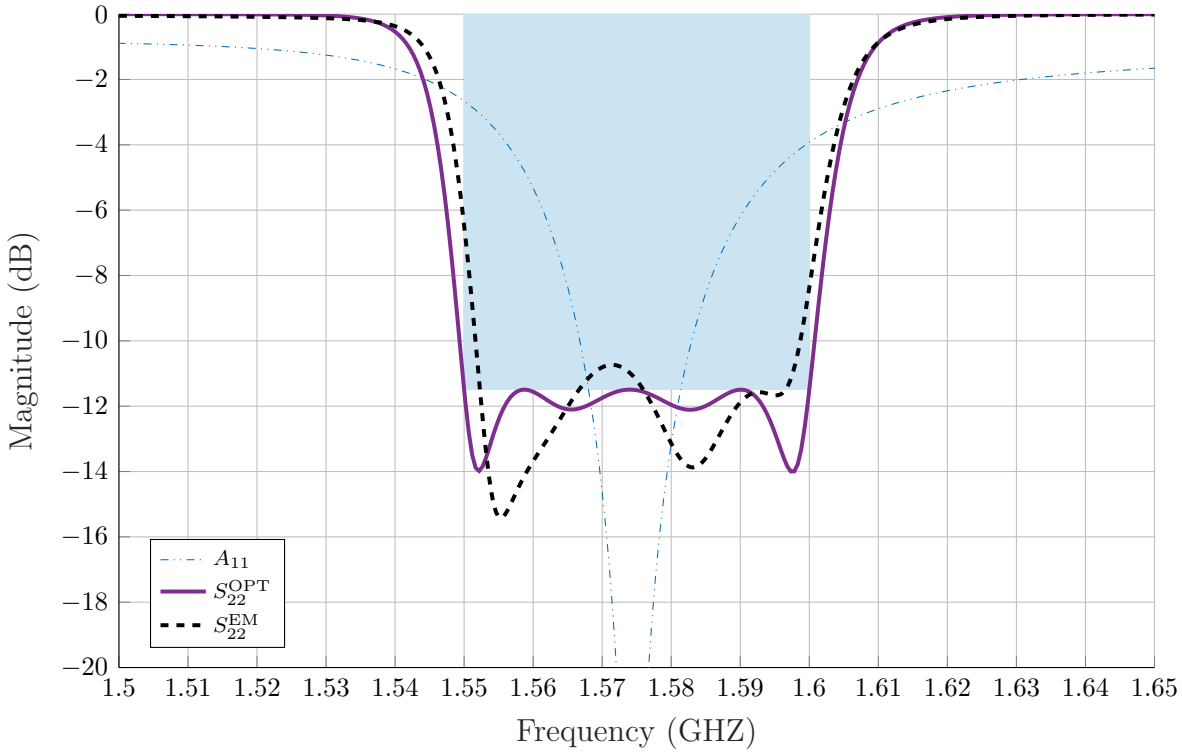


Figure 6.14: Reflection: synthesis and lossless EM simulation.

case. This lower reflection is possible because of the power being dissipated inside the structure, which is not reflected. Nevertheless this response is not optimal anymore when losses are considered. In fact minimising the global reflection has no sense in the case of lossy devices as it would be possible to obtain zero reflection by dissipating everything inside the matching filter and still no transmission would be obtained.

Nevertheless even in the lossy case, where dissipation inside the matching filter is taken into account, the lower bound provided for the global reflection level ψ_{opt} computed by the proposed algorithm, still provides fundamental a upper bound $1-|\psi_{opt}|^2$ on the maximum transmission level that can be achieved. Namely

$$\min_{\omega\in\mathbb{I}} |S_{21}(\omega)|^2 \leq 1-|\psi_{opt}|^2.$$

In this case, some alternative criteria should be considered to perform, starting from the optimal filter in the lossless case, a final re-optimisation where losses are considered. This alternative criterium might be the amount of power transmitted by the global system or, in the case of an antenna, the radiated efficiency. Note that, for the moment we have not said anything about the transmission of the global system or the radiation efficiency of the antenna, only the matching criterion has been considered in this example. However in chapter 9 we return to this topic considering the efficiency of the global system in terms of dissipation and taking into account that criterion in the design of the matching filter.

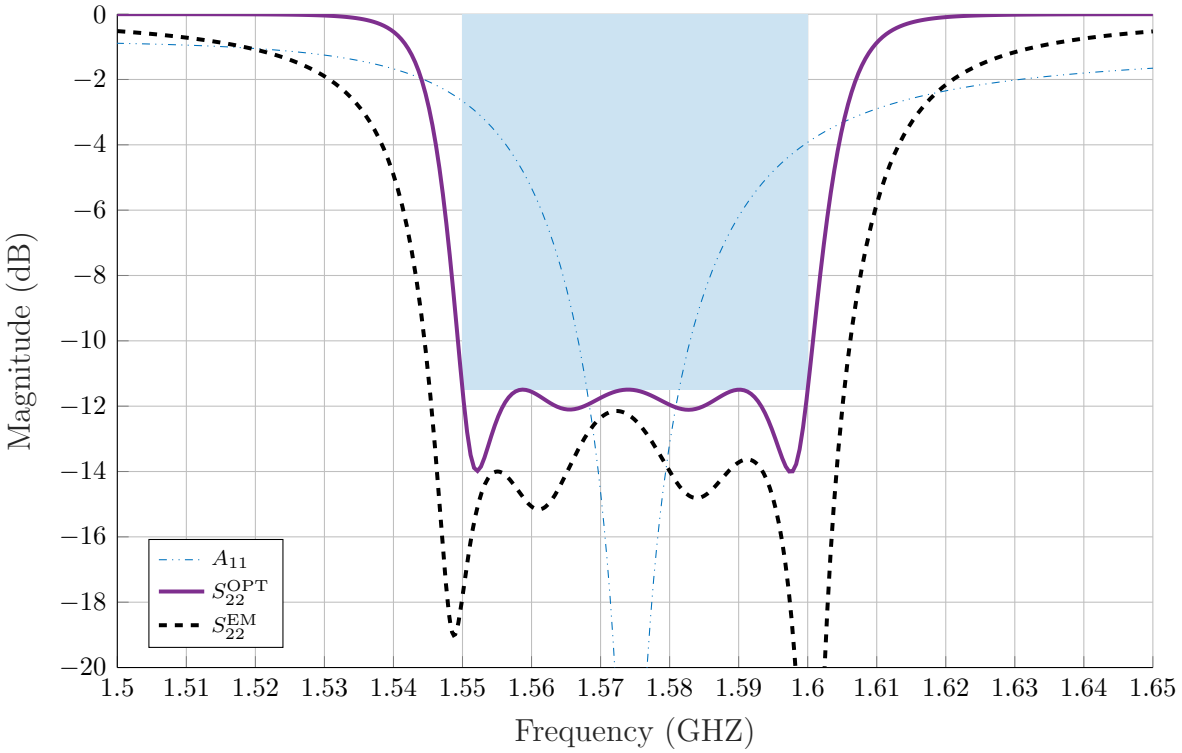


Figure 6.15: Reflection: synthesis and EM simulation considering dissipation losses in the structure shown in fig. 6.12.

6.11 Further applications: multi-band matching

As a second example consider an antenna of type inverted-F (IFA) of the type presented in [38]. In the statement of the matching problem, we consider each frequency band designated for the standard LORA, namely the band for the European standard, America and the band that is planned to be used for LORA communications in China. The frequency specifications of each band are listed in table 6.3.

Band	*F. min* (MHz)	*F. max* (MHz)
EU	863	870
US	902	928
China	779	787

Table 6.3: Lora frequency bands

As in the previous example, we set the transmission polynomial of the matching filter $R_F = 1$ and calculate the optimal matching level based on the McMillan K degree of this filter from 1 to 13. This result is shown in fig. 6.16. In this case the load is also of degree 1, therefore the bound shown in fig. 6.16 is sharp for all values of K.

The first detail that stands out is that the curve shown in fig. 6.16 does not have the smooth shape we saw earlier in fig. 6.2. Indeed the difference in the level of matching

obtained for two adjacent values of K is negligible in certain cases, as for $K = 6$ and $K = 7$ while for different adjacent degrees as $K = 7$ and $K = 8$ the decrease in the matching level is considerably higher. This effect is intimately linked to the fact that the application is multi-band, $\mathbb{I}$ is composed of three different bands. In this case, increasing the McMillan degree of the matching filter only by 1 may not be enough to improve the level of matching in all bands, since one additional pole or reflection zero of the matching filter in one of the bands have a negligible effect on another band that is far enough away.

This fact also indicates that the optimal solution for the matching filter will not be Max McMillan's degree in certain cases. For example for $K = 7$ the solution is practically the same as for $K = 6$ so the matching filter of degree 6 is almost optimal for degree 7 as well. The same applies to the filters of degree $K = 8$ and $K = 10$ which provide almost the same result as for $K = 9$ and $K = 11$. In fact, in fig. 6.17 we show the response of the global system that reaches the optimal matching level for $K = 13$. Note that this response not only satisfies the optimality criterion for $K = 13$, namely to reach the optimal reflection level at $N + 1$ points, but also satisfies that criterion for $K = 14$ since, as can be seen in fig. 6.17, the level of optimal matching is reached at 15 different points within the three bands.

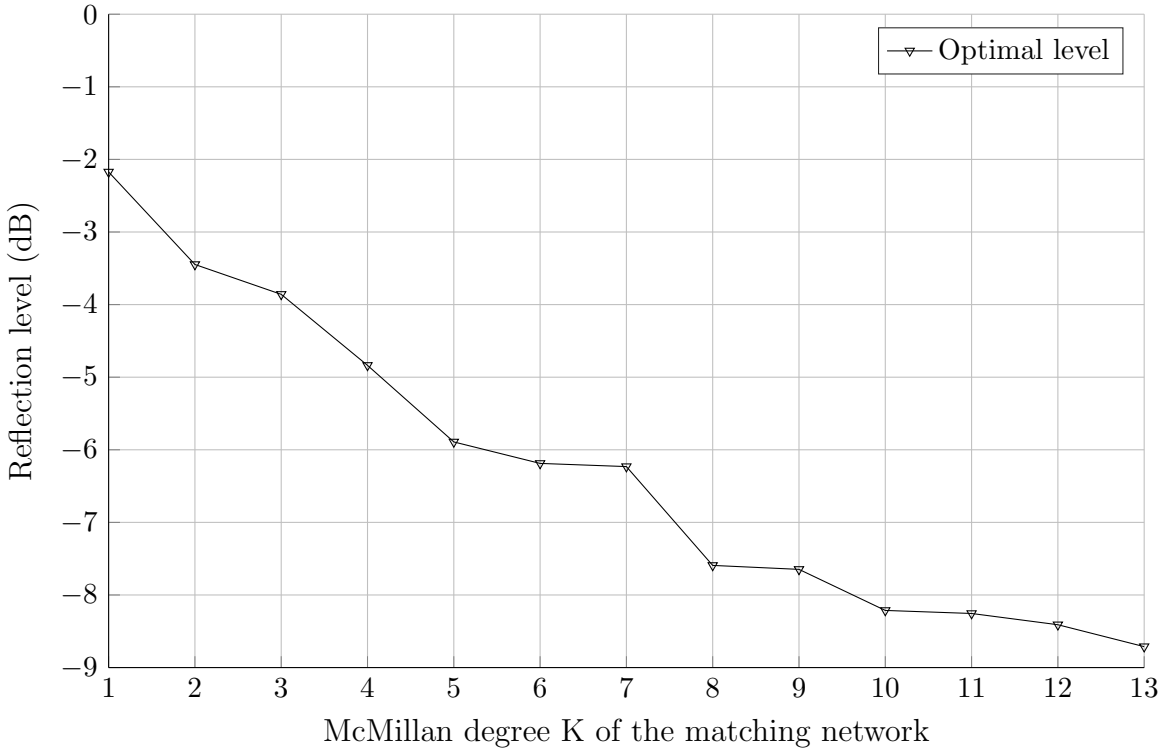


Figure 6.16: Optimal matching level considering the frequency bands in table 6.3

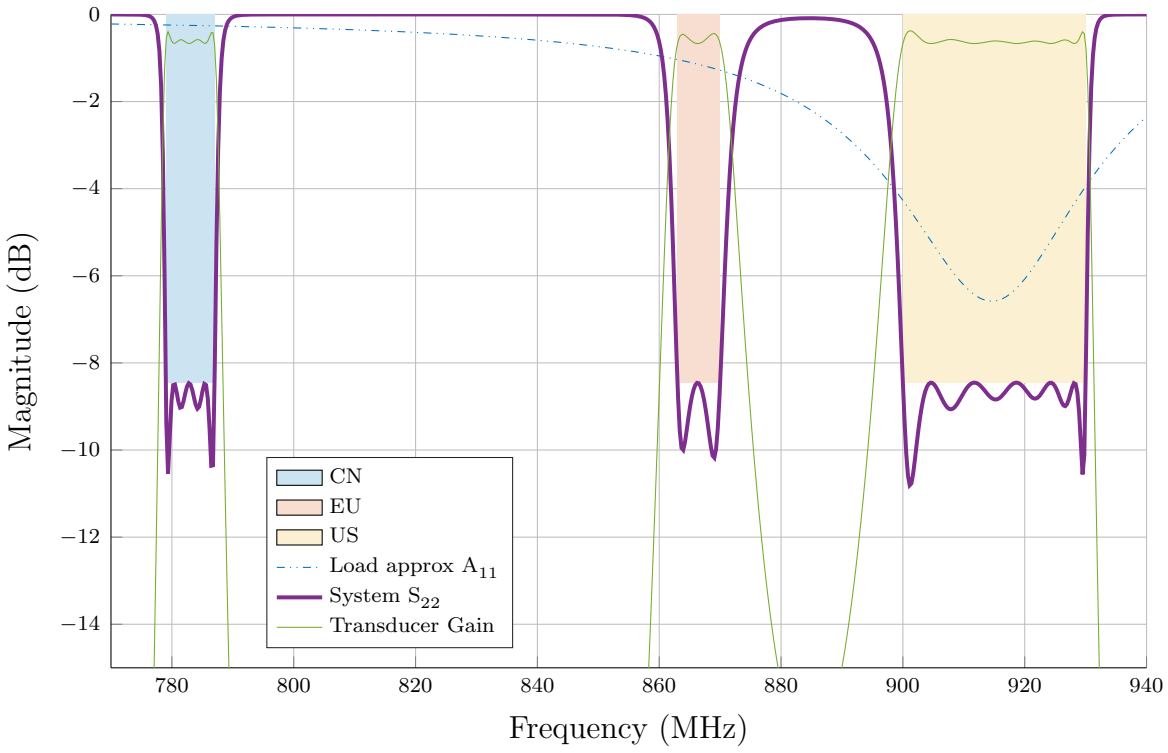


Figure 6.17: Result of matching an antenna with a matching filter of McMillan degree $K = 13$ in all *LORA* bands (table 6.3)

6.12 The bandwidth problem.

The potential bandwidth of an antenna is closely related to broadband matching theory, in the sense that in order to reach a certain bandwidth with a given antenna, it is necessary to obtain conjugate impedance matching over that bandwidth within a prescribed tolerance. In previous sections, we deal with the problem of minimising the global reflection S_{22} in a prescribed interval $\mathbb{I}$. However a problem that arises, even more often in practice, is, given a prescribed reflection level L_{opt} (the required level to guarantee the optimal operation of the system) find the largest interval $\mathbb{I}$ such that

$$\max_{\tau\in\mathbb{I}} |S_{22}(\tau)|^2 \leq L_{opt}.$$

We introduce now the definition of *potential bandwidth* as it is used in this section.

Definition 6.12.1 (Potential bandwidth)**.** *Given the rational function L_{11} of degree 1 and considering function $S_{22} \in \mathbb{C}_R^N$ of finite degree N feasible for such load. We denote by potential bandwidth the maximum frequency τ_c such that the supremum, of $|S_{22}(\tau)|^2$ with $\tau \in [-\tau_c, \tau_c]$ is below a prescribed level L_{opt}.*

$$\tau_{opt} = \max_{S_{22}\in\mathbb{C}_R^N} \tau_c \qquad s.t. \qquad \max_{\tau\in\mathbb{I}} |S_{22}(\tau)|^2 \leq L_{opt} \qquad \mathbb{I} = [-\tau_c, \tau_c].$$

Therefore, the problem is reformulated in order to obtain the best bandwidth attainable with a given load and a matching network of a fixed finite degree when the reflection level is prescribed. Note that an iterative version of problem 4.6.1 where no selectivity constraint is considered namely

Problem 6.12.1 (Direct problem)**.**

$$\begin{aligned} &\textit{Find:} && L_{opt} = \min_{P\in\mathbb{P}_+^{2N}} \max_{\omega\in\mathbb{I}} \frac{P(\tau)}{R(\omega)} \\ &\textit{Subject to:} && f(P) \leq K && K \in \mathbb{R}_+. \end{aligned}$$

By adjusting the limits of the interval $\mathbb{I}$ at each iteration, for instance following a dichotomy algorithm to attain the desired L_{opt}, the bandwidth problem could be solved. Nevertheless, the bandwidth problem can be stated in a more *direct* and elegant form. With this aim let us state a dual version of problem 6.12.1.

Problem 6.12.2 (Dual problem)**.**

$$\begin{aligned} &\textit{Find:} && K_{opt} = \min_{P\in\mathbb{P}_+^{2N}} f(P) \\ &\textit{Subject to:} && \frac{P(\omega)}{R(\omega)} \leq L && \forall\omega \in \mathbb{I}. \end{aligned}$$

We have following lemma

Lemma 6.12.1. *problem 6.12.2 is equivalent to problem 4.6.1 in the following sense: if P_{opt} solves problem 4.6.1 with optimal criterion L_{opt} then it is also the solution of problem 6.12.2 with $L = L_{opt}$ and optimal criterion $K_{opt} = K$.*

Proof. Let P_{opt} be the optimal solution to problem 6.12.1 with criterium L_{opt} given the value $K \in \mathbb{R}$. Now assume P_{opt} is not the solution to problem 6.12.2. This implies there exists a polynomial $P_A \in \mathbb{P}_+^{2N}$ such that

$$\begin{aligned} &f(P_A) < f(P_{opt}) \leq K \\ &\frac{P_A(\omega)}{R_A(\omega)} \leq L_{opt} && \forall\omega \in \mathbb{I}. \end{aligned}$$

Therefore there exist a positive value $\epsilon < 1$ such that the polynomial ϵP_A provides a better criterium in problem 6.12.1 than P_{opt}. This contradict the optimality of P_{opt} for problem 6.12.1.

Conversely, assume now P_{opt} is the solution problem 6.12.2 with criterium K_{opt} and any value $L \geq 0$. Suppose further P_{opt} is not the optimal solution to problem 6.12.1. Therefore there exist $P_B \in \mathbb{P}_+^{2N}$ such that

$$\begin{aligned} &\max_{\omega\in\mathbb{I}} \frac{P_B(\omega)}{R(\omega)} < \max_{\omega\in\mathbb{I}} \frac{P_{opt}(\omega)}{R(\omega)} \leq L \\ &f(P_B) \leq K_{opt}. \end{aligned}$$

In this case we can multiply P_B by a constant $\varepsilon > 1$ such that $f(\varepsilon P_B) < f(P_{opt})$ what contradict the optimality of P_{opt}. This concludes the proof. □

Problem 6.12.2 provides direct information about the feasibility of the desired bandwidth by comparing the obtained value K_{opt} with K:

$$\begin{aligned} &K_{opt} \leq K \longrightarrow \text{feasible}, \\ &K_{opt} > K \longrightarrow \text{not feasible}. \end{aligned}$$

Note if we obtain strict feasibility ($K_{opt} < K$) in the previous test, then the bandwidth can be increased until K_{opt} matches K. Here we are still using an iterative procedure to determine the potential bandwidth. Nevertheless it is possible to compute how the K_{opt} will be modified after modifying the interval $\mathbb{I}$ allowing us to determine the optimal bandwidth without recomputing K_{opt}.

After solving problem 6.12.2 with a reference passband, (i.e. taking $\tau_c = 1$), we apply the change of variable $t = \tau_{opt} \cdot \tau$ where τ_{opt} is the sought potential bandwidth. We compute now how the function $f(P)$ is modified with the proposed change of variable.

$$\hat{f}(P) = \int_{\mathbb{R}} \frac{\log\left(1 + \frac{R(\tau_{opt}\cdot x)}{P(\tau_{opt}\cdot x)}\right)}{(\tau_{opt}\cdot x - \alpha)^2} dx.$$

Now define

$$t = \tau_{opt} \cdot x \qquad\qquad dt = \tau_{opt} \cdot dx.$$

The function becomes

$$\hat{f}(P) = \int_{\mathbb{R}} \frac{\log\left(1 + \frac{R(t)}{P(t)}\right)}{(t-\alpha)^2} \frac{dt}{\tau_{opt}} = \frac{f(P)}{\tau_{opt}}.$$

This provides us with the following lemma

Lemma 6.12.2. *Given the problem 6.12.2 with an interval $\mathbb{I} = [-\tau_c, \tau_c]$ the optimal criterium K_{opt} is inverse proportional to the value τ_c. Precisely $K_{opt} = \frac{K_0}{\tau_c}$ where K_0 is the optimal criterion with $\tau_c = 1$.*

Finally compute τ_{opt} such that $\frac{K_{opt}}{\tau_{opt}} = K$. We have

$$\tau_{opt} = \frac{K_{opt}}{K_1}.$$

Therefore the potential bandwidth is obtained by solving the following problem once:

Problem 6.12.3 (Bandwidth problem)**.**

$$\begin{aligned} &\textit{Find:} && \tau_{opt} = \min_{P \in \mathbb{P}_+^{2N}} f_K(P) \\ &\textit{Subject to:} && P(\tau) \leq L_{opt}\cdot R(\tau) \qquad L_{opt} \geq 0 \qquad \forall \tau \in [-1,1] \end{aligned}$$

where L_{opt} is the desired reflection level within the band and

$$f_K(P) = \frac{\int_{\mathbb{R}} \log\left(1 + \frac{R(x)}{P(x)}\right)(x-\alpha)^{-2} dx}{2\pi j \left[\frac{d}{d\tau} \log L_{22}(\tau)\right]_{\tau=\alpha}}.$$

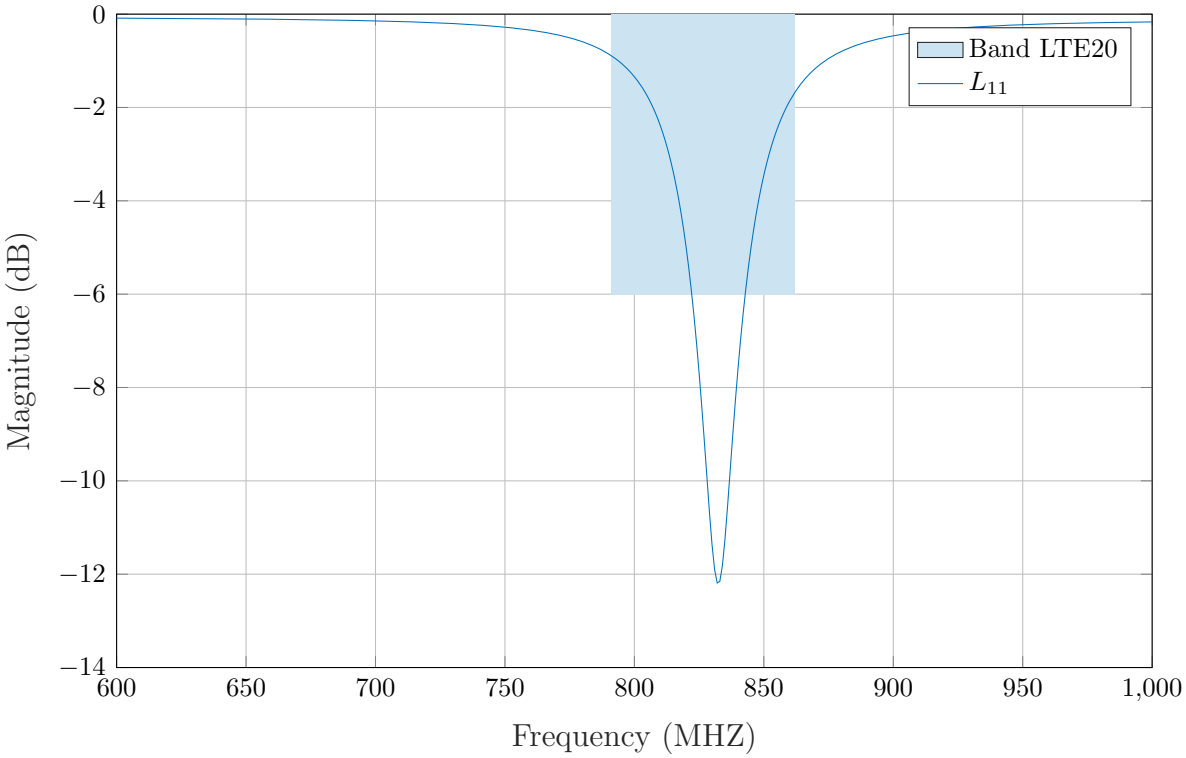


Figure 6.18: Measured reflection L_{11} of the antenna with 50Ω reference.

6.12.1 Practical example

Next we show as an illustrative example an antenna designed for the LTE-20 band:

$$\text{LTE-20 Band: } 791 - 862\ MHz.$$

A bandwidth of at least $70MHz$ and a reflection level smaller than $-6dB$ is required. This antenna present the input reflection shown in fig. 6.18. The load has no finite transmission zero, therefore we choose $R = 1$ as transmission polynomial. The potential bandwidth can be obtained from the data issue of problem 6.12.3. Results are shown in fig. 6.19, as a function of the degree of the matching network N and the reflection level l in dB such that

$$L = \left(l^{-1} - 1\right)^{-1}.$$

If we fix now $l = -6dB$ we obtain the result shown in fig. 6.20. It can be seen that the desired bandwidth of 70 MHz can only be obtained with a matching network of degree 10, which in practice is not a realistic solution since a matching network of that complexity would introduce an important amount of losses and reduce drastically the system efficiency, apart from the footprint increment.

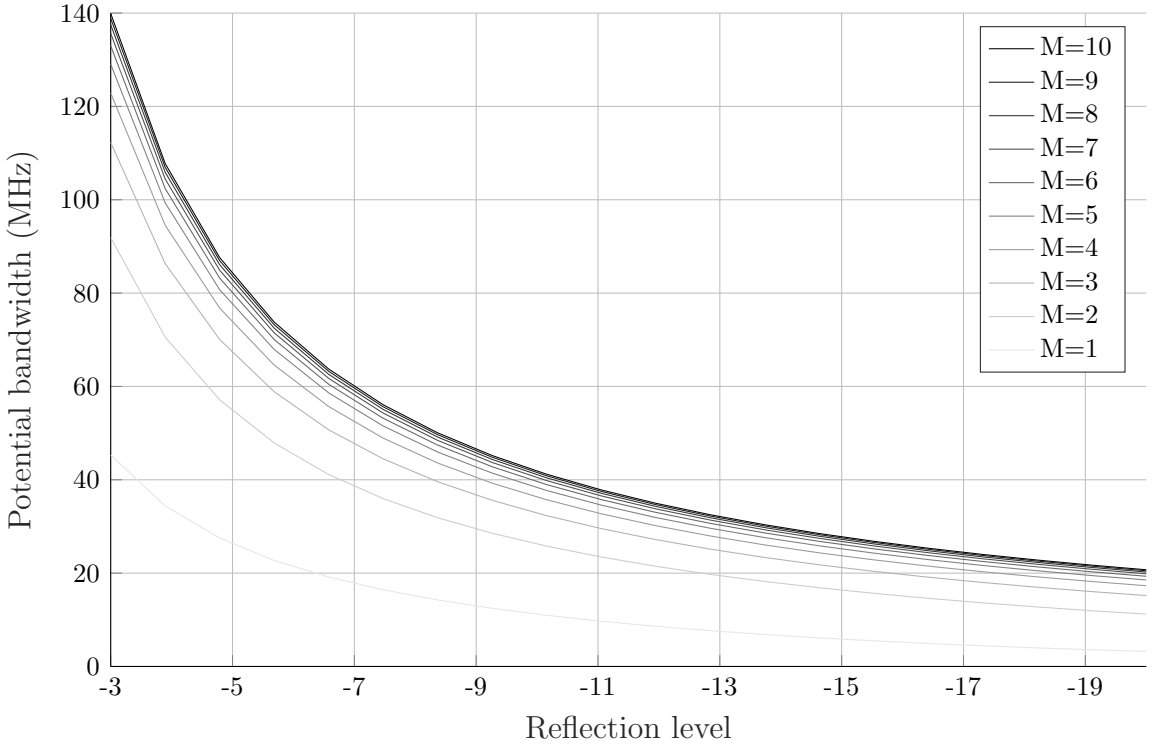


Figure 6.19: Potential bandwidth of the presented antenna.

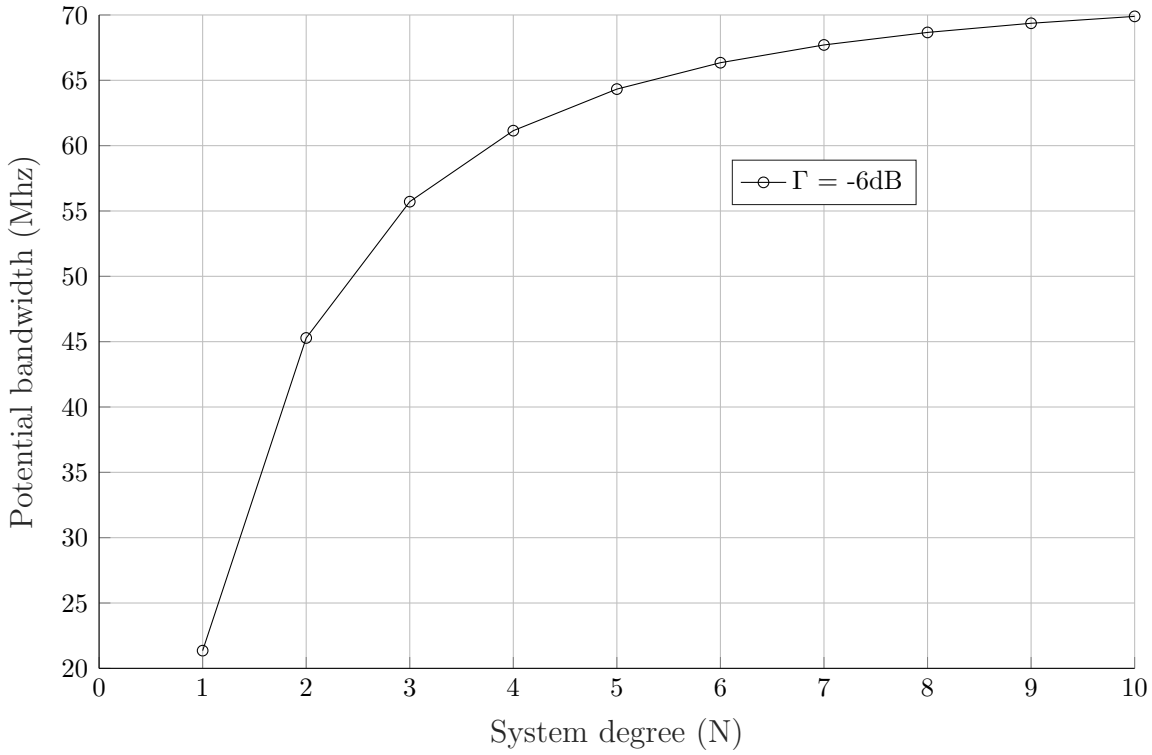


Figure 6.20: Potential bandwidth with a reflection level $l = -6dB$.

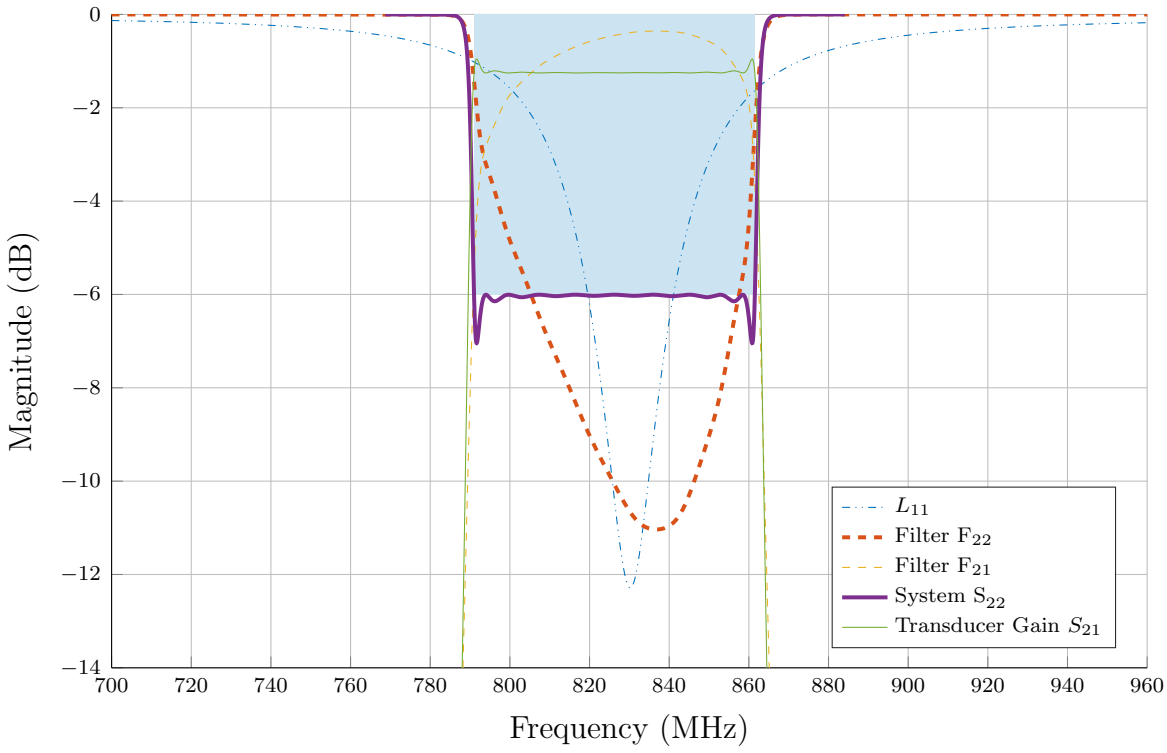


Figure 6.21: Matching results with a degree $N = 10$ achieving a reflection level of $-6dB$ over a bandwidth of 70 MHz.

References

[34] R. M. Fano, *Theoretical Limitations on the Broadband Matching of Arbitrary Impedances*, ser. Technical report: Research Laboratory of Electronics. MIT Res. Lab. of Electronics, 1947.

[35] R. J. Cameron, R. Mansour, and C. M. Kudsia, *Microwave Filters for Communication Systems: Fundamentals, Design and Applications*. Wiley, 2007.

[36] Ansoft, "HFSS 12.0," www.ansoft.com, 2010.

[37] F. Seyfert. (2014) Software Presto-HF. Available: https://project.inria.fr/presto-hf/.

[38] L. Lizzi and F. Ferrero, "Impact of Ground Plane Reduction in Antennas for Compact Terminals," in *2018 2nd URSI Atlantic Radio Science Meeting, AT-RASC 2018*, 2018.

Part III

Numerical implementation

Chapter 7:

Formulation as a Semi-Definite Program

After all the theory developed in chapter 4 and considering the parametrisation of the set of admissible polynomials obtained at the end of chapter 5, it is time to do a little more practical work. Now we are finally able to formulate the matching problem in a more concise way. This formulation fits on the framework of one of the problems that are well known in the field of convex optimisation. We are speaking about a non-linear semi-definite program (NLSDP) which might be considered to be among the most complex problems in numeric optimisation that can be optimally solved. This reason motivates us to perform a numerical implementation of the developed theory. Furthermore, the numerical implementation allow us to deal with different matching scenarios coming from several external organizations which are facing the matching problem.

In chapter 4 the problem of matching has been formulated as a minimax problem equivalent to the classical synthesis problem of transfer functions with an additional constraint that carries over the set of polynomials among which we seek the optimal solution to the problem. This condition to determine the admissibility of a polynomial P, which might seem a little abstract in chapter 4, can be expressed in terms of a matrix inequality by comparing a matrix $\mathbf{U}(P)$ depending on P with another matrix $\mathbf{\Pi}$ which depends solely on the load under consideration. This is the main result obtained in chapter 5 which states the positive polynomial P is admissible if and only if

$$\mathbf{U}(P) \succeq \mathbf{\Pi}.$$

In addition, we believe it is convenient to highlight the fact that the different applications of the theory developed in chapter 4, together with the relevant results, are not found in this chapter. These results have their own chapter which follows the present one, similarly as what we did in chapter 6. In this way we obtain a certain atomicity that facilitates the task to a reader not so interested in the numerical aspects of the problem, allowing him to go directly to chapter 10 to review these results.

7.1 Statement of the general problem

We state now the problem we deal with in this chapter issue of the theory developed in chapters 4 and 5. This is a problem that contains a matrix inequality and that will be progressively transformed throughout the present chapter into a problem that only involves matrix inequalities. As you can already guess, matrix algebra is an important pillar of the implementation of the matching problem. For this reason, and with the aim of achieving greater clarity in the text, the variables that represent matrices will be indicated in capital letters and bold from this chapter.

Consider the load introduced in chapter 4 of McMillan degree M with rational scattering matrix $\mathbf{L}$ in the Belevitch form (see eq. (2.38))

$$\mathbf{L} = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix} = \frac{1}{q_L} \begin{pmatrix} p_L^\star & -r_L^\star \\ r_L & p_L \end{pmatrix}.$$

Again we denote $R_L = r_L r_L^\star$ the transmission polynomial of $\mathbf{L}$ and $\alpha_1, \alpha_2, \cdots, \alpha_M \in \mathbb{C}^-$ the transmission zeros of $\mathbf{L}$ which are assumed to have simple multiplicity.

Let us state now the generalised form of the matching problem introducing the characterisation of the admissible set obtained from theorem 5.4.1.

Problem 7.1.1 (General matching problem)**.** *Minimize the reflection level,* Γ*, in the passband* Ψ *under some specified rejection level in the stop-band.*

$$\begin{aligned} \textit{Find:} \qquad & \Psi = \min_{P\in\mathbb{P}_+^{2N}} \max_{\omega\in\mathbb{I}} \frac{P(\omega)}{R(\omega)}, \\ \textit{Subject to:} \qquad & P(\omega) \le R(\omega)\cdot\Gamma \qquad\qquad \omega\in\mathbb{J}, \\ & \mathbf{U}(P) \succeq \mathbf{\Pi}, \end{aligned}$$

where $\mathbb{I}$ *and* $\mathbb{J}$ *stand for the passband and stop-band intervals respectively with the matrix* $\mathbf{U}(P)$ *and* $\mathbf{\Pi}$ *defined as in eqs.* (5.10) *and* (5.11)*.*

As the first step toward the implementation, 7.1.1 is restated by introducing the slack variable Ψ which becomes the criterium to be minimised:

Problem 7.1.2 (General problem)**.**

$$\begin{aligned} \textit{Find:} \qquad & \min_{(\Psi,P)} \Psi \qquad\qquad (\Psi,P)\in\mathbb{R}_+\times\mathbb{P}_+^{2N}, & \\ \textit{Subject to:} \qquad & P(\omega) \le \Psi\cdot R(\omega) \qquad \omega\in\mathbb{I}, & (7.1)\\ & P(\omega) \ge \Gamma\cdot R(\omega) \qquad \omega\in\mathbb{J}, & (7.2)\\ & \mathbf{U}(P) \succeq \mathbf{\Pi}. & (7.3) \end{aligned}$$

As a result, we obtain a problem of minimization of the slack variable Ψ over the pair (Ψ, P). Note Ψ also represents the toss level on the passband $\mathbb{I}$, as it is ensured by eq. (7.1) meanwhile eq. (7.2) imposes some additional rejection constraint on the stop-band $\mathbb{J}$. Finally, eq. (7.3) guarantees the admissibility of the polynomial P.

7.2 Positive polynomials

Note in problem 7.1.2 we several types of positivity constraints are imposed. On the one hand we have the constraint $P(\omega)\ge 0$ for all $\omega\in\mathbb{R}$ and on the other hand if we define the polynomials $W_\Psi, W_\Gamma \in \mathbb{P}^{2N}$ as

$$\begin{aligned} W_\Psi &= \Psi R - P, \\ W_\Gamma &= P - \Gamma R, \end{aligned}$$

then we obtain different additional positivity constraint which holds only on a finite interval of the real line, namely

$$\begin{aligned} P(\omega) &\ge 0 \qquad \forall\omega\in\mathbb{R}, \\ W_\Psi(\omega) &\ge 0 \qquad \forall\omega\in\mathbb{I}, \\ W_\Gamma(\omega) &\ge 0 \qquad \forall\omega\in\mathbb{J}. \end{aligned}$$

Constraints on the positivity of a polynomial $P(\omega)\ge 0$ is recurrent in optimisation as positive polynomials form a convex set, although such constraint is difficult to express over the coefficients of the polynomial P. Note the condition $P(\omega)\ge 0$ represent in

fact a set of an infinite amount of linear constraints $P(\omega_i) \geq 0$ for every point $\omega_i \in \mathbb{R}$. The space of polynomials $P \in \mathbb{P}^{2N}$ under the said set of infinite linear constraints is indeed a convex set. The same argument can be applied to polynomials positive on an interval $\mathbb{I}$ with $\mathbb{I} \subset \mathbb{R}$. However in practice, an infinite amount of constraints is hardly ever implemented as a finite amount of resources are available in the machines used to perform that computation. One reasonable approximation is the discretisation of the intervals $\mathbb{I}, \mathbb{J}$ with a finite number of points $x_1, x_2, \cdots, x_{n_\mathbb{I}} \in \mathbb{I}$ and $y_1, y_2, \cdots, y_{n_\mathbb{J}} \in \mathbb{J}$ as well as the sampling of the real line by choosing a set of control points $\omega_i \in \mathbb{R}$ with $i \in [1, n_\mathbb{R}]$, which allows us to impose the positivity constrain on a finite amount of points. We have

$$\begin{aligned} P(\omega_i) &\geq 0 && 1 \leq i \leq n_\mathbb{I}, \\ W_\Psi(x_i) &\geq 0 && 1 \leq i \leq n_\mathbb{J}, \\ W_\Gamma(y_i) &\geq 0 && 1 \leq i \leq n_\mathbb{R}. \end{aligned}$$

As the number of points $n_\mathbb{I}, n_\mathbb{J}, n_\mathbb{R}$ tends to infinity, the previous set of constraints converges to the desired positivity constraint on the corresponding intervals.

7.2.1 Parametrisation by means of linear matrix inequalities

We introduce in this section a more elegant formulation that ensures the positivity of a polynomial uniformly on the real line or an interval $\mathbb{I} \subset \mathbb{R}$. This is done by means of a SDP problem, namely a semi-definite program, where some constraint in term of the positivity of a matrix is present. Problems of type SDP have gained popularity in recent years with the apparition of interior point methods. Interior point methods allows to solve SDP optimally and efficiently, what was not the case before, even in the case of a convex problem. This fact comes from the difficulty of ensuring the positive-definiteness of a matrix, rendering complicated the numerical solution of an optimisation problem over a set of positive definite matrices, regardless of the convexity of such set.

Positive polynomials can be associated to positive definite matrices such that to each square matrix of size $N \times N$ correspond a unique polynomial in $\mathbb{P}_+^{2N}$. We have that a matrix $\mathbf{V}$ is positive definite $\mathbf{V} \succeq 0$, then there exist a unique polynomial associated to it, which is positive. Note the relation is not one to one as the space of $N \times N$ matrices is bigger than $\mathbb{P}^{2N}$. This property, which is detailed later on, allows to state problem with constraints on the positivity of a polynomial in the form of a SDP, involving the positivity of one or more matrices.

7.2.2 Trace

Before providing the parametrisation of polynomials by means of square symmetric matrices, some definitions are required. We begin by defining the trace of a matrix.

Definition 7.2.1 (Trace)**.** *Given a square matrix* $\mathbf{V}$ *of size* $(N+1) \times (N+1)$*, define* $\mathrm{tr}_k(\mathbf{V})$ *with* $-N \leq k \leq N$ *as the sum of the elements in the k-th diagonal of* $\mathbf{V}$*.*

Diagonals of $\mathbf{V}$ *are numbered such that the index 0 corresponds to the main diagonal, positive indexes correspond to diagonals on the upper triangle and negative indexes to the diagonals in the lower triangle as it is indicated in fig. 7.1.*

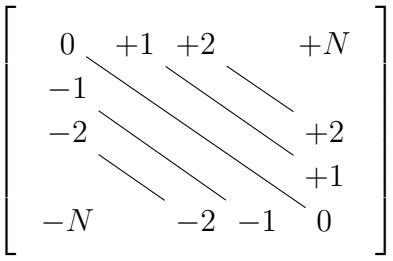


Figure 7.1: Numbering of matrix diagonals

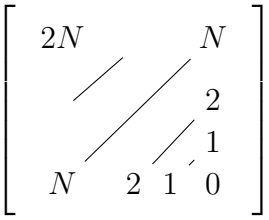


Figure 7.2: Numbering of matrix anti-diagonals

Additionally, we provide a second definition referring to the trace along the anti-diagonal of a matrix $\mathbf{V}$.

Definition 7.2.2 (Anti-diagonal trace)**.** *Given a square matrix* $\mathbf{V}$ *of size* $(N+1)\times(N+1)$*, define* $\mathrm{atr}_k(\mathbf{V})$ *with* $0 \leq k \leq 2N$ *as the sum of the elements in the k-th anti-diagonal of* $\mathbf{V}$*.*

Anti-diagonals of $\mathbf{V}$ *are numbered from the bottom-right element that corresponds to index 0 to the top-left element of* $\mathbf{V}$ *(index 2N) as in fig. 7.2.*

Finally we introduce the function tri that associates to each matrix $\mathbf{V} \in \mathbb{H}^N$, the vector of coefficients in the lower triangle.

Definition 7.2.3 (Lower triangle elements)**.** *Define* tri$\mathbf{V}$ *with* $\mathbf{V} \in \mathbb{H}^N$ *the linear map*

$$\mathrm{tri}\mathbf{V} : \mathbb{H}^N \longrightarrow \mathbb{C}^{(N^2+N)/2},$$

where $\mathrm{tri}(\mathbf{V})$ *is the vector of elements column-wise in the lower triangle of* $\mathbf{V}$*. Similarly we also stablish a correspondence between real symmetric matrices, namely* $\mathbb{S}^N$ *and the elements of* $\mathbb{R}^{(N^2+N)/2}$ *obtained by the function* tri*.*

7.2.3 The Gram matrix

We discuss next the polynomial parametrisation used in this work. A detailed review of this parametrisation can be found throughout [39]. Note that the set $\mathbb{P}^N$ of polynomials with real coefficients identifies with $\mathbb{R}^{N+1}$, namely the column vector of coefficients, associated to a particular basis. Consider the canonic basis B_N of the space $\mathbb{P}^N$ defined as

$$B_N(\lambda) = [\lambda^N, \lambda^{N-1}, \cdots, \lambda^0]^T.$$

We denote by $P_B \in \mathbb{R}^{N+1}$ the vector of coefficients of the polynomial $P \in \mathbb{P}^N$ such that

$$P(\lambda) = B_N(\lambda)^T \cdot P_B \qquad\qquad \forall \lambda \in \mathbb{C}.$$

Consider now the set of polynomial $P \in \mathbb{P}^{2N}$ with real coefficients $P_B \in \mathbb{R}^{2N+1}$ such that

$$P(\lambda) = B_{2N}(\omega)^T \cdot P_B,$$

where $B_{2N}(\omega)$ is the canonic basis of degree $2N$. Each polynomial P of this class can be associated to a matrix $\mathbf{V} \in \mathbb{S}^{N+1}$, which is not unique, such that

$$P(\lambda) = B_{2N}(\lambda)^T \cdot P_B = B_N(\lambda)^T \cdot \mathbf{V} \cdot B_N(\lambda) \qquad \forall \lambda \in \mathbb{C}. \tag{7.4}$$

Definition 7.2.4 (Gram matrix)**.** *The set of matrices $\mathbf{V} \in \mathbb{S}^{N+1}$ satisfying eq. (7.4) is the set of Gram matrices associated to P.*

This parametrisation provides us with a theorem [39, Theorem 2.5] that allows handling of positivity constraints on polynomials by means of SDP .

Theorem 7.2.1 (Polynomial positivity)**.** *The polynomial $P \in \mathbb{P}^{2N}$ is positive, namely $P \in \mathbb{P}^{2N}_+$ if and only if there exist a matrix $\mathbf{V} \in \mathbb{S}^{N+1}$, with $\mathbf{V} \succeq 0$ associated to P. Additionally, we have $P(\lambda) > 0$ for all $\omega \in \mathbb{R}$ if and only if there exist a matrix $\mathbf{V} \in \mathbb{S}^{N+1}$, associated to P such that $\mathbf{V} \succ 0$.*

Proof of necessity. Necessity follows from the fact that $P(\lambda) = B_N(\lambda)^T \cdot \mathbf{V} \cdot B_N(\lambda)$ is a quadratic form. Therefore if $\mathbf{V} \succeq 0$ ($\mathbf{V} \succ 0$), then the quadratic form $P(\lambda)$ satisfies $P(\lambda) \geq 0$ ($P(\lambda) > 0$) for all $B_N(\lambda) \neq 0$. □

Proof of sufficiency. Consider $P \in \mathbb{P}^{2N}_+$. In this case we can express $P = p \cdot p^\star$ with $p \in \mathbb{P}^N$. If we denote by $p_B \in \mathbb{R}^{N+1}$ the vector of coefficients of the polynomial p with respect to the basis $B_N(\lambda)$, namely $p(\lambda) = B_N(\lambda) \cdot p_B$ for all $\lambda \in \mathbb{C}$ then

$$P(\lambda) = {B_N}^T(\lambda) \cdot p_B \cdot \left({B_N}^T(\lambda) \cdot p_B\right)^\star = {B_N}^T(\lambda) \cdot p_B p_B^\star \cdot B_N(\lambda).$$

Defining $\mathbf{V} = p_B p_B^\star$ we have $\mathbf{V} \succeq 0$. Therefore there exist a Gram matrix $\mathbf{V} \succeq 0$ associated to P.

Consider now a polynomial $P \in \mathbb{P}^{2N}_+$ such that $P(\lambda) > 0$ for all $x \in \mathbb{R}$. Let define the polynomial $s \in \mathbb{P}^{2N}_+$ as follows

$$s(\lambda) = \lambda^{2N} + \lambda^{2N-2} + \cdots + \lambda^2 + \lambda^0.$$

Note that the $(2N+1) \times (2N+1)$ identity matrix $\mathbf{I}_{2N+1}$ is a Gram matrix associated to s. Let us define as well the polynomial $\hat{P} = P - s$. In this case we can take $\epsilon > 0$ small enough such that $\hat{P} \in \mathbb{P}^{2N}_+$. Additionally we have for all $\lambda \in \mathbb{C}$

$$\hat{P}(\lambda) = {B_N}^T(\lambda) \cdot \left(\hat{\mathbf{V}} + \epsilon \mathbf{I}_{2N+1}\right) \cdot B_N(\lambda).$$

where $\hat{\mathbf{V}} \succeq 0$ is a Gram matrix of $P(\lambda)$ computed as before. Therefore the matrix $\mathbf{V} = \hat{\mathbf{V}} + \epsilon \mathbf{I}_{2N+1}$ is a Gram matrix associated to P which satisfies $\mathbf{V} \succ 0$. □

We can then consider the set of real symmetric matrices of size $(N+1) \times (N+1)$ which are positive definite, denoted here by $\mathbb{S}^{N+1}_+$ instead of the set of positive polynomials $\mathbb{P}^{2N}_+$. The motivation behind this choice is the fact that the positivity constraint is easier to

ensure over $\mathbb{S}^{N+1}$ than over $\mathbb{P}^{2N}$. Nevertheless note there exist a one to one correspondence between the set $\mathbb{S}^{N+1}$ and $\mathbb{R}^{(N+2)(N+1)/2}$. Similarly each element of $\mathbb{P}^{2N}$ can be mapped to an element of $\mathbb{R}^{2N+1}$. Therefore this ease for handling positivity comes at the expenses of an optimisation space of higher dimension since if we replace the set $\mathbb{R}^{2N+1}$ by $\mathbb{S}^{N+1}$, the number of coefficient that parametrise the optimisation set increases from $2N+1$ to $(N+2)(N+1)/2$.

7.2.4 Basis of Tchebyshev polynomials

We define in this section a kind of polynomials which are associated to a basis composed of the Tchebyshev polynomials of degree less or equal to N. We need to define first the Tchebyshev polynomial of degree N. Note that a Tchebyshev polynomial is always associated to a given interval.

The Tchebyshev polynomial of degree N associated to an interval I is the polynomial of degree N whose absolute value remains bounded by 1 on the interval I and grows the fastest outside the given interval. If we consider the interval $[-1, 1]$, the Tchebyshev polynomial of degree N can be obtained by the following recurrence

$$\mathcal{T}_N(\omega) = 2\omega\, \mathcal{T}_{N-1}(\omega) - \mathcal{T}_{N-2}(\omega),$$

where $\mathcal{T}_0(\omega) = 1$ and $\mathcal{T}_1(\omega) = \omega$. We define then the Tchebyshev basis $\mathcal{B}_N(\lambda)$ of order N as

$$\mathcal{B}_N(\lambda) = [\,\mathcal{T}_N(\lambda),\, \mathcal{T}_{N-1}(\lambda), \cdots,\, \mathcal{T}_0(\lambda)]^{\,T}.$$

This choice of basis is made to overcome some numerical issues faced in the optimisation procedures which are introduced later on. These issues arises from the fact that Tchebyshev polynomials are recurrent in the field of filter synthesis as they are the solution of a Zolotarev problem of the first kind. However the coefficients of Tchebyshev polynomials easily become badly conditioned when the degree increases, which motivates the choice of the Tchebyshev basis for our purposes.

The Tchebyshev polynomials of degree from 0 to N form an orthonormal basis of the space $\mathbb{P}^N$ such that any polynomial $P \in \mathbb{P}^N$ can be expressed as

$$P(\lambda) = \mathcal{B}_N(\lambda)^T \cdot P_{\mathcal{B}},$$

where $P_{\mathcal{B}} \in \mathbb{R}^{N+1}$ is the column vector with the coefficients of P with respect to the Tchebyshev basis.

It should be noted that the vectors $P_B, P_{\mathcal{B}} \in \mathbb{R}^{N+1}$ are different representation of the polynomial $P \in \mathbb{P}^N$ with respect to the canonic and Tchebyshev basis respectively, namely

$$P(\lambda) = B_N(\lambda) \cdot P_B = \mathcal{B}_N(\lambda) \cdot P_{\mathcal{B}} \qquad \forall \lambda \in \mathbb{C}.$$

We have $\mathcal{B}_N(\lambda) = \mathbf{C}_N \cdot B_N(\lambda)$ where $\mathbf{C}_N$ is a non-singular $(N+1) \times (N+1)$ matrix with real coefficients. Furthermore if $\mathbf{V}$ is a Gram matrix of $P \in \mathbb{P}^{2N}$, we have

$$P(\lambda) = B_N(\lambda)^T \mathbf{V} B_N(\lambda) = \mathcal{B}_N(\lambda)^T \mathbf{C}_N^{-1T} \mathbf{V} \mathbf{C}_N^{-1} \mathcal{B}_N(\lambda).$$

Since $\mathbf{C}_N$ is non-singular, the matrix $\mathbf{T} = \mathbf{C}_N^{-1T}\mathbf{V}\mathbf{C}_N^{-1}$ is also a Gram matrix associated to P. Hence from theorem 7.2.1, we have $P \in \mathbb{P}_+^{2N}$ if and only if $\mathbf{T} \succeq 0$.

7.3 Positivity on an interval

In this section we study the class of polynomials that are positive on an interval. Without loss of generality we consider the interval $[-1, 1]$. Note that a polynomial $P_1(\lambda)$ positive on any arbitrary interval $[a, b]$ can be mapped to a polynomial $P_2(x)$ positive on the unit interval by the variable change $\lambda = (2x - (b + a))(b - a)^{-1}$.

To ensure the positivity of a polynomial $P \in \mathbb{P}^{2N}$ on an interval, we use the following theorem derived from [39, Theorem 1.11]

Theorem 7.3.1 (Positivity on an interval)**.** *Given the polynomial $P \in \mathbb{P}^{2N}$, we have $P(\omega) \geq 0$ for all $\omega \in [-1, 1]$ if and only if*

$$P(\lambda) = F(\lambda) + (1 - \lambda^2)G(\lambda),$$

where $F \in \mathbb{P}_+^{2N}$ and $G \in \mathbb{P}_+^{2N-2}$.

Polynomials F, G in theorem 7.3.1 are positive on the real line. Therefore there exist $\mathbf{T}_F \in \mathbb{S}_+^{N+1}$ and $\mathbf{T}_G \in \mathbb{S}_+^N$ such that for all $\lambda \in \mathbb{C}$

$$F(\lambda) = B_N(\lambda)^T \mathbf{T}_F B_N(\lambda),$$
$$G(\lambda) = B_{N-1}(\lambda)^T \mathbf{T}_G B_{N-1}(\lambda).$$

Corollary 7.3.1. *The polynomial $P \in \mathbb{P}^{2N}$ is positive on the unit interval if and only if there exist two Gram matrices $(\mathbf{T}_F, \mathbf{T}_G) \in \mathbb{S}_+^{N+1} \times \mathbb{S}_+^N$ such that*

$$P(\lambda) = B_N(\lambda)^T \mathbf{T}_F B_N(\lambda) + (1 - \lambda^2) B_{N-1}(\lambda)^T \mathbf{T}_G B_{N-1}(\lambda) \qquad \forall \lambda \in \mathbb{C}. \tag{7.5}$$

7.3.1 Dealing with several intervals

Once we are able to characterise the positivity of a polynomial on an interval we can use such characterisation to ensure the positivity of a polynomial $P \in \mathbb{P}^{2N}$ on a finite union of compact intervals $\mathbb{I} \subset \mathbb{R}$.

Next we argue toward parametrisation of the polynomials $P \in \mathbb{P}^{2N}$ such that

$$P(\omega) \geq 0 \qquad \forall \omega \in \mathbb{I}.$$

To obtain such parametrisation, we use the tool that we have already introduced, this is the parametrisation of the positive polynomials on the unit interval. Let define $\mathbb{I} = \bigcup_{i=1}^n I_i$ with $I_i = [a_i, b_i]$. Then consider the set of distinct polynomials $\{P^{(1)}, P^{(2)}, \cdots, P^{(n)}\}$ with $P^{(i)} \in \mathbb{P}_+^{2N}$. The polynomial $P^{(i)}$ is positive on the unit interval if and only if there exist $(\mathbf{T}_F^{(i)}, \mathbf{T}_G^{(i)}) \in \mathbb{S}_+^{N+1} \times \mathbb{S}_+^N$ that parametrise the polynomial $P^{(i)}$ as in eq. (7.5). Now we apply to each polynomial $P^{(i)}$ the change of variable that sends the unit interval onto the interval $I_i = [a_i, b_i]$. Thus take the polynomial $\hat{P}(\omega) = P^{(i)}(\Phi_i(\omega))$ with

$$\Phi_i(\omega) = \frac{2\omega + a_i + b_i}{b_i - a_i}.$$

The polynomial $\hat{P}(\omega)$ is positive for all $\omega \in I_i$ if and only if the polynomial $P^{(i)}(\omega)$ is positive on the unit interval. Hence we obtain a set of polynomials $\{\hat{P}^{(1)}, \hat{P}^{(2)}, \cdots, \hat{P}^{(n)}\}$ which are positive on the corresponding interval I_i.

Now we use an interesting trick to guarantee the positivity in the set $\mathbb{I}$. First note the following lemma

Lemma 7.3.1. *Consider two polynomials $P^{(1)}, P^{(2)} \in \mathbb{P}^{2N}$. Assume there exist a set of $2N+1$ different points ω_i such that $P^{(1)}(\omega_i) = P^{(2)}(\omega_i)$ for all $i \in [1, 2N+1]$. Then the polynomials $P^{(1)}$ and $P^{(2)}$ are necessarily the same.*

Proof. Suppose $P_1, P_2 \in \mathbb{P}^{2N}$ are distinct polynomials and $P_1(\omega_i) = P_2(\omega_i)$ for $1 \leq i \leq 2N+1$ with $\omega_i \neq \omega_j$ if $i \neq j$. The polynomial $P \in \mathbb{P}^{2N}$ defined as $P = P_1 - P_2$ vanishes in at least $2N+1$ points. This is not possible unless $P = 0$, and therefore $P_1 = P_2$. □

Now we pick a set $\mathbb{X}$ of distinct $2N+1$ control points $\mathbb{X} = \{\omega_1, \omega_2, \cdots, \omega_{2N+1}\}$. Then we impose the equality

$$\hat{P}^{(1)}(\omega) = \hat{P}^{(2)}(\omega) = \cdots = \hat{P}^{(n)}(\omega) \qquad \forall \omega \in \mathbb{X}, \tag{7.6}$$

obtaining that every polynomials $\hat{P}^{(i)}$ with $i \in [1, n]$ is the same. Therefore the polynomial $P = \hat{P}^{(1)} = \hat{P}^{(2)} = \cdots = \hat{P}^{(n)}$ verifies

$$P(\omega) \geq 0 \qquad \forall \omega \in \mathbb{I},$$

as long as there exist $(\mathbf{T}_F^{(i)}, \mathbf{T}_G^{(i)}) \in \mathbb{S}_+^{N+1} \times \mathbb{S}_+^N$ associated to each polynomial $P^{(i)}(\omega)$ for all $i \in [1, n]$. Note further that eq. (7.6) can be stated over the polynomials $P^{(i)}$ with $1 \leq i \leq n$ as

$$P^{(1)}(\Phi_i(\omega)) = P^{(2)}(\Phi_i(\omega)) = \cdots = P^{(n)}(\Phi_i(\omega)) \qquad \forall \omega \in \mathbb{X}.$$

We have developed here the theory behind the characterisation of a polynomial $P \in \mathbb{P}^{2N}$ such that $P(\omega) \geq 0$ for all $\omega \in \mathbb{I}$. Let us now provide the formal theorem that state the results. Denote first by $\mathbb{X}_i$ the image of the set $\mathbb{X}$ under the application $\Phi_i : \mathbb{X} \mapsto \mathbb{X}_i$. Therefore we have the set of points where each polynomial $P^{(i)}$ is evaluated

$$\mathbb{X}_i = \left\{ x_1^{(i)}, x_2^{(i)} \cdots, x_{2N+1}^{(i)} \right\} \qquad x_k^{(i)} = \Phi_i(\omega_k).$$

Consider now the $(N+1) \times (2N+1)$ matrix $\mathbf{B}_N^{(i)}$ defined as

$$\mathbf{B}_N^{(i)} = \begin{pmatrix} B_N(x_1^{(i)}) & B_N(x_2^{(i)}) & \cdots & B_N(x_{2N+1}^{(i)}) \end{pmatrix} \quad x_k^{(i)} \in \mathbb{X}_i \quad \forall k \in [1, 2N+1],$$

and the diagonal matrix $\mathbf{X}^{(i)}$ with the evaluation of $1 - \left(x_k^{(i)}\right)^2$ on its diagonal

$$\mathbf{X}^{(i)} = \begin{bmatrix} 1 - \left(x_1^{(i)}\right)^2 & 0 & \cdots & 0 \\ 0 & 1 - \left(x_2^{(i)}\right)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 - \left(x_{2N+1}^{(i)}\right)^2 \end{bmatrix}.$$

Then we state the following theorem

Theorem 7.3.2 (Positivity on a finite union of intervals)**.** *Consider the union of compact intervals* $\mathbb{I} = \bigcup_{i=1}^{n} I_i$ *with* $\mathbb{I} \in \mathbb{R}$*. The polynomial* $P \in \mathbb{P}^{2N}$ *is positive in* $\mathbb{I}$

$$P(\omega) \geq 0 \qquad\qquad \forall \omega \in \mathbb{I}$$

if and only if there exist $(\mathbf{T}_F^{(i)}, \mathbf{T}_G^{(i)}) \in \mathbb{S}_+^{N+1} \times \mathbb{S}_+^N$ *for each* $i \in [1, n]$ *such that*

$$\mathbf{Б}_N^{(i)^T} T = \mathbf{Б}_N^{(i)^T} \mathbf{T}_F^{(i)} \mathbf{Б}_N^{(i)} + \mathbf{X}^{(i)} \mathbf{Б}_{N-1}^{(i)^T} \mathbf{T}_G^{(i)} \mathbf{Б}_{N-1}^{(i)} \qquad\qquad \forall i \in [1, n],$$

where T *contains the coefficients* $T \in \mathbb{R}^{2N+1}$ *with respect to the Tchebyshev basis. The matrix* $\mathbf{X}^{(i)}$ *is defined as*

$$\mathbf{X}^{(i)} = \begin{bmatrix} 1 - \left(x_1^{(i)}\right)^2 & 0 & \cdots & 0 \\ 0 & 1 - \left(x_2^{(i)}\right)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 - \left(x_{2N+1}^{(i)}\right)^2 \end{bmatrix},$$

where $x_k^{(i)} = \Phi_i(\omega_k)$ *and* ω_k *with* $k \in [1, 2N+1]$ *belong to a set of* $2N+1$ *distinct points arbitrarily distributed in the complex plane. Finally the matrix* $\mathbf{Б}_N^{(i)}$ *contains the evaluation of the Tchebyshev polynomials of degree up to* N *at those points* $x_k^{(i)}$

$$\mathbf{Б}_N^{(i)} = \begin{bmatrix} Ч_N\left(x_1^{(i)}\right) & Ч_N\left(x_2^{(i)}\right) & \cdots & Ч_N\left(x_{2N-1}^{(i)}\right) \\ \vdots & \vdots & \iddots & \vdots \\ Ч_1\left(x_1^{(i)}\right) & Ч_1\left(x_2^{(i)}\right) & \cdots & Ч_1\left(x_{2N-1}^{(i)}\right) \\ Ч_0\left(x_1^{(i)}\right) & Ч_0\left(x_2^{(i)}\right) & \cdots & Ч_0\left(x_{2N-1}^{(i)}\right) \end{bmatrix}.$$

Proof. Follows from the previous theory. □

7.4 Matrix parametrisation

After introducing the characterization of polynomial positivity in terms of a matrix inequality, as well as a generalized characterization of positive polynomials in a finite union of compact subintervals of the real line, it is time to apply these concepts to problem 7.1.2.

First we formulate the problem as a minimization of the slack variable Ψ under the condition that a set of polynomials dependent on Ψ are positive in certain subsets of $\mathbb{R}$. The problem is as follows.

Problem 7.4.1 (General matching problem with slack polynomials)**.**

$$\begin{aligned} \textit{Find:} \qquad & \min_{(\Psi, P)} \Psi && (\Psi, P) \in \mathbb{R}_+ \times \mathbb{P}_+^{2N}, \\ \textit{Subject to:} \qquad & W_\Psi(\omega) \geq 0 && \omega \in \mathbb{I}, \\ & W_\Gamma(\omega) \geq 0 && \omega \in \mathbb{J}, \\ & \mathbf{U}(P) \succeq \mathbf{Л}, \\ \textit{and the equalities:} \qquad & W_\Psi(\omega) = \Psi R(\omega) - P(\omega), && (7.7) \\ & W_\Gamma(\omega) = P(\omega) - \Gamma R(\omega). && (7.8) \end{aligned}$$

7.4.1 Positivity on the real axis

In order to introduce the results obtained in the previous section, the optimization variable $P \in \mathbb{P}_+^{2N}$ is to be replaced by a Gram matrix $\mathbf{T}_P \in \mathbb{S}^{N+1}$ that parametrizes this polynomial P, so that the positivity of P in the entire real axis can be replaced by a matrix inequality $\mathbf{T}_P \succeq 0$.

Remark 7.4.1. *It is important to note, as already mentioned above, that the fact of performing the optimization on the set of symmetric matrices $\mathbb{S}^{N+1}$ instead of on the set of polynomials $P \in \mathbb{P}_+^{2N}$, leads to an increase in the dimensionality of the problem since the number of parameters increases $2N+1$, which corresponds to the number of coefficients in P, up to $(N+2)(N+1)/2$, namely the number of coefficients in the lower triangle of the matrix $\mathbf{T}_P$ (dimension of $\mathrm{tri}\mathbf{T}_P$). However, this increase in the number of parameters eliminates an even higher number of constraints, providing a more accurate result and a more efficient calculation.*

With respect to the conditions of positivity in each of the corresponding intervals, theorem 7.3.2 can be used, replacing the W_Ψ, W_Γ polynomials with a set of positive definite matrices. In particular if

$$\mathbb{I} = \bigcup_{i=1}^{n} I_i \qquad I_i \subset \mathbb{R},$$
$$\mathbb{J} = \bigcup_{i=1}^{m} J_i \qquad J_i \subset \mathbb{R},$$

we consider the set of matrices

$$(\mathbf{M}_\Psi^{(i)}, \mathbf{N}_\Psi^{(i)}) \in \mathbb{S}_+^{N+1} \times \mathbb{S}_+^N \qquad \forall 1 \leq i \leq n \tag{7.9}$$

that characterises the polynomial W_Ψ and

$$(\mathbf{M}_\Gamma^{(i)}, \mathbf{N}_\Gamma^{(i)}) \in \mathbb{S}_+^{N+1} \times \mathbb{S}_+^N \qquad \forall 1 \leq i \leq m \tag{7.10}$$

to characterise the polynomial W_Γ.

7.4.2 Positivity on a closed subset of $\mathbb{R}$

It is important to note that, once the polynomial P has been parametrized by the Gram matrix $\mathbf{T}_P$ which easily provides the aforementioned polynomial P and therefore also the W_Ψ, W_Γ polynomials, it is very expensive to calculate the Gram matrices that parametrize the polynomials W_Ψ, W_Γ which are already known.

Therefore, what we do in this case is to add each of the Gram matrices that parametrize the polynomials W_Ψ, W_Γ introduced in eqs. (7.9) and (7.10) as variables in the optimization problem. Therefore we have

$$\mathbf{T}_P \succeq 0,$$
$$\mathbf{M}_\Psi^{(i)}, \mathbf{N}_\Psi^{(i)} \succeq 0 \qquad \forall 1 \leq i \leq n,$$
$$\mathbf{M}_\Gamma^{(k)}, \mathbf{N}_\Gamma^{(k)} \succeq 0 \qquad \forall 1 \leq k \leq m.$$

Remark 7.4.2. *It is important to note that only the lower triangle of each matrix is necessary. Therefore we have a polynomial $P \in \mathbb{P}_+^{2N}$, obtained from the matrix $\mathbf{T}_P$. Additionally we obtain a set of polynomials $W_\Psi^{(i)}$ and $W_\Gamma^{(k)}$ for all $i \in [1,n]$ and $k \in [1,m]$, which are positive in the intervals $\mathbb{I}$ and $\mathbb{J}$ respectively.*

However, when doing this the connection between $W_\Psi^{(i)}$, $W_\Gamma^{(k)}$ and P is lost as we allows P, W_Ψ and W_Γ to be independent. Hence it is necessary to explicitly include the equalities in eqs. (7.7) and (7.8) into the problem to ensure the matrix T_P and the set of matrices $(\mathbf{M}_\Psi^{(i)}, \mathbf{N}_\Psi^{(i)})$ with $1 \leq i \leq n$ and $(\mathbf{M}_\Gamma^{(k)}, \mathbf{N}_\Gamma^{(k)})$ where $1 \leq k \leq m$ parametrise indeed the same polynomial P. This is done, for instance, by imposing

$$\begin{aligned} W_\Psi(\omega_i) &= \Psi R(\omega_i) - P(\omega_i), \\ W_\Gamma(\omega_i) &= P(\omega_i) - \Gamma R(\omega_i), \end{aligned}$$

in a set of at least $2N+1$ points ω_i, which can be taken to be equal to the set $\mathbb{X} = \{\omega_1, \omega_2, \cdots, \omega_{2N+1}\}$ introduced above. Therefore we use the map $\mathbf{Б}_N$ and $\mathbf{Б}_{2N}$ defined as

$$\begin{aligned} \mathbf{Б}_N &= \begin{pmatrix} Б_N(\omega_1) & Б_N(\omega_2) & \cdots & Б_N(\omega_{2N+1}) \end{pmatrix} \qquad \omega_k \in \mathbb{X} \quad \forall k \in [1, 2N+1], \\ \mathbf{Б}_{2N} &= \begin{pmatrix} Б_{2N}(\omega_1) & Б_{2N}(\omega_2) & \cdots & Б_{2N}(\omega_{2N+1}) \end{pmatrix} \quad \omega_k \in \mathbb{X} \quad \forall k \in [1, 2N+1], \end{aligned}$$

we have

$$\begin{pmatrix} P(\omega_1) - \Psi R(\omega_1) \\ P(\omega_2) - \Psi R(\omega_2) \\ \vdots \\ P(\omega_{2N+1} - \Psi R(\omega_{2N+1})) \end{pmatrix} = \mathbf{Б}_N{}^T \mathbf{T}_P \mathbf{Б}_N - L\mathbf{Б}_{2N} T_R,$$

where T_R is the coefficient vector of R with the Tchebyshev basis. Therefore we obtain the set of linear equalities

$$\forall i \in [1,n]:$$
$$L\mathbf{Б}_{2N} T_R - \mathbf{Б}_N{}^T \mathbf{T}_P \mathbf{Б}_N = \mathbf{Б}_N^{(i)T} \mathbf{M}_\Psi^{(i)} \mathbf{Б}_N^{(i)} + \mathbf{X}^{(i)} \mathbf{Б}_{N-1}^{(i)T} \mathbf{N}_\Psi^{(i)} \mathbf{Б}_{N-1}^{(i)}, \tag{7.11a}$$

$$\forall k \in [1,m]:$$
$$\mathbf{Б}_N{}^T \mathbf{T}_P \mathbf{Б}_N - \Gamma\mathbf{Б}_{2N} T_R = \mathbf{Б}_N^{(k)T} \mathbf{M}_\Gamma^{(k)} \mathbf{Б}_N^{(k)} + \mathbf{X}^{(k)} \mathbf{Б}_{N-1}^{(k)T} \mathbf{N}_\Gamma^{(k)} \mathbf{Б}_{N-1}^{(k)}. \tag{7.11b}$$

7.4.3 Parametrisation of $\mathbf{U}(P)$

In this section, a re-parameterisation of the problem of matching based on positive defined matrices has been carried out. However, it is still necessary to calculate the original P polynomial since the matrix $\mathbf{U}(P)$ depends on P.

The evaluation of the polynomial $P(\lambda)$ which has been parametrized by the positive definite matrix $\mathbf{T}_P$ can be obtained as $Б_N(\lambda)^T \mathbf{T}_P Б_N(\lambda)$. Thus we re-define the function

$u_{\mathbf{T}}(\lambda)$ with $\mathbf{T} \in \mathbb{S}_+^{N+1}$ as the minimum phase function of λ such that

$$|u_{\mathbf{T}}(\omega)|^2 = \frac{B_N(\omega)^T \mathbf{T} B_N(\omega)}{B_N(\omega)^T \mathbf{T} B_N(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R},$$

$$\Im(u_{\mathbf{T}}(\lambda)) = 0 \qquad \lambda = -j,$$

where R is the fixed transmission polynomial of degree at most $2N$. Now we can also re-define the matrix $\mathbf{U}(\mathbf{T})$ as

$$\mathbf{U}(\mathbf{T}) = \frac{1}{j} \begin{bmatrix} \frac{u_{\mathbf{T}}(\alpha_1)\overline{u_{\mathbf{T}}(\alpha_1)}}{\alpha_1 - \overline{\alpha_1}} & \frac{u_{\mathbf{T}}(\alpha_1)\overline{u_{\mathbf{T}}(\alpha_2)}}{\alpha_1 - \overline{\alpha_2}} & \cdots & \frac{u_{\mathbf{T}}(\alpha_1)\overline{u_{\mathbf{T}}(\alpha_N)}}{\alpha_1 - \overline{\alpha_N}} \\ \frac{u_{\mathbf{T}}(\alpha_2)\overline{u_{\mathbf{T}}(\alpha_1)}}{\alpha_2 - \overline{\alpha_1}} & \frac{u_{\mathbf{T}}(\alpha_2)\overline{u_{\mathbf{T}}(\alpha_2)}}{\alpha_2 - \overline{\alpha_2}} & \cdots & \frac{u_{\mathbf{T}}(\alpha_2)\overline{u_{\mathbf{T}}(\alpha_N)}}{\alpha_2 - \overline{\alpha_N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{u_{\mathbf{T}}(\alpha_M)\overline{u_{\mathbf{T}}(\alpha_1)}}{\alpha_M - \overline{\alpha_1}} & \frac{u_{\mathbf{T}}(\alpha_M)\overline{u_{\mathbf{T}}(\alpha_2)}}{\alpha_M - \overline{\alpha_2}} & \cdots & \frac{u_{\mathbf{T}}(\alpha_M)\overline{u_{\mathbf{T}}(\alpha_N)}}{\alpha_M - \overline{\alpha_N}} \end{bmatrix}.$$

7.4.4 Formulation of the general matching SDP

We can now formulate a SDP version of problem 7.4.1 over the matrices T_P, $(\mathbf{M}_\Psi^{(i)}, \mathbf{N}_\Psi^{(i)})$ for all $i \in [1, n]$ and $(\mathbf{M}_\Gamma^{(k)}, \mathbf{N}_\Gamma^{(k)})$ for all $k \in [1, m]$.

Problem 7.4.2 (Generalised matching SDP)**.**

$$\begin{aligned}
\textit{Find:} \quad & \min_{\substack{\Psi \ \mathbf{T}_P \\ \mathbf{M}_\Psi^{(i)}, \mathbf{N}_\Psi^{(i)} \\ \mathbf{M}_\Gamma^{(k)}, \mathbf{N}_\Gamma^{(k)}}} \Psi, \\
\textit{where:} \quad & \Psi \in \mathbb{R}_+ \qquad (\mathbf{M}_\Psi^{(i)}, \mathbf{N}_\Psi^{(i)}) \in \mathbb{S}^{N+1} \times \mathbb{S}^N \qquad \forall i \in [1, n], \\
& \mathbf{T}_P \in \mathbb{S}^{N+1} \qquad (\mathbf{M}_\Gamma^{(k)}, \mathbf{N}_\Gamma^{(k)}) \in \mathbb{S}^{N+1} \times \mathbb{S}^N \qquad \forall k \in [1, m], \\
\textit{Subject to:} \quad & \mathbf{T}_P \succeq 0, && (7.12) \\
& \mathbf{M}_\Psi^{(i)} \succeq 0 \qquad \forall i \in [1, n], && (7.13) \\
& \mathbf{N}_\Psi^{(i)} \succeq 0 \qquad \forall i \in [1, n], && (7.14) \\
& \mathbf{M}_\Gamma^{(k)} \succeq 0 \qquad \forall k \in [1, m], && (7.15) \\
& \mathbf{N}_\Gamma^{(k)} \succeq 0 \qquad \forall k \in [1, m], && (7.16) \\
& \mathbf{U}(\mathbf{T}_P) \succeq \mathbf{J}\mathbf{I}, && (7.17)
\end{aligned}$$

with the additional equality constraints in eqs. (7.11a) *and* (7.11b).

At this point, we have obtained a formulation of the matching problem in the form of a SDP. However, this problem is a non-linear SPD, since the matrix $\mathbf{U}(\mathbf{T}_P)$ depends non-linearly on the parameters of the problem, namely the coefficients of the $\mathbf{T}_P$ matrix.

7.5 Computation of the minimum phase factor u_P

To calculate this matrix $\mathbf{U}(\mathbf{T})$, it is necessary to evaluate the function $u_{\mathbf{T}}(\omega)$ in the points α_i with $1 \leq i \leq N$ located inside the domain of analyticity. This section is devoted to the development of the procedure necessary for the calculation of the minimum phase function $u_{\mathbf{T}}$.

7.5.1 Polynomial coefficients

We introduce next a linear map $\mathcal{P}_N$ that associated the vector of coefficients $P_B \in \mathbb{R}^{(N+2)(N+1)/2}$ to each symmetric matrix $\mathbf{T}_P \in \mathbb{S}^N$. Note the set $\mathbb{S}^N$ and $\mathbb{R}^{(N+2)(N+1)/2}$ are isomorphisms. We define the map $\mathcal{P}_N$ which maps the set $\mathbb{R}^{(N+2)(N+1)/2}$ onto $\mathbb{R}^{2N+1}$.

Definition 7.5.1 (Computation of the polynomial coefficients from a Gram matrix associated to it.)**.** *Define $\mathcal{P}_N$ as the linear map*

$$\mathcal{P}_N : \mathbb{R}^{(N+2)(N+1)/2} \longrightarrow \mathbb{R}^{2N+1},$$

that given a Gram matrix $\mathbf{T}_P \in \mathbb{S}^N$, associate to the vector $\text{tri}(\mathbf{T}_P) \in \mathbb{R}^{(N+2)(N+1)/2}$ the column vector $P_B \in \mathbb{R}^{2N+1}$ with the coefficients vector $P_B = [t_{2N}, \cdots, t_0]^T$ with respect to the Tchebyshev basis such that

$$B_{2N}(\lambda)^T \cdot P_B = B_N(\lambda)^T \mathbf{T}_P B_N(\lambda) \qquad \forall \lambda \in \mathbb{C},$$

where $B_N(\lambda)$ is the Tchebyshev basis of degree N evaluated at λ. The elements t_k are obtained as

$$\begin{aligned} 2t_k &= \text{atr}_k(\mathbf{T}_P) && N \leq k \leq 2N, \\ 2t_k &= \text{atr}_k(\mathbf{T}_P) + 2\text{tr}_k(\mathbf{T}_P) && 1 \leq k < N, \\ 2t_0 &= \text{atr}_0(\mathbf{T}_P) + \text{tr}_0(\mathbf{T}_P). \end{aligned}$$

Furthermore if we denote $\mathbf{\Xi}(\mathcal{P}_N)$ the $(2N+1) \times (N+2)(N+1)/2$ matrix associated to the map $\mathcal{P}_N$ we have

$$B_N(\lambda)^T \mathbf{T}_P B_N(\lambda) = B_{2N}(\lambda)^T \cdot \mathbf{\Xi}(\mathcal{P}_N) \cdot \text{tri}(\mathbf{T}_P) \qquad \forall \lambda \in \mathbb{C}. \tag{7.18}$$

Example 7.5.1. *Consider the matrix:*

$$\mathbf{T}_P = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ a_2 & a_5 & a_6 & a_7 \\ a_3 & a_6 & a_8 & a_9 \\ a_4 & a_7 & a_9 & a_{10} \end{pmatrix},$$

then the vector $P_B = \mathcal{P}_4(\mathbf{T}) = [t_6, t_5, t_4, t_3, t_2, t_1, t_0]^T$ is defined as:

$$\begin{aligned} 2t_6 &= a_1, & t_3 &= 0.5(a_4 + a_6 + a_6 + a_4) + a_4, \\ 2t_5 &= a_2 + a_2, & t_2 &= 0.5(a_7 + a_8 + a_7) + a_3 + a_7, \\ 2t_4 &= a_3 + a_5 + a_3, & t_1 &= 0.5(a_9 + a_9) + a_2 + a_6 + a_9, \\ & & t_0 &= 0.5(a_{10} + a_1 + a_5 + a_8 + a_{10}). \end{aligned}$$

As an illustrating example, in this case we have:

$$\mathbf{\Xi}(\mathcal{P}_4) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 4 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 4 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 2 \end{pmatrix}.$$

It is important to highlight the fact that the vector of coefficients of the polynomial $P \in \mathbb{P}_+^{2N}$ can be obtained from the matrix $\mathbf{T}_P$ by means of the application $\mathcal{P}_N$ as

$$P_B = \mathbf{\Xi}(\mathcal{P}_N) \cdot \text{tri}\mathbf{T}_P.$$

7.5.2 Polynomial factorisation

Once in disposition of the polynomial $P \in \mathbb{P}_+^{2N}$, it is necessary to obtain the polynomial $p \in \mathbb{P}^N$ such that

$$p^\star p = P.$$

What we could think about trivially is the computation of the polynomial p as the factorization of P by calculating its roots as

$$P = \prod_{i=1}^{N} (\lambda - \xi_i)(\lambda - \overline{\xi_i})$$

to later select those N roots inside analyticity domain $\xi_i \in \overline{\mathbb{C}^-}$. However, this procedure has several drawbacks. The first one appears due to the choice of the Tchebyshev basis to represents the polynomials and computing roots of a polynomial defined in terms of the basis of Tchebyshev polynomials of degree up to $2N$ is not immediate, the second one is the fact that roots computation is an unstable operation for polynomials of high degree.

As an alternative to the roots computation, we apply the method proposed in [40], where a newton optimisation is used to compute the polynomial $p(\omega)$. We are looking for the polynomial $p \in \mathbb{P}^N$ of minimum phase such that $p^\star p = P$ with $P \in \mathbb{P}_+^{2N}$. Hence, given the polynomial p_i of minimum phase, we define the function $f(p_i)$ as

$$[f(p_i)](\omega) = p_i^\star(\omega) p_i(\omega) = P_i(\omega),$$

at each frequency ω. Compute now the best linear approximation of the function f given by the derivative $Df(p_i)$ with respect to the coefficients of p_i at any frequency point ω. We have

$$Df(p_i)\left[p_{i+1} - p_i\right] = \left[f(p_{i+1}) - f(p_i)\right],$$

when $p_{i+1} \to p_i$. Additionally we have the following lemma

Lemma 7.5.1. *Let $p_i \in \mathbb{P}^N$ be a polynomial of minimum phase with $p_i(\omega) \neq 0$ for all $\omega \in \mathbb{R}$, and $P_i = f(p_i) = p_i^\star p_i$. Given a positive polynomial $P \in \mathbb{P}_+^{2N}$, there exist a polynomial P_{i+1} in the form*

$$P_{i+1} = \kappa P + (1-\kappa)P_i,$$

with $0 < \kappa \leq 1$ such that the polynomial p_{i+1} defined as

$$p_{i+1} = p_i + Df(p_i)^{-1}(P_{i+1} - P_i), \tag{7.19}$$

where is of minimum phase as well

Proof. Let us write eq. (7.19) in the form

$$p_{i+1} = p_i + Df(p_i)^{-1}(\kappa P_{i+1} - \kappa P_i) = p_i + \kappa Df(p_i)^{-1}(P_{i+1} - P_i).$$

Then the proof follows directly from Rouche's theorem [41, corollary to theorem 18] as if $p_i(\omega) \neq 0$ for all $\omega \in \mathbb{R}$, we can take a value κ small enough such that

$$|p_i(\omega)| \geq \kappa |Df(p_i)^{-1}(\omega)(P_{i+1}(\omega) - P_i(\omega))| \qquad \forall \omega \in \mathbb{R}.$$

Hence the function $p_i(\omega)$ has the same number of zeros inside the analyticity domain $\mathbb{C}^-$ as the function $p_i(\omega) + \kappa Df(p_i)^{-1}(\omega)(P_{i+1}(\omega) - P_i(\omega))$. Note that since both functions are polynomials of the same degree, if $p_i(\omega)$ is of minimum phase, then p_{i+1} is of minimum phase as well. □

Using the iterative algorithm indicated in lemma 7.5.1 we can obtain the minimum phase factorisation of a polynomial $P \in \mathbb{P}_+^{2N}$. Furthermore it should be noted that even if the polynomial P vanishes at a point $\omega_0 \in \mathbb{R}$, the previous algorithm provides as with a polynomial $p_i \in \mathbb{P}^N$ such that $P_i = p_i^\star p_i$ is arbitrarily close to the polynomial P.

Note that in the decomposition $P = p^\star p$, polynomial p is unique up to an unimodular constant. To disambiguate we consider polynomial p is normalised at a reference frequency $\lambda_0 \in \mathbb{C}^-$ such that $\Im p(\lambda_0) = \eta$. Denote now by $p_B \in \mathbb{R}^{2N+1}$ the vector $[a_N, a_{N-1}, \cdots, a_0, b_N, b_{N-1}, \cdots, b_0]^T$ with the coefficients of the polynomial $p \in \mathbb{P}^N$ in the Tchebyshev basis such that

$$p(\lambda) = \sum_{i=0}^{N} (a_i + jb_i)\ \Psi_i(\omega). \tag{7.20}$$

We can then define a map between the vector $p_B \in \mathbb{R}^{2N+2}$ and the pair $(P_B, \eta) \in \mathbb{R}^{2N+1} \times \mathbb{R}$ where $\eta = \Im p(\lambda_0)$

$$\mathcal{Q} : p_B \in \mathbb{R}^{2N+1} \longrightarrow (P_B, \eta) \in \mathbb{R}^{2N+1} \times \mathbb{R}. \tag{7.21}$$

7.5.2.1 Computing derivatives

It should be noted $\mathcal{Q}$ is not a linear application. However, we can compute the matrix associated to the linearisation of the application $\mathcal{Q}$ at the point p_B, namely the Jacobian

matrix $\mathbf{J}_{\mathcal{Q}}(p_B)$. This Jacobian matrix presents the structure shown in eq. (7.22).

$$\mathbf{J}_{\mathcal{Q}}(p_B) = \begin{pmatrix} a_N & & & & & b_N & & & & \\ a_{N-1} & a_N & & & & b_{N-1} & b_N & & & \\ \vdots & \vdots & \ddots & & & \vdots & \vdots & \ddots & & \\ a_1 & a_2 & \cdots & a_N & & b_1 & b_2 & \cdots & b_N & \\ 2a_0 & a_1 & \cdots & a_{N-1} & a_N & 2b_0 & b_1 & \cdots & b_{N-1} & b_N \\ a_1 & 2a_0 & \cdots & a_{N-2} & a_{N-1} & b_1 & 2b_0 & \cdots & b_{N-2} & b_{N-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{N-1} & a_{N-2} & \cdots & 2a_0 & a_1 & b_{N-1} & b_{N-2} & \cdots & 2b_0 & b_1 \\ a_N & a_{N-1} & \cdots & a_1 & 2a_0 & b_N & b_{N-1} & \cdots & b_1 & 2b_0 \\ 0 & 0 & \cdots & 0 & 0 & c_N & c_{N-1} & \cdots & c_1 & c_0 \end{pmatrix} \tag{7.22}$$

where $c_i = \Upsilon_i(\lambda_0)$ for all $i \in [0, N]$.

From [42] we can deduce that the application $\mathcal{Q}$ in eq. (7.21) is continuous and has differentiable inverse $\mathcal{Q}^{-1}$

$$\mathcal{Q}^{-1} : (P_B, \eta) \mapsto p_B : \Im(p_B(\lambda_0)) = \eta.$$

We apply now the iterative procedure introduced in lemma 7.5.1. Let us denote by $p^{(i)} \in \mathbb{P}^N$ the minimum phase factor obtained in the i-th iteration and define $P^{(i)} = p^{(i)\star} p^{(i)}$. Additionally denote $\eta^{(i)} = \Im p^{(i)}(\lambda_0)$. The factorisation of the polynomial P is performed by iterating over the following formula

$$p_B^{(i+1)} = p_B^{(i)} + \kappa \mathbf{J}_{\mathcal{Q}}(p_B)^{-1} \begin{bmatrix} , P_B - P_B^{(i)} \\ 0 - \eta^{(i)}, \end{bmatrix},$$

where the quantity

$$E = \begin{bmatrix} , P_B - P_B^{(i)} \\ 0 - \eta^{(i)}, \end{bmatrix},$$

represents the error made in the i-th iteration. Note that the termination criterium can be given either by reaching an error smaller than a prescribed target value or by the conditioning of the matrix $\mathbf{J}_{\mathcal{Q}}(p_B)$ in the case that the polynomial P vanishes at some point $\omega_0 \in \mathbb{R}$. We can now compute the minimum phase factor of the function $u_P(\omega)$ whose modulus squared is defined as

$$|u_P(\omega)|^2 = \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R}, \tag{7.23}$$

by selecting the minimum phase factor of both numerator and denominator of eq. (7.23).

7.5.3 Derivatives of the spectral factorisation

Let us consider now the application that associates to each polynomial $P \in \mathbb{P}_+^{2N}$, the minimum phase polynomial $q \in \mathbb{P}^N$ such that $q^\star q = P + R$ and $\Im q(\lambda_0) = 0$, namely $\mathcal{Q}(P + R, 0)$.

After performing the factorisation with the previous or any other method, we obtain the function $u_P(\omega)$ as

$$u_P(\omega) = \frac{p(\omega)}{q(\omega)},$$

where $q^\star(\omega)q(\omega) = p^\star(\omega)p(\omega) + R(\omega)$. Note the derivatives of $P = p^\star p$ with respect to the k-th coefficient of p are given by $\mathbf{J}_\mathcal{Q}(p_B)$. Similarly if we denote by q_B the vector of coefficient of q in the Tchebyshev basis, the derivatives of the positive polynomial $q^\star q$ are given by $\mathbf{J}_\mathcal{Q}(q_B)$. Additionally we have

$$\mathrm{D}(q^\star q) = \mathrm{D}(P + R) = \mathrm{D}P.$$

Therefore

$$\begin{aligned} \mathbf{J}_\mathcal{Q}(q_B)\mathrm{D}_k q &= \mathrm{D}_k P, \\ \mathrm{D}_k q &= \mathbf{J}_\mathcal{Q}(q_B)^{-1}\mathrm{D}_k P. \end{aligned} \tag{7.24}$$

The previous formulation shown that the Jacobian matrix of the map $\mathcal{Q}^{-1}(P + R, 0)$ is given by $\mathbf{J}_\mathcal{Q}(q_B)^{-1}$.

Similarly, we can compute the second derivative of q by derivating again the polynomial P with respect to its l-th coefficient we have

$$\mathrm{D}^2 P = \mathrm{D}_l p \mathrm{D}_k p^\star + p\mathrm{D}^2_{k,l} p^\star + \mathrm{D}_l p^\star \mathrm{D}_k p + p^\star \mathrm{D}^2_{l,k} p = 2\Re\left(\mathrm{D}_l p^\star \mathrm{D}_k p + p^\star \mathrm{D}^2_{l,k} p\right).$$

Thus

$$2\Re\left(\mathrm{D}_l q^\star \mathrm{D}_k q + q^\star \mathrm{D}^2_{l,k} q\right) = 0.$$

Remark now that the matrix $\mathbf{J}_\mathcal{Q}(p)$ represents the linearisation of the application $p \mapsto p^\star p$ in a neighbourhood of the point p. Therefore, if we denote now by $\mathrm{D}_k p$ and $\mathrm{D}_l p$ the derivative of the polynomial p with respect to the k-th and l-th coefficients respectively, then the coefficients of the product $\mathrm{D}_l p^\star \mathrm{D}_k p$ are obtained by the matrix operation $\mathbf{J}_\mathcal{Q}(\mathrm{D}_l p)\cdot \mathrm{D}_k p_B$. Similarly at the point p, the product by $q^\star$ is computed by means of the matrix $\mathbf{J}_\mathcal{Q}(p)$. Thus

$$2\left[\mathbf{J}_\mathcal{Q}(\mathrm{D}_l q)\mathrm{D}_k q + \mathbf{J}_\mathcal{Q}(\mathrm{D}^2_{l,k} q)\right] = 0.$$

Moreover introducing eq. (7.24) we have

$$\begin{aligned} \mathbf{J}_\mathcal{Q}(\mathrm{D}_l q)\mathbf{J}_\mathcal{Q}(q_B)^{-1}\mathrm{D}_k P + 2\mathbf{J}_\mathcal{Q}(p)\mathrm{D}^2_{l,k} q = 0, \\ \mathrm{D}^2_{l,k} q = -\frac{1}{2}\mathbf{J}_\mathcal{Q}(p)^{-1}\mathbf{J}_\mathcal{Q}(\mathrm{D}_l q)\mathbf{J}_\mathcal{Q}(q_B)^{-1}\mathrm{D}_k P. \end{aligned}$$

Finally, we can evaluate such derivatives at the points α_i by means of the basis vector $E_N^C(\alpha_i)$ which contains real and imaginary coefficients as

$$E_N^C(\alpha_i) = \begin{bmatrix} E_N(\alpha_i) \\ jE_N(\alpha_i) \end{bmatrix}.$$

We obtain the following expressions referring to a scalar quantity, namely the derivative of the evaluation $q(\alpha_i)$ with respect to the k-th and the l-th coefficients of P.

$$\begin{aligned} \mathrm{D}_k q(\alpha_i) &= \frac{E_N^C(\alpha_i)^T}{2} \mathbf{J}_{\mathcal{Q}}(q_E)^{-1} \mathrm{D}_k p, \\ \mathrm{D}_{l,k}^2 q(\alpha_i) &= -\frac{E_N^C(\alpha_i)^T}{2} \mathbf{J}_{\mathcal{Q}}(p)^{-1} \mathbf{J}_{\mathcal{Q}}(\mathrm{D}_l q) \mathbf{J}_{\mathcal{Q}}(q_E)^{-1} \mathrm{D}_k P. \end{aligned}$$

Note that the derivative of the vector of coefficients of p with respect to the k-th coefficient is just the vector with 1 in the k-th position and zeros everywhere else. Additionally if we consider the derivative with respect of each of the coefficients of p, namely we set $\mathrm{D}_k P$ equal to the identity matrix of size $(2N+2) \times (2N+2)$ we obtain the gradient of the evaluation $q(\alpha_i)$

$$G_{q(\alpha_i)} = \frac{E_N^C(\alpha_i)^T}{2} \mathbf{J}_{\mathcal{Q}}(q_E)^{-1}. \tag{7.25}$$

Similarly the Hessian matrix of $q(\alpha_i)$ is obtained by taking every element $\mathrm{D}_l p$ and $\mathrm{D}_l q$

$$\mathbf{H}_{q(\alpha_i)} = -\begin{bmatrix} \frac{E_N^C(\alpha_i)^T}{2} \mathbf{J}_{\mathcal{Q}}(p)^{-1} \mathbf{J}_{\mathcal{Q}}(\mathrm{D}_1 q) \mathbf{J}_{\mathcal{Q}}(q_E)^{-1} \\ \frac{E_N^C(\alpha_i)^T}{2} \mathbf{J}_{\mathcal{Q}}(p)^{-1} \mathbf{J}_{\mathcal{Q}}(\mathrm{D}_2 q) \mathbf{J}_{\mathcal{Q}}(q_E)^{-1} \\ \vdots \\ \frac{E_N^C(\alpha_i)^T}{2} \mathbf{J}_{\mathcal{Q}}(p)^{-1} \mathbf{J}_{\mathcal{Q}}(\mathrm{D}_{2N+2} q) \mathbf{J}_{\mathcal{Q}}(q_E)^{-1} \end{bmatrix}.$$

Note we have just computed the derivatives of the map that associates to each polynomial P the corresponding stable polynomial q such that $q^\star q = P + R$ with respect to P while the derivatives with respect to the vector $\mathrm{tri}\mathbf{T}_P$ are required instead. Nevertheless the coefficient vector P_E is obtained as $P_E = \mathbf{\Xi}(\mathcal{P}_N)\mathrm{tri}\mathbf{T}_P$. Therefore if we denote by $G_{\mathbf{T}}$ and $\mathbf{H}_{\mathbf{T}}$ the gradient vector and the Hessian matrix respectively of $q(\alpha_i)$ with respect to the vector $\mathrm{tri}\mathbf{T}_P$, then we have

$$\begin{aligned} G_{\mathbf{T}} q(\alpha_i) &= G_{q(\alpha_i)} \cdot \mathbf{\Xi}(\mathcal{P}_N), \\ \mathbf{H}_{\mathbf{T}} q(\alpha_i) &= \mathbf{\Xi}(\mathcal{P}_N)^T \cdot \mathbf{H}_{q(\alpha_i)} \cdot \mathbf{\Xi}(\mathcal{P}_N). \end{aligned}$$

Finally, if we repeat the previous computation for the derivatives of p with respect to the coefficients of $P = p^\star p$ we obtain

$$\begin{aligned} G_{\mathbf{T}} p(\alpha_i) &= G_{p(\alpha_i)} \cdot \mathbf{\Xi}(\mathcal{P}_N), \\ \mathbf{H}_{\mathbf{T}} p(\alpha_i) &= \mathbf{\Xi}(\mathcal{P}_N)^T \cdot \mathbf{H}_{p(\alpha_i)} \cdot \mathbf{\Xi}(\mathcal{P}_N), \end{aligned}$$

where

$$G_{p(\alpha_i)} = \frac{\cancel{B}_N^C(\alpha_i)^T}{2}\mathbf{J}_{\mathcal{Q}}(p)^{-1}$$

$$\mathbf{H}_{p(\alpha_i)} = -\begin{bmatrix} \frac{\cancel{B}_N^C(\alpha_i)^T}{2}\mathbf{J}_{\mathcal{Q}}(p)^{-1}\mathbf{J}_{\mathcal{Q}}(\mathrm{D}_1 p)\mathbf{J}_{\mathcal{Q}}(p)^{-1} \\ \frac{\cancel{B}_N^C(\alpha_i)^T}{2}\mathbf{J}_{\mathcal{Q}}(p)^{-1}\mathbf{J}_{\mathcal{Q}}(\mathrm{D}_2 p)\mathbf{J}_{\mathcal{Q}}(p)^{-1} \\ \vdots \\ \frac{\cancel{B}_N^C(\alpha_i)^T}{2}\mathbf{J}_{\mathcal{Q}}(p)^{-1}\mathbf{J}_{\mathcal{Q}}(\mathrm{D}_{2N+2} p)\mathbf{J}_{\mathcal{Q}}(p)^{-1} \end{bmatrix}.$$

7.5.4 Derivatives of the matrix $\mathbf{U_T}$

We compute next the first and second derivatives of the function $u_P(\alpha_i)$ with respect to polynomial P. Note that we have already computed the derivatives of $p(\alpha_i)$ and $q(\alpha_i)$ with respect of the vector tri$\mathbf{T}_P$ that we can use together with the chain rule to obtain the derivatives of $u_P(\alpha_i)$ with respect to tri$\mathbf{T}_P$. We have

$$\begin{aligned} \mathrm{D}_k u_P &= \frac{\mathrm{D}_k p}{q} - u_P \mathrm{D}_k q, \\ \mathrm{D}^2_{k,l} u_P &= \frac{\mathrm{D}^2_{k,l} p - u_P \mathrm{D}^2_{k,l} q}{q} - \frac{\mathrm{D}_k p \mathrm{D}_l q + \mathrm{D}_l p \mathrm{D}_k q}{q^2} + 2u_P \frac{\mathrm{D}_k q \mathrm{D}_l q}{q^3}. \end{aligned} \tag{7.26}$$

If we consider now the element (i, h) of the matrix $\mathbf{U_T}$ with $i, h \in [1, N]$ we have

$$[\mathbf{U_T}]_{i,h} = \frac{u_{\mathbf{T}}(\alpha_i)\overline{u_{\mathbf{T}}(\alpha_h)}}{\alpha_i - \overline{\alpha_h}},$$

$$\mathrm{D}_k[\mathbf{U_T}]_{i,h} = \frac{\mathrm{D}_k u_{\mathbf{T}}(\alpha_i)\overline{u_{\mathbf{T}}(\alpha_h)} + u_{\mathbf{T}}(\alpha_i)\overline{\mathrm{D}_k u_{\mathbf{T}}(\alpha_h)}}{\alpha_i - \overline{\alpha_h}},$$

$$\begin{aligned} \mathrm{D}^2_{k,l}[\mathbf{U_T}]_{i,h} = &\frac{\mathrm{D}^2_{k,l} u_{\mathbf{T}}(\alpha_i)\overline{u_{\mathbf{T}}(\alpha_h)} + \mathrm{D}_k u_{\mathbf{T}}(\alpha_i)\overline{\mathrm{D}_l u_{\mathbf{T}}(\alpha_h)}}{\alpha_i - \overline{\alpha_h}} + \\ &+ \frac{\mathrm{D}_l u_{\mathbf{T}}(\alpha_i)\overline{\mathrm{D}_k u_{\mathbf{T}}(\alpha_h)} + u_{\mathbf{T}}(\alpha_i)\overline{\mathrm{D}^2_{k,l} u_{\mathbf{T}}(\alpha_h)}}{\alpha_i - \overline{\alpha_h}}. \end{aligned}$$

The computed formulas provides us with the first and second derivatives of each element of the matrix $\mathbf{U_T}$ with respect to the vector tri$\mathbf{T}_P$ namely the coefficients in the lower triangle of the matrix $\mathbf{T}_P$. These derivatives result of vital importance in the next chapter for the numerical solution of problem 7.4.2, specially given the non-linearity of the matrix $\mathbf{U_T}$.

References

[39] B. Dumitrescu, *Positive Trigonometric Polynomials and Signal Processing Applications*, 2017.

[40] H. J. Orchard and A. N. Willson, "On the computation of a minimum-phase spectral factor," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 50, no. 3, pp. 365–375, 2003.

[41] L. V. Ahlfors, *Complex analysis*, 3rd ed. McGraw-Hill Education, 1966. [Online]. Available: https://books.google.fr/books?id=RfYK28TcZEwC

[42] L. Baratchart, M. Olivi, and F. Seyfert, "Boundary Nevanlinna-Pick interpolation with prescribed peak points. Application to impedance matching," *SIAM Journal on Mathematical Analysis*, 2017.

Chapter 8:

Computation of the optimal solution

Semi-definited programs have become popular in recent years as a result of the appearance of interior-point optimization methods. Interior-point optimisation allows to solve problems with matrix inequalities constraints by guaranteeing the feasibility of the problem throughout the optimization. In addition, if the function to be minimized or maximized, as well as the feasible region, are convex; interior-point optimization methods represent a practical tool to compute numerically the optimal solution to the problem, which was not evident before the appearance of such methods.

Among the positive definite problems that have gained attention with the appearance of the interior-point methods out-stand linear matrix inequality programs or LMI. These programs consist of the minimization or maximization of a linear criterion under a series of matrix inequalities. In addition, these matrices must depend linearly on the parameters of the problem. To obtain the solution of these problems, barrier functions come to the rescue. In particular, logarithmic barrier functions are of special importance in the case of LMI programs.

We return now to the semi-definited program obtained at the end of chapter 7. This program is characterized as a convex optimization problem, therefore any local solution is a global solution. Particularly, we are facing a problem of the type NLSDP, namely a non-linear semi-definited program. NLSDP are still among the types of problems that can be solved optimally in practice by numerical methods. However, this resolution becomes substantially more complex than in the linear case. Indeed, non-linear semi-definite programs might represent, within the field of convex problems, the most complex type of problem that can, at the present time, be solved numerically with guarantee of optimality.

In this chapter we provide an overview of the numerical implementation of the NLSDP problem derived in chapter 7, including a review of interior-point optimization methods. For this matter, we begin with a motivation for the use of the said methods in the resolution of convex problems subject to matrix inequalities.

An important remark before continuing, is the fact that the rest of this chapter is dedicated solely to the numerical resolution of the optimization problem formulated at the end of the previous section. This resolution could be done with the help of one of the commercial solvers available, although only a couple of them exist for the time being due to the non-linearity of the matrix inequality in eq. (8.11). In any case, in the following sections we provide an overview of the optimization methods used to solve the aforementioned type of problems.

In addition, although we did not perform any specific study on, for example, the order of the number of operations necessary to obtain the said solution with a certain precision, or the calculation time with respect to the size of the problem or the number of inequalities, what is provided is a detailed review of the algorithm implemented to calculate the desired solution. This information will be appreciated by a non-conformist reader who does not settle for simply pressing a red button on a black box and waiting for the solution to appear on screen but needs to know what is actually happening inside.

For the reader convenience, let us restate here problem 7.4.2. This problem, already

formulated as a non-linear SDP, presents different matrix constraints of every possible type that can be imposed while keeping the convexity property. As a result, the complexity of the problem is maximised as a barrier function of different nature is conceived to handle each one of those constraints. Furthermore, equality constrains are also imposed in problem 7.4.2, providing a complete illustrating example of non-linear SDP.

Problem 7.4.2 (Generalised matching SDP)**.**

$$
\begin{aligned}
\textit{Find:} \quad & \min_{\substack{\Psi\ \mathbf{T}_P \\ \mathbf{M}_\Psi^{(i)},\mathbf{N}_\Psi^{(i)} \\ \mathbf{M}_\Gamma^{(k)},\mathbf{N}_\Gamma^{(k)}}} \Psi, \\
\textit{where:} \quad & \Psi \in \mathbb{R}_+ \qquad (\mathbf{M}_\Psi^{(i)},\mathbf{N}_\Psi^{(i)}) \in \mathbb{S}^{N+1}\times\mathbb{S}^{N} \qquad \forall i \in [1,n], \\
& \mathbf{T}_P \in \mathbb{S}^{N+1} \qquad (\mathbf{M}_\Gamma^{(k)},\mathbf{N}_\Gamma^{(k)}) \in \mathbb{S}^{N+1}\times\mathbb{S}^{N} \qquad \forall k \in [1,m], \\
\textit{Subject to:} \quad & \mathbf{T}_P \succeq 0, && (7.12)\\
& \mathbf{M}_\Psi^{(i)} \succeq 0 \qquad \forall i \in [1,n], && (7.13)\\
& \mathbf{N}_\Psi^{(i)} \succeq 0 \qquad \forall i \in [1,n], && (7.14)\\
& \mathbf{M}_\Gamma^{(k)} \succeq 0 \qquad \forall k \in [1,m], && (7.15)\\
& \mathbf{N}_\Gamma^{(k)} \succeq 0 \qquad \forall k \in [1,m], && (7.16)\\
& \mathbf{U}(\mathbf{T}_P) \succeq \mathbf{\Pi}, && (7.17)
\end{aligned}
$$

with the additional equality constraints in eqs. (7.11a) *and* (7.11b).

$$
\forall i \in [1,n]: \quad L\mathbf{Б}_{2N}T_R - \mathbf{Б}_N{}^T\mathbf{T}_P\mathbf{Б}_N = \mathbf{Б}_N^{(i)^T}\mathbf{M}_\Psi^{(i)}\mathbf{Б}_N^{(i)} + \mathbf{X}^{(i)}\mathbf{Б}_{N-1}^{(i)^T}\mathbf{N}_\Psi^{(i)}\mathbf{Б}_{N-1}^{(i)}, \tag{7.11a}
$$

$$
\forall k \in [1,m]: \quad \mathbf{Б}_N{}^T\mathbf{T}_P\mathbf{Б}_N - \Gamma\mathbf{Б}_{2N}T_R = \mathbf{Б}_N^{(k)^T}\mathbf{M}_\Gamma^{(k)}\mathbf{Б}_N^{(k)} + \mathbf{X}^{(k)}\mathbf{Б}_{N-1}^{(k)^T}\mathbf{N}_\Gamma^{(k)}\mathbf{Б}_{N-1}^{(k)}. \tag{7.11b}
$$

It should be noted that the revision of the computational procedure provided here is supposed to be self-contained, namely no additional information should be required. Nevertheless, if the reader is further interested on other more advance aspects of interior points methods, I must recommend some other lectures which are more appropriate for this topic as [43, 44] which we find particularly enlightening. As you can already guess, in the rest of this chapter we do not continue with the study of the problem of matching itself or provide any results with respect to the aforementioned problem. We are exclusively dealing with the solution of this particular optimization problem, having to wait until chapter 9 to return to the problem of matching again.

Next, there is the list of each of the constraints present in problem 7.4.2 and treated in this chapter.

- Linear equalities (section 8.2): we use the elimination method to ensure that equal-

ities are satisfied reducing at the same time the number of parameter (dimension) of the problem. This technique ensures eqs. (7.11a) and (7.11b).

- LMI: linear matrix inequalities are handled by the classical logarithmic barrier function. We can distinguish between two different kind of these constraints
  - Strict feasible matrix inequalities (section 8.4.1): strict feasible inequalities need to be satisfied at every point during the whole optimisation process. This is the case of eq. (7.12) as if $\mathbf{T}_P \neq 0$ the matrix function $\mathbf{U}_{\mathbf{T}_P}$ is not defined and therefore admissibility can not be determined.
  - Non-strict matrix inequalities (section 8.4.2): eqs. (7.13) to (7.16) are constraints that could be violated during the optimisation process as the convexity of the problem does not required these to be satisfied. Note that even if any of the mentioned inequalities is not satisfied, the problem remain well defined. Nevertheless, non-strict inequalities still need to be satisfied by the optimal solution.
- NLMI (section 8.4.3): these are matrix inequalities where the matrix function depends non-linearly on the parameters of the problem as it is the case, in this problem, of eq. (7.17). It must be assumed that non-linear matrix inequalities are non-strict. With the last assumption, this case still represents the most complicated kind of constraint that one can consider in SDP.

Note we have not spoken here about non-linear equality constraints. However, since no practical benefit can be obtained from these, conversely to the case of linear equalities which allow to reduce the dimensionality of the problem. Therefore we can simply implement each of the non-linear equalities such as

$$B = A^T \cdot X,$$

by means of two non-linear inequalities as

$$B \geq A^T x,$$
$$B \leq A^T x.$$

Next we review in more detail each of the listed constraints, but not without stating before problem 7.4.2 in an standard form.

8.1 Semi-definite program statement: criterium and objective function

We restate now the previous SDP problem as a minimisation over a vector $X \in \mathbb{R}^K$ of a convex function $f(X)$. This minimisation is constrained by a set of matrix inequalities where each matrix function depends on the vector X. Furthermore we have a given number of linear equality constraints in the form $\mathbf{A}X = B$.

First, we parametrise each of the linear matrices in problem 7.4.2 by the elements in the lower triangle, using the symmetry property to reduce the number of parameters.

We obtain a minimisation problem with respect to the column vector $X \in \mathbb{R}^K$ defined in eq. (8.1).

$$X = \begin{bmatrix} \Psi \\ \text{tri}\mathbf{T}_P \\ \text{tri}\mathbf{M}_\Psi^{(1)} \\ \vdots \\ \text{tri}\mathbf{M}_\Psi^{(n)} \\ \text{tri}\mathbf{N}_\Psi^{(1)} \\ \vdots \\ \text{tri}\mathbf{N}_\Psi^{(n)} \\ \text{tri}\mathbf{M}_\Gamma^{(1)} \\ \vdots \\ \text{tri}\mathbf{M}_\Gamma^{(m)} \\ \text{tri}\mathbf{N}_\Gamma^{(1)} \\ \vdots \\ \text{tri}\mathbf{N}_\Gamma^{(m)} \end{bmatrix} \tag{8.1}$$

Note if we have a matrix $\mathbf{T}_P \in \mathbb{S}^{N+1}$ then the number of coefficients in the lower triangle is equal to $(N+2)(N+1)/2$. Hence we have $\text{tri}\mathbf{T}_P \in \mathbb{R}^{(N+2)(N+1)/2}$. Applying the same relation to each of the matrices composing the vector X through the tri function, we can compute the size K of such vector $X \in \mathbb{R}^K$.

$$\begin{aligned} K &= 1 + \frac{(N+2)(N+1)}{2} + (n+m)\frac{(N+2)(N+1)}{2} + (n+m)\frac{(N+1)N}{2}, \\ &= 1 + \frac{(N+2)(N+1)}{2} + (n+m)(N+1)^2. \end{aligned}$$

Having all matrices parametrised by X, the criterium of problem 7.4.2 becomes

$$\min_{\substack{\Psi\ \mathbf{T}_P \\ \mathbf{M}_\Psi^{(i)},\mathbf{N}_\Psi^{(i)} \\ \mathbf{M}_\Gamma^{(k)},\mathbf{N}_\Gamma^{(k)}}} \Psi = \min_{X \in \mathbb{R}^K} f(X),$$

where, in this case $f(X)$ is a linear function, namely the value of Ψ, and can be expressed as

$$f(X) = C^T X,$$

with $C, X \in \mathbb{R}^K$. The vector C is the vector with a value 1 in the first position and zeros everywhere else.

$$C = \begin{bmatrix} 1 \\ 0_{K-1} \end{bmatrix}.$$

8.2 Linear equalities

We rewrite now the equality constraints eqs. (7.11a) and (7.11b) in function of the vector X. First note that we can express the evaluation of the gram matrix $\mathbf{T}_P$ in terms of the vector $\mathrm{tri}\mathbf{T}_P$ by using the map $\mathcal{P}_N$ as in eq. (7.18).

$$\mathbf{Б}_N{}^T \mathbf{T}_P \mathbf{Б}_N = \mathbf{Б}_{2N}{}^T \mathbf{Ξ}(\mathcal{P}_N) \cdot \mathrm{tri}\mathbf{T}_P.$$

Then we have for all $i \in [1, n]$ and for all $k \in [1, m]$

$$\begin{aligned}
\mathbf{Б}_{2N} T_R \Psi - \mathbf{Б}_{2N}{}^T \mathbf{Ξ}(\mathcal{P}_N) \mathrm{tri}\mathbf{T}_P &= \mathbf{Б}_{2N}^{(i)^T} \mathbf{Ξ}(\mathcal{P}_N) \mathrm{tri}\mathbf{M}_\Psi^{(i)} + \mathbf{X}_\Psi^{(i)} \mathbf{Б}_{2N-2}^{(i)^T} \mathbf{Ξ}(\mathcal{P}_{N-1}) \mathrm{tri}\mathbf{N}_\Psi^{(i)}, \\
\mathbf{Б}_{2N}{}^T \mathbf{Ξ}(\mathcal{P}_N) \mathrm{tri}\mathbf{T}_P - \Gamma \mathbf{Б}_{2N} T_R &= \mathbf{Б}_{2N}^{(k)^T} \mathbf{Ξ}(\mathcal{P}_N) \mathrm{tri}\mathbf{M}_\Gamma^{(k)} + \mathbf{X}_\Gamma^{(k)} \mathbf{Б}_{2N-2}^{(k)^T} \mathbf{Ξ}(\mathcal{P}_{N-1}) \mathrm{tri}\mathbf{N}_\Gamma^{(k)}.
\end{aligned}$$

Let us denote $\mathbf{Ш} = \mathbf{Б}_{2N}{}^T \mathbf{Ξ}(\mathcal{P}_N)$ as well as

$$\begin{aligned}
\mathbf{Ш}_\Psi^{(i)} &= \mathbf{Б}_{2N}^{(i)^T} \mathbf{Ξ}(\mathcal{P}_N) & \mathbf{Ц}_\Psi^{(i)} &= \mathbf{X}_\Psi^{(i)} \mathbf{Б}_{2N-2}^{(i)^T} \mathbf{Ξ}(\mathcal{P}_{N-1}) & \forall i \in [1, n], \\
\mathbf{Ш}_\Gamma^{(k)} &= \mathbf{Б}_{2N}^{(k)^T} \mathbf{Ξ}(\mathcal{P}_N) & \mathbf{Ц}_\Gamma^{(k)} &= \mathbf{X}_\Gamma^{(k)} \mathbf{Б}_{2N-2}^{(k)^T} \mathbf{Ξ}(\mathcal{P}_{N-1}) & \forall k \in [1, m],
\end{aligned}$$

and then write it in the form $\mathbf{A}X = B$ to obtain eq. (8.2).

Problem 8.2.1 (Non-linear SDP)**.**

$$\begin{aligned}
\textit{Find:} \quad & \min_{X \in \mathbb{R}^K} C^T X, \\
\textit{Subject to:} \quad & \mathbf{T}_P(X) \succeq 0, \\
& \mathbf{M}_\Psi^{(i)}(X) \succeq 0 && \forall i \in [1, n], \\
& \mathbf{N}_\Psi^{(i)}(X) \succeq 0 && \forall i \in [1, n], \\
& \mathbf{M}_\Gamma^{(k)}(X) \succeq 0 && \forall k \in [1, m], \\
& \mathbf{N}_\Gamma^{(k)}(X) \succeq 0 && \forall k \in [1, m], \\
& \mathbf{U}(X) \succeq \mathbf{Л}, \\
& \mathbf{A}X = B.
\end{aligned}$$

Problem 8.2.1 is already stated in a form allowing for the application of standard SDP solvers. Nevertheless it can still be further simplified for a matter of computational efficiency. This simplification is achieve by the elimination of the equality constraints $\mathbf{A}X = B$. Such elimination can be performed as long as an initial point X_0 satisfying the equality constraints $\mathbf{A}X_0 = B$ is known.

8.2.1 Initial point

It should be noted first that the initial point X_0 must also satisfy the strict feasible matrix inequalities, namely $\mathbf{T}_P(X_0) \succ 0$. Hence, even if problem 8.2.1 is convex, a fully arbitrary initial point cannot be used. Nevertheless assume the vector X can be divided in two sub-vectors $X_F \in \mathbb{R}^{K_F}, X_U \in \mathbb{R}^{K_U}$ where $K = K_F + K_U$ and

$$X = \begin{bmatrix} X_F \\ X_U \end{bmatrix}$$

$$
\underbrace{\begin{bmatrix}
-\text{Б}_{2N}T_R & \text{Ш} & \text{Ш}_\Psi^{(1)} & \cdots & \mathbf{0}_M & \mathbf{\Pi}_\Psi^{(1)} & \cdots & \mathbf{0}_N \\
\cdots & \cdots & \cdots & \ddots & \cdots & \cdots & \ddots & \cdots \\
-\text{Б}_{2N}T_R & \text{Ш} & \mathbf{0}_M & \cdots & \text{Ш}_\Psi^{(n)} & \mathbf{0}_N & \cdots & \mathbf{\Pi}_\Psi^{(n)} \\
0_{2N+1} & -\text{Ш} & \multicolumn{6}{c}{\multirow{3}{*}{$\mathbf{\Pi}\text{Ш}_\Gamma^{(1)} \quad \cdots \quad \mathbf{0}_M \quad \mathbf{\Pi}_\Gamma^{(1)} \quad \cdots \quad \mathbf{0}_N$}} \\
\cdots & \cdots & \multicolumn{6}{c}{} \\
0_{2N+1} & -\text{Ш} & \multicolumn{6}{c}{$\mathbf{0}_M \quad \cdots \quad \mathbf{\Pi}\text{Ш}_\Gamma^{(m)} \quad \mathbf{0}_N \quad \cdots \quad \mathbf{\Pi}_\Gamma^{(m)}$}
\end{bmatrix}}_{\mathbf{A}}
\underbrace{\begin{bmatrix}
\Psi \\
\text{triT}_P \\
\text{triM}_\Psi^{(1)} \\
\cdots \\
\text{triM}_\Psi^{(n)} \\
\text{triN}_\Psi^{(1)} \\
\cdots \\
\text{triN}_\Psi^{(n)} \\
\text{triM}_\Gamma^{(1)} \\
\cdots \\
\text{triM}_\Gamma^{(m)} \\
\text{triN}_\Gamma^{(1)} \\
\cdots \\
\text{triN}_\Gamma^{(m)}
\end{bmatrix}}_{X}
=
\underbrace{\begin{bmatrix}
0_{2N+1} \\
\cdots \\
0_{2N+1} \\
-\Gamma\text{Б}_{2N}T_R \\
\cdots \\
-\Gamma\text{Б}_{2N}T_R
\end{bmatrix}}_{B}
$$

$\mathbf{A}_F$ — $\mathbf{A}_U$ — X_F — X_U

$$\mathbf{0}_M = \mathbf{0}_{2N+1,(N+2)(N+1)/2}$$
$$\mathbf{0}_N = \mathbf{0}_{2N+1,(N+1)N/2}$$

Equation 8.2: Linear equalities

such that the strict matrices can be computed only from X_F. For any vector X_F such that strict inequalities are satisfied, $\mathbf{T}_P(X_F) \succ 0$, there exist $X_U \in \mathbb{R}^{K_U}$ such that

$$\mathbf{A} \cdot \begin{bmatrix} X_F \\ X_U \end{bmatrix} = B$$

as long as $K_U \geq (n+m)(2N+1)$. Dividing also the matrix $\mathbf{A} = [\mathbf{A_F}, \mathbf{A_U}]$ with $\mathbf{A_F}$ of size $(2N+1) \times (K_F)$ and $\mathbf{A_U}$ of size $(2N+1) \times (K_U)$ the linear equation system

$$\mathbf{A_U} \cdot X_U = B - \mathbf{A_F} \cdot X_F \tag{8.3}$$

is under-determined. Therefore we can just take any solution X_U to eq. (8.3).

Example 8.2.1. *Consider problem 8.2.1, we can take*

$$\Psi = 1,$$
$$\mathbf{T}_P = \mathbf{I}_M \succ 0,$$

with $\mathbf{I}_M$ *the* $(N+1) \times (N+1)$ *identity matrix. Then we have*

$$X_F = \begin{bmatrix} \Psi \\ \text{tri}\mathbf{T}_P \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tri}\mathbf{I}_M \end{bmatrix},$$

which provides us the equation system in eq. (8.4).

8.2.2 Equality elimination

We apply elimination method to ensure equalities constraints $\mathbf{A}X = B$ are satisfied while reducing the number of variables in the problem. Note we already dispose of an initial point $X_0 \in \mathbb{R}^K$ such that $\mathbf{A}X_0 = B$. If we now denote $Y = X - X_0$ we have the homogeneous system

$$\mathbf{A}Y = \mathbf{A}(X - X_0) = B - B = 0.$$

Considering the vector Y as the new problem variable, the equality constrains become

$$Y \in \ker(\mathbf{A}),$$

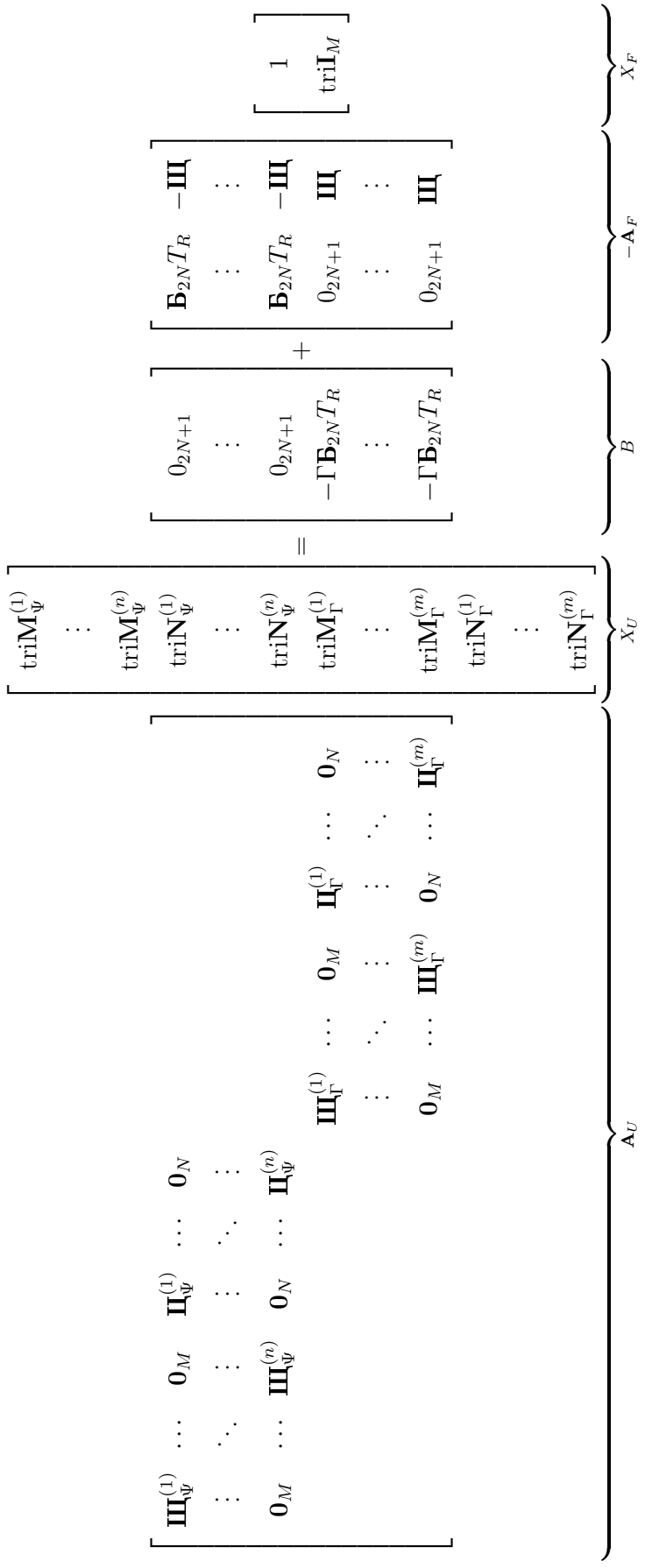
where $\ker(\mathbf{A})$ denotes the null-space of the linear application $Y \mapsto \mathbf{A}Y$. Moreover the vector Y belongs to $\ker \mathbf{A}$ if and only if it can be written as

$$Y = K_A W \qquad\qquad W \in \mathbb{R}^Z, \tag{8.5}$$

where K_A is a basis of $\ker \mathbf{A}$ and Z denotes the nullity of $\mathbf{A}$. Introducing the expression of Y in eq. (8.5) we have the parametrisation of the vector X

$$X = X_0 + K_A W,$$

which ensures the equality constraints $\mathbf{A}X = B$ are satisfied for every $W \in \mathbb{R}^Z$. Now we need an initial point W_0 such that the strict feasible matrix inequalities are satisfied. This is now trivial as the vector $X = X_0$ has been computed to satisfied those strict



Equation 8.4: Linear system for the computation of a feasible initial point

inequalities. Therefore we choose $W_0 = 0_Z$. With this new parametrisation in W, and assuming that the matrix $\mathbf{A}$ is of full rank, the dimension of the problem is reduced by a quantity equal to the number of equalities, namely $(n+m)(2N+1)$. Therefore we have

$$\begin{aligned} Z &= K - (n+m)(2N+1), \\ &= 1 + \frac{(N+2)(N+1)}{2} + (n+m)(N+1)^2 - (n+m)(2N+1), \\ &= 1 + \frac{(N+2)(N+1)}{2} + (n+m)N^2. \end{aligned}$$

We can now restate problem 8.2.1 without equality constraints.

Problem 8.2.2 (Non-linear SDP without equality constraints)**.**

$$\begin{aligned} \textit{Find:} \quad & \min_{W \in \mathbb{R}^Z} C^T K_A W, \\ \textit{Subject to:} \quad & \mathbf{T}_P(X_0 + K_A W) \succeq 0, && (8.6) \\ & \mathbf{M}_\Psi^{(i)}(X_0 + K_A W) \succeq 0 \qquad \forall i \in [1,n], && (8.7) \\ & \mathbf{N}_\Psi^{(i)}(X_0 + K_A W) \succeq 0 \qquad \forall i \in [1,n], && (8.8) \\ & \mathbf{M}_\Gamma^{(k)}(X_0 + K_A W) \succeq 0 \qquad \forall k \in [1,m], && (8.9) \\ & \mathbf{N}_\Gamma^{(k)}(X_0 + K_A W) \succeq 0 \qquad \forall k \in [1,m], && (8.10) \\ & \mathbf{U}(X_0 + K_A W) \succeq \mathbf{\Pi}. && (8.11) \end{aligned}$$

Next we begin with the introduction to the interior-point methods. Problem 8.2.2 is the minimization of a linear function, namely the function $f(W) = C^T K_A W$, with respect to the vector $W \in \mathbb{R}^Z$. This minimization is done within a subspace of $\mathbb{R}^Z$ determined by the positivity of a set of matrices which depend on the vector W. To determine the aforementioned subspace can be used many optimization techniques of varied nature.

In our case, we choose to use interior point methods what implies that every point W during the optimisation process is located inside a determined region, which as we show later on, does not necessarily coincide with the feasible region to the problem.

8.3 Introduction to interior point methods

Consider now a subspace $\mathbb{K} \subset \mathbb{R}^Z$ and suppose we are interested on minimising the function $f(W)$ within that subspace. This problem can be solved by folding the constraints of belonging to the set $\mathbb{K}$ into the criterium and then computing the solution to an unconstrained optimisation problem. The said unconstrained problem refers to the minimisation of the function $f(W) + h(W)$ where $h(W)$ is the hard barrier function that takes the value 0 if W belong to the inside of $\mathbb{K}$ and infinity if W is exactly on the boundary of $\mathbb{K}$. Therefore the solution to this unconstrained problem is also the solution to the problem of minimising $f(W)$ with $W \in \mathbb{K}$ since when W reaches the boundary of the feasible set the criterium grows infinitely what represents without any doubt a non

decreasing criterium.

This concept is analogous to having a *numeric barrier* on the boundary of the feasible set, where *numeric barrier* stands for an infinity criterium value. Nevertheless a function $h(W)$ that only takes the values 0 or ∞ depending on the vector W is absolutely non-differentiable and does not facilitates in any aspect the solution of the constrained problem. In order to overcome this issue we apply a trick used by mathematician each time this kind of unfriendly functions are faced, we approach it with a differentiable function. Hence, instead of the hard barrier function $h(W)$ we use a different function which is as flat as possible in the interior of the feasible domain while smoothly growing toward infinity as long as W approaches the boundary of the domain. This function is continuous on the domain $\mathbb{K}$ and at least twice differentiable, allowing for an efficient numerical computation of its fist and second derivatives, namely the gradient and the Hessian matrix with respect to the vector W.

Additionally, this smooth barrier function also depends on a parameter $t \in \mathbb{R}_+$ used to control the error in the approximation of the function $h(W)$. Let us denote here this function by $\beta_t(W)$. We have $\lim_{t\to\infty} \beta_t(W) = h(W)$ for all $W \in \mathbb{K}$ meanwhile for a small value of t the function function $\beta_t(W)$ is extremely well behaved. Therefore the barrier function is easy to minimize with a small value of t, say $t = 0$, for example by the classical Newton method, while, as the value of t increases, the said minimization becomes more complicated, requiring a greater number of iterations in the aforementioned Newton method.

Nevertheless note that since the solution to the optimization problem is easy to calculate for $t = 0$, if the value of t is increased by a small quantity ϵ, we have a new optimisation problem that is extremely close to the previous one. Then if the optimal solution to the problem with $t = 0$ is taken as initial point, we can make the problem with $t = \epsilon$ as easy to solve as necessary by taking the value of ϵ arbitrary small. Indeed the proofs showing the convergence of the interior points methods are based on the fact that if a value W_{opt} is the optimal solution to the problem for a given value of t, then you can always take ϵ small enough such that the point W_{opt} is still in the region of quadratic convergence of the classical Newton method. This process of gradually transforming the problem by means of a parameter t from a version of the problem that is trivially solvable to a final problem which is much more complicated is called homotopy.

We discuss next another vital property of the barrier function $\beta_t(W)$. Note when we are facing a highly not convex problem as it could be the case of our matching problem, we could consider such problem as solved from the moment when we are able to express such problem by using a convex formulation. Particularly, if you are able to formulate a convex problem and prove the unique local solution to the convex problem is global optimal solution to the original one, then you are done. The reason comes from the fact that in theory, the optimal solution to a convex problem can be easily found by any optimisation method as there is no local minima where we could get stuck.

However, if a non-convex barrier function $\beta_t(W)$ is used, the convexity of the problem we just came with is lost. Hence the work carried out to transform the original problem

onto a convex one was all for nothing. With this argument we try to emphasize the importance of the convexity of the barrier function $\beta_t(W)$. This is indeed a crucial property of the barrier function.

8.3.1 Lagrangian function

As mentioned above, barrier functions are used to include the constraints of the problem in the function to be optimized in such a way that the constrained problem becomes a non-constrained optimization problem. To provide a first notion of the type of problem we face, let us consider a simple inequality constrained problem.

Problem 8.3.1 (Constrained optimisation problem)**.**

$$\begin{aligned} \textit{Find:} \quad & \min_W f(W), \\ \textit{Subject to:} \quad & g(W) \geq 0. \end{aligned}$$

We denote W_{opt} the value of W providing the optimal criterium in problem 8.3.1. Now let us introduce the Lagrange function which is defined as

$$\Lambda(W, y) = f(W) - y \cdot g(W). \tag{8.12}$$

where $y \geq 0$ is called the Lagrange multiplier. We also define $W_{opt}^{(y)}$ by means of the minimisation of the Lagrangian $\Lambda(W, y)$ with respect to W

$$W_{opt}^{(y)} = \arg\min_W \Lambda(W, y).$$

Computing now the gradient of $\Lambda(W, y)$ with respect to W we have

$$\mathrm{D}_W \Lambda(W, y) = \mathrm{D}_W f(W) - y \mathrm{D}_W g(W),$$

which vanishes at the optimal point $W_{opt}^{(y)}$

$$\mathrm{D}_W \Lambda(W_{opt}^{(y)}, y) = 0 \qquad \forall y \geq 0.$$

The first thing to be noted here is the fact that, if the point $W_{opt}^{(y)}$ is feasible for problem 8.3.1, namely $g(W_{opt}^{(y)}) \geq 0$, we have $y g(W_{opt}^{(y)}) \geq 0$ and hence the value of the Lagrangian at the point $W_{opt}^{(y)}$ verifies

$$\Lambda(W_{opt}^{(y)}, y) \leq f(W_{opt}) \qquad \forall y \geq 0.$$

This can be seen intuitively for any $y \geq 0$. For instance, take $y = 0$, then $\Lambda(W_{opt}^{(0)}, 0)$ is the solution to problem 8.3.1 without any constraint, and therefore the criterium $f(W_{opt})$ including the constraint $g(W) \geq 0$ is necessarily worse. Similarly, if we take $y > 0$, simply by introducing the value W_{opt} into eq. (8.12), since $y \cdot g(W_{opt}^{(y)}) \geq 0$ we have

$$\Lambda(W_{opt}, y) = f(W_{opt}) - y \cdot g(W_{opt}) \leq f(W_{opt}), \tag{8.13}$$

where equality holds if and only if $g(W_{opt}) = 0$. Therefore the minimisation of the Lagrangian function in eq. (8.12) with any value $y \geq 0$ provides a lower bound on the solution to problem 8.3.1 as long as the obtained minimiser $W_{opt}^{(y)}$ is feasible for problem 8.3.1.

At this point, we find the classical Lagrangian dual problem which simply consists on looking for the best lower bound to problem 8.3.1. As the said lower bound is given by $\Lambda(W_{opt}^{(y)}, y)$, then we just maximise this quantity with respect to y.

Problem 8.3.2 (Lagrange dual problem)**.**

$$\begin{aligned} &\textit{Find:} && \max_{y} \Lambda(W_{opt}^{(y)}, y), \\ &\textit{Subject to:} && y \geq 0. \end{aligned}$$

This dual problem provides the best lower bound with respect to the Lagrange multiplier value y. In addition, often the dual problem is solved at hands of the original problem, commonly called primal, since it allows to include the constraints within the objective function. In fact, in many cases it is possible to show that this better lower bound is indeed sharp. Therefore, the solution to Lagrange's dual problem is also the optimal solution to the original problem.

Additionally note that the previous concept can be easily generalised to matrix inequalities. For instance if have the problem

$$\begin{aligned} &\text{Find:} && \min_{W} f(W), && \\ &\text{Subject to:} && \mathbf{A}(W) \succeq 0 && \mathbf{A}(W) \in \mathbb{S}^N, \end{aligned}$$

then we can just define the Lagrangian function $\Delta(W, \mathbf{Y})$

$$\Delta(W, \mathbf{Y}) = f(W) - \langle \mathbf{Y}, \mathbf{A}(W) \rangle, \tag{8.14}$$

where $\mathbf{Y} \in \mathbb{S}_+^N$ is the matrix version of the Lagrange multiplier and $\langle \mathbf{Y}, \mathbf{A} \rangle = \operatorname{tr}(\mathbf{Y}, \mathbf{A})$ denotes the Frobenious product, namely an inner product in the space $\mathbb{S}^N$.

After this introduction to Lagrange dual problem, which we refer conveniently later on, let us move on to the barrier functions.

8.4 Barrier functions

Barrier functions represents the main interest of the present chapter, providing a slightly different approach to include inequality constraints into the criterium. Consider again problem 8.3.1 with the inequality constraint $g(W) \geq 0$. This time, instead of simply adding the function $yg(W)$ into the criterium, we define a different version of the Lagrangian function as

$$\Lambda(W, t) = f(W) + \frac{1}{t}\beta_t(g(W)),$$

where the function $\beta_t : \mathbb{R}_+ \mapsto \mathbb{R}$ satisfies

$$\lim_{\substack{x\to 0\\ x>0}} \beta_t(W) = \infty \qquad\qquad \forall t > 0.$$

Note that the domain of $\Lambda(W,t)$ is the feasible region for problem 8.3.1. Furthermore, as t tends to infinity, the function $\Lambda(W,t)$ tends to $f(W)$ in the region where $g(W) > 0$.

$$\lim_{t\to\infty} \Lambda(W,t) = f(W) \qquad\qquad \forall W :\ g(W) > 0.$$

The barrier function $\beta_t(g(W))$ ensures that if the minimiser of the Lagrangian $\Lambda(W,t)$, denoted here by $W_{opt}^{(t)}$

$$W_{opt}^{(t)} = \arg\min_W \Lambda(W,t)$$

is a feasible point for problem 8.3.1. This is so since when $g(W)$ approaches 0, the value of the function $\Lambda(W,t)$ grows to infinity, which is certainly not the minimum of $\Lambda(W,t)$. Additionally note we have only defined $\Lambda(W,t)$ for $t > 0$, where we give more importance to the function $f(W)$ over the barrier $\beta_t(g(W))$ as long as $t \to \infty$. The value $t = 0$ implies then that only the barrier function is considered. Hence we can define

$$\Lambda(W,0) = \beta_0(g(W)).$$

The minimisation of the function $\Lambda(W,0)$ provides thereby a point $W_{opt}^{(0)}$ as far as possible from the boundary of the domain of $\Lambda(W,0)$ and therefore *centered* in the feasible region of problem 8.3.1.

Next we come back to the semi-definite program in problem 8.2.2 and apply the notion of barrier functions introduced above. For this matter, we introduce different choices of the function β_t, each one suitable for one of the constraints presented in problem 8.2.2.

8.4.1 The log barrier

The first barrier considered in this work is the pure logarithmic barrier β^L

$$\begin{aligned} \beta^L : \mathbb{S}^N &\mapsto \mathbb{R}, \\ \mathbf{T} &\mapsto \beta^L(\mathbf{T}) = -\log\det\mathbf{T}. \end{aligned}$$

This barrier function is widely used in the literature of interior-points methods. The log barrier is used in this work to ensure the positive-definiteness of the matrix $\mathbf{T}_P$ in eq. (8.6). Therefore consider the problem

Problem 8.4.1 (Simple SDP)**.**

$$\begin{aligned} &\textit{Find:} && \min_{X\in\mathbb{R}^{N^2}} f(W), \\ &\textit{Subject to:} && \mathbf{T}(W) \succeq 0. \end{aligned}$$

It should be noted that when the matrix $\mathbf{T}$ evolves toward a singular matrix during the optimisation process, the function $\beta^L(\mathbf{T})$ grows to infinity. At the limit we have

$$\beta_t^L(\mathbf{T}) = \infty \qquad \forall \mathbf{T} \succeqq 0.$$

Additionally, the convexity of the function β^L is easily verified as shown for instance in [44, Chapter 3, section 1].

Theorem 8.4.1 (Convexity of the log barrier)**.** *The function $\beta^L(\mathbf{A})$ is convex.*

Proof. The function $\beta^L(\mathbf{A})$ is defined for all $\mathbf{A} \succ 0$. Then we take a line $\mathbf{A} = \mathbf{Z} + t\mathbf{Y}$ in the domain of $\beta^L(\mathbf{A})$. We consider $\mathbf{Z} \succ 0$ and $\mathbf{Z} + t\mathbf{Y} \succ 0$. We have

$$\mathbf{Z} + t\mathbf{Y} = \sqrt{\mathbf{Z}}(\mathbf{I}_N + t\sqrt{\mathbf{Z}^{-1}}\mathbf{Y}\sqrt{\mathbf{Z}^{-1}})\sqrt{\mathbf{Z}}.$$

Taking the determinant of the previous expression

$$\det(\mathbf{Z} + t\mathbf{Y}) = \det \mathbf{Z} + \det\left(\mathbf{I}_N + t\sqrt{\mathbf{Z}^{-1}}\mathbf{Y}\sqrt{\mathbf{Z}^{-1}}\right).$$

Now compute the Schur decomposition of the matrix $\sqrt{\mathbf{Z}^{-1}}\mathbf{Y}\sqrt{\mathbf{Z}^{-1}}$ as

$$\sqrt{\mathbf{Z}^{-1}}\mathbf{Y}\sqrt{\mathbf{Z}^{-1}} = \mathbf{V}^T\mathbf{T}\mathbf{V}$$

where $\mathbf{T}$ is a triangular matrix with the eigenvalues of $\sqrt{\mathbf{Z}^{-1}}\mathbf{Y}\sqrt{\mathbf{Z}^{-1}}$ as diagonal elements and $\mathbf{V}$ is unitary, namely $\mathbf{V}^T\mathbf{V} = \mathbf{I}_N$. Therefore

$$\det(\mathbf{I}_N + t\mathbf{V}^T\mathbf{T}\mathbf{V}) = \det(\mathbf{V}^T(\mathbf{I}_N + t\mathbf{T})\mathbf{V}) = \prod_{k=1}^{N}(1 + t\upsilon_k),$$

where υ_k denotes the eigenvalues of the matrix $\sqrt{\mathbf{Z}^{-1}}\mathbf{Y}\sqrt{\mathbf{Z}^{-1}}$. Now we have

$$\begin{aligned}
\beta^L(\mathbf{Z} + t\mathbf{Y}) &= -\log\det(\mathbf{Z} + t\mathbf{Y}) \\
&= -\log\det(\mathbf{Z}) - \log\left(\prod_{k=1}^{N}(1 + t\upsilon_k)\right) \\
&= -\log\det(\mathbf{Z}) - \sum_{k=1}^{N}\log(1 + t\upsilon_k).
\end{aligned}$$

If we compute the derivatives of $\beta^L(\mathbf{Z} + t\mathbf{Y})$ with respect to t we have

$$\begin{aligned}
\mathrm{D}_t\beta^L(\mathbf{Z} + t\mathbf{Y}) &= -\sum_{k=1}^{N}\frac{\upsilon_k}{1 + t\upsilon_k}, \\
\mathrm{D}_t^2\beta^L(\mathbf{Z} + t\mathbf{Y}) &= \sum_{k=1}^{N}\frac{\upsilon_k^2}{(1 + t\upsilon_k)^2} \geq 0.
\end{aligned}$$

We have $\mathrm{D}_t^2\beta^L(\mathbf{Z} + t\mathbf{Y}) \geq 0$ and hence the convexity of β^L. □

If we denote now by $X \in \mathbb{R}^{N^2}$ the vector containing the coefficients of the matrix $\mathbf{T}$ such that $\mathbf{T} = \mathbf{T}(W)$, then we have the scalar function

$$\begin{aligned} \beta^L : \mathbb{R}^{N^2} &\mapsto \mathbb{R}, \\ X &\mapsto \beta^L(W) = -\log\det \mathbf{T}(W). \end{aligned}$$

We include in fig. 8.1 a graphical example to illustrate the function $\frac{1}{t}\beta^L(\mathbf{T})$ namely

$$\frac{1}{t}\beta^L(\mathbf{T}) = \frac{1}{t}\log\det \mathbf{T}$$

for different values of the parameter t. This illustration shows how the barrier function evolves from a soft barrier (with a small value for the parameter t) to the hard barrier function $h(\mathbf{T})$ such that

$$h(\mathbf{T}) = \begin{cases} 0 & \mathbf{T} \succ 0 \\ \infty & \mathbf{T} \succeq 0 \end{cases}.$$

8.4.1.1 Gradient and Hessian matrix

The definition of the function $\beta^L(W)$ as a scalar function with a vector variable simplifies the task of writing its gradient $\nabla\beta^L(W)$ and Hessian matrix $\mathbf{H}\beta^L(W)$, which becomes a N^2 column vector and a $N^2 \times N^2$ symmetric matrix respectively. Nevertheless, before providing the expressions of $\nabla\beta^L(W)$ and $\mathbf{H}\beta^L(W)$, we should introduce the function $\mathrm{inv}(W)$.

Definition 8.4.1 (Function inv)**.** *We define* $\mathrm{inv} : \mathbb{R}^{N^2} \mapsto \mathbb{R}^{N^2}$ *as the function that associates to each vector* $X \in \mathbb{R}^{N^2}$*, the column vector with the coefficients of* $\mathbf{T}(W)^{-1}$

$$\mathrm{inv}(W) = \mathrm{v}(\mathbf{T}(W)^{-1})^T.$$

Let us now compute the Jacobian and Hessian matrix of the function $\mathrm{inv}(W)$ since they are required later on.

Theorem 8.4.2 (Jacobian of the inverse matrix)**.** *The Jacobian matrix of the function* $\mathrm{inv}(W)$ *defined before takes the expression*

$$\mathbf{J}_{\mathrm{inv}}(W) = \mathbf{T}(W)^{-1} \otimes \mathbf{T}(W)^{-1},$$

where $\mathbf{A} \otimes \mathbf{B}$ *represents the Kronecker product of the matrices* $\mathbf{A}$ *and* $\mathbf{B}$*.*

Note that the matrix function $\mathbf{J}_{\mathrm{inv}}(W)$ keeps the symmetry property of $\mathbf{T}(W)$. In other words, if $\mathbf{T}(W) \in \mathbb{S}^N$, then $\mathbf{J}_{\mathrm{inv}}(W) \in \mathbb{S}^{N^2}$. We provide next the derivatives of the function $F(W) = \langle \mathbf{Y}, \mathbf{T}(W)^{-1} \rangle$ where $\mathbf{Y} \in \mathbb{S}^N$ and $\langle A, B \rangle$ denotes the Frobenious product. This product is defined as

$$\langle A, B \rangle = \mathrm{v}(A)^T \mathrm{v}(B).$$

Theorem 8.4.3 (Hessian matrix of the inv function)**.** *Consider the scalar function* $F : \mathbb{R}^{N^2} \mapsto \mathbb{R}$ *defined as*

$$F(W) = \mathrm{v}(\mathbf{Y})^T \mathrm{inv}(W).$$

The Hessian matrix of the function $F(W)$ *can be computed as*

$$\mathbf{H}_F(W) = \mathrm{inv}(W)\mathrm{v}(\mathbf{Y})^T \nabla_{\mathrm{inv}}(W) + \nabla_{\mathrm{inv}}(W)^T \mathrm{v}(\mathbf{Y})\mathrm{inv}(W)^T.$$

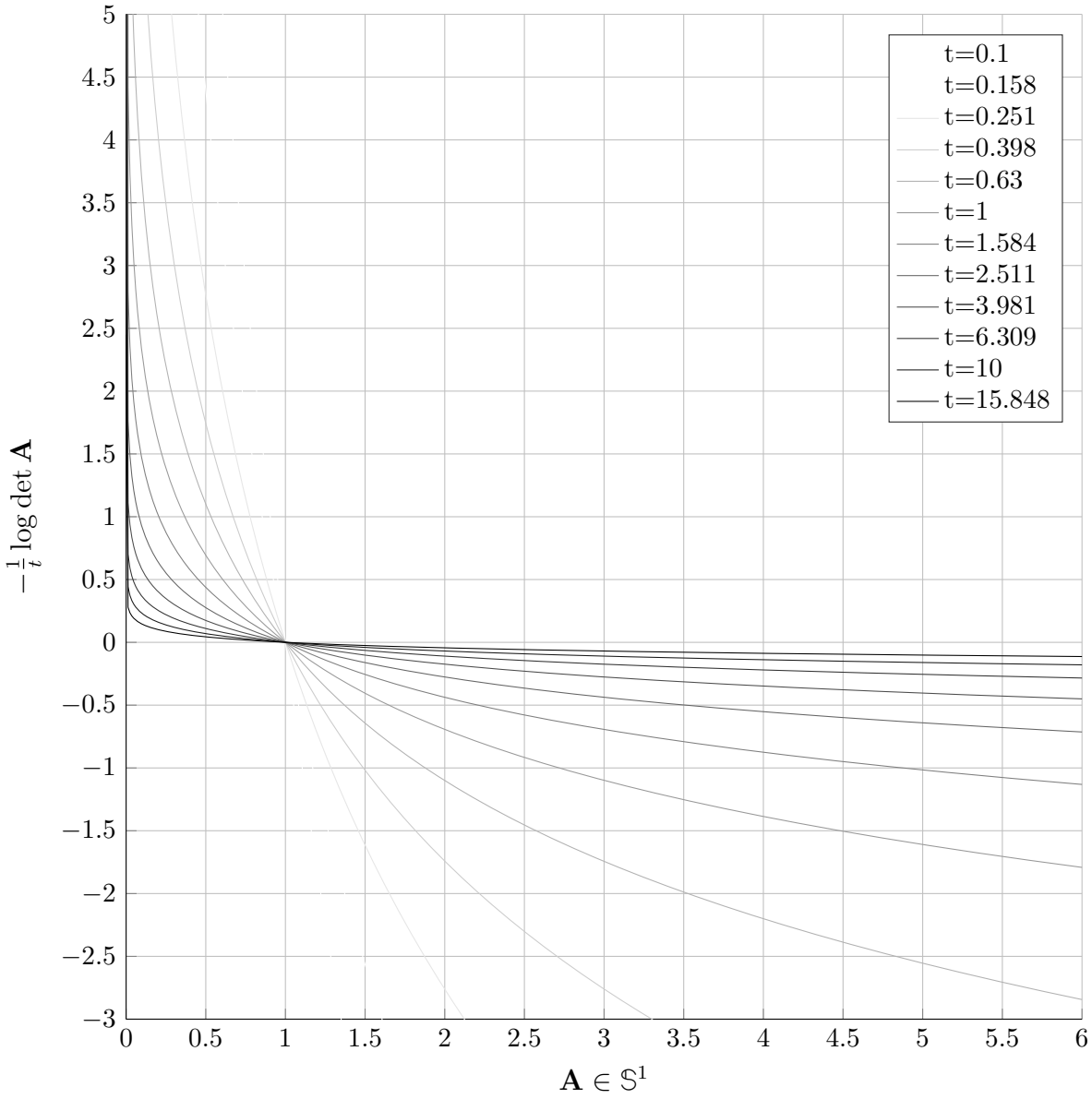


Figure 8.1: Log barrier

Proof of theorem 8.4.2. Denote $\mathbf{A} = \mathbf{T}^{-1}$. To prove the theorem, we compute the derivative of the entry $\mathbf{A}_{i,h}$ with respect to the element $\mathbf{T}_{k,l}$ in the original matrix. Let us compute the derivative of the product $\mathbf{AT}$

$$\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{AT} = (\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A})\mathbf{T} + \mathbf{A}(\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{T}) = 0.$$

Therefore we have

$$(\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A})\mathbf{T} = -\mathbf{A}(\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{T}). \tag{8.15}$$

Multiplying at the left both sides of eq. (8.15) by $\mathbf{A} = \mathbf{T}^{-1}$ we get

$$\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A} = -\mathbf{A}(\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{T})\mathbf{A}.$$

Notice the derivative $\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{T}$ with respect to every element k, l is the zero matrix of size $N \times N$ with 1 in the entry (k, l). Therefore the derivative of each coefficient $\mathbf{A}_{i,h}$ with respect to the (k, l) entry in $\mathbf{T}$, namely $\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{A}_{k,l}$ is given by the element (i, h) of the matrix $\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A}$, which takes the expression

$$\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{A}_{k,l} = -\mathbf{A}_{i,k}\mathbf{A}_{h,l}.$$

The previous quantities correspond to the entries in the matrix resulting from the Kroneker product

$$\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{A}_{k,l} = [\mathbf{A} \otimes \mathbf{A}]_{iN+h,kN+l}\,.$$

The proof is completed. □

Proof of theorem 8.4.3. Denote again $\mathbf{A} = \mathbf{T}^{-1}$. Now let us compute the second derivative of the inverse matrix. We have $\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{A} = \mathbf{A}\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{T}\mathbf{A}$. Computing further the second derivative with respect to $\mathbf{T}_{k,l}$ of the previous expression.

$$\begin{aligned}\mathrm{D}^2_{\mathbf{T}_{k,l}\mathbf{T}_{i,h}}\mathbf{A} &= (\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A})(\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{T})\mathbf{A} + \mathbf{A}(\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{T})(\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A}),\\ &= (\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A})(\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{T})\mathbf{A} + \Big((\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A})(\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{T})\mathbf{A}\Big)^T.\end{aligned}$$

The matrix $\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{T}$ contains the value 1 in the (i, h) entry, which is the only non-zero entry. Therefore, we have as before

$$[(\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A})(\mathrm{D}_{\mathbf{T}_{i,h}}\mathbf{T})\mathbf{A}]_{p,q} = (\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A}_{p,i})\mathbf{A}_{h,q} \qquad k, l, i, h, p, q \in [1, N].$$

If we denote now $\mathbf{A} = \mathbf{A}(W)$ where $X \in \mathbb{R}^{N^2}$ is the column vector with the coefficients of $\mathbf{A}$, for every pair $(k, l) \in \mathbb{R}^N \times \mathbb{R}^N$ and $(p, i) \in \mathbb{R}^N \times \mathbb{R}^N$ the derivatives $\mathrm{D}_{\mathbf{T}_{k,l}}\mathbf{A}_{p,i}$ are given by the Jacobian matrix of the function $\mathrm{inv}(W) = \mathrm{v}(\mathbf{T}(W)^{-1})$ computed before. Then the Hessian matrix of F can be computed row-wise as

$$\mathbf{H}_F(W) = \begin{bmatrix} \mathbf{A}_{1,1}(W)\mathrm{v}(\mathbf{Y})^T\mathbf{J}_{\mathrm{inv}}(W) \\ \mathbf{A}_{2,1}(W)\mathrm{v}(\mathbf{Y})^T\mathbf{J}_{\mathrm{inv}}(W) \\ \vdots \\ \mathbf{A}_{N,N}(W)\mathrm{v}(\mathbf{Y})^T\mathbf{J}_{\mathrm{inv}}(W) \end{bmatrix} + \begin{bmatrix} \mathbf{A}_{1,1}(W)\mathrm{v}(\mathbf{Y})^T\mathbf{J}_{\mathrm{inv}}(W) \\ \mathbf{A}_{2,1}(W)\mathrm{v}(\mathbf{Y})^T\mathbf{J}_{\mathrm{inv}}(W) \\ \vdots \\ \mathbf{A}_{N,N}(W)\mathrm{v}(\mathbf{Y})^T\mathbf{J}_{\mathrm{inv}}(W) \end{bmatrix}^T.$$

Hence $\mathbf{H}_F(W) = \mathrm{inv}(W)\mathrm{v}(\mathbf{Y})^T\mathbf{J}_{\mathrm{inv}}(W) + \big(\mathrm{inv}(W)\mathrm{v}(\mathbf{Y})^T\mathbf{J}_{\mathrm{inv}}(W)\big)^T$. □

The function $\text{inv}(W)$ together with the expression of its Jacobian $\mathbf{J}_{\text{inv}}(W)$ allows us to easily obtain the derivatives of the barrier function β^L.

Theorem 8.4.4 (Gradient and Hessian matrix of the log barrier)**.** *The gradient and the Hessian matrix of the function $\beta^L(W)$ with respect to W are obtained as*

$$\nabla\beta^L(W) = -\text{inv}(W), \tag{8.16}$$

$$\mathbf{H}\beta^L(W) = \mathbf{T}(W)^{-1} \otimes \mathbf{T}(W)^{-1}. \tag{8.17}$$

Notice that, apart from the product operations, one matrix inversion is required to compute $\nabla\beta_t^L(W)$ and $\mathbf{H}\beta_t^L(W)$ what entails a high computational efficiency.

Let us provide next the proof of eqs. (8.16) and (8.17). First notice the *Jacobi formula*

Theorem 8.4.5 (Jacobi formula)**.** *The derivative of* $\det \mathbf{T}$ *with respect to the* (i, h) *entry of the matrix* $\mathbf{T}$ *equals the* i, h *adjugate of the matrix* $\mathbf{T}$*.*

$$\mathrm{D}_{\mathbf{T}_{i,h}} \det \mathbf{T} = \text{adj}_{i,h}\mathbf{T}.$$

Proof of eq. (8.16)*.* Using theorem 8.4.5 we write

$$\mathrm{D}_{\mathbf{T}_{i,h}} \log\det \mathbf{T} = \frac{\mathrm{D}_{\mathbf{T}_{i,h}} \det\mathbf{T}}{\det \mathbf{T}} = \frac{\text{adj}_{i,h}\mathbf{T}}{\det \mathbf{T}},$$

which is the expression of the (i, h) entry in the matrix $\mathbf{A} = \mathbf{T}^{-1}$. □

Proof of theorem 8.4.4. The proof of eq. (8.16) follows from theorem 8.4.5. For the proof of eq. (8.17) note that the Hessian matrix of β^L corresponds to the Jacobian matrix of the function $\text{inv}(W)$. Hence eq. (8.17) follows from theorem 8.4.2. □

8.4.1.2 Lagrangian function

We can now write the Lagrangian function associated to problem 8.4.1 as

$$\Lambda^L(W, t) = f(W) - \frac{1}{t}\log\det \mathbf{T}(W),$$

which is a convex function in X. The barrier β_t^L guarantees the feasibility of the problem for each of the matrices $\mathbf{T}$ obtained during the optimization process. This barrier function allows us to obtain an approximation of the solution to problem 8.4.1 by solving the following sequence of convex problems.

Problem 8.4.2 (Problem SDP(t) with strict feasible barrier)**.**

$$\textit{Find:} \qquad \min_{X\in\mathbb{R}^{N^2}} \Lambda^L(W, t).$$

Since problem 8.4.2 has no constraints, the optimal solution W^{opt} satisfies

$$\nabla_{\Lambda^L}(W^{opt}) = \nabla_f(W^{opt}) - \frac{1}{t}\left\langle \mathbf{T}(W^{opt})^{-1}, \nabla\mathbf{T}\right\rangle = 0. \tag{8.18}$$

Remark that eq. (8.18) can be compared to the derivative of eq. (8.14) by taking

$$\mathbf{Y} = \frac{1}{t}\mathbf{T}(W^{opt})^{-1}.$$

Therefore the point W^{opt} minimises the Lagrangian

$$\Lambda(W, \mathbf{Y}) = f(W) - \langle \mathbf{Y}, \mathbf{T}(W) \rangle.$$

As in eq. (8.13), the minimisation of the Lagrangian provides us a lower bound on the optimal criterium of problem 8.4.1 where the quantity

$$G = \langle \mathbf{Y}, \mathbf{T}(W) \rangle = \frac{1}{t} \text{tr} \left(\mathbf{T}(W^{opt})^{-1}, \mathbf{T}(W^{opt}) \right) = \frac{N}{t}$$

represents the duality gap. Hence we have

$$f(W^{opt}) - \frac{N}{t} < \min_{X \in \mathbb{R}^{N^2}} f(W) \qquad \mathbf{T}(W) \succeq 0. \tag{8.19}$$

Note that the solution W^{opt} to problem 8.4.2 is only N/t sub-optimal for problem 8.4.1. Therefore as the value of the parameter t increases in problem 8.4.2, we approach the solution to problem 8.4.1. However, since this barrier function is not defined for $\mathbf{T} \not\succ 0$ a feasible initial matrix $\mathbf{T} \succ 0$ is required. This is the reason why the function β is only used to ensure the positivity of $\mathbf{T}_P$ in eq. (8.6).

If a feasible initial point is not known, we can redefine this barrier, to handle any arbitrary starting point as it is done in the following section.

8.4.2 Shifted logarithm barrier

This function is defined similarly to the logarithmic barrier function introduced in the previous section. However, given an initial matrix $\mathbf{T}$, we perform a scaling and displacement so that the matrix $\mathbf{T}$ is within the domain of this function. The cited function is

$$\begin{aligned} \beta_s^S : \mathbb{R}^{N^2} &\mapsto \mathbb{R} \\ X &\mapsto \beta_s^S(W) = -\log\det\left(s\mathbf{T}(W) + \mathbf{I}_N\right), \end{aligned}$$

where $\mathbf{I}_N$ represents the $N \times N$ identity matrix. Notice the function $\beta_s^S(W)$ with $s > 0$ is only defined for $\mathbf{T}(W) \succ -\mathbf{I}_N/t$. Additionally a barrier is present at the points where $\mathbf{T}(W) \succeq -\mathbf{I}_N/t$.

It can also be verified that the function β_s^S converges as well towards the ideal barrier $h(\mathbf{T})$ when $t \to \infty$. We show in fig. 8.2 the evolution of the function $\frac{1}{s}\beta_s^S(\mathbf{T}) = \frac{1}{s}\log\det(s\mathbf{T} + \mathbf{I}_N)$ where in order to allow for a 2D visualisation, we choose the matrix $\mathbf{T} \in \mathbb{S}^1$, namely a scalar value.

8.4.2.1 Gradient and Hessian matrix

Note that only a linear transformation as been applied to the matrix $\mathbf{T}(W)$ with respect to the previous function β_t^L, therefore we can easily state the following theorem

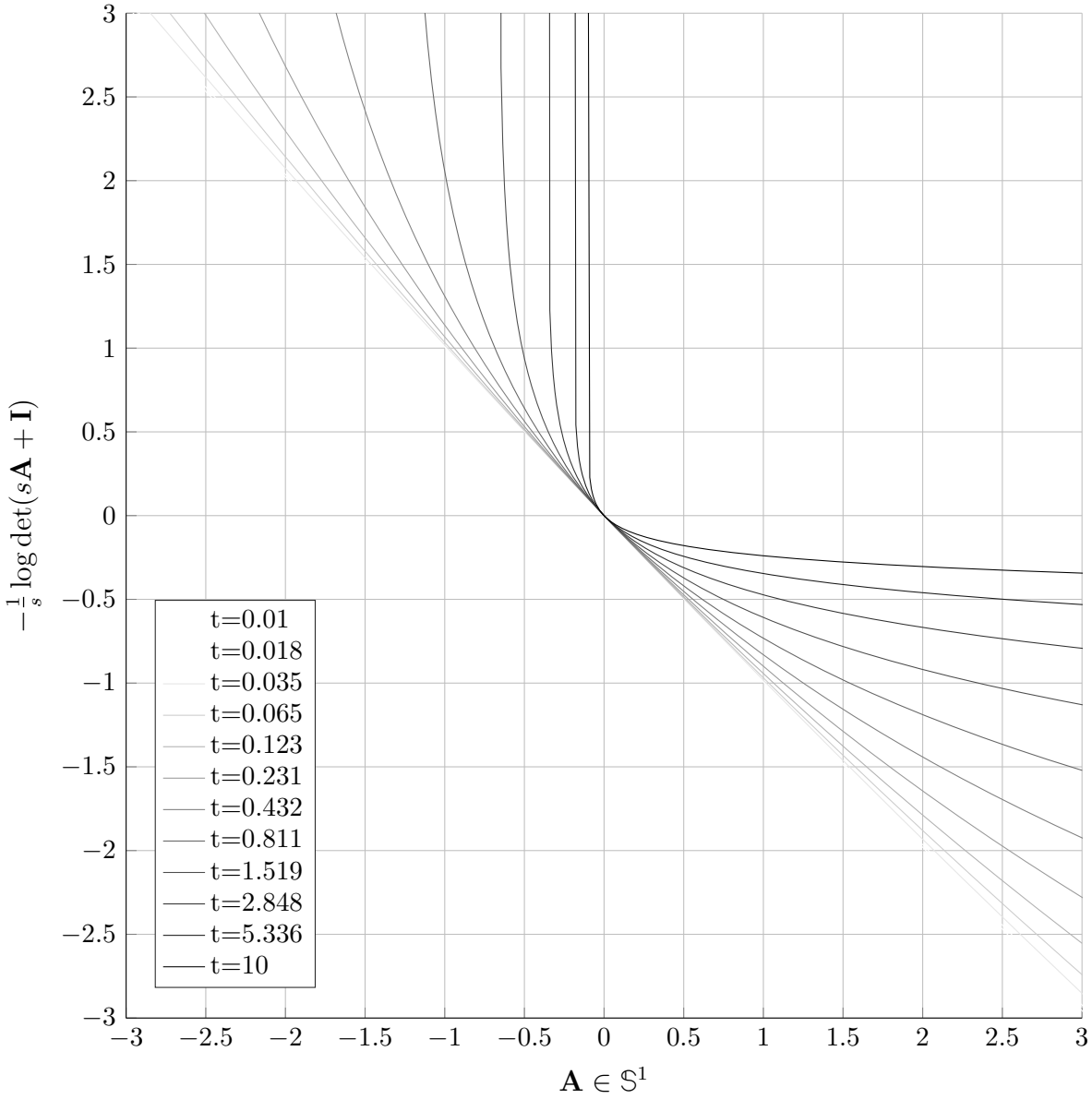


Figure 8.2: Shifted log barrier

Theorem 8.4.6 (Gradient and Hessian matrix of the shifted logarithmic barrier)**.** *The gradient and the Hessian matrix of the function $\beta_S^S(W)$ with respect to W are obtained as*

$$\nabla\beta_s^S(W) = -s \cdot \mathrm{v}((s\mathbf{T}(W) + \mathbf{I}_N)^{-1})^T,$$
$$\mathbf{H}\beta_s^S(W) = s^2 \cdot (s\mathbf{T}(W) + \mathbf{I}_N)^{-1} \otimes (s\mathbf{T}(W) + \mathbf{I}_N)^{-1}.$$

Proof. Follows from eqs. (8.16) and (8.17) □

8.4.2.2 Lagrangian function and weight parameter update

This time we solve a secuence of non constrained problem SDP(ν, s) where the value of s is also increased in each iteration.

Problem 8.4.3 (Problem SDP(ν,s) with non-strict feasible barrier)**.**

$$\textit{Find:} \qquad \min_{X\in\mathbb{R}^{N^2}} \Lambda^S(W, s),$$
$$\textit{where:} \qquad \Lambda^S(W, s) = f(W) - \frac{1}{\nu s}\log\det(s\mathbf{T}(W) + \mathbf{I}_N).$$

We consider further that the objective function is linear, namely $f(W) = \mathbf{C}^T X$ with $\mathbf{C} \in \mathbb{R}^{N^2}$. In addition to the aforementioned update of s, notice the presence of the value $\nu \in \mathbb{R}_+$ which plays the role of a Lagrange multiplier. This parameter is used to *push* the matrix $\mathbf{T}(W)$ toward the feasible domain if $\mathbf{T}(\hat{X})$, where $\hat{X}$ denotes the optimal point X in the previous iteration, is infeasible. Similarly if $\mathbf{T}(\hat{X})$ is inside the feasible region, the value of ν is increased allowing for $\mathbf{T}(W)$ to approach the boundary of the admissible domain. In particular, we take advantage of the linear dependence of the function $f(W)$ and the matrix $\mathbf{T}(W)$ with respect to X and perform, at each iteration a linsearch toward the boundary of the feasible region. We compute the value $\mu \in \mathbb{R}$ that satisfies

$$\mathbf{T}(\hat{X}) + \mu\mathbf{C} \succeqq 0.$$

Then we update the parameter ν such that

$$\|\nabla f(\hat{X} + \mu\mathbf{C})\| = \frac{1}{\nu s}\|\nabla\beta_s^S(\hat{X} + \mu\mathbf{C})\|,$$
$$\|\mathbf{C}\| = \frac{1}{\nu} \cdot \|\mathrm{v}((s\mathbf{T}(\hat{X}) + s\mu\mathbf{C} + \mathbf{I}_N)^{-1})\|.$$

Therefore the value ν is updated according to the following rule

$$\nu = \frac{\|\mathrm{v}((s(\mathbf{T}(\hat{X}) + \mu\mathbf{C}) + \mathbf{I}_N)^{-1})\|}{\|\mathbf{C}\|}.$$

This update decreases the value of ν if the point $\hat{X}$ obtained in the previous iteration was not feasible while producing bigger values of ν if the previous optimal point is feasible. At the point $\hat{X}$ we have

$$\nabla_{\Lambda^S}(\hat{X}) = \nabla_f(\hat{X}) - \frac{1}{\nu}\left\langle (s\mathbf{T}(\hat{X}) + \mathbf{I}_N)^{-1}, s\nabla\mathbf{T}(\hat{X})\right\rangle = 0,$$

which corresponds to the derivative of the function

$$\Lambda(W, \mathbf{Y}) = f(W) - \langle \mathbf{Y}, \mathbf{T}(W) \rangle,$$

with the choice of $\mathbf{Y}$

$$\mathbf{Y} = \frac{1}{\nu} \left(\mathbf{T}(\hat{X}) + \frac{1}{s} \mathbf{I}_N \right)^{-1}.$$

Therefore similarly to eq. (8.19) we have

$$f(\hat{X}) - \frac{1}{\nu} \left\langle \left(\mathbf{T}(\hat{X}) + \frac{1}{s} \mathbf{I}_N \right)^{-1}, \mathbf{T}(\hat{X}) \right\rangle < \min_{X \in \mathbb{R}^{N^2}} f(W) \qquad \mathbf{T}(W) \succeq 0.$$

Note that

$$\lim_{s \to \infty} \left\langle \left(\mathbf{T}(\hat{X}) + \frac{1}{s} \mathbf{I}_N \right)^{-1}, \mathbf{T}(\hat{X}) \right\rangle = \left\langle \mathbf{T}(\hat{X})^{-1}, \mathbf{T}(\hat{X}) \right\rangle = N.$$

Obtaining that, if $\mathbf{T}(\hat{X}) \succ 0$, the duality gap is approached by $\frac{1}{\nu} N$ and decreases as the value of ν increases.

8.4.3 A barrier function for non-linear constraints

As we have demonstrated in previous sections, the convexity property is one of the benefits of the logarithm barrier to guarantee the feasibility of linear matrix inequalities. However when non-linear matrices come into play, things are a bit more complex. This is the case of the matrix constraint in eq. (8.11).

Lemma 8.4.1. *Let f_1 and f_2 be convex uni-variate functions with f_2 non-decreasing over its domain. Then the composition $f_2(f_1)$ is a convex function.*

Remark 8.4.1. *Note that the convexity property does not follows from the composition of two convex functions without any additional assumption.*

Remark 8.4.2. *Note further that in theorem 8.4.1 we have shown the convexity of the logarithm barrier $\beta^L(\mathbf{T})$ with respect to the entries of the matrix $\mathbf{T}$. Nevertheless, the convexity of $\beta^L(\mathbf{T})$ does not holds in general if the entries of $\mathbf{T}$ are only supposed to be convex functions.*

Remark 8.4.3. *Furthermore in problem 8.2.2, we have not stated anything about the convexity of the function $X_0 + K_A W$ in eq.* (8.11)*. The convexity property carries over the set of polynomials $P \in \mathbb{P}_+^{2N}$, parametrised in this case as a function of the variable $W \in \mathbb{R}^Z$ such that $\mathbf{U}(P) \succeq \mathbf{J}\mathbf{I}$. This condition, although is enough to guarantee the absence of local minima in the primal problem, it is not sufficienty to ensure the convexity of the barrier $\log\det \mathbf{U}(P)$. Therefore local minima might appears uppon the minimisation of the Lagrangian function. As a result the convexity property obtained in part II is lost by the choice of an innapropiate barrier function.*

In order to overcome the issue discused in remark 8.4.3, instead of the previously introduced logarithmic barrier, we apply the barrier function studied in [45]. This barrier function is defined as

$$\begin{aligned}\beta_r^Y : \mathbb{S}^N &\mapsto \mathbb{S}^N,\\ \mathbf{U} &\mapsto \beta_r^Y(\mathbf{U}) = (r\mathbf{U} + \mathbf{I}_N)^{-1} - \mathbf{I}_N.\end{aligned}$$

Notice the β_r^Y is a matrix function instead of a scalar one. This matrix function is combined with a matricial Lagrange multiplier $\mathbf{Y} \in \mathbb{S}_+^N$ to compose the following Lagrangian function

$$\Lambda(W, r, \mathbf{Y}) = f(W) + \frac{1}{r}\langle \mathbf{Y}, \beta_r^Y(\mathbf{U}(W))\rangle \qquad \mathbf{U}(W) = \mathbf{U}(W) - \mathbf{\Pi}. \tag{8.20}$$

Additionally, the matrix multiplier $\mathbf{Y}$ is updated in each iteration according to

$$\mathbf{Y} = (r\mathbf{U} + \mathbf{I}_N)^{-1}\,\hat{\mathbf{Y}}\,(r\mathbf{U} + \mathbf{I}_N)^{-1}.$$

where $\hat{\mathbf{Y}}$ represent the matrix $\mathbf{Y}$ used in previous iteration.

Lemma 8.4.2. *Consider a matrix function $\mathbf{U}(P)$ with the convex property in the matrix sense, namely for all $P_1, P_2 \in \mathbb{P}_+^{2N}$ and $0 \leq \kappa \leq 1$ we have*

$$\mathbf{U}(\kappa P_1 + (1-\kappa)P_2) \succeq \kappa\mathbf{U}(P_1) + (1-\kappa)\mathbf{U}(\mathbb{P}_2).$$

Then the Lagrangian function defined in eq. (8.20) *is convex.*

Therefore this special choice of barrier provides us with a convex augmented lagrangian. However the convexity is obtained at the expenses of imposing an additional property on the matrix $\mathbf{U}(P)$, with is stronger than the convexity of the set of polynomials $P \in \mathbb{P}_+^{2N}$ such that $\mathbf{U}(P) \succeq \mathbf{\Pi}$. This property is the convexity of the matrix function $\mathbf{U}(P)$ which has been studied in chapter 5.

Finally, as in the previous cases, it is interesting to show a graphical representation of the function β_r^Y in the case where $\mathbf{U} \in \mathbb{S}^1$. This function is plotted in fig. 8.3 for different values of the parameter t.

8.4.3.1 Gradient and Hessian matrix

Let us now state

Theorem 8.4.7 (Derivatives of the barrier function β_r^Y)**.** *Consider the vector function $B(W) : \mathbb{R}^{N^2} \mapsto \mathbb{R}^{N^2}$ where*

$$B(W) = \mathrm{v}(\mathbf{Y})^T \mathrm{v}\left((r\mathbf{U}(W) + \mathbf{I}_N)^{-1} - \mathbf{I}_N\right).$$

The gradient and Hessian matrix of B are expressed as

$$\begin{aligned}\nabla_B(W) &= \mathrm{v}(Y)^T\left[(r\mathbf{U}(W) + \mathbf{I}_N)^{-1} \otimes (r\mathbf{U}(W) + \mathbf{I}_N)^{-1}\right],\\ \mathbf{H}_B(W) &= \mathrm{v}\left[(r\mathbf{U}(W) + \mathbf{I}_N)^{-1}\right]\nabla_B(W) + \nabla_B(W)^T.\mathrm{v}\left[(r\mathbf{U}(W) + \mathbf{I}_N)^{-1}\right]^T\end{aligned}$$

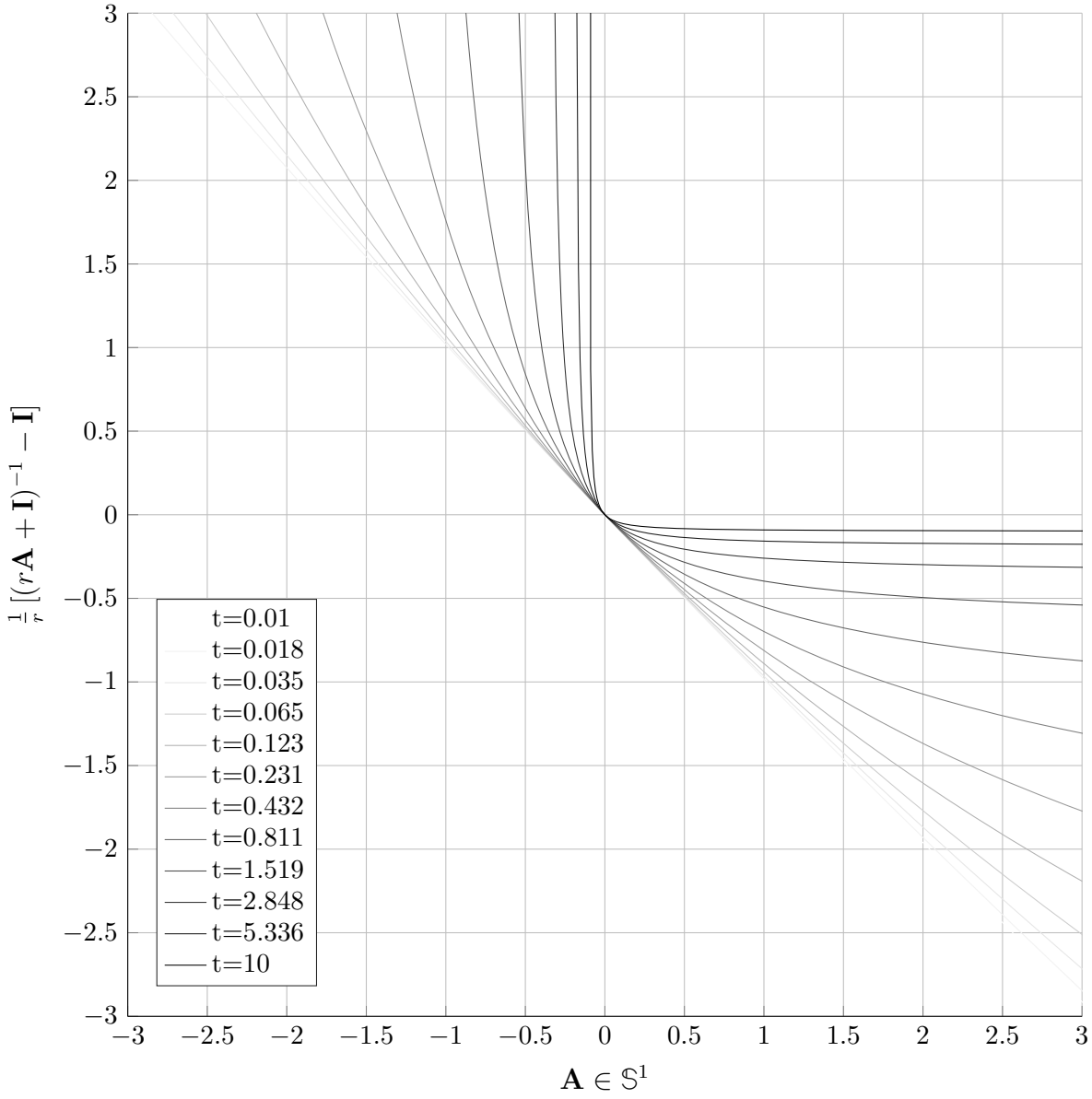


Figure 8.3: Nonlinear barrier

Proof. Follows from theorems 8.4.2 and 8.4.3. □

Again, it is important to remark that, apart from the operations required to compute the product, only one matrix inversion is required to compute ∇_B and $\mathbf{H}_B$, namely the inversion of the matrix $(r\mathbf{U}(W) + \mathbf{I}_N)$. This implies, as in the case of the previous barrier functions, a significant computational efficiency.

8.4.3.2 Lagrangian function

The function in eq. (8.20), which has the form in eq. (8.14) is the Lagrangian associated to the following optimisation problem

Problem 8.4.4 (Unconstrained SDP(k) with non linear matrix inequalities)**.**

$$\begin{aligned} \textit{Find:} \qquad & \min_{X\in\mathbb{R}^{N^2}} f(W), \\ \textit{Subject to:} \qquad & -\beta_r^Y(W) \succeq 0, \end{aligned}$$

where $\beta_r^Y(W) \preceq 0$ *if and only if* $\mathbf{U}(W) \succeq 0$.

Therefore if we denote $\hat{X} = \arg\min_{X\in\mathbb{R}^{N^2}} \Lambda(W, r, \mathbf{Y})$, assuming $\mathbf{U}(\hat{X}) \succeq 0$ we obtain the bound

$$f(\hat{X}) + \frac{1}{r}\langle \mathbf{Y}, \beta_r^Y(\hat{X})\rangle \leq \min_{X\in\mathbb{R}^{N^2}} f(W) \qquad\qquad \mathbf{U}(W) \geq 0,$$

with the duality gap given by $-\frac{1}{r}\langle \mathbf{Y}, \beta_r^Y(\hat{X})$.

8.5 A non-constrained convex problem

Combining each of the barrier functions presented here, namely aplying β^L to eq. (8.6), β_s^S to eqs. (8.7) to (8.10) and β_r^Y to eq. (8.11), we obtain the following Lagrangian function that includes every matrix inequality in problem 8.2.2

$$\begin{aligned} \Lambda(W, t, s, r, \nu, \mathbf{Y}) = C^T K_A W & \\ & + \frac{1}{t}\beta_t^L(\mathbf{T}_P(W)) \\ & + \frac{1}{\nu s}\sum_{l=1}^{n} \beta_s^S(\mathbf{M}_\Psi^{(l)}(W)) + \beta_s^S(\mathbf{N}_\Psi^{(l)}(W)) \\ & + \frac{1}{\nu s}\sum_{k=1}^{m} \beta_s^S(\mathbf{M}_\Gamma^{(k)}(W)) + \beta_s^S(\mathbf{N}_\Gamma^{(k)}(W)) \\ & + \frac{1}{r}\langle \mathbf{Y}, \beta_r^Y(\mathbf{U}(W) - \mathbb{I})\rangle. \end{aligned} \tag{8.21}$$

This Lagrangian function allows us to obtain an approximation of the solution of the original problem by solving a series of convex optimization problems without constraints.

Problem 8.5.1 (Unconstrained dual Lagrange problem in iteration i: SDP(i))**.**

$$\textit{Find:} \qquad \min_W \Lambda(W, t_i, s_i, r_i, \nu_i, \mathbf{Y}_i).$$

where in each iteration the values t_i, s_i, r_i are increased in a given amount meanwhile the multipliers $\nu_i, \mathbf{Y}_i$ are updated as indicated previously in the correspondent sections. In addition assuming all matrices are positive semi-definite, the vector $W_{opt}^{(i)}$ providing the optimal solution in the i-th iteration satisfies

$$C^T K_A W_{opt}^{(i)} - \phi^{(i)}(W_{opt}^{(i)}) < C^T K_A W_{opt},$$

where W_{opt} is the minimizer of the original problem and

$$\phi^{(i)}(W_{opt}^{(i)}) = \frac{N+1}{t_i} + \frac{(n+m)(2N+1)}{\nu_i} - \frac{1}{r_i}\left\langle \mathbf{Y}_i, \beta_{r_i}^Y(W_{opt}^{(i)}) \right\rangle.$$

Then we can ensure that the solution is at most $\phi^{(i)}(W_{opt}^{(i)})$ sub-optimal. In addition $\phi^{(i)}(W_{opt}^{(i)}) \to 0$ when $t, \nu, r \to \infty$. Therefore the solution to the optimal problem is approached at each iteration.

8.5.1 The newton solver

The Newton method is one of the most efficient methods used in optimization. Especially in the search for the extreme values of a multivariate scalar function $f(W) : \mathbb{R}^Z \mapsto \mathbb{R}$. Although it is very probable that said method is already more than known considering the profile of a potential reader, it always deserves a small discussion, even more considering its importance in this chapter. Newton's method is based on the minimization of the quadratic approximation of the function $f(W)$ around an initial point $W^{(i)}$. We define the quadratic approximation of f around W_i as $F_i(W)$, and denote the gradient of F_i by

$$\nabla_{F_i}(W) : \mathbb{R}^Z \mapsto \mathbb{R}^Z.$$

The function $\nabla_{F_i}(W)$ is linear in W and therefore its Jacobian matrix $\mathbf{J}_{\nabla F_i}$ does not depends on the variable W. Hence

$$\mathbf{J}_{\nabla F_i}(W_0 - W) = \nabla_{F_i}(W_0) - \nabla_{F_i}(W).$$

The point W_i^{opt} such that $\nabla_{F_i}(W_i^{opt}) = 0_Z$ is then expressed as

$$W_i^{opt} = W - \mathbf{J}_{\nabla F_i}^{-1} \nabla_{F_i}(W)$$

for any $W \in \mathbb{R}^Z$. The Jacobian matrix of the gradient of $F_i(W)$ corresponds to the Hessian matrix $\mathbf{H}_f(W)$ of the function $f(W)$ at the point W_i. Newton's method allows the displacement in the i-th iteration to the point W_{i+1} which minimizes the quadratic approximation of the function $f(W)$ at W_i. The operation performed at each iteration is given by

$$W_{i+1} = W_i - \mathbf{H}_f(W_i)^{-1} \nabla_f(W_i). \tag{8.22}$$

This algorithm provides a quadratic convergence rate in the region where the second order approximation of the f function is good. Notice if the function to be minimised $f(W)$ is quadratic, previous method provides the optimal minimiser W^{opt} in only one iteration as as the quadratic approximation F coincides with the function f itself. Nevertheless, a quadratic approximation might not be good for the Lagrangian function in eq. (8.21),

specially as the values of t, s, r grows. Furthermore, note that the vector W in eq. (8.21) must always remain within the feasible region for the strict feasible inequalities, namely $\mathbf{T}_P(W) \succ 0$. To account for this restriction, we add a parameter ξ to eq. (8.22) such that

$$W_{i+1} = W_i + \xi \text{dir}(W_i), \tag{8.23}$$

where $\text{dir}(W_i)$ denotes the direction given by $-\mathbf{H}_f(W_i)^{-1}\nabla_f(W_i)$. This method of updating the variable W corresponds to the modified Newton method, which can be interpreted as a variable displacement in the direction indicated by $\text{dir}(W_i)$ which points towards the minimum of the quadratic approximation of $f(W)$ at $W = W_i$. This prevents the increment given by $\text{dir}(W_i)$ from being too large or too short, providing for instance a vector W_i outside the region where $TT_P(W) \succ 0$. Finally note that the exception where the f function has a good quadratic approximation and the previous method fails is the case where f is linear, in which case the Hessian matrix of f is zero. Nevertheless in this case the descent direction of the function $f(W)$ is simply provided by the gradient $-\nabla_f(W)$. Therefore we choose the direction $\text{dir}(W)$ as

$$\text{dir}(W) = \begin{cases} -\mathbf{H}_f(W)^{-1}\nabla_f(W) & \text{if} \quad \mathbf{H}_f(W) \succ 0 \\ -\nabla_f(W) & \text{if} \quad \mathbf{H}_f(W) \succeq 0 \end{cases}.$$

8.5.2 The linear search

The last step for the determination of the minimizing vector W^{opt} in iteration i -th is the determination of the optimal value of the parameter ξ in eq. (8.23). Considering the point W_i and the direction given by $\text{dir}(W_i)$, the parameter ξ is chosen in an interval $[\xi_{min}, \xi_{max}]$ where $\mathbf{T}(W_i + \xi \text{dir}(W_i)) \succ 0$. The interval $[\xi_{min}, \xi_{max}]$ is determined a priori by a dichotomy procedure. By the convexity of the feasible set of (W), if it is verified that $W_i + \xi \text{dir}(W_i)$ for the values ξ_{max} and ξ_{min} are inside such set, it is also verified for every $\xi \in [\xi_{min}, \xi_{max}]$. The optimal ξ is obtained by solving the following optimisation problem.

Problem 8.5.2 (Linear search)**.**

$$\textit{Find:} \qquad \xi_{opt} = \arg\min_{\xi} g(\xi) \qquad \xi \in [\xi_{min}, \xi_{max}],$$

where

$$g(\xi) = f(W_i + \xi \text{dir}(W_i)).$$

Problem 8.5.2 is a convex optimisation problem in one single variable which can be solved easily with any classical tool. Nevertheless since we disposed of the analytical expression of the gradient ∇_f and Hessian matrix $\mathbf{H}_f$ corresponding to the function $f(W)$, we can also compute ∇_g and $\mathbf{H}_g$ as the projection of the former ones into the space (of dimension one) spanned by the vector $\text{dir}(W_i)$

$$\begin{aligned} \nabla_g(\xi_i) &= \nabla_f(W_i + \xi_i \text{dir}(W_i))^T \text{dir}(W_i), \\ \mathbf{H}_g(\xi_i) &= \text{dir}(W_i)^T \mathbf{H}_f(W_i + \xi_i \text{dir}(W_i)) \text{dir}(W_i). \end{aligned}$$

Therefore simplest method to minimise the function $g(\xi)$ is by applying the Newton algorithm again. We start by an initial guess $\xi_0 \in [\xi_{min}, \xi_{max}]$ which is refined iteratively by the update formula

$$\xi_{i+1} = \begin{cases} \xi_i - \mathbf{H}_g(\xi_i)^{-1}\nabla_g(\xi_i)^{-1} & \text{if} \quad \mathbf{H}_g(\xi_i) > 0 \\ \xi_{max} & \text{if} \quad \mathbf{H}_g(\xi_i) = 0 \end{cases}.$$

This lineseach algorithm is carried on until the gradient of $g(\xi)$ is sufficiently small $\nabla_g(\xi) \approx 0$ or a value ξ_{i+1} located outside the interval $[\xi_{min}, \xi_{max}]$ is obtained, in which case the closest value to ξ_{i+1} in the interval $[\xi_{min}, \xi_{max}]$ is taken.

References

[43] E. de Klerk, *Aspects of semidefinite programming : interior point algorithms and selected applications*. Kluwer Academic Publishers, 2002.

[44] S. Boyd and L. Vandenberghe, *Convex Optimization*, ser. Berichte über verteilte messysteme. Cambridge University Press, 2004. [Online]. Available: https://books.google.fr/books?id=mYm0bLd3fcoC

[45] M. Stingl, "On the solution of nonlinear semidefinite programs by augmented Lagrangian methods," *Outlook*, 2006.

Chapter 9:

Hard bounds and sub-optimal functions

Throughout the theory developed in chapter 4, and thanks to the characterisation of the set of admissible polynomials obtained in chapter 5, we have obtained a convex formulation of the matching problem. This formulation consists of the minimization on the set of positive polynomials $P \in \mathbb{P}_+^{2N}$ of the maximum over ω of the filtering function $P(\omega)/R(\omega)$ with ω belonging to a band $\mathbb{I}$ contained in the real axis. Furthermore, as is usual in the classical synthesis, it is possible to add an arbitrary number of restrictions on the minimum selectivity allowed within another frequency band $\mathbb{J}$, whose intersection with the set $\mathbb{I}$ is empty.

Introducing the variable slack Ψ so that $\Psi \geq P(\omega)/R(\omega)$ for all $\omega \in \mathbb{I}$, the matching problem was formulated in problem 7.1.2 as follows

Problem 7.1.2 (General problem)**.**

$$\begin{aligned} \textit{Find:} \qquad & \min_{(\Psi,P)} \Psi && (\Psi, P) \in \mathbb{R}_+ \times \mathbb{P}_+^{2N}, && \\ \textit{Subject to:} \qquad & P(\omega) \leq \Psi \cdot R(\omega) && \omega \in \mathbb{I}, && (7.1) \\ & P(\omega) \geq \Gamma \cdot R(\omega) && \omega \in \mathbb{J}, && (7.2) \\ & \mathbf{U}(P) \succeq \mathbf{\Pi}. && && (7.3) \end{aligned}$$

This problem has been conveniently reformulated in this part III of the thesis until the form of a non-linear semi-defined program is obtained. The solution to this problem can be computed by conventional interior-point techniques. However, due to the relaxation introduced in chapter 4 with the concept of admissibility, in the majority of cases problem 7.1.2 does not produce a filtering function which can be implemented with a matching network of fixed McMillan degree K. However, problem 7.1.2 provides, in any case, hard lower bounds to the original problem with McMillan degree K, which may or not be sharp. In this chapter, we make an analysis of these bounds considering a McMillan degree K for the matching network.

9.1 Blaschke product and feasible function

Equation (7.3) guarantees the existence of an admissible function $u_{P_{opt}}(\lambda)$, therefore according to lemma 4.2.1, there exists a function $f \in \mathbb{\Sigma}^M$ such that $S_{22}(\lambda) = f(\lambda)u_{P_{opt}}(\lambda)$ is of degree as most $M+N$ and feasible for the load, namely $S_{22}(\lambda) \in \mathbb{F}^{N+M}$. This function $f(\lambda)$ can be computed by the Schur recursion procedure reviewed in appendix B as the interpolant function $f \in \mathbb{E}(u_{P_{opt}})$ solution to the interpolation problem

$$f(\alpha_i) = \frac{L_{22}(\alpha_i)}{u_{P_{opt}}(\alpha_i)} \qquad \forall i \in [1, M]. \tag{9.1}$$

Additionally from corollary 4.4.1 we have that if eq. (7.3) is binding, namely $\mathbf{U}(P_{opt}) \succeq \mathbf{\Pi}$, then $\mathbb{E}(u_{P_{opt}})$ contains only a Blaschke product $b(\lambda)$ of degree $\mathit{\Lambda}$ which equals the rank of $\mathbf{U}(P_{opt}) - \mathbf{\Pi}$. Thus

$$|b(\omega)u_P(\omega)| = |u_P(\omega)| \qquad \forall \omega \in \mathbb{R}.$$

Therefore we have a function $S_{22} = b \cdot u_P \in \mathbb{F}^{N+\mathcal{A}}$ which satisfies

$$|S_{22}(\omega)|^2 \geq \gamma \qquad \forall \omega \in \mathbb{J}, \tag{9.2}$$

$$|S_{22}(\omega)|^2 \leq \psi \qquad \forall \omega \in \mathbb{I}, \tag{9.3}$$

where $\psi = (\Psi^{-1} - 1)^{-1}$ and $\gamma = (\Gamma^{-1} - 1)^{-1}$. Namely the constraints that correspond to eqs. (7.1) and (7.2) stated over the function $|u_{P_{opt}}(\lambda)|^2$. On the contrary if eq. (7.3) is not binding in problem 7.1.2, namely $\mathbf{U}(P_{opt}) \succ \mathbf{\Pi}$, then the set of functions $f \in \mathbb{E}(u_{P_{opt}})$ is not a singleton. Note that by corollary 4.4.1, this case only occurs if eq. (7.2) is saturated as well, and therefore eq. (9.2) can not be guaranteed for a function $S_{22} = f \cdot u_{P_{opt}}$ with $f \in \mathbb{\Sigma}^M$. Nevertheless if $\mathbb{E}(u_{P_{opt}})$ is not a singleton, the functions $f \in \mathbb{E}(u_{P_{opt}})$ are parametrised by theorem B.1.3 as

$$f(\lambda) = \frac{A(\lambda) + B(\lambda)f(\lambda)}{C(\lambda) + D(\lambda)f(\lambda)} \qquad A, B, C, D \in \mathbb{P}^M,$$

where the polynomials A, B, C, D are given by the interpolation data and computed by means of the Schur recursion. If we now choose $f(\lambda) = c$ with c a uni-modular constant, then the function $b(x)$ takes the expression

$$b(\lambda) = \frac{A(\lambda) + cB(\lambda)}{C(\lambda) + cD(\lambda)} \qquad A, B, C, D \in \mathbb{P}^M. \tag{9.4}$$

The function $b(\lambda)$ defined by eq. (9.4) is a Blaschke product of degree M satisfying eq. (9.1). Therefore we have $S_{22} = b \cdot u_P \in \mathbb{F}^{N+M}$ satisfying eqs. (9.2) and (9.3).

The possibility of obtaining this function S_{22} from the optimal solution P_{opt} implies that problem 7.1.2 not only gives us a lower limit to the solution of the problem of matching with a fixed degree, but also allows to calculate a function $S_{22} \in \mathbb{F}^{N+\mathcal{A}}$ with $\mathcal{A}$ the rank of the matrix $\mathbf{U}(P_{opt}) - \mathbf{\Pi}$. In other words, the mentioned S_{22} is a function of degree $N + \mathcal{A}$ that can be expressed in absolute value as the magnitude of the reflection at the input of a matching network connected to the load, namely $S_{22} = F_{22} \circ L$. This function also verifies the selectivity constraints imposed on the original problem. Nevertheless the McMillan degree of the function F_{22} is increased with respect to the desired McMillan degree K for the matching filter. Indeed we obtain a function $F_{22} \in \mathbb{\Sigma}^{K+\mathcal{A}}$ (of McMillan degree $K + \mathcal{A}$) .

9.1.1 De-embedding of the load

In possession of the Blaschke product $b(\lambda)$ it is possible to obtain the function S_{22} which allows the extraction of the load to recover by means of eq. (3.13) the reflection function of the matching network F_{22}. The function $b(\lambda)$ in eq. (9.4) can be written as

$$b(\lambda) = \frac{p_b(\lambda)}{q_b(\lambda)} = \prod_{k=1}^{\mathcal{A}} \frac{\lambda - \xi_k}{\lambda - \overline{\xi_k}} \qquad \xi_k \in \mathbb{C}^-.$$

Moreover we denote

$$u_{P_{opt}}(\lambda) = \frac{p(\lambda)}{q(\lambda)} \qquad p, q \in \mathbb{P}^N,$$

where p, q are polynomials of minimum phase satisfying $p^\star p = P_{opt}$ and $q^\star q = P_{opt} + R$. Now we compute $S_{22} \in \mathbb{F}^{N+\mathcal{I}}$ as

$$S_{22}(\lambda) = \frac{p_S(\lambda)}{q_S(\lambda)} = \frac{p(\lambda)}{q(\lambda)} \frac{p_b(\lambda)}{q_b(\lambda)}.$$

Nevertheless it should be remarked that $S_{22} \notin \mathbb{F}_R^{N+\mathcal{I}}$, this can be verified after computing the transmission polynomial which corresponds to the function S_{22} obtained for the global system, namely

$$\begin{aligned} R_S = q_S^\star q_S - p_S^\star p_S &= (qq_b)^\star qq_b - (pp_b)^\star pp_b \\ &= (q^\star q - p^\star p) p_b^\star p_b = R \cdot p_b^\star p_b. \end{aligned} \tag{9.5}$$

Notice that eq. (9.5) entails that $\mathcal{I}$ additional transmission zeros are introduced in the global system at the positions where p_b vanishes, namely the points ξ_k with $1 \leq k \leq \mathcal{I}$.

Let us now compute the reflection coefficient F_{22} of the matching filter by the de-embedding of the scattering matrix of the load, which is defined as

$$L(\lambda) = \frac{1}{q_L(\lambda)} \begin{pmatrix} p^\star(\lambda) & -r^\star(\lambda) \\ r(\lambda) & p(\lambda) \end{pmatrix},$$

with q_L the stable polynomial satisfying $q_L^\star q_L = p_L^\star p_L + r_L^\star r_L$. The de-embedding operation is expressed as

$$\begin{aligned} F_{22} &= \frac{L_{22} - S_{22}}{\det L - S_{22} L_{11}} \\ &= \frac{\dfrac{p_L}{q_L} - \dfrac{p_S}{q_S}}{\dfrac{p_L p_L^\star + r_L r_L^\star}{q_L^2} - \dfrac{p_S}{q_S}\dfrac{p_L^\star}{q_L}} \\ &= \frac{\dfrac{p_L}{q_L} - \dfrac{p_S}{q_S}}{\dfrac{q_L^\star}{q_L} - \dfrac{p_S}{q_S}\dfrac{p_L^\star}{q_L}} \\ &= \frac{p_L q_S - p_S q_L}{q_L^\star q_S - p_L^\star p_S} \end{aligned}$$

, where numerator and denominator are polynomials of degree $N + \mathcal{I} + M$. Nevertheless at the transmission zeros α_i we can express

$$\begin{aligned} F_{22}(\alpha_i) &= \frac{L_{22}(\alpha_i) - S_{22}(\alpha_i)}{L_{11}(\alpha_i) L_{22}(\alpha_i) - S_{22}(\alpha_i) L_{11}(\alpha_i)} \\ &= \frac{L_{22}(\alpha_i) - S_{22}(\alpha_i)}{L_{11}(\alpha_i)(L_{22}(\alpha_i) - S_{22}(\alpha_i))} \qquad \forall i \in [1, M]. \end{aligned}$$

Note that we have $L_{12}(\alpha_i) L_{21}(\alpha_i) = 0$, obtaining a pole-zero cancellation at each point α_i. Additionally since the points α_i are transmission zeros of both the global system and the load we also have

$$\begin{aligned} q_S^\star(\alpha_i) q_S(\alpha_i) &= p_S^\star(\alpha_i) p_S(\alpha_i), \\ q_L^\star(\alpha_i) q_L(\alpha_i) &= p_L^\star(\alpha_i) p_L(\alpha_i). \end{aligned}$$

Therefore conjugating both sides we obtain

$$q_S(\overline{\alpha_i})\overline{q_S(\alpha_i)} = p_S(\overline{\alpha_i})\overline{p_S(\alpha_i)},$$
$$q_L(\overline{\alpha_i})\overline{q_L(\alpha_i)} = p_L(\overline{\alpha_i})\overline{p_L(\alpha_i)}.$$

Therefore

$$S_{22}(\overline{\alpha_i}) = \left(\overline{S_{22}(\alpha_i)}\right)^{-1} \qquad \forall i \in [1, M],$$
$$L_{22}(\overline{\alpha_i}) = \left(\overline{L_{22}(\alpha_i)}\right)^{-1} \qquad \forall i \in [1, M],$$

what implies that M additional simplifications occur at the points $\overline{\alpha_i}$. Hence after cancelling out the common zeros in numerator and denominator of F_{22} a rational function

$$F_{22}(\lambda) = \frac{p_F(\lambda)}{q_F(\lambda)}$$

is obtained with p_F and q_F polynomials of degree $N + \mathit{Я} - M = K + \mathit{Я}$.

Remark 9.1.1. *Note that the additional transmission zeros at the points ξ_k with $k \in [1, \mathit{Я}]$ are not present in the transmission polynomial of the load and therefore they are not simplified after the load is de-embedded. This fact entails two important consequences*

1. *The degree of the matching filter reflection F_{22} is increased by $\mathit{Я}$ with respect to the desired degree K. We obtain $F_{22} \in \mathbb{\Sigma}^{K+\mathit{Я}}$.*

2. *The filter has $\mathit{Я}$ additional transmission zeros at the positions ξ_k with $1 \leq k \leq \mathit{Я}$, namely the roots of p_b. These transmission zeros can be arbitrarily distributed between the transmission coefficients F_{21} and F_{12}. Therefore we have $F_{22} \in \mathbb{\Sigma}^{K+\mathit{Я}}_{R_F \cdot p_b}$.*

This increase in degree with respect to the desired degree K in the original problem is the price to pay for the convex formulation studied in the preceding chapters.

Remark 9.1.2. *Note that the degree increase $\mathit{Я}$ corresponds to the range of the matrix $\mathbf{U}(P_{opt}) - \mathbf{Л}$, which is bounded by the McMillan degree of the load L. Therefore we have $\mathit{Я} \leq M$ which indicates that the maximum degree of the obtained function F_{22} can be computed with the information of the load L.*

9.1.2 Degree of the Blaschke product

Let us now study the link between the degree of the Blaschke product ($\mathit{Я}$) and the degree of the load (M). With this goal, we begin by reformulating, once again, problem 7.1.2. It is important to remember that the purpose of eq. (7.3) is to guarantee the admissibility of the polynomial P. Note that from definition 4.2.2 we have that P is admissible if and only if $\mathbb{E}(u_P)$ is not empty. As we commented already, the set $\mathbb{E}(u_P)$ is a singleton if and only if $\mathbf{U}(P) \trianglerighteq \mathbf{Л}$. Therefore,

$$\mathbf{U}(P) \trianglerighteq \mathbf{Л} \iff \mathbb{E}(u_P) \neq \emptyset.$$

From lemma 4.3.5 we can parametrise the the interior of the set $\mathbb{A}_R^N$ in terms of $\mathbb{E}(u_P)$ as

$$\overset{\circ}{\mathbb{A}}{}_R^N = \left\{ P \in \mathbb{P}_+^{2N} \mid \text{card}\mathbb{E}(u_P) > 1; P(\omega) > 0 \ \forall \omega \in \mathbb{R} \right\},$$

where card denotes the cardinality number. We can now use theorem B.1.2 to characterise the polynomials P such that the set $\mathbb{E}(u_P)$ contains at least two functions

Theorem 9.1.1 (Characterisation of the admissible set by scalar inequalities)**.** *Given the polynomial $P \in \mathbb{P}_+^{2N}$ which does not vanishes on the real line, we have $P \in \overset{\circ}{\mathbb{A}}{}_R^N$ if and only if*

$$\left| \frac{L_{22}(\alpha_i)}{u_P(\alpha_i)} \right| < 1 \qquad \forall i \in [1, M], \tag{9.6}$$

$$\delta(\gamma_k^{(k)}(P), \gamma_{k-1}^{(k)}(P)) < \left| \frac{\alpha_k - \alpha_{k-1}}{\alpha_k - \overline{\alpha_{k-1}}} \right| \qquad \forall k \in [2, M], \tag{9.7}$$

where

$$\gamma_k^{(1)}(P) = \frac{L_{22}(\alpha_k)}{u_P(\alpha_k)} \qquad \forall k \in [1, M],$$

$$\gamma_k^{(l+1)}(P) = \frac{\gamma_k^{(l)}(P) - \gamma_l^{(l)}(P)}{1 - \overline{\gamma_l^{(l)}(P)}\gamma_k^{(l)}(P)} \frac{\alpha_k - \alpha_l}{\alpha_k - \overline{\alpha_l}} \qquad \forall l \in [1, M-1] \qquad \forall k \in [l+1, M]. \tag{9.8}$$

Therefore the matrix inequality $\mathbf{U}(P) \succ \mathbf{\Pi}$ can be replaced by the set of scalar inequalities shown in eqs. (9.6) and (9.7). We use now a limiting argument to characterise also the boundary $\partial\mathbb{A}_R^N$ of the set $\mathbb{A}_R^N$, namely the polynomials $P \in \mathbb{P}_+^{2N}$ such that $\mathbf{U}(P) \succeq \mathbf{\Pi}$.

Remark 9.1.3. *It should be noted that, assuming $\alpha_k \neq \alpha_l$ for all $k \neq l$, parameters $\gamma_k^{(l+1)}(P)$ are not defined if $\gamma_l^{(l)}(P) = \gamma_k^{(l)}(P)$ and $\left|\gamma_l^{(l)}(P)\right| = \left|\gamma_k^{(l)}(P)\right| = 1$. This comes from the fact that the pseudo-hyperbolic distance $\delta(a, b)$ is only defined for $|a|$ and $|b|$ smaller than 1.*

However $\delta(a, a) = 0$ for all $a \in \mathbb{D}$ and the limit of $\delta(a, b)$ when a, b approach $t_0 \in \mathbb{T}$ equals 0 provided that $|a| < 1$ and $|b| < 1$. It is not the case, for instance, if tangential directions to the unit disc are taken. Nevertheless if interior-point methods are used, the feasibility of the polynomial P can be ensured at every iteration. Therefore P can only tend to the boundary $\partial\mathbb{A}_R^N$ at the end of the optimisation and not tangentially. In addition, if the feasibility of P is ensured, only the case for $i = 1$ in eq. (9.6) is not-redundant, as if the P tends toward a polynomial P_{opt} that saturates eq. (9.6) with $i > 1$, then it also saturates eq. (9.7) with $k = i$.

9.1.3 Alternative characterisation of the admissible polynomials by means of scalar inequalities.

Note from remark 9.1.3 that given the set of points $\alpha_1, \alpha_2 \cdots \alpha_M \in \mathbb{C}^-$ and the interpolation values $\gamma_1(P), \gamma_2(P) \cdots \gamma_M(P) \in \mathbb{D}$ and provided that the set of interpolating functions $\mathbb{E}^M$ is not empty, a limiting case can be considered where the set $\mathbb{E}^M$ degenerates to a singleton containing only a Blaschke product. To include this case, we complete

the function in eq. (9.8) with the non-tangential limit at the boundary of the admissibility domain as

$$\gamma_k^{(l+1)}(P) = \begin{cases} \dfrac{\gamma_k^{(l)}(P) - \gamma_l^{(l)}(P)}{1 - \overline{\gamma_l^{(l)}(P)}\gamma_k^{(l)}(P)} \dfrac{\alpha_k - \overline{\alpha_l}}{\alpha_k - \alpha_l} & \gamma_l^{(l)}(P) \neq \gamma_k^{(l)}(P) \\ 0 & \gamma_l^{(l)}(P) = \gamma_k^{(l)}(P) \end{cases}, \tag{9.9}$$

for all $l \in [1, M-1]$ and for all $k \in [l+1, M]$. We replace now eq. (7.3) with eqs. (9.6) and (9.7) to restate a new version of problem 7.1.2 involving only scalar inequalities.

Problem 9.1.1 (General matching problem with scalar inequalities)**.**

$$\begin{aligned} \textit{Find:} \quad & \min_{(\Psi,P)} \Psi & & (\Psi, P) \in \mathbb{R}_+ \times \mathbb{P}_+^{2N}, \\ \textit{Subject to:} \quad & P(\omega) \leq \Psi \cdot R(\omega) & & \omega \in \mathbb{I}, \\ & P(\omega) \geq \Gamma \cdot R(\omega) & & \omega \in \mathbb{J}, \\ & \delta(\gamma_k^{(k)}(P), \gamma_{k-1}^{(k)}(P)) \leq \left| \frac{\alpha_k - \alpha_{k-1}}{\alpha_k - \overline{\alpha_{k-1}}} \right| & & \forall k \in [1, M], \end{aligned} \tag{9.10}$$

where $\gamma_k^{(1)} = \frac{L_{22}(\alpha_k)}{u_P(\alpha_k)}$*, and* $\gamma_0^{(k)} = \alpha_0 = 0$ *for all* $k \in [1, M]$*.*

We state now a theorem on the number of constraints in eq. (9.10) that are saturated at the optimal point.

Lemma 9.1.1. *The polynomial* P_{opt} *solution to problem 9.1.1 can only saturate at most one of the constraints in eq.* (9.10)*.*

The proof is based on the fact that if any constraint in eq. (9.10) is saturated, namely $\delta(\gamma_i^{(i)}(P), \gamma_{i-1}^{(i)}(P)) = \left| \frac{\alpha_i - \alpha_{i-1}}{\alpha_i - \overline{\alpha_{i-1}}} \right|$ with $i \in [1, M-1]$, the remaining ones are zero from the definition in eq. (9.9). In this case the only interpolating function is a Blaschke product of degree $i-1$.

Proof. Suppose that $|\gamma_1^{(1)}(P_{opt})| = 1$, then since the interpolation is feasible we have $\gamma_1^{(1)}(P_{opt}) = \gamma_k^{(1)}(P_{opt})$, for all $k \in [1, M]$. Computing now the value of eq. (9.7) we obtain

$$0 \leq \left| \frac{\alpha_k - \alpha_{k-1}}{\alpha_k - \overline{\alpha_{k-1}}} \right| \qquad \forall k \in [2, M],$$

which is clearly satisfied. In this case the set $\mathbb{E}(u_{P_{opt}})$ contains only the function $\gamma_1^{(1)}$. Conversely, suppose that $|\gamma_1^{(1)}(P_{opt})| < 1$ and

$$\delta(\gamma_i^{(i)}(P_{opt}), \gamma_{i-1}^{(i)}(P_{opt})) = \left| \frac{\alpha_i - \alpha_{i-1}}{\alpha_i - \overline{\alpha_{i-1}}} \right| \qquad i \in [2, M].$$

Then we have

$$|\gamma_i^{(i+1)}(P_{opt})| = \delta(\gamma_i^{(i)}(P_{opt}), \gamma_{i-1}^{(i)}(P_{opt})) \left| \frac{\alpha_i - \overline{\alpha_{i-1}}}{\alpha_i - \alpha_{i-1}} \right| = 1.$$

Once again, since we supposed $\mathbb{E}(u_{P_{opt}})$ is not empty

$$\gamma_k^{(i+1)} = \gamma_i^{(i+1)} \qquad i+1 < k \leq M.$$

Additionally, we from eq. (9.7) with $k \in [i+1, M-1]$ we obtain $0 \leq \left| \frac{\alpha_k - \alpha_{k-1}}{\alpha_k - \overline{\alpha_{k-1}}} \right|$. Therefore when the i-th constraint is saturated we obtain $\delta(\gamma_k^{(k)}, \gamma_{k-1}^{(k)}) = 0$ for all k in $[i+1, M]$. Hence at the optimal point only one of the constrains in 9.7 can be binding. □

9.1.4 Reducible matching problem

Note now that, if the k-th constraint of the condition set given by eq. (9.10) is binding, the set $\mathbb{E}(u_{P_{opt}})$ is a singleton containing a Blaschke product of degree $k-1$. In this case the inequalities in eq. (9.10) with $i \in [k+1, M]$ are not saturated and therefore the can be removed without modifying the optimal solution P_{opt} to problem 9.1.1. In other words, the interpolation conditions in eq. (9.1) with $k+1 > i \geq M$ are redundant since they are not used in any of the binding constraints. Thus the problem can be reduced by removing the interpolation conditions at $\alpha_{k+1}, ..., \alpha_M$.

This simplification of the matching problem by eliminating the interpolation conditions at the points α_i with $i > k$ corresponds to a different problem with a load $\hat{L}$ of lower degree k since it only has k transmission zeros in the points α_i with $i \in [1, k]$. Therefore the solution P_{opt} of the problem of matching with the load L of degree M and with a fixed degree N for the global system is also the optimal solution to the reduced problem with global degree N and a load $\hat{L}$ of degree $k < M$. This implies that the Blaschke product obtained from P_{opt} and the load $\hat{L}$ is still of degree $k-1$, namely the degree of the load minus 1.

The previous argument can also be applied to problem 7.1.2 depending on the range of the array $\mathbf{U}(P_{opt}) - \mathbf{J\!I}$ as stated next.

Theorem 9.1.2 (Degree of the Blaschke product)**.** *Let P_{opt} be the optimal polynomial for problem 7.1.2 and $\mathcal{R}$ the rank of the matrix $\mathbf{U}(P_{opt}) - \mathbf{J\!I}$ with a load L of McMillan degree M. If $\mathcal{R} < M-1$ then there exist a reduced load $\hat{L}$ of degree $\mathcal{R}+1$ such that P_{opt} is also the optimal solution to the problem of matching the reduced load $\hat{L}$.*

Corollary 9.1.1. *Problem 7.1.2 is not reducible if and only if $\mathcal{R} \geq M-1$.*

Remark 9.1.4. *Theorem 9.1.2 gives us the necessary motivation to consider the case where $\mathcal{R} \geq M-1$ as the general case of the matching problem. This case corresponds to the set of problems of the form problem 7.1.2 that are not reducible, which can be characterized by a matrix $\mathbf{U}(P_{opt}) - \mathbf{J\!I}$ of rank at least $M-1$.*

9.1.5 Example of reducible matching problem

To facilitate the correct understanding of the concept of reducibility introduced above, we show below how, from a non-reducible matching problem, we can obtain another equivalent and reducible problem which provides the same optimal solution as the first. To begin, we consider the load $\hat{L}$ of degree $\mathcal{R}+1$ and set $N \geq \mathcal{R}+1$ in problem 7.1.2.

Consider the problem of matching a load L of degree M within a passband $\mathbb{I} \subset \mathbb{R}$. We pick a degree for the global system $N > M$ and a transmission polynomial $R = R_F R_L$ where R_L is the transmission polynomial of the load and $R_F \in \mathbb{P}_+^{2N-2M}$. We denote by $\mathbb{A}_R^N$ the set of admissible polynomials corresponding to the load. We have

$$P_{opt} = \arg \min_{P \in \mathbb{A}_R^N} \max_{\omega \in \mathbb{I}} \frac{P(\omega)}{R(\omega)}.$$

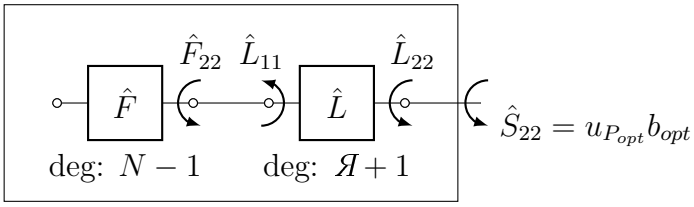


Figure 9.1: Optimal global system of McMillan degree $N + \mathcal{A}$.

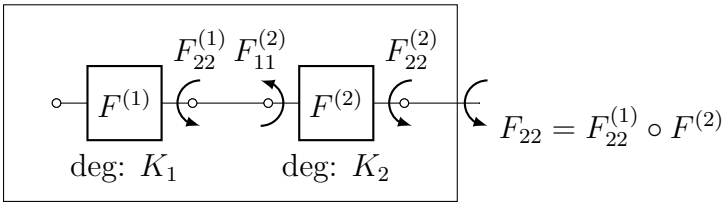


Figure 9.2: Decomposition of the matching filter $\hat{F}$ of McMillan degree $K = K_1 + K_2$ into the sub-devices $F^{(1)}$ and $F^{(2)}$ with McMillan degree K_1 and K_2 respectively.

Assume that once the optimal solution P_{opt} is calculated, we have a Blaschke product b_{opt} of degree $\mathcal{A}$. Therefore the reflection $S_{22} = u_P \cdot b_{opt}$ is of degree $N + \mathcal{A}$ and after performing the extraction of the load of degree $\mathcal{A} + 1$, the matching network obtained presents a degree $N + \mathcal{A} - (\mathcal{A} + 1) = N - 1$.

Theorem 9.1.3 (Equivalent reducible problem)**.** *The polynomial P_{opt} is also solution to a reducible problem with a load of degree $\hat{M} > M$ and same global degree and transmission polynomial. Denoting the admissible set for this new load by $\mathbb{A}_R^{\hat{N}}$ we have*

$$P_{opt} = \arg \min_{P \in \mathbb{A}_R^{\hat{N}}} \max_{\omega \in \mathbb{I}} \frac{P(\omega)}{R(\omega)}.$$

We illustrate in fig. 9.1 the obtained global system, which is composed of a matching filter of degree $N - 1$ cascaded with the load of degree $\mathcal{A} + 1$.

Proof. The modulus squared of the function $\hat{S}_{22}$ in fig. 9.1 can be expressed as

$$|\hat{S}_{22}(\omega)|^2 = \frac{P_{opt}(\omega)}{P_{opt}(\omega) + R(\omega)} \qquad \forall \omega \in \mathbb{R}.$$

Let us express the matching filter $\hat{F}$ in fig. 9.1 as the cascade of two sub-devices $\hat{F} = F^{(1)} \circ F^{(2)}$ where $F^{(1)}$ is of McMillan degree $K_1 < N - 1$ while $F^{(2)}$ has McMillan degree K_2 with $1 \leq K_2 \leq N - 1$ as shown if fig. 9.2. This decomposition can be done for any matching filter $\hat{F}$ of any arbitrary degree $K > 0$ since the sub-device $F^{(1)}$ could be of degree 0.

Now are disposed to state the reducible version of the former problem by considering the load L constructed as $L = F^{(2)} \circ \hat{L}$. Note that the device $F^{(2)}$ is part of the optimal filter obtained with the non-reducible problem. Let us again illustrate this problem in fig. 9.3. We are looking for the best matching filter F such that the squared modulus of the global reflection $S_{22} = F_{22} \circ L$ can be expressed as

$$|S_{22}(\omega)|^2 = \frac{P(\omega)}{P(\omega) + R(\omega)} \qquad P \in \mathbb{P}_+^{2N} \qquad \forall \omega \in \mathbb{R}.$$

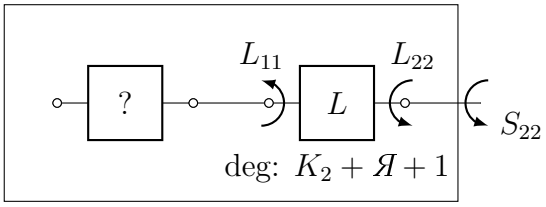


Figure 9.3: Reducible problem considering the load $L = F^{(2)} \circ \hat{L}$.

However, the best filter F can not be other than the sub-device $F^{(1)}$ since the existence of a better matching network for the problem outlined in fig. 9.3 would contradict optimality of the matching filter F obtained in the preceding problem. We have

$$F = F^{(1)},$$
$$S_{22} = \hat{S}_{22} = u_{P_{opt}} b_{opt}.$$

This implies that the Blaschke product for the problem illustrated in fig. 9.3 is of degree $\mathcal{A} - 1$. With the above procedure, we have constructed a problem of matching a load of degree $M = K_2 + \mathcal{A} + 1$ such that the matrix $\mathbf{U}(P_{opt}) - \mathbf{\Pi}$ is of rank $\mathcal{A} < M - 1$. Therefore, according to corollary 9.1.1 this problem is reducible. □

Remark 9.1.5. *Note that, in this case the reducibility of the obtained problem is trivial since by extracting $F^{(2)}$ from the load, we obtain the simplified problem that we started with.*

9.2 Sub-optimal feasible function

In this section we consider exclusively the case of a matching problem formulated as in problem 7.1.2 which is not reducible. Obviously, in practice there are also reducible problems which can be addressed directly or through the convenient prior simplification. However, from a theoretical point of view, non-reducible problems present a greater interest.

Furthermore, we consider the case where the optimal polynomial P_{opt} verifies

$$\mathbf{U}(P_{opt}) \succeq \mathbf{\Pi}.$$

The motivation for this assumption comes from the fact that if we have $\mathbf{U}(P_{opt}) \succ \mathbf{\Pi}$, then the constraint eq. (7.3) is not binding and the polynomial P_{opt} is the optimal solution to the classical filter synthesis problem, namely problem 2.13.4, which has already been studied in depth. With those assumptions we have

$$\operatorname{rank}(\mathbf{U}(P_{opt}) - \mathbf{\Pi}) = \mathcal{A} = M - 1.$$

In this case the evolution of the optimal lower bound with respect to the degree of the load is clear since the obtained function S_{22} is of degree $N + M - 1$ while the matching network F has McMillan degree $N - 1$. As have already discussed, problem 7.1.2 provides

lower hard bounds for the optimal criterium in problem 4.1.1 in the following sense: if l_B denotes the criterium provided by the polynomial P_{opt} in problem 7.1.2

$$\psi_{opt} = \max_{\omega \in \mathbb{I}} \frac{P_{opt}(\omega)}{P_{opt}(\omega) + R(\omega)}$$

and ψ_{best} the optimal criterium in problem 4.1.1 as

$$\psi_{best} = \min_{S_{22} \in \mathbb{F}_R^N} \max_{\omega \in \mathbb{I}} |S_{22}(\omega)|^2.$$

subject to $\gamma \leq |S_{22}(\omega)|^2$ for all $\omega \in \mathbb{J}$. Then we have the lower bound

$$\psi_{opt} \leq \psi_{best}.$$

Additionally, problem 7.1.2 provides us with a function $S_{22} \in \mathbb{F}^{N+\mathcal{A}}$ such that

$$\max_{\omega \in \mathbb{I}} |S_{22}(\omega)|^2 = l_B.$$

However, we are still missing a function $S_{22} \in \mathbb{F}_R^N$. Our goal now is, firstly, to obtain a function $S_{22} \in \mathbb{F}_R^N$; and secondly to minimise the level of reflection $\psi = \max_{\omega \in \mathbb{I}} |S_{22}|^2$ approaching the optimal bound ψ_{opt} as closely as possible while satisfying the selectivity constraints $\gamma \leq |S_{22}(\omega)|^2$ for all $\omega \in \mathbb{J}$.

With respect to the calculation of a feasible function $S_{22} \in \mathbb{F}_R^N$, we can easily find different possibilities, each of them providing different results in terms of optimality or computational efficiency. Next we are going to review two procedures of completely different nature but nevertheless, both related to the thematic or algorithms treated in this thesis.

9.2.1 Forced Blaschke simplification. A matching problem with prescribed reflection zeros.

Here is a rather different technique for obtaining a function $S_{22} \in \mathbb{F}_R^N$ by solving a series of problems similar to problem 7.1.2. It is important to note that this time we are calculating a feasible function for the global system S_{22} and not for the matching filter as in the previous section. However, if this function S_{22} presents a McMillan N, the function F_{22} obtained after the extraction of the load is of degree $K = N - M$. To obtain the function $S_{22} \in \mathbb{F}_R^N$ we impose a McMillan degree of N on the function S_{22} calculated as

$$S_{22} = u_{P_{opt}} \cdot b_{opt} \tag{9.11}$$

where $u_{P_{opt}}$ is a minimum phase function and McMillan degree N while b_{opt} is a Blaschke product of degree $\mathcal{A}$ in the form

$$b_{opt} = \prod_{k=1}^{\mathcal{A}} \frac{\lambda - \xi_k}{\lambda - \overline{\xi_k}}. \tag{9.12}$$

Note that we are considering in this section only the case where the Blaschke product b_{opt} is of degree $\mathcal{A}$, therefore a necessary condition to obtain a function S_{22} of degree

N is that $\mathit{Я}$ simplifications occur between the function $u_{P_{opt}}$ and the Blaschke product b_{opt}.

Since b_{opt} does not vanish in $\mathbb{C}^+$, namely where $u_{P_{opt}}$ has poles, then $\mathit{Я}$ pole-zero cancellation shall occurs at the points $\overline{\xi_k}$ with $1 \leq k \leq \mathit{Я}$. Therefore we impose the fact that the function $u_{P_{opt}}$ vanishes at the points ξ_k. Note that we do not know a priory the points $\overline{\xi_k}$, namely the poles of the Blaschke. Nevertheless we try to enforce the simplification by imposing some of the roots of the polynomial P in problem 7.1.2 to given positions in the complex plane. Then we have

Problem 9.2.1 (General problem with prescribed reflection zeros)**.**

$$\begin{array}{lll} \textit{Find:} & \min\limits_{(\Psi,P)} \Psi & (\Psi, P) \in \mathbb{R}_+ \times \mathbb{P}_+^{2N}, \\ \textit{Subject to:} & P(\omega) \leq \Psi \cdot R(\omega) & \omega \in \mathbb{I}, \\ & P(\omega) \geq \Gamma \cdot R(\omega) & \omega \in \mathbb{J}, \\ & P(\xi_k) = 0 & 1 \leq k \leq \mathit{Я}, \\ & \mathbf{U}(P) \succeq \mathbf{\Pi}, & \end{array}$$

where the points ξ_k are fixed by the user as the estimated poles of the Blaschke product. To fix the points ξ_k, we take the poles of the optimal Blaschke product b_{opt} obtained after solving problem 7.1.2.

Similarly to problem 7.1.2, we can state the convexity of this enhanced problem

Theorem 9.2.1 (Convexity of the matching problem with prescribed reflection zeros)**.** *Problem 9.2.1 is convex.*

Proof. The proof follows directly from the fact that the set P_+^{2N} is a vector space, and therefore the set of polynomials $P \in \mathbb{P}_+^{2N} | P(\xi_k) = 0 \ \forall k \in [1, \mathit{Я}\,]$ is a convex sub-space of $\mathbb{P}_+^{2N}$. □

Therefore, this problem is solved as explained for the case of problem 7.1.2 since these new conditions of equality can be ensured by the elimination method discussed in section 8.2.2. Let us now provide a reformulated version of theorems 4.5.1 and 4.5.3 which holds for problem 9.2.1.

Theorem 9.2.2 (Number of extremal point of P_{opt})**.** *Let P_{opt} be the polynomial providing the optimal criterium Ψ_{opt} to problem 9.2.1. Denote $x_i \in \mathbb{I}$ with $i \in [1, n \leq N]$ all the roots of the polynomial $P_{opt} - \Psi_{opt} R$ within the interval $\mathbb{I}$. Considering the multiplicity function $\mu(x_i)$ defined in definition 4.5.1, we have*

$$\frac{1}{2} \sum_{i=1}^{n} \mu(x_i) \geq N - \mathit{Я} + 1.$$

Proof. We can now just adapt the proof of theorem 4.5.1 which is very similar to the proof of theorem 9.2.2. Indeed the proof follows almost identically with two exception. The first one is that this time we assume

$$\frac{1}{2} \sum_{i=1}^{n} \mu(x_i) \leq N - \mathit{Я}.$$

The second one comes from the fact that the positive polynomial $\Phi(\lambda) \in \mathbb{P}_+^{2N}$ must have zeros at the points ξ_k to ensure that the obtained polynomial $\hat{P}$ satisfies $\hat{P}(\xi_k) = 0$ for all $1 \leq k \leq \mathit{Я}$. Since Φ is a positive polynomial, it also vanishes at the points $\overline{\xi_k}$. Therefore, counting multiplicity, $\Phi(\lambda)$ has at most $2N - 2\mathit{Я}$ zeros at the points x_i and $2\mathit{Я}$ additional zeros at the points $\xi_k, \overline{\xi_k}$, thus obtaining a polynomial of degree $2N$. □

Furthermore with the assumption that $\mathit{Я} = M - 1$ and denoting as in previous sections by $K = N - M$ the McMillan degree of the matching network F such that $S_{22} = F_{22} \circ L$, we have the following corollary

Corollary 9.2.1. *If all points x_i have multiplicity less or equal to 2, then the optimal criterium Ψ_{opt} is attained by the function $P_{opt}(\omega)/R(\omega)$ at least $K + 2$ times within the interval $\omega \in \mathbb{I}$.*

With respect to the remaining theorems and properties enunciated in chapter 4, they still hold for problem 9.2.1. This can be easily verified if we reformulate problem 9.2.1 in the following form similar to problem 7.1.2 allowing us to reuse the already provided proofs of the theorems stated in previous chapters, with the exception of theorems 4.5.1 and 4.5.3.

Problem 9.2.2 (General problem with prescribed reflection zeros)**.**

$$\begin{array}{rcl} \textit{Find:} & \min\limits_{(\Psi,P)} \ \Psi & (\Psi, P) \in \mathbb{R}_+ \times \mathbb{P}_+^{2(N-\mathit{Я})}, \\ \textit{Subject to:} & \Xi(\omega)P(\omega) \leq \Psi \cdot R(\omega) & \omega \in \mathbb{I}, \\ & \Xi(\omega)P(\omega) \geq \Gamma \cdot R(\omega) & \omega \in \mathbb{J}, \\ & \mathbf{U}(P) \succeq \boldsymbol{\mathit{Л}}, & \end{array}$$

with $\Xi(\lambda) = \prod_{k=1}^{\mathit{Я}} (\lambda - \xi_k)\left(\lambda - \overline{\xi_k}\right)$.

Remark 9.2.1. *Note that we assume $N \geq \mathit{Я}$ in problem 9.2.2.*

9.2.2 A fixed-point version of the optimisation algorithm

We have just introduced a version of the matching problem where some reflection zeros, namely roots of the polynomial P, are fixed. However, it is important to bear in mind that the Blaschke product b_{opt} depends on both the problem data, namely the load, and the polynomial P_{opt}. Then, once the optimal solution to problem 9.2.2 is obtained, the Blaschke product b_{opt} and also the set of points ξ_k are modified and therefore the solution obtained P_{opt} is not optimal for the new set of points ξ_k with $1 \leq k \leq \mathit{Я}$. Additionally, since pole-zero cancellation in eq. (9.11) might not occur anymore, the obtained solution is again of the class $\mathbb{F}^{N+\mathit{Я}}$, namely we have $F_{22} \in \mathbb{\Sigma}^{K+\mathit{Я}}$.

Thus, in order to obtain a matching filter of degree K, we propose the implementation of a fixed-point algorithm consisting of the resolution of a series of problems such as the one formulated in problem 9.2.2 where the polynomial $S(\lambda)$ in each iteration is obtained from the zeros and poles of the Blaschke product obtained in the previous iteration. With this algorithm we attempt to reach with this procedure a fix point where the solution of this series of problem converges toward S_{22} in the form given by eq. (9.11) where $\mathit{Я}$

pole-zero cancellations occurs providing a function $S_{22} \in \mathbb{F}_R^N$. This fixed-point algorithm has been used in all examples of matching filter synthesis provided in next chapter.

We proceed as follows

- Iteration 0. The algorithm starts with the solution to problem 7.1.2 with a global degree $N \geq M$ is computed. We obtain the polynomial $P_{opt}^{(0)}$ solution to problem 7.1.2 as well as the minimum phase function $u_{P_{opt}^{(0)}}(\lambda)$. We compute the Blaschke product $b_{opt}^{(0)}$ such that $S_{22}^{opt} = u_{P_{opt}^{(0)}} \cdot b_{opt}^{(0)}$ is feasible, particularly $S_{22}^{opt} \in \mathbb{F}^{N+\mathcal{I}}$ with $\mathcal{I} = M-1$. The function $b_{opt}^{(0)}$ is in the form given by eq. (9.12), namely

$$b_{opt}^{(0)} = \prod_{k=1}^{\mathcal{I}} \frac{\lambda - \xi_k^{(0)}}{\lambda - \overline{\xi_k^{(0)}}}.$$

  The points $\xi_k^{(0)}$ with $k \in [1, \mathcal{I}]$ are used to initialise the fixed-point algorithm.

- Iteration i-th. With the points $\xi_k^{(0)}$ obtained in the previous iteration define

$$\Xi^{(i)}(\lambda) = \prod_{k=1}^{\mathcal{I}} \left(\lambda - \xi_k^{(0)}\right)\left(\lambda - \overline{\xi_k^{(0)}}\right).$$

  Now we state and solve problem 9.2.2 with the same global degree N and using the function $S^{(0)}$. We have $N \geq M$ and $\mathcal{I} = M-1$, therefore the condition $N \geq \mathcal{I}$ holds. We obtain the polynomial $P_{opt}^{(i)}$ solution to the problem and a Blaschke product $b_{opt}^{(i)}$ such that $b_{opt}^{(i)} \cdot u_{\left(S^{(i)} P_{opt}^{(i)}\right)}$ is feasible. The function $b_{opt}^{(i)}$ has the form

$$b_{opt}^{(i)} = \prod_{k=1}^{\mathcal{I}} \frac{\lambda - \xi_k^{(i)}}{\lambda - \overline{\xi_k^{(i)}}}.$$

- Iterarion $i+1$-th. We update the function $\Xi(\lambda)$ and solve again problem 9.2.2. Note that if the function $\Xi^{(i+1)}$ is not heavily modified, then the optimal polynomial in the iteration $k+1$, namely $P_{opt}^{(i+1}$, is not far from the polynomial $P_{opt}^{(i)}$. Therefore the polynomial $P_{opt}^{(i)}$ is a remarkably good starting point in the iteration $i+1$. To avoid big changes from the polynomial $\Xi^{(i)}$ to $\Xi^{(i+1)}$ we use the update formula

$$\Xi^{(i+1)}(\lambda) = (1-\kappa)\Xi^{(i)}(\lambda) + \kappa \prod_{k=1}^{\mathcal{I}} \left(\lambda - \xi_k^{(i)}\right)\left(\lambda - \overline{\xi_k^{(i)}}\right) \qquad 0 < \kappa \leq 1,$$

  where the value of κ can be modified to adjust the convergence rate. In the implementation done in this work, we have selected $\kappa = 0.5$.

- Termination criterium. We define the error function $E(i+1)$ in the $i+1$-th iteration as

$$E(i+1) = \max_{\omega \in \mathbb{I}} \left| 1 - \frac{u_{\left(S^{(i)} P_{opt}^{(i)}\right)}(\omega)}{u_{\left(S^{(i+1)} P_{opt}^{(i+1)}\right)}(\omega)} \right|.$$

The algorithm is then stopped when for a small value of ε we have $E(i) < \varepsilon$.

It should be noted that we have not provided any convergence proof of the presented fixed-point algorithm. Indeed although this algorithm has shown an outstanding convergence rate in practice, we can still find some cases where the algorithm fails to converge, this might happens, for instance when some of the points ξ_k are located close to the real interval $\mathbb{I}$. For this reason we also stop the algorithm if it happens that $E(i+1) > E(i)$. Assuming the algorithm is stopped at the t-th iteration, we compute the global reflection parameter S_{22}^{best} as

$$S_{22}^{best} = u_{\left(S^{(t)} P_{opt}^{(t)}\right)}.$$

With this fixed-point algorithm, we try to obtain at the final iteration, $\mathcal{Я}$ pole-zero simplifications in the function S_{22}^{best}, such that $S_{22}^{best} \in \mathbb{F}_R^N$. In this case the Darlington equivalent of the load L can be de-embedded from the function S_{22}^{best} obtaining a function $F_{22} \in \mathbb{\Sigma}_{R_F}^K$ such that $F_{22} \circ L = S_{22}^{best}$. Nevertheless these simplifications are never exact in practice, even assuming the convergence of the algorithm, and the task of removing the common factors in the numerator and denominator of S_{22}^{best} together with the posterior de-embedding of the load can become problematic.

To overcome this issue, we introduce next a numeric algorithm which allows us to de-embed the load from the function S_{22}^{best} without need of removing the common poles and zeros, and obtaining a function F_{22} with the desired degree, namely $F_{22} \in \mathbb{\Sigma}_{R_F}^K$.

9.2.3 Load extraction to obtaining a matching network of degree K. A different application of the point-wise matching algorithm.

The algorithm proposed in this section is a direct application of the point-wise matching procedure proposed in [46]. This method has already been introduced in section 3.4.2 and is based on theorem 3.4.1. The aforementioned procedure allows us to determine the unique rational Schur function F_{22} of degree K which solves the interpolation problem

$$F_{22}(x_i) = \nu_i \qquad \forall i \in [0, K].$$

Therefore we can just distribute a set of points $x_i \in \mathbb{I}$ with $0 \leq i \leq K$ and compute the interpolation values ν_i from eq. (3.13) as the values of the filter reflection F_{22} at those points. We have

$$\nu_i = \frac{L_{22}(x_i) - S_{22}(x_i)}{\det(L(x_i)) - S_{22}(x_i) L_{11}(x_i)} \qquad \forall i \in [0, K].$$

The procedure introduced in section 3.4.2 allows us to obtain a function $F_{22}(\omega)$ of McMillan degree K which perfectly interpolates the values ν_i on a given set of $K+1$ distinct frequency points $x_0, x_1 \cdots x_K \in \mathbb{R}$. Moreover this procedure allows for the transmission polynomial R_F to be prescribed, therefore the obtained rational function F_{22} is of the class $\mathbb{\Sigma}_{R_F}^K$, namely

$$F_{22}(\omega) = \frac{p_F}{q_F}, \tag{9.13}$$

with $p_F \in \mathbb{P}^K$ and q_F the stable polynomial satisfying $q_F^\star q_F = p_F^\star p_F + R_F$. Additionally, as it has been proved in [46], the application that associates to each polynomial $p_F \in \mathbb{P}^K$ the evaluation of the rational function F_{22} at each point x_i is twice differentiable and has differentiable inverse. If we denote by $\Theta_F \in \mathbb{R}^{2N+1}$ the vector $[\theta_{2N+1}, \theta_{2N-1}, \cdots, \theta_0]^T$ with the coefficients of the polynomial p_F with respect to the Tchebyshev basis as in eq. (7.20) we can consider the function $[q(p_F)](x_i)$ with $i \in [0, K]$, whose gradient with respect to the vector Θ_p denoted here by $\nabla_{q(x_i)}(\Theta_p)$ has already been calculated in eq. (7.25). Therefore the derivatives of the function $[q(p_F)](x_i)$ at each point x_i with respect to the coefficient θ_k of p_F takes the expression derived in eq. (7.26), namely

$$\mathrm{D}_k F_{22}(x_i) = \frac{\mathrm{D}_k p_F(x_i)}{[q(p_F)](x_i)} - F_{22}(x_i)\mathrm{D}_k q_F(x_i) \qquad \forall i \in [0, K]. \tag{9.14}$$

If we consider now the function $F : \mathbb{R}^{2K+2} \longrightarrow \mathbb{C}^{K+1}$ defined as

$$F(\Theta_p) = \begin{bmatrix} F_{22}(x_0) & F_{22}(x_1) & \cdots & F_{22}(x_K) \end{bmatrix}^T,$$

then the Jacobian matrix of F with respect to the vector Θ_p can be computed as

$$\mathbf{J}_F(\Theta_p) = \mathrm{diag}(q_F(x_i))^{-1}\mathbf{J}_p(\Theta_p) - \mathrm{diag}(\hat{F}_{22}(x_i))\mathbf{J}_q(\Theta_p),$$

where $\mathbf{J}_p(\Theta_p)$ and $\mathbf{J}_q(\Theta_p)$ represent the Jacobian matrices of $p_F(x_i)$ and $q_F(x_i)$ respectively, at each point x_i and with respect to Θ_p. Therefore for a point $\Theta_p^{(l)}$ in a neighbourhood of the vector $\Theta_p^{(0)}$ we have

$$\mathbf{J}_F(\Theta_p^{(l)})\left(\Theta_p^{(l+1)} - \Theta_p^{(l)}\right) = F(\Theta_p^{(l+1)}) - F(\Theta_p^{(l)}).$$

Finally, the function F_{22} can be calculated by means of the homotopy which consists in deforming an initial solution $F(\Theta_p^{(0)})$ in small increments $\Delta F = F(\Theta_p^{(l+1)}) - F(\Theta_p^{(l)})$ so that in each iteration the correspondent vector Θ_p^{l+1} can be approximated by

$$\Theta_p^{l+1} = \Theta_p^l + \mathbf{J}_F(\Theta_p^{(l)})^{-1}\Delta F.$$

At each time the Jacobian matrix $\mathbf{J}_F(\Theta_p^{(l)})$ is inverted. However as it has been shown in the literature, the matrix $\mathbf{J}_F(\Theta_p^{(l)})$ is well conditioned and accidents are not encountered in the general case. This process is then iterated until to obtain the vector Θ_p satisfying

$$F(\Theta_p) = \begin{bmatrix} \nu_0 & \nu_1 & \cdots & \nu_K \end{bmatrix}^T.$$

This vector Θ_p contains the coefficients of the polynomial $p_F \in \mathbb{P}^N$ such that the function $F_{22}(\omega) = \frac{p_F(\omega)}{q_F(\omega)}$ interpolates the values ν_i at the points x_i with $i \in [0, K]$.

9.2.4 Local optimisation of the matching network

With the previous algorithm, we try to obtain a function $F_{22} \in \mathbb{F}_R^K$ which provides a criterion in problem 4.1.1 as close as possible to the optimal bound ψ_{opt} obtained from problem 7.1.2. However, none of the presented algorithms guarantees the optimality of the function F_{22} for problem 4.1.1. Therefore, as a final step, we perform a local optimization of the solution obtained. This minimization can be expressed as

Problem 9.2.3 (Local minimisation)**.**

$$\begin{aligned} \textit{Find:} \qquad & \psi_{best} = \min_{F_{22}\in\mathbb{\Sigma}_R^N} \max_{\omega\in\mathbb{I}} \delta\left(F_{22}(\omega), L_{11}^\star(\omega)\right), \\ \textit{Subject to:} \qquad & \delta\left(F_{22}(\omega), L_{11}^\star(\omega)\right)^2 \geq \gamma. \end{aligned}$$

To perform such minimisation, the function $F_{22} \in \mathbb{\Sigma}_R^N$ is again parametrised as a function of the polynomial $p_F \in \mathbb{P}^K$ as in eq. (9.13), namely

$$F_{22}(\omega) = \frac{p_F(\omega)}{q_F(\omega)},$$

with $p_F, q_F \in \mathbb{P}^K$ and q_F the stable polynomial satisfying $q_F q_F^\star = p_F p_F^\star + R_F$. Note that given the pair of polynomials p_F, q_F the analytical expressions for the gradient and Hessian matrix of F_{22} with respect to the coefficients of p_F have already been computed. Therefore we can make use of these analytical formulas to write an efficient solver for problem 9.2.3 by means of the already recurrent Newton method.

9.3 Summary

Next we provide an overview of the matching algorithm used in conjunction with loads of degree $M > 1$. This algorithm is analogous to the procedure summarised in section 6.9 for a load of degree 1, nevertheless a couple of additional steps are required.

1. Computation of the Darlington equivalent of the load. This step is unchanged with respect to section 6.9. We obtain in this case a rational 2×2 scattering matrix L of McMillan degree M in the form

   $$L(\omega) = \begin{pmatrix} p_L^\star(\omega) & -r^\star(\omega) \\ r(\omega) & p_L(\omega) \end{pmatrix}.$$

   We consider here that the transmission polynomial of the load $R_L = r_L r_L^\star$ does not vanish on the real axis, namely $R_L(\omega) \neq 0$ for all $\omega \in \mathbb{R}$. Additionally we assume all roots of R_L to have simple multiplicity. Therefore there exists M points $\alpha_i \in \mathbb{C}^-$ such that $R_L(\alpha_i) = 0$ for all $i \in [1, M]$.

2. Determine lower bounds for the system reflection. We fix here a transmission polynomial R_F for the matching filter and a passband $\mathbb{I}$. Now with the points α_i computed before from the Darlington equivalent of the load, the solution to problem 7.1.2 with a global degree $N \geq M$ is computed. We obtain the polynomial P_{opt} solution to problem 7.1.2 as well as the minimum phase function $u_{P_{opt}}(\lambda)$. This function provides us with a lower hard bound ψ_{opt} for the reflection for the global system reflection within the band $\mathbb{I}$. If we consider a reflection $F_{22} \in \mathbb{\Sigma}_{R_F}^K$ We have for all S_{22} in the form $S_{22} = F_{22} \circ L$

   $$\max_{\omega\in\mathbb{I}} |S_{22}(\omega)|^2 \geq \psi_{opt}.$$

3. Fixed-point algorithm. We perform a fixed-point algorithm as described in section 9.2.2 by solving iteratively a sequence of problems in the form of problem 9.2.2. As the result of the fixed-point algorithm we obtain a function $S_{22}^{best}(\omega) \in \mathbb{F}^{N+\mathit{Я}}$ where $\mathit{Я}$ pole-zero cancellations occurs.

4. De-embedding of the load. We use the interpolation procedure discussed in section 9.2.3 to de-embed the Darlington equivalent of the load L by computing the filter reflection $F_{22} \in \mathbb{\Sigma}_{R_F}^K$ which solves the interpolation problem

$$F_{22}(x_i) = \frac{L_{22}(x_i) - S_{22}(x_i)}{\det(L(x_i)) - S_{22}(x_i)L_{11}(x_i)}.$$

in a set of $K+1$ points x_i distributed within the passband $\mathbb{I}$. The function F_{22} provides an upper bound ψ_{best} for the reflection level ψ solution to problem 4.1.1. We have

$$\psi_{opt} \leq \psi \leq \psi_{best}. \tag{9.15}$$

5. Local optimisation. Finally we perform a local optimisation of the filter reflection F_{22} to compute

$$\psi = \min_{F_{22} \in \mathbb{\Sigma}_{R_F}^K} \max_{\omega \in \mathbb{I}} |F_{22} \circ L|,$$

where we already dispose of the information that the computed value ψ must satisfy eq. (9.15).

References

[46] L. Baratchart, M. Olivi, and F. Seyfert, "Boundary Nevanlinna-Pick interpolation with prescribed peak points. Application to impedance matching," *SIAM Journal on Mathematical Analysis*, 2017.

Part IV

Numerical results and practical applications of the matching filter synthesis

Chapter 10:

Practical examples and results

In this chapter we present some practical applications of the theory developed previously in the thesis, particularly in chapter 4. In this section we focus on antenna matching so we consider as load the reflection of an antenna. The matching filter, therefore, should provide a global reflection as weak as possible throughout the working band.

Note that in this chapter the load is generally a device with a single port, however in these cases we obtain, using the Darlington equivalent, a two-port device without losses, and with the reflection of the load as a reflection in port 1. Then we can use this Darlington equivalent to build the global system and solve the problem of matching as we have done so far. In this case the element S_{11} of the global system corresponds to the reflection at the input of the matching filter when it is connected to the antenna. On the other hand the element S_{21} of the global system represents an equivalent transmission, namely $|S_{21}|^2 = 1 - |S_{11}|^2$ which, as the load and the generator are terminated by the reference impedance, we denote *transducer gain* following the notation introduced in [47].

The goal is to minimize the reflection level ψ such that $|S_{11}(\omega)|^2 \leq \psi$ in the whole band $\omega \in \mathbb{I}$ by means of a matching filter of fixed McMillan degree K, or what is equivalent due to the absence of losses, maximize the *transducer gain*. It is also important to note that with the algorithm presented in [48] the obtaining of a matching filter of McMillan degree K is guaranteed.

With the theory developed in the preceding chapters, on the other hand, we obtain a lower limit for the best level of matching ψ_{opt} reached with a McMillan degree filter K, but without obtaining in all cases a filter of that degree that reaches that level of matching ψ_{opt}. However, through the process presented in this chapter it is possible to obtain a sub-optimal filter of degree K which is locally optimal for the matching problem and provides a level of matching $\psi_{best} \geq \psi_{opt}$.

In addition, for each of the antennas presented here, we make a theoretical study comparing the lower limit ψ_{opt} with respect on the McMillan degree K of the matching filter with the reflection level ψ_{best} achieved through a sub-optimal filter of the same McMillan degree K. This sub-optimal filter is calculated by the fix-point algorithm described in chapter 9 followed by a local optimisation.

Finally, it should be noted that, in addition to this sub-optimal filter, the formulation of the matching problem developed in this thesis also provides us with another matching filter through which the optimal matching level ψ_{opt} is reached. Nevertheless this filter has a McMillan $K + M - 1$ degree,where M is the McMillan degree of the load, or the Darlington equivalent in this case, so it can not be considered as a solution to the original problem.

10.1 Small superdirective antenna

As a first and simple example, we consider the problem of matching the small superdirective antenna presented in [49] in the interval $\mathbb{I}$ defined as

$$\mathbb{I} = [870, 900]\ MHz.$$

The reflection of this antenna L_{11} appears in fig. 10.1 and corresponds to the reflection at the input of a two-ports Darlington equivalent of McMillan degree $M = 2$. In addition we also set the polynomial R_F to have no finite transmission zeros, in particular $R_F = 1$.

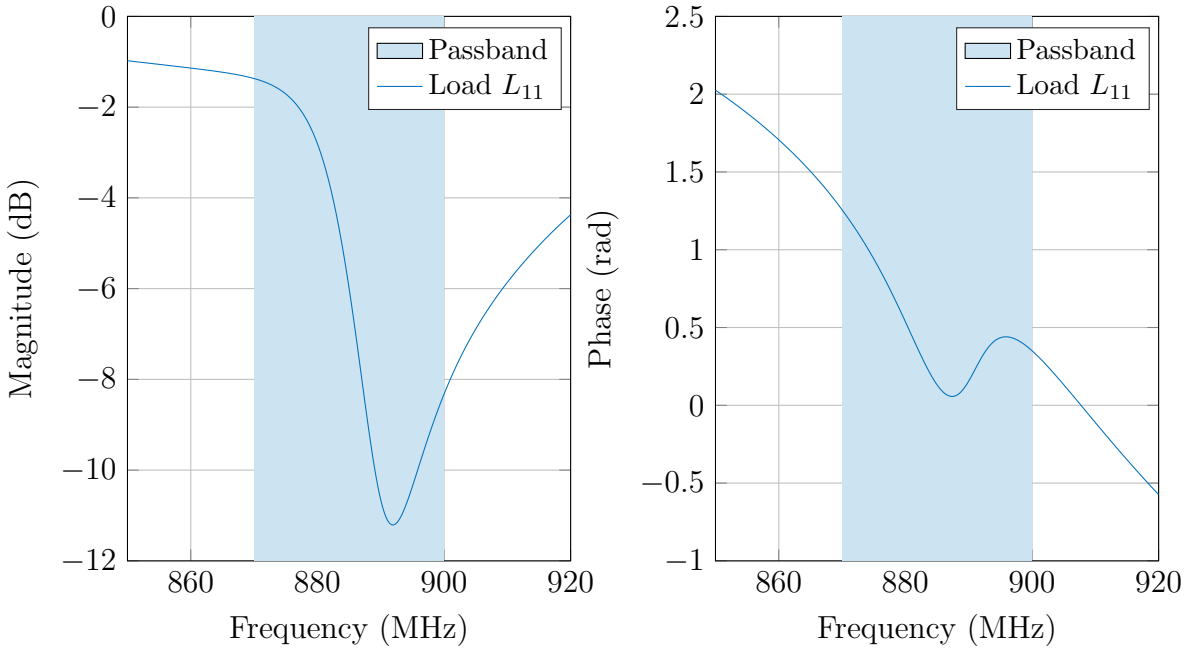


Figure 10.1: Superdirective small antenna

In table 10.1 we list the lower bound ψ_{opt} obtained for the reflection level of the global system as a function of the McMillan degree K of the matching filter from $K = 1$ to $K = 12$. Additionally we compute, for each degree K in table 10.1, by means of the fixed-point algorithm proposed in section 9.2.2, the function $S_{22}^{best} \in \mathbb{F}_R^N$. The function S_{22}^{best} allows us to obtain a function $F_{22} \in \mathbb{\Sigma}_{R_F}^K$, namely a matching filter of degree K having the polynomial R_F as transmission polynomial. We denote by ψ_{best} the reflection level provided by the function S_{22}^{best} which is attained with a sub-optimal filter of degree K and transmission polynomial R_F.

In fig. 10.2 we show the obtained level ψ_{best} with the sub-optimal matching filter as well as the lower limit ψ_{opt}. We start with the solution to problem 7.1.2. Note the mismatch of the reflection L_{11} around 870 MHz and the significant improvement for any value of K obtaining a matching level between -6.5 and -9 dB. It is also interesting to note the proximity of the obtained level ψ_{best} to the lower limit ψ_{opt}. Indeed note in fig. 10.3 the extremely small optimality gap, which quickly converges towards zero. This fact together with the local optimality of the filter that provides the matching level ψ_{best} certify the obtained result.

10.1.1 Example of matching filter synthesis

Next we study with higher detail the case of degree $K = 5$. Let us start with the solution to problem 7.1.2. We show in fig. 10.4 the reflection coefficient L_{11} compared to the

Degree (K)	ψ_{best} dB	ψ_{opt} dB
1	-6.5028	-7.0389
2	-7.4389	-7.7777
3	-7.9916	-8.1925
4	-8.3218	-8.4512
5	-8.5351	-8.6237
6	-8.6815	-8.7444
7	-8.7859	-8.8320
8	-8.8629	-8.8977
9	-8.9212	-8.9480
10	-8.9663	-8.98745
11	-9.0020	-9.0188
12	-9.0306	-9.0442

Table 10.1: Obtained matching level vs lower bound.

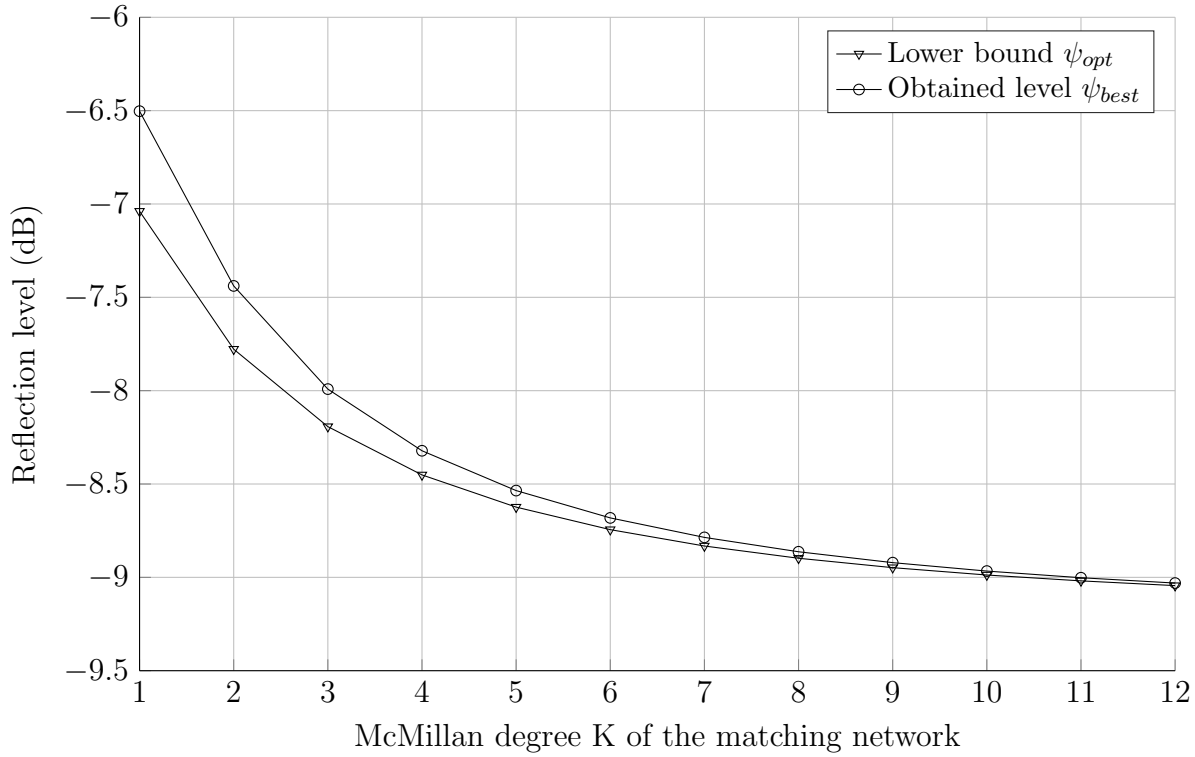


Figure 10.2: Lower bounds and obtained reflection level

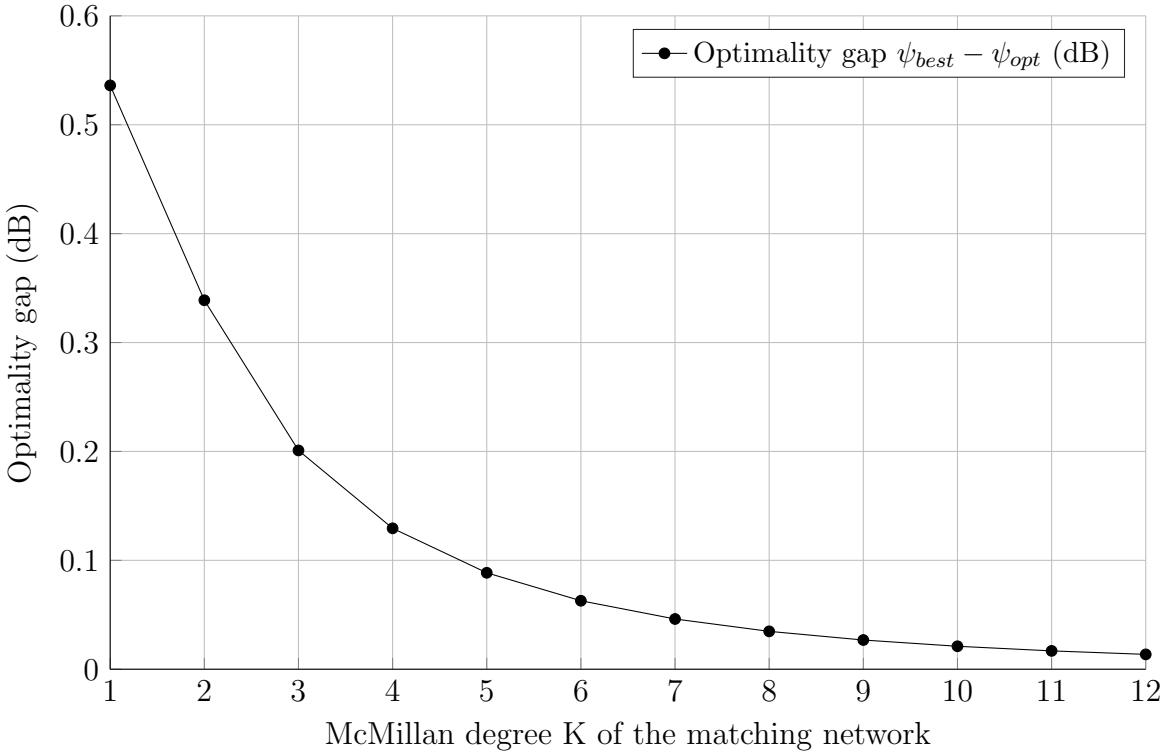


Figure 10.3: Optimality gap.

function S_{22}^{opt} as well as the transducer gain defined as $1 - |S_{22}^{opt}|^2$. The maximum level reached by the reflection parameter S_{22}^{opt} provides the lower hard bound ψ_{opt} attainable in the matching problem considered in this section.

We also compare the poles and zeros of the load reflection L_{22} (in fig. 10.5a) to the poles and zeros of S_{22}^{opt} (in fig. 10.5b). We can also see the transmission zeros α_i with $i \in [1, 2]$ in fig. 10.5a and the transmission zeros of the system in fig. 10.5b. It can be noted that the points α_i are also tranmission zeros of the global system. Nevertheless an additional transmission zero appears in the global system. This transmission is introduced by the Blaschke product and can be spotted in fig. 10.5b as it coincides with a zero of S_{22}^{opt}.

10.1.1.1 Fixed-point algorithm

We carry out now the fixed point algorithm presented in section 9.2.2, which in the case of $K = 2$ provides us the function S_{22}^{best} whose modulus is traced in fig. 10.6.

Remark 10.1.1. *Note that, as commented already, this function reaches a reflection level ψ_{best} that is extremely close to the lower bound ψ_{opt}. It can be remarked from fig. 10.6 that the functions S_{22}^{best} and S_{22}^{opt} achieve nearly the same reflection level within the passband.*

We can also see in fig. 10.7a the poles and zeros of this function S_{22}^{best}. It can be verified that the additional pole introduced by the Blaschke product is being cancelled with a zero of S_{22}^{best}. Thanks to this pole-zero simplification, we are able to obtain, once the load is de-embedded, a function $F_{22} \in \mathbb{\Sigma}_{R_F}^K$ whose poles and zeros are indicated in

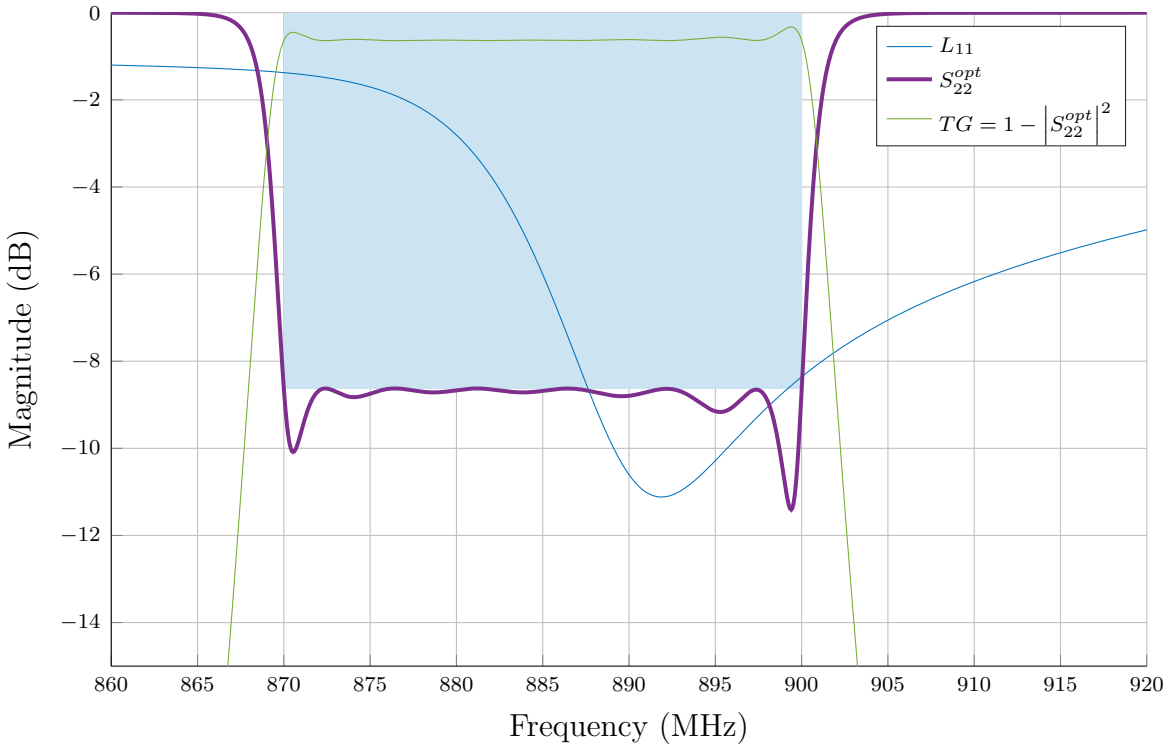
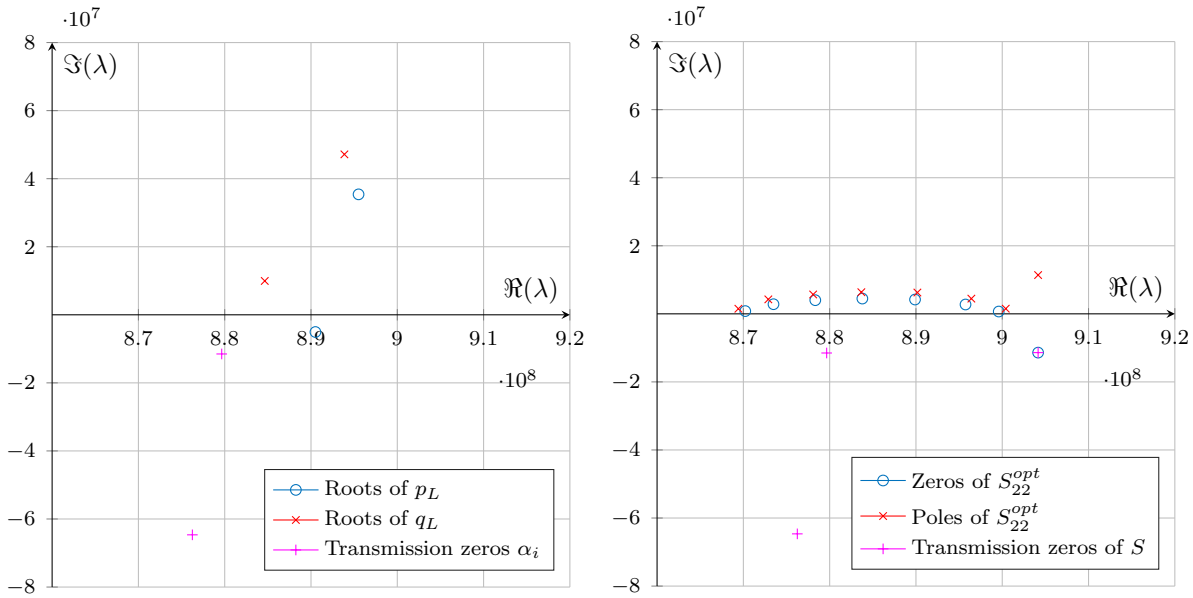


Figure 10.4: Response S_{22}^{opt} and transducer gain obtained by solving problem 7.1.2



(a) Poles/zeros of L_{11} and T.Z. α_i.

(b) Poles/zeros of S_{22}^{opt} and T.Z. of S.

Figure 10.5: Poles/zeros of the functions L_{11} and S_{22}^{opt}, and tranmission zeros.

fig. 10.7b.

We have obtained an lower bound $\psi_{opt} = -8.62\ dB$ as well as an upper bound $\psi_{best} = -8.54\ dB$ for the solution to problem 4.1.1, namely the original problem. We can therefore bound the reflection level ψ solution to problem 4.1.1 as

$$\psi_{opt} \leq \psi \leq \psi_{best}$$

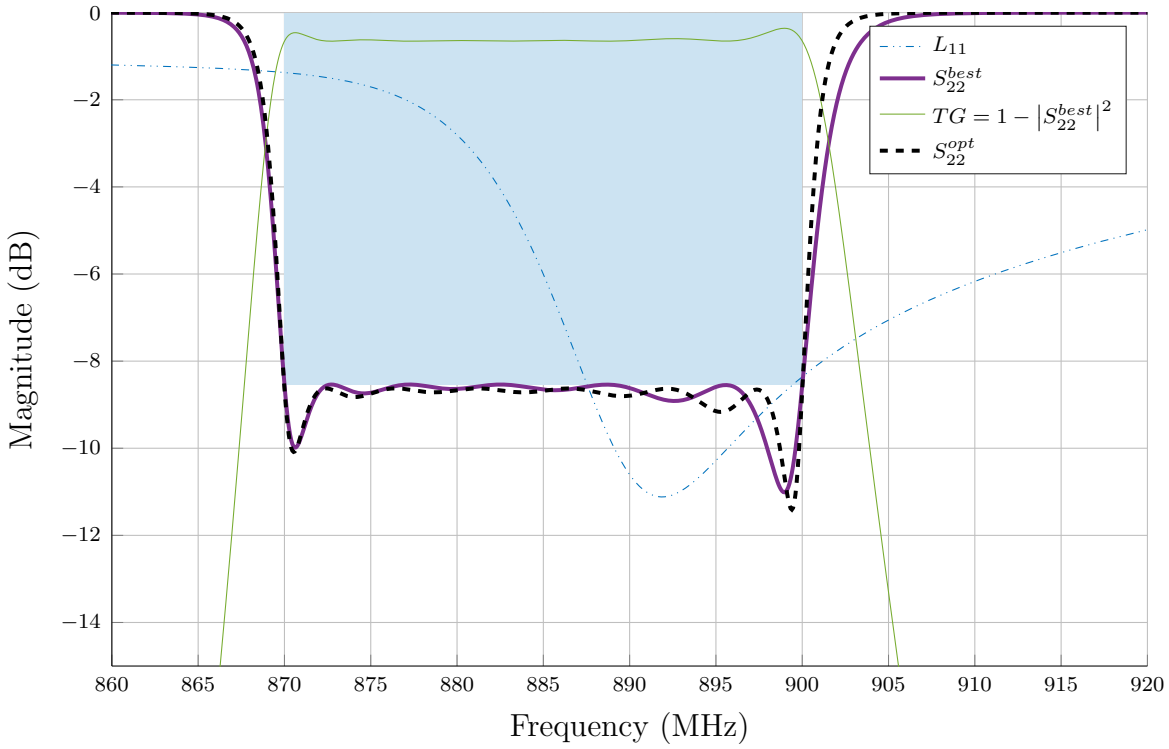


Figure 10.6: Sub-optimal function S_{22}^{best}.

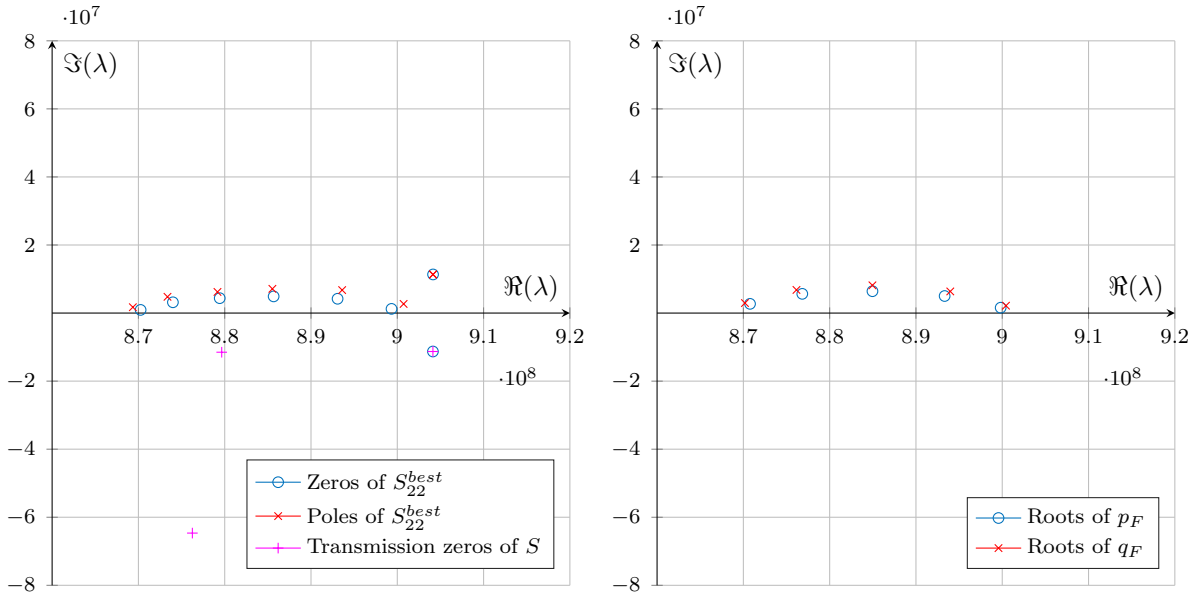
10.1.1.2 Load de-embedding and matching filter optimisation

We are now in disposal of an exceptionally accurate bound for the solution to problem 4.1.1 with the load, passband and other parameters considered in this section. Additionally we have computed an initial point which achieves the upper bound $\psi_{best} = -8.54\ dB$ for the solution to problem 4.1.1. We can then use this starting point to perform a final optimisation over the function F_{22}. The optimisation problem is stated as

$$\text{Find:} \qquad \psi = \min_{F_{22} \in \mathbb{\Sigma}_{R_F}^K} \max_{\omega \in \mathbb{I}} |S_{22}|^2,$$

where $S_{22} = F_{22} \circ L$.

It can be noted that the best possible improvement that can be obtained by means of this final optimisation is less than 0.1 dB as we also have the lower bound $\psi_{opt} = -8.62\ dB$. Nevertheless this optimisation allows us to guarantee the local optimality of



(a) Poles/zeros of S_{22}^{best}

(b) Poles/zeros of F_{22}

Figure 10.7: Poles and zeros of S_{22}^{best}, F_{22} and transmission polynomials.

the obtained solution. We show in fig. 10.8 the modulus of the function S_{22} issue of the local optimisation and compared to the initial point S_{22}^{best} in absolute value. As we can remark in fig. 10.8 both functions coincide. Note that although the system reflection is not improved, this final optimisation provides us with outstanding information about the function S_{22}^{best}. It tell us that the global reflection computed by the fixed-point algorithm from section 9.2.2 is already a local optimum for problem 4.1.1. Finally plot in fig. 10.9 the transmission and reflection parameters F_{21} and F_{22} of the optimum matching filter obtained for the load in fig. 10.1

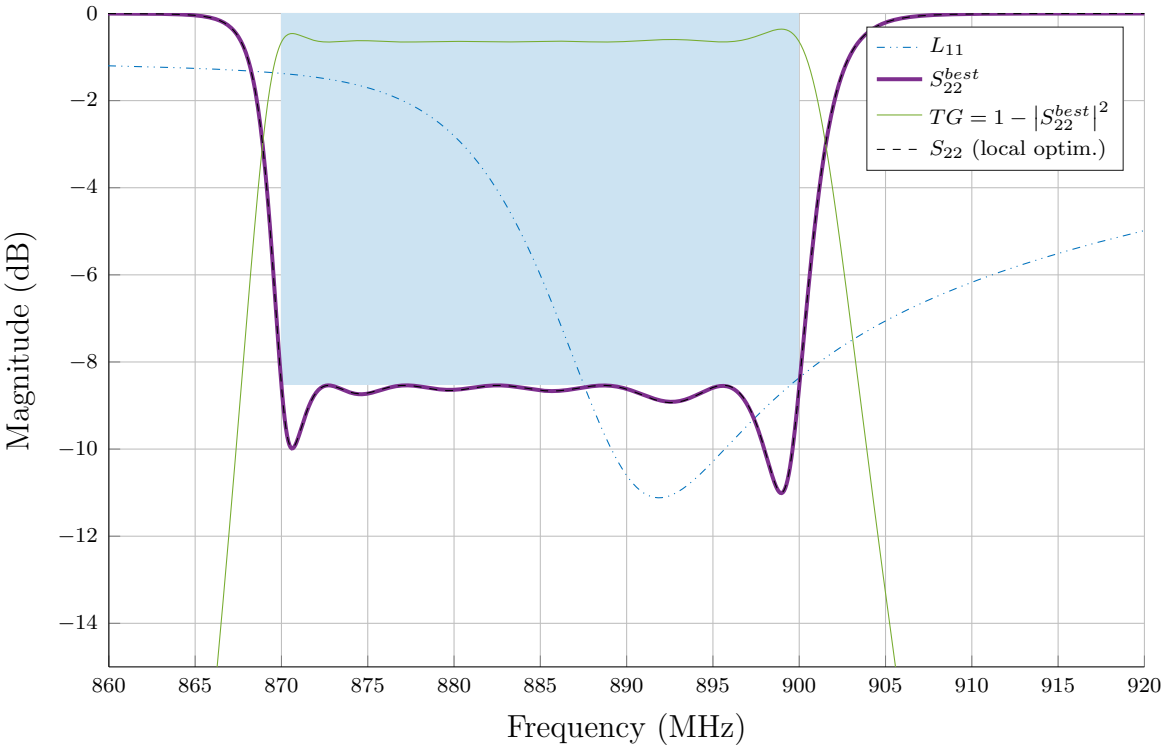


Figure 10.8: Result of the local optimisation and comparison with the starting point S_{22}^{best}.

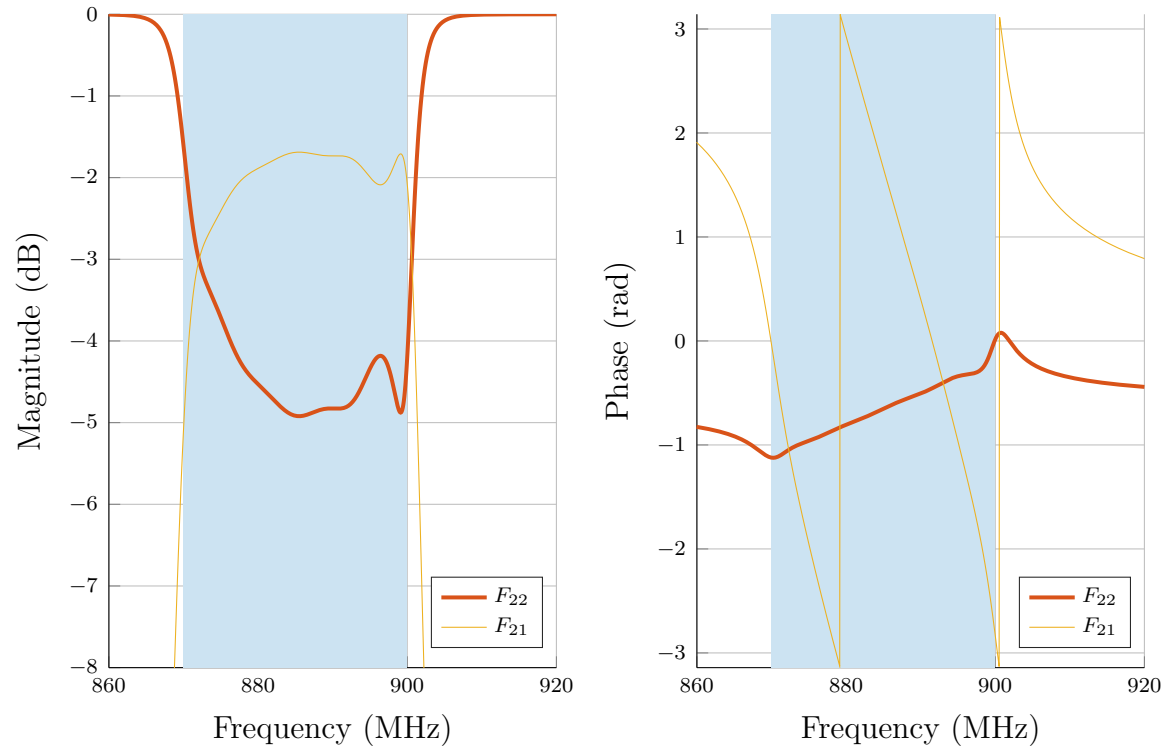


Figure 10.9: Scattering parameter of the optimum matching filter.

10.2 High degree antenna

Here we provide a more complex example of matching filter synthesis. This time we consider a load with the input reflection L_{11} shown in fig. 10.10 where the passband has been normalised to the interval $\mathbb{I} = [-1, 1]$. The reflection coefficient L_{11} in fig. 10.10 is obtained as the $(1, 1)$ element of a Darlington equivalent L with McMillan degree $M = 5$

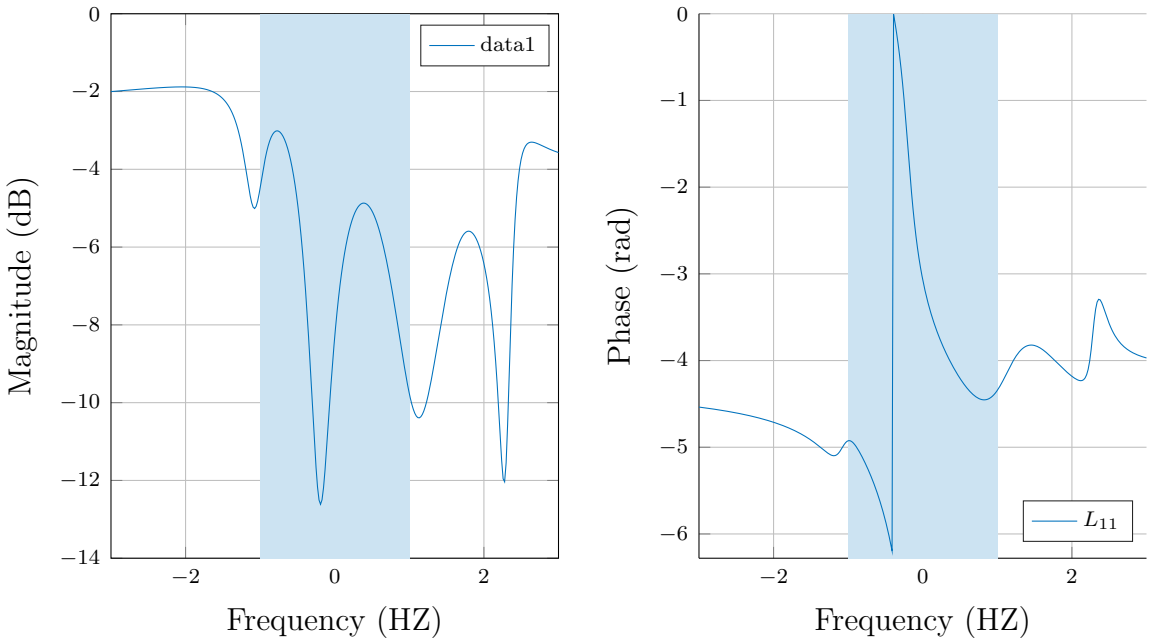


Figure 10.10: Input reflection of a load of degree 5.

$$L = \frac{1}{q_L} \begin{pmatrix} p_L^\star & -r_L^\star \\ r_L & p_L \end{pmatrix},$$

with $r_L, p_L \in \mathbb{P}^5$ and q_L the stable polynomial satisfying $q_L q_L^\star = p_L p_L^\star = r_L r_L^\star$. As before we compute the points $\alpha_i \in \mathbb{C}^-$ with $i \in [1, M]$ as the roots of the tranmission polynomial $R_L = r_L r_L^\star$ in the lower half plane.

10.2.1 Global system optimisation

In contrast to the previous example, this time we consider a fairly complicated load, as the one whose reflection is shown in fig. 10.10 and perform the synthesis of a matching filter of low degree. In particular we chose a filter of McMillan degree $K = 1$. Therefore the degree of the global system remains $N = K + M = 6$. Additionally we fix the tranmission polynomial for the matching filter $R_F = 1$. We can then state the matching problem over the interval $\mathbb{I}$.

Next we solve problem 7.1.2 to obtain an optimal polynomial $P_{opt} \in \mathbb{P}_+^{2N}$. This polynomial allows us to compute the minimum phase function $u_{P_{opt}}$ which provides us with a lower bound ψ_{opt} for the matching level along with a blaschke product b_{opt} such

that the function $S_{22}^{opt} = u_{P_{opt}} \cdot b_{opt}$ is feasible. Note that we obtain a Blaschke product of degree $\mathcal{A} = M - 1 = 4$. The function S_{22}^{opt} is plotted in absolute value in fig. 10.11 where it is compared to the input reflection of the load L_{11} without matching filter. Additionally we also show the transducer gain of the global system which is computed as $TG = 1 - |S_{22}^{opt}|^2$. It can be noticed that the load reflection has been considerably improved at the centre of the passband where the function L_{11} is about $-3dB$ in absolute value. Indeed we obtain in this case a reflection level of -7.43 dB which corresponds to the lower bound ψ_{opt}.

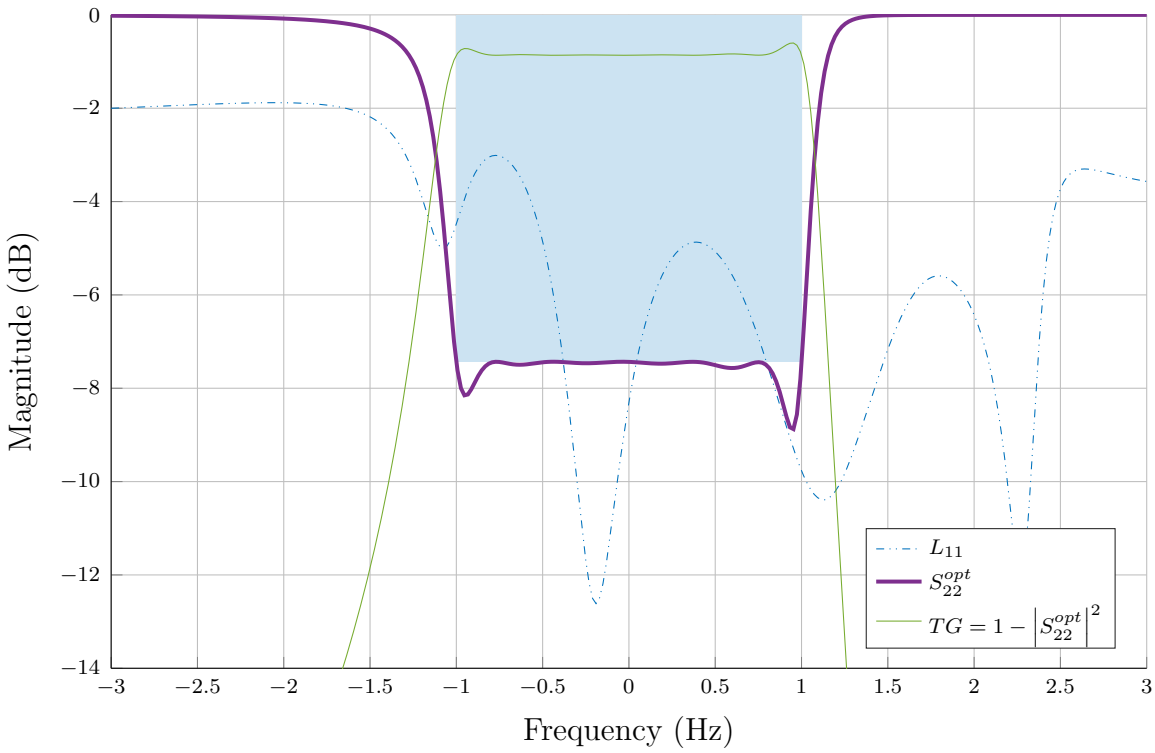
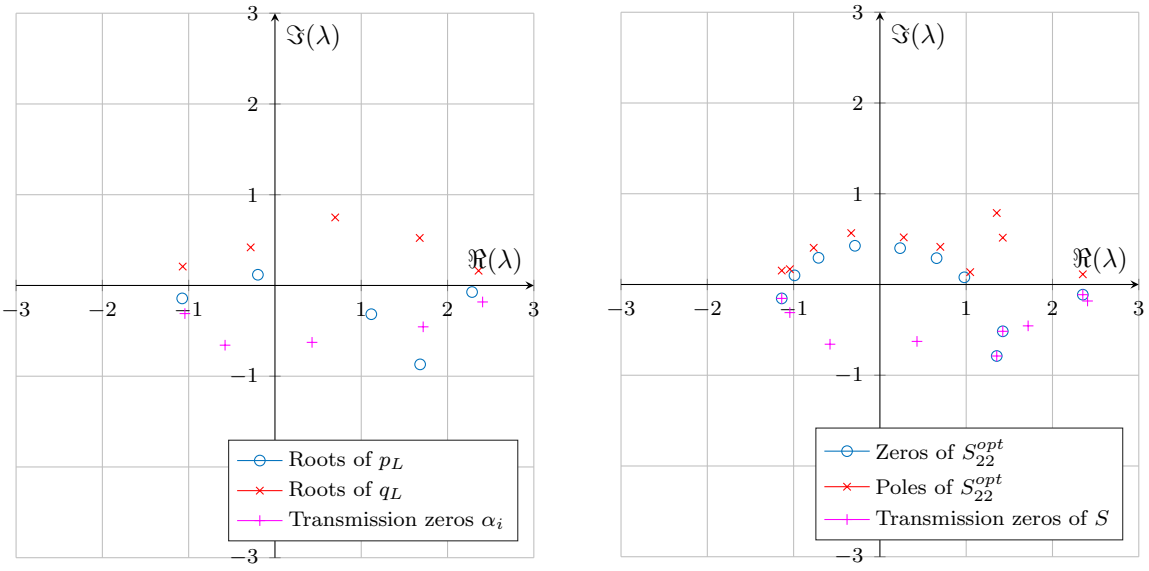


Figure 10.11: Function S_{22}^{opt} and transducer gain $TG = 1 - |S_{22}^{opt}|^2$ obtained by solving problem 7.1.2.

To further compare the obtained result to the load L, we show in fig. 10.12a the complex roots of the polynomials p_L and q_L, namely the poles and zeros of the function L_{22}. Additionally we indicate in fig. 10.12a the location of the 5 tranmission zeros $\alpha_i \in \mathbb{C}^-$ of the load. Similarly we show in fig. 10.12b the poles and zeros of the function S_{22}^{opt} along with the tranmission zeros of the system S. We first observe the increase in degree of the function S_{22}^{opt}, which is of degree $N + \mathcal{A} = 10$, with respect to the desired degree $N = 6$. We can further verify that the points α_i are also tranmission zeros of the global system, along with $\mathcal{A}$ additional transmission zeros corresponding to the Blaschke product b_{opt}. These tranmission zeros introduced by the Blaschke product can be spotted in fig. 10.12b as the tranmission zeros that coincide with a zero of the function S_{22}^{opt}.

Remark 10.2.1. *Similarly to the previous example, it should be noted that since we have $S_{22}^{opt} \in \mathbb{F}^{N+\mathcal{A}}$, it is possible to extract now the Darlington equivalent of the load L from the function S_{22}^{opt} to recover a function $F_{22} \in \mathbb{\Sigma}^{K+\mathcal{A}}$. Nevertheless this function F_{22} would*

not belong to the space $\mathbb{\Sigma}_{R_F}^K$*, namely the set of functions where the filter reflection* F_{22} *is sought for.*



(a) Poles/zeros of L_{22} and T.Z. of L.

(b) Poles/zeros of S_{22}^{opt} and T.Z. of S.

Figure 10.12: Poles and zeros of the function L_{22} and S_{22}^{opt} and transmission zeros.

10.2.2 Fixed-point algorithm and de-embedding of the load

In this section we show the results of the fixed-point algorithm proposed in section 9.2.2. We plot in fig. 10.15 the absolute value of the function S_{22}^{best} compared to the function S_{22}^{opt} obtained above. We have

$$\psi_{best} = \max_{\omega \in \mathbb{I}} |S_{22}^{opt}|^2 = -6.78 \; dB$$

Note that in this case, we obtain a relative larger optimality gap as the difference between the reflection levels provided by the functions S_{22}^{opt} and S_{22}^{best} is of 2.2 dB. This is the price to pay in order to have a function $S_{22}^{best} \in \mathbb{F}_R^N$ instead of the function S_{22}^{opt} which is of degree $N + \mathit{\Lambda} = 10$. Furthermore, as we have also done in the previous example we have computed the optimal bound ψ_{opt} and the reflection level ψ_{best} as a function of the filter degree K, which are plotted in fig. 10.13. It can be noted that the value of the optimal bound ψ_{opt} is almost constant with respect to the degree K even from $K = 1$. However this is not unexpected as the function S_{22}^{opt} corresponding to $K = 1$ is of degree $K + M = 6$ and shows already a rather flat absolute value within the passband as shown in fig. 10.11. Additionally we illustrate in fig. 10.14 the reflection level ψ_{best} achieved by the function S_{22}^{best} issue of the fixed point algorithm. This function exibit a faster variation with respect to the degree K, with a considerably large optimality gap (see fig. 10.14) for $1 \leq K \leq 4$ and quickly converging towards ψ_{opt} for $K \geq 5$. We provide in table 10.2 a list of the values plotted in fig. 10.13 and fig. 10.14.

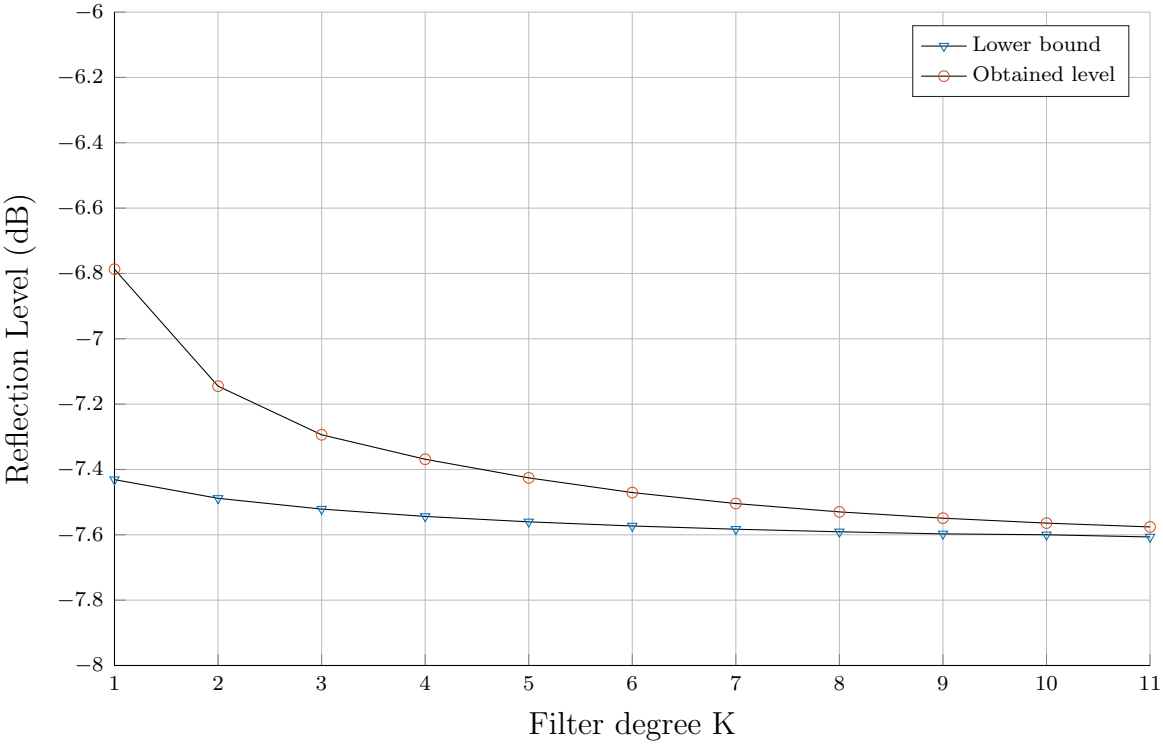


Figure 10.13: Lower bounds and obtained reflection level

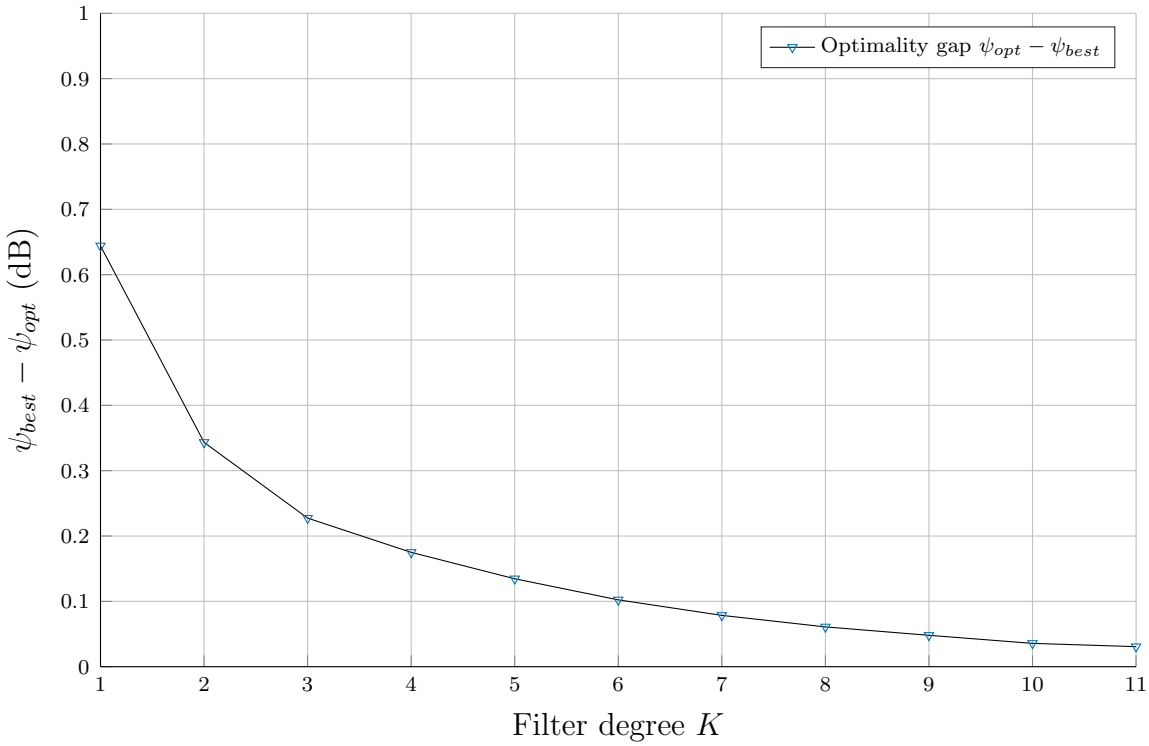


Figure 10.14: Optimality gap.

Degree K	ψ_{opt} (dB)	ψ_{best} (dB)	Opt. Gap (dB)
1	-7.431	-6.787	0.644
2	-7.488	-7.144	0.343
3	-7.521	-7.293	0.227
4	-7.543	-7.368	0.174
5	-7.560	-7.425	0.134
6	-7.572	-7.470	0.102
7	-7.582	-7.504	0.078
8	-7.590	-7.529	0.060
9	-7.597	-7.549	0.048
10	-7.599	-7.564	0.035
11	-7.606	-7.575	0.030

Table 10.2: Lower reflection bound and achieved value.

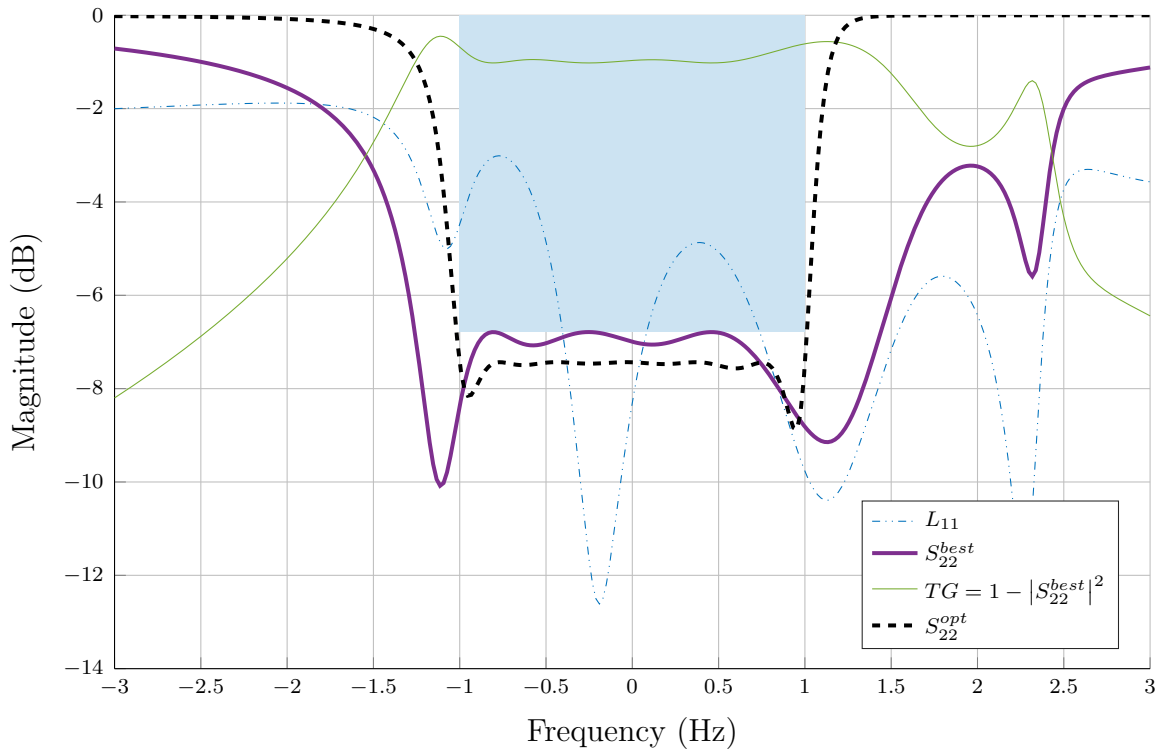
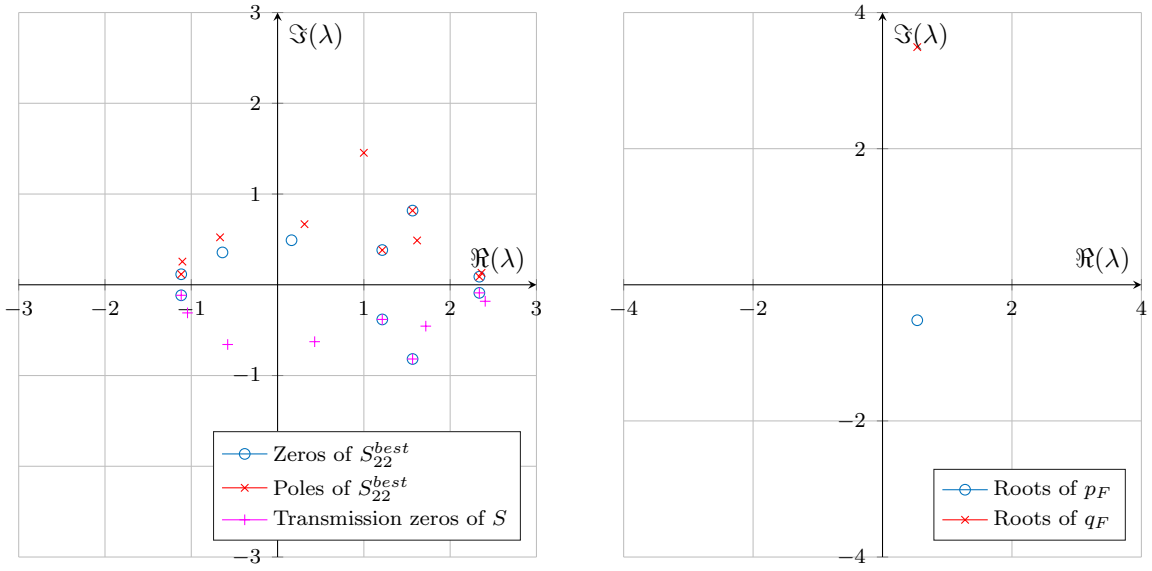


Figure 10.15: Function S_{22}^{best} obtained by the fixed-point algorithm and comparison with the function S_{22}^{opt}.

Figure 10.16a shows the position in the complex plane of the zeros and poles of the function S_{22}^{best}. We indicate as well the tranmission zeros of the system, namely the points α_i with $1 \leq i \leq M$ along with the $\mathcal{N}$ additional tranmission zeros introduced by the Blaschke product. Note in fig. 10.16a that all poles of the Blaschke product coincide now with a zero of the function S_{22}^{best} producing in total 4 simplifications. These simplifications allow us to de-embed the Darlington equivalent of the load from the function S_{22}^{best}, obtaining a function $F_{22} \in \mathbb{\Sigma}_{R_F}^K$. We show in fig. 10.16b the single pole and zero of the filter reflection F_{22}, which is indeed of degree $K = 1$.



(a) Poles/zeros of S_{22}^{best} and T.Z. of S.

(b) Poles/zeros of F_{22}.

Figure 10.16: Poles and zeros of the global system and the matching filter after the fixed-point algorithm.

Finally we perform the local optimisation of the matching filter to compute

$$\min_{F_{22} \in \mathbb{\Sigma}_{R_F}^K} \max_{\omega \in \mathbb{I}} |F_{22} \circ L|.$$

We trace in fig. 10.17 the function $|S_{22}| = |F_{22} \circ L|$ obtained from the previous optimisation along with the initial point given by S_{22}^{best}.

Remark 10.2.2. *It can be noted again how the result from the local optimisation has barely moved away from the initial point. Indeed in all examples provided in this thesis, the result of the fixed-point algorithm coincides with the optimum point computed by means of the local optimisation made a posteriori. This result opens the path for a prospective work that could link the result computed by the fixed-point algorithm discuses in section 9.2.2 and the optimal points of problem 4.1.1.*

To conclude this section, we provide in fig. 10.18 the scattering parameters of the matching filter, which are computed as the 2-port extension of the function F_{22}. These parameters correspond to a device of McMillan degree $K = 1$ which can be easily implemented with any simple resonant structure.

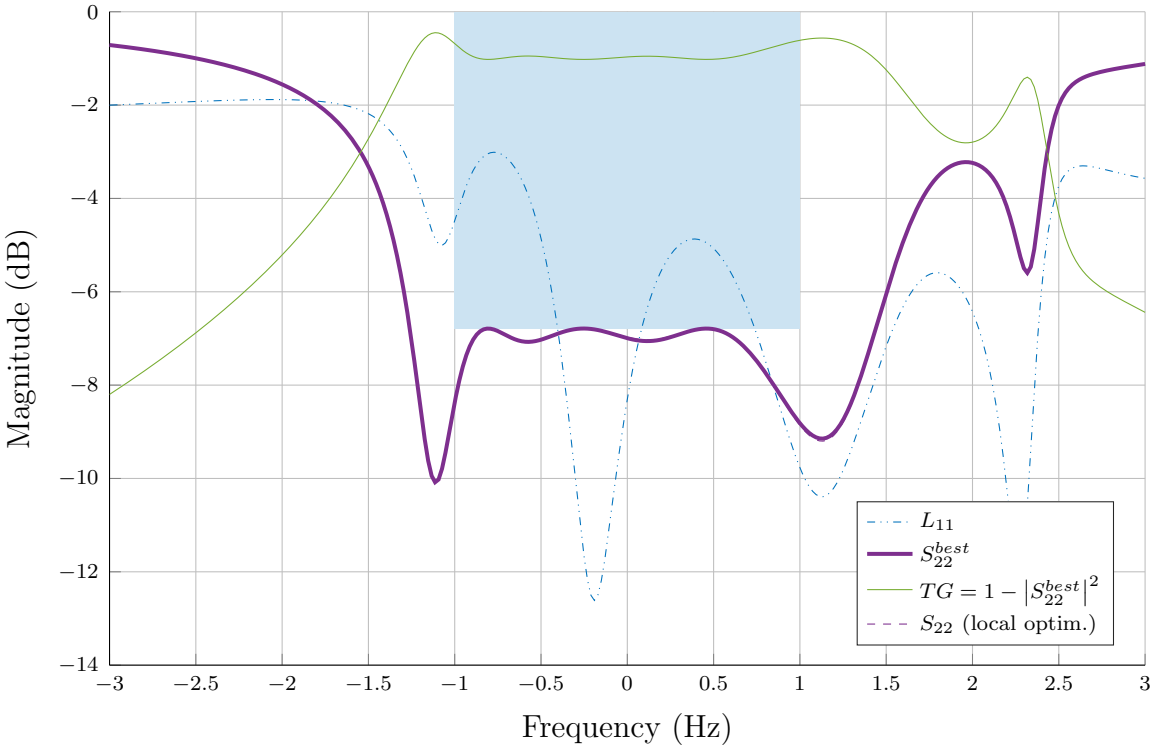


Figure 10.17: Result of the local optimisation and comparison with the starting point S_{22}^{best}.

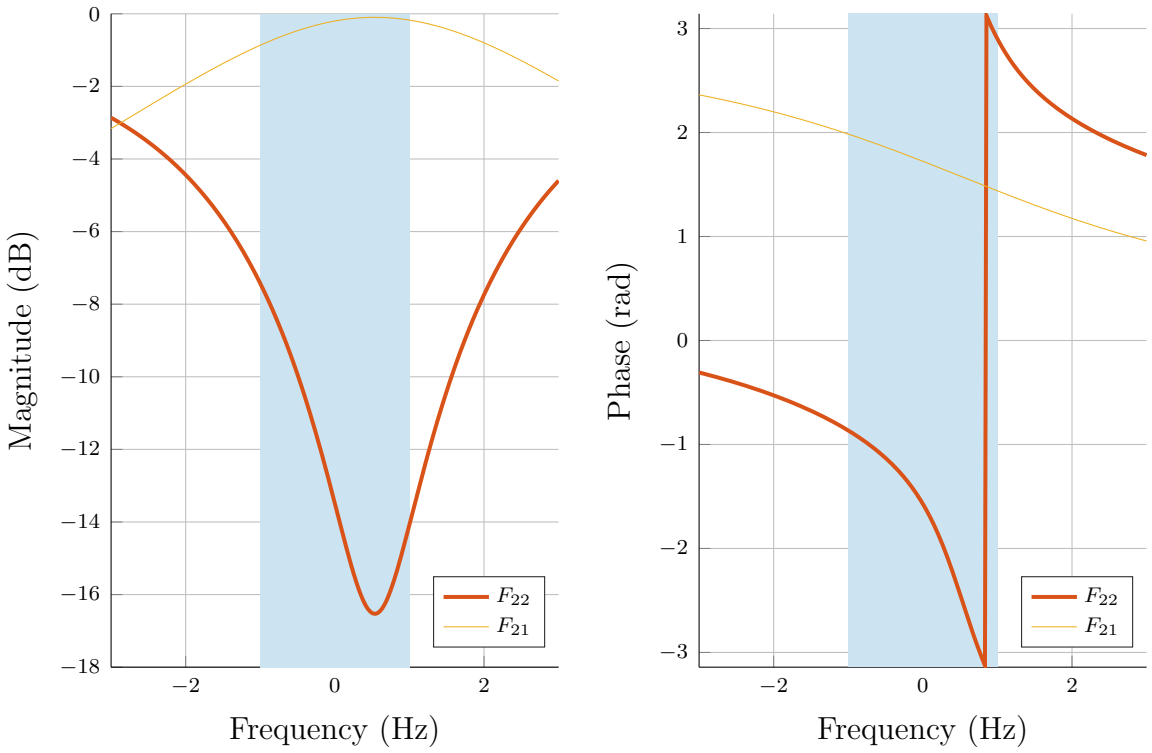


Figure 10.18: Scattering parameters of the matching filter of degree $K = 1$.

10.3 Yagi antenna

Next, we return to the reflection of the yagi-type antenna shown in fig. 3.4 and used in section 3.4.2 to exemplify the type of result obtained with the point-wise matching algorithm introduced in [48]. This reflection is shown again in fig. 10.19 where the passband interval $\mathbb{I}$ is indicated, namely from 2.2GHz to 2.5GHz. This load has a Darlington equivalent of minimal McMillan degree $M = 3$. At this point, it is interesting to compare the results obtained with the algorithm cited in section 3.4.2 and by the theory presented in this thesis. Furthermore, to properly compare with the example provided in section 3.4.2, two transmission zeros are imposed at the frequencies $\nu_1 = 2.17$ GHz and $\nu_2 = 2.53$ GHz. Therefore we have $R = R_F R_L = (\lambda - 2.53)^2(\lambda - 2.17)^2 R_L$.

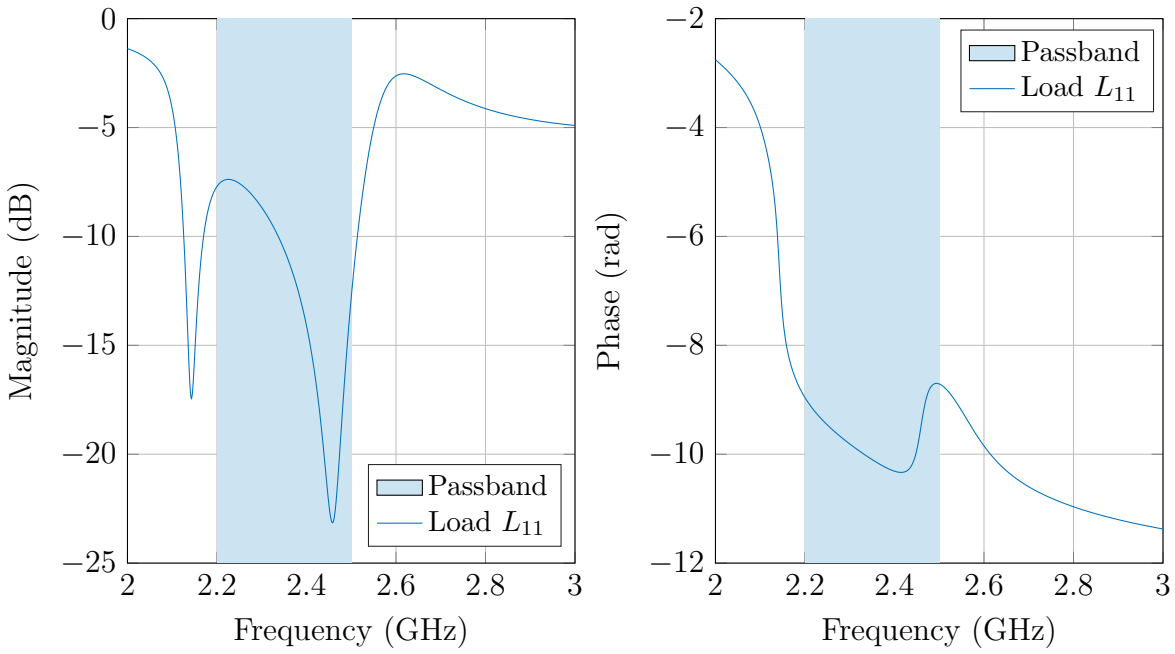


Figure 10.19: Reflection of yagi antenna

Additionally, we impose a third transmission zero at infinity, this means that the degree of the polynomial r_F is strictly lower than the degree of the polynomial p_F. Therefore we have $deg(p_F) > 2$. Assuming that p_F and r_F have no common roots, we conclude that a McMillan degree K greater or equal to 3 is required. We perform now a comparison, for each McMillan degree $K \geq 3$, between the reflection level obtained by the point-wise matching algorithm and the level ψ_{best} provided by the sub-optimal filter computed as explained in previous section. At the same time, both results are compared to the lower bound ψ_{opt} for the matching level with a filter of McMillan degree K which is computed by means of the convex relaxation of the problem. This comparison can be seen in fig. 10.20. The first fact that should be highlighted in fig. 10.20 is the difference between the level of matching provided by the point-wise matching algorithm $\psi_{pointwise}$(green square markers) and the level obtained by the uniform matching procedure ψ_{best} (red circular markers) developed in this thesis. This result is not surprising since it is known that the type of global responses provided by the point-wise matching method, namely with perfect

matching points on the frequency axis, are generally not optimal for the uniform matching problem. Indeed as can be seen in fig. 10.20, the level of matching obtained by the point-wise matching method presented in [48] is clearly not optimal in terms of matching. Furthermore, if we compare both results with the lower limit ψ_{opt} for each value of K (blue triangular markers), we can verify that the result obtained by the uniform matching algorithm tends towards the lower limit ψ_{opt} as K increases. This does not happen with the level of matching obtained through point-wise matching which, although it also tends to a certain limit as K increases, this limit does not coincide, generally with the lower bound for degree K. We have in general

$$\lim_{K\to\infty} \psi_{best}(K) = \psi_{opt}(K).$$

With respect to the result obtained with low values of K, we can see that the obtained level ψ_{best} is further from the limit ψ_{opt} when K decreases. Indeed the duality gap, defined here as the difference between the ψ_{best} level in dB and the ψ_{opt} limit in dB, is maximum for small values of K and tends to zero when K increases as can be checked in fig. 10.21. In fig. 10.21 we can see that the duality gap is considerably small, with a difference between ψ_{best} and ψ_{opt} of around 2dB for $K = 3$ and less than $1dB$ from $K = 4$. This information, namely the fact that the optimality gap is small, gives us a certification of the optimality of the sub-optimal matching filter which reaches a level of matching ψ_{best}.

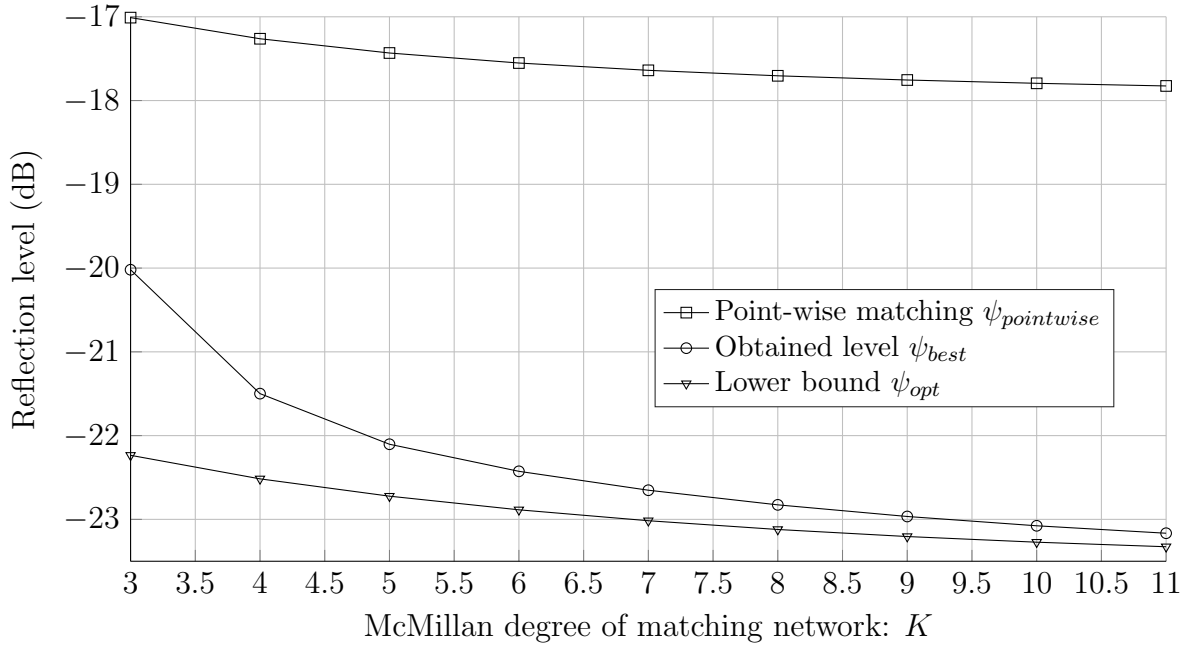


Figure 10.20: Lower bounds and obtained reflection level

Finally, we have decided to show two examples of the type of result obtained with our uniform matching algorithm with two particular values of K.

Firstly we choose $K = 5$, namely the same degree used in section 3.4.2 to be able to perform a comparison with the former. We see in fig. 10.22a a comparison between the result obtained with point-wise matching and with uniform matching, both in term of

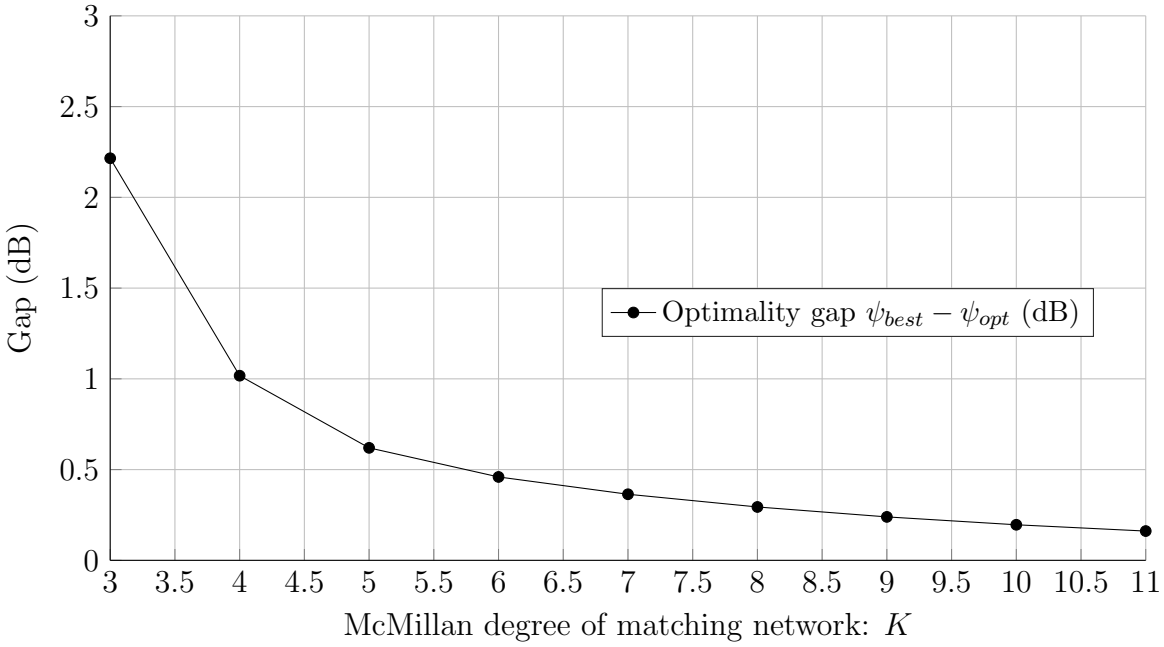


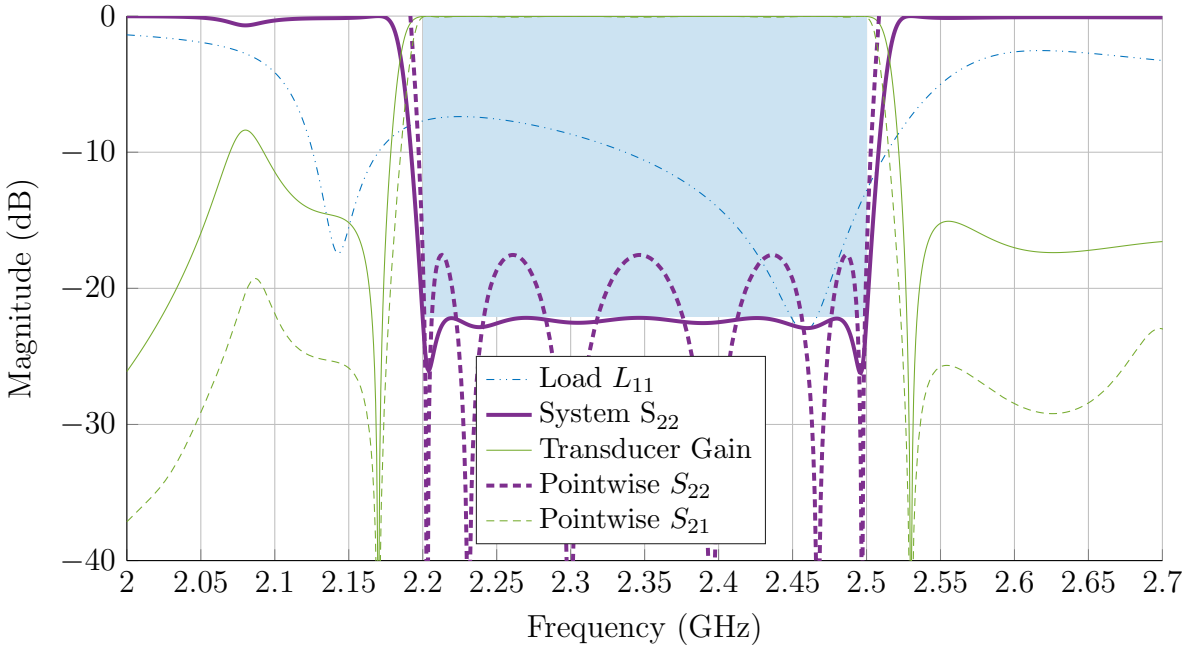
Figure 10.21: Optimality gap.

the global reflection, parameter S_{22} in the figure, and transducer gain S_{21}. Figure 10.22a shows that the matching level of obtained by the uniform matching algorithm is lower than that provided by the response obtained through point-wise matching. It is also necessary to emphasize that the point-wise matching algorithm provides a more selective response, as it can be clearly seen when comparing the transducer gain outside the band.

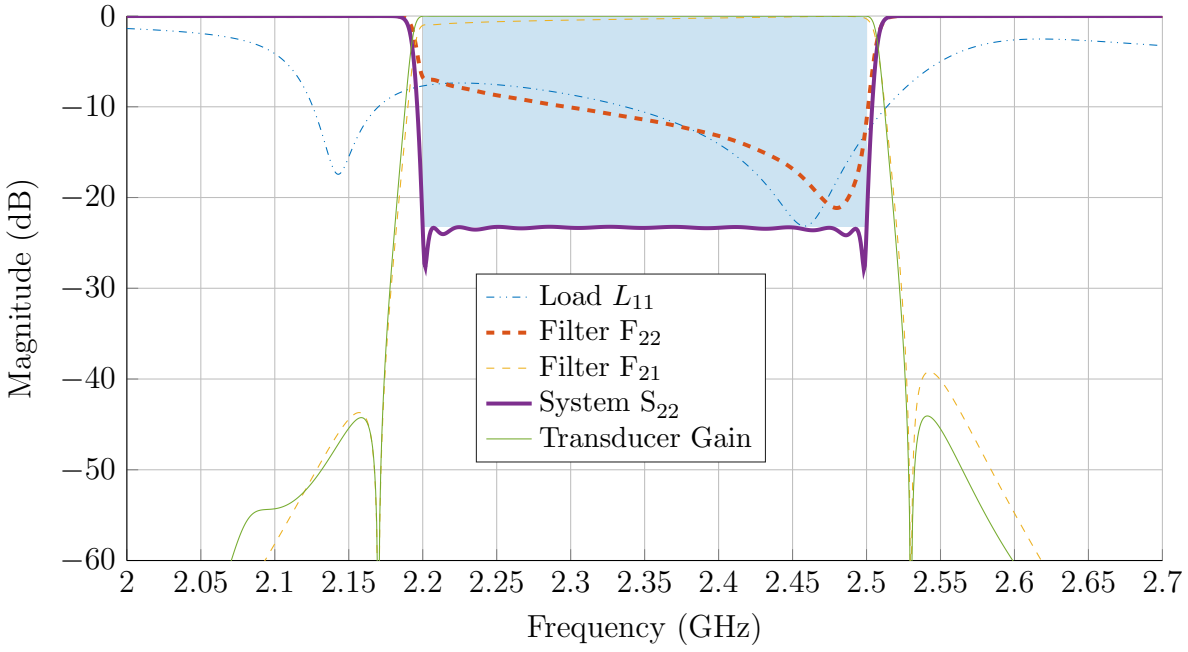
Remark 10.3.1. *Note that although in the problem of uniform matching we can add a restriction in terms of selectivity, in this case no requirement has been imposed in that regard. Therefore we can compare both results when the objective is to achieve the best possible matching within the band. Nevertheless, if we add a certain selectivity level to the requirements in the uniform matching problem, the response obtained would progressively become closer to the response obtained by the point-wise matching method.*

In the second example we show a high degree result, namely $K = 10$ to provide an idea of the type of limit response to which the presented algorithm converges upon increasing the degree. We can see, in fig. 10.22b, on the one hand the sub-optimal filter of McMillan degree $K = 10$ represented by dashed lines and on the other hand the global reflection and transducer gain obtained with this filter.

Remark 10.3.2. *Note that the filter reflection F_{22} in module approaches the function L_{11}. This fact corresponds to the expected result since when the global reflection, whose module is expressed as the pseudo-hyperbolic distance $|S_{22}(\omega)| = \delta(F_{22}(\omega), \overline{L_{11}(\omega)})$, tends to zero, the module $|F_{22}(\omega)|$ approaches the absolute value of the reflection of the load $|L_{11}|$.*



(a) Matching network of McMillan degree $K = 5$



(b) Matching network of McMillan degree $K = 10$

Figure 10.22: Result of matching a yagi antenna

10.4 Dual-band RHCP Antenna

In this section we present a dual-band antenna implemented by microstrip patch which is used to illustrated the second part of the theory provided in this chapter. This antenna, which features two orthogonal excitation ports and four slots, provides a right-hand circular polarised radiated wave when the excitation use for both ports presents a 90 degrees phase shift.

The aforementioned structure is shown in fig. 10.23a. A more detailed description of this antenna can be found in [50]. This antenna is part of a receiver for GNSS which has as objective the reception in the frequency bands of the GPS and GALILEO system indicated in table 10.3. We can see a top view of the manufactured path antenna in section 10.4.

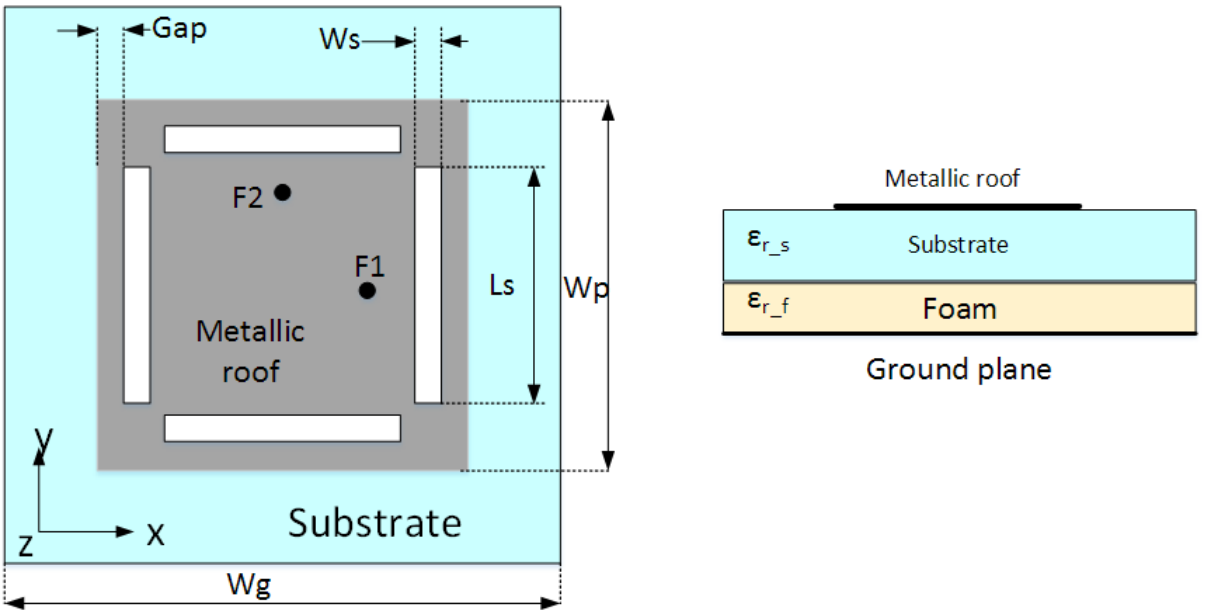
Band	*F. min (GHz)*	*F. max (GHz)*
GPS L2	1.21	1.24
GPS L1	1.55	1.60
GALILEO E6	1.26	1.30

Table 10.3: GNSS bands used for the matching problem

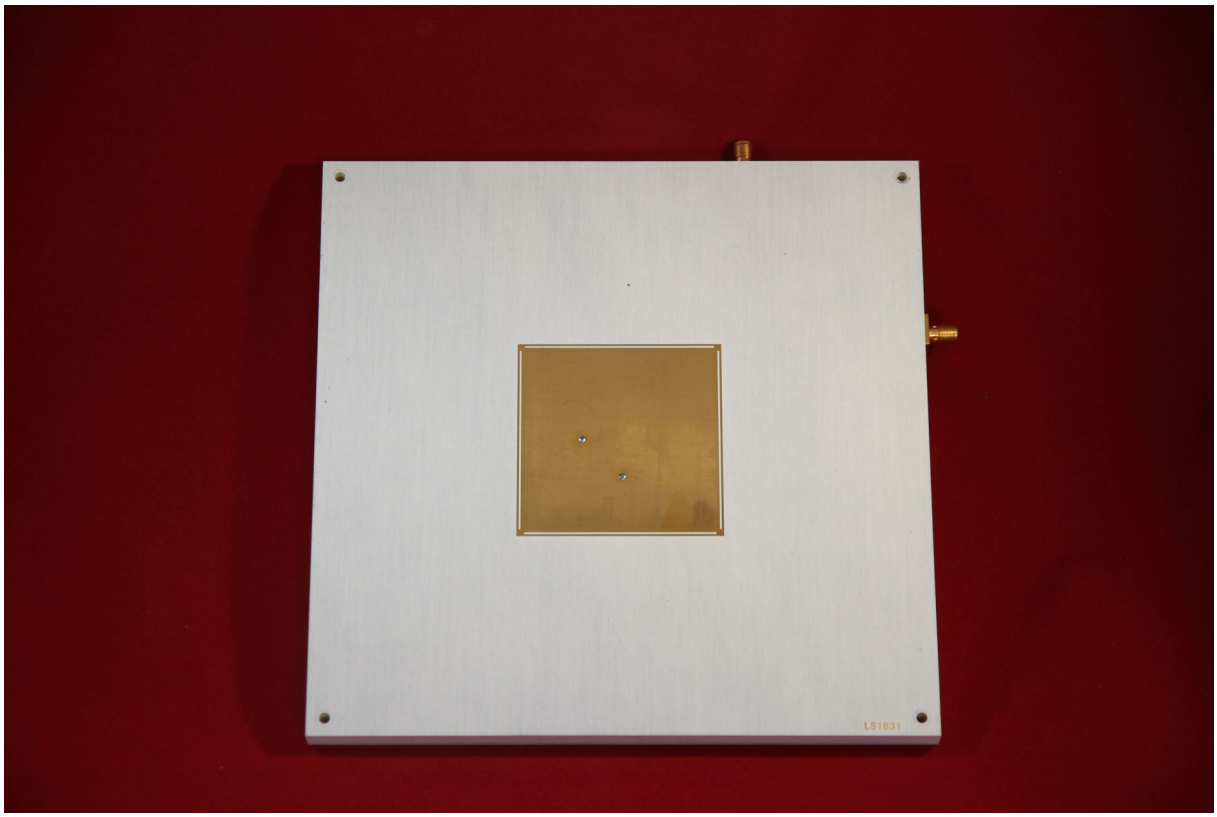
Since in this case the antenna has 2 ports, one for each polarization, we obtain the scattering matrix A of size 2×2 plotted in function of the frequency in fig. 10.24 along with the relevant frequency bands. In this case the elements A_{11} and A_{22} correspond to the reflection in ports 1 and 2 respectively. These are the reflections to be matched in this example by means of the introduced broadband matching algorithm. The transmission parameter A_{21} however, is not related to the transmission of the antenna but represents the coupling between the two input ports.

To excite each port with the right phase, the excitation is performed through a 90 degree hybrid coupler which provides the appropriate signals in phase and quadrature configuration as it is illustrated in fig. 10.25. This coupler is excited by port 1 with the input signal while port 4 is loaded by a matched load $R_0 = 50\Omega$.

To solve the problem of mismatch, a matching filter is used to match each port of the antenna. Note that although the antenna is excited by ports 1 and 2 simultaneously, the coupling between both ports is weak enough, below 20 dB in all passband as it can be seen in fig. 10.24, to consider each one separately. Therefore we are facing two independent matching problems, one with the load of reflection $L_{11}^{(1)} = A_{11}$ and the other with the reflection $L_{11}^{(2)} = A_{22}$. However, both reflections are almost identical due to the symmetry of the antenna, therefore the same matching filter F is used for both ports as indicated in fig. 10.26.



(a) Schematic view



(b) Manufactured prototype

Figure 10.23: Dual-band RHCP antenna integrated a microstrip slotted patch

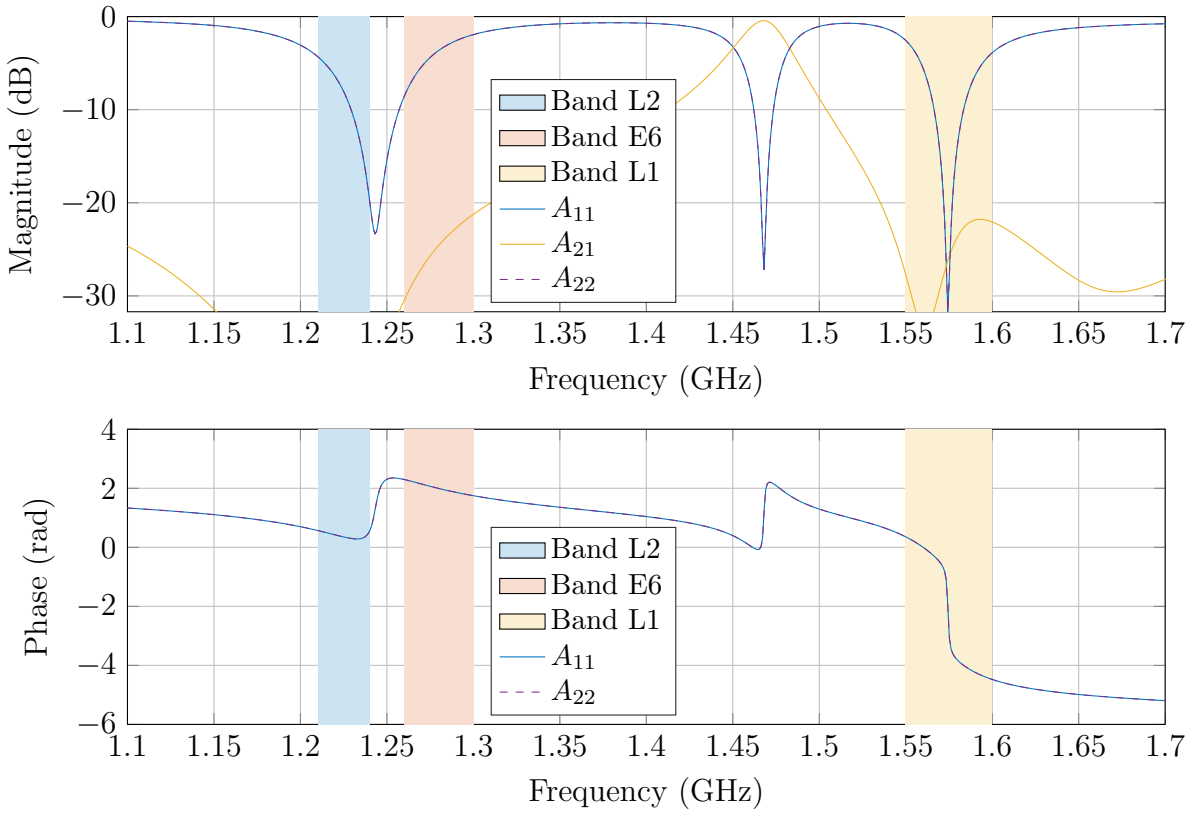


Figure 10.24: Scattering parameters of the dual-band antenna

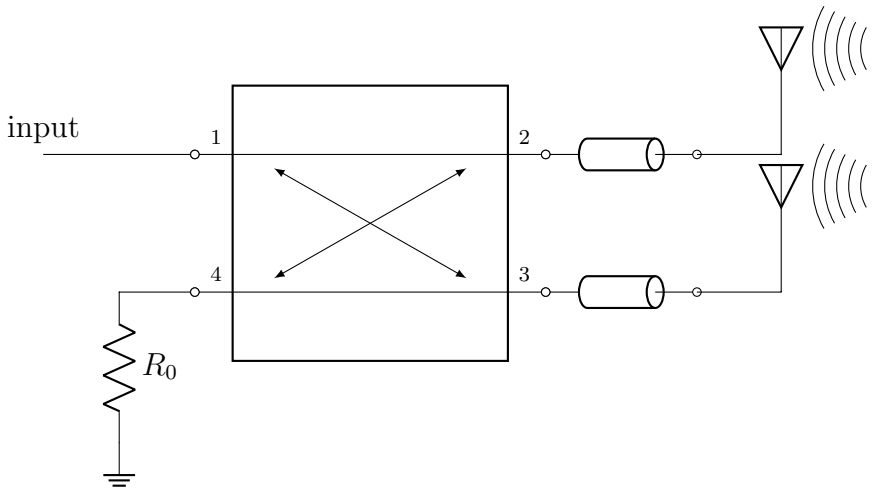


Figure 10.25: Schematic of the RHCP antenna connected to a hybrid which provides the appropriate feeding signals to each input port with 90 phase difference.

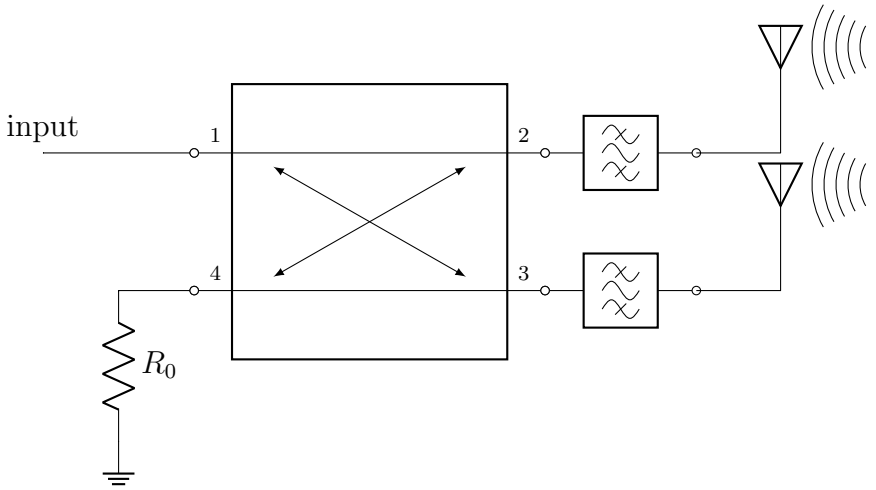


Figure 10.26: Schematic of the RHCP antenna with both matching filters inserted.

10.4.1 Matching bounds

Following the same procedure as in the examples shown previously, below we perform a study of the optimal matching level in the three bands with the antenna introduced in fig. 10.24. For this we set the transmission polynomial of the matching filter $R_F = 1$ obtaining $R = R_F R_L = R_L$ a polynomial of degree 3, and we vary the McMillan degree of the filter K between 1 and 9. Each of the reflection parameters A_{11} and A_{22} can be extended to obtain a two-port device of McMillan degree $M = 3$ by means of the Darlington equivalent. We compute, for each value of K, the optimal solution to problem 7.1.2 considering degree $N = K + M$ for the global system.

By means of the fixed-point algorithm presented in chapter 9, we obtain in each case a sub-optimal matching filter of McMillan degree K which is locally optimal for problem 4.1.1, namely the original matching problem. In fig. 10.27 we show the lower bound obtained as the optimal criterium for problem 7.1.2 along with the level of matching provided by the sub-optimal filter of degree K. It should be noted that, as it was the case in fig. 6.16, the levels shown in fig. 10.27 do not present a smooth decrease in function of K as the derivative of both levels vary strongly as K increases. This is due to the multi-band character and it may occurs, for instance, that the optimal matching filter of McMillan degree K is in fact not of full degree, namely of degree less than K.

Furthermore, in this case we also obtain an extremely high optimality gap for matching filters of degree less than 3 ($K \leq 2$) as it can be seen in fig. 10.28. It is usual for the optimality gap to be maximum for $K = 1$ and then to tend to zero when the value of K increases. However in this case the value of the lower bound obtained for $K \leq 2$ does not provide much information about the optimality of the sub-optimal solution. Nevertheless, in this case the lower bound given by the blue line in fig. 10.28 gives us information of different nature and even more valuable than the optimality of the sub-optimal filter obtained. This additional information refers to the optimal placement of the tranmission zeros, which allows to improve the global reflection level, until reaching the value ψ_{opt} with a matching filter of degree $K + M - 1 = K + 2$.

Remark 10.4.1. *It is important to remember now that the matching value corresponding to the lower limit can be obtained through a filter of degree $K+M-1$, namely $F_{22}^{opt} \in \mathbb{F}^{K+2}$.*

Nevertheless note that this reflection F_{22}^{opt} does not belong to $\mathbb{F}_R^{K+2}$ since two additional transmission zeros are introduced. Therefore fig. 10.27 also serves as a comparison between the matching level obtained with a filter of degree K and no tranmission zeros and a filter of degree $K+2$ with 2 finite transmission zeros.

For a proper comparison, we plot both levels for the same value of K in fig. 10.29. We can verify in fig. 10.29 that the optimal matching level, which is attain using a filter presenting two finite transmission zeros is not reached in any case by means of a filter which does not include those transmission zeros. This does not happen, for instance, in fig. 10.2 where the filter obtained with degree K provides a matching level lower than the hard bound computed for degree $K-1$. Therefore in the previous case, the filter of degree $K+1$ which provides the matching level in blue in fig. 10.2 at the expense of including a finite transmission zero is of no interest as for degree $K+1$ a better matching filter is obtained without the need of such transmission zero.

The result shown in fig. 10.29 not only informs us of the improvement in the level of matching obtained when considering a polynomial $r_F \in \mathbb{P}^2$, without also providing us with information about the position of said transmission zeros (the roots of $r_F r_F^\star$ in the complex plane) so that the indicated level is reached. Taking as an example the case $K=3$, the problem of matching as it has been formulated in this thesis provides us with the information collected in table 10.4.

Degree (K)	F_{22}^{best} (no TZ)	F_{22}^{opt} (2 TZ)	Bound (no TZ)
1	-3.4188	N/A	-8.6438
2	-3.5220	N/A	-8.6470
3	-7.8482	-8.6438	-10.1902
4	-8.1115	-8.6470	-10.4057
5	-9.4450	-10.1902	-11.0517
6	-10.1423	-10.4057	-11.4547
7	-10.7360	-11.0517	-11.6029
8	-11.3288	-11.4547	-12.0888
9	-11.4656	-11.6029	-12.1966

Table 10.4: Matching results (in dB) provided by the presented algorithm

10.4.2 Results

Below we analyse more deeply three of the cases listed in table 10.4. In fig. 10.30b with dashed lines we plot the scattering parameters of the Filter of McMillan 3 which gives us

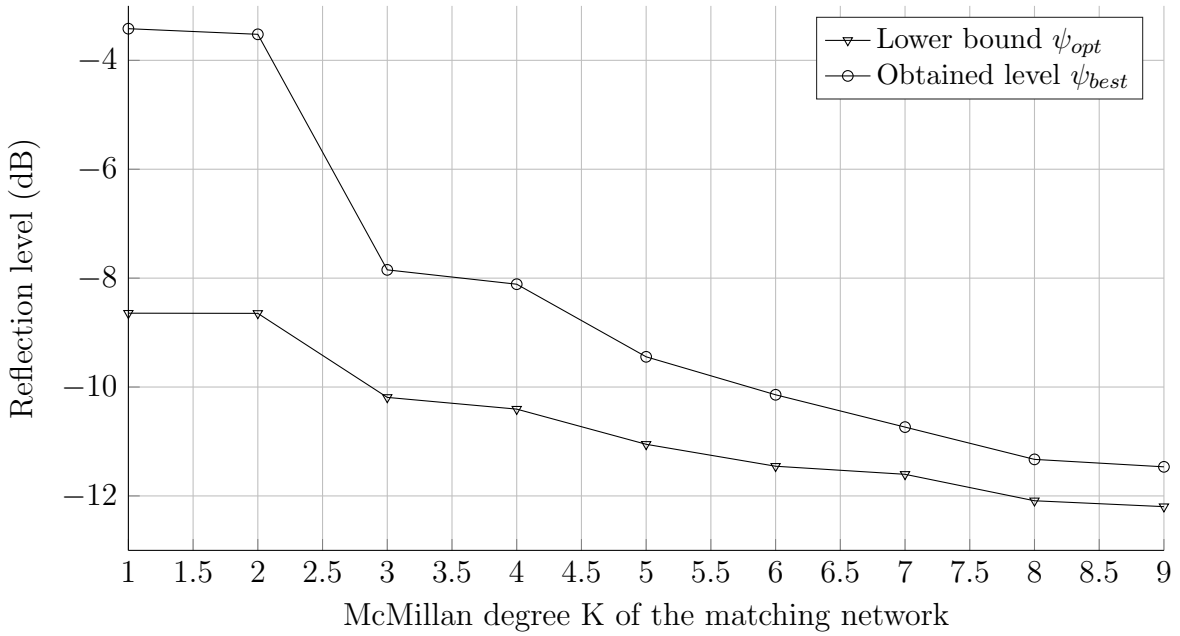


Figure 10.27: Lower bounds and obtained reflection level

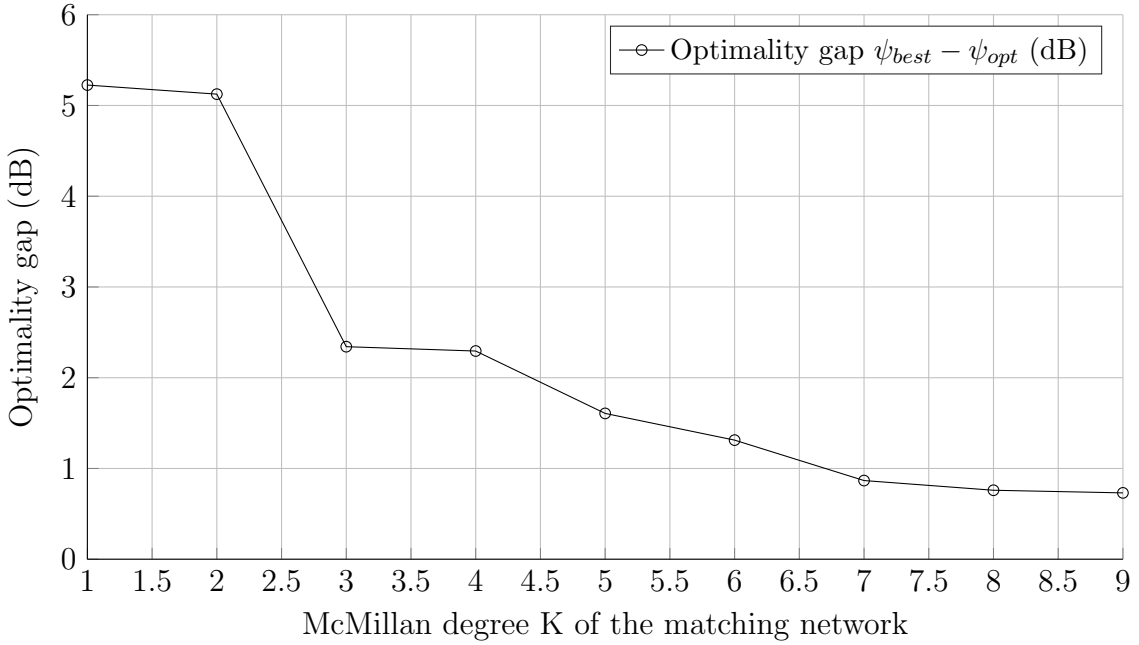


Figure 10.28: Optimality gap

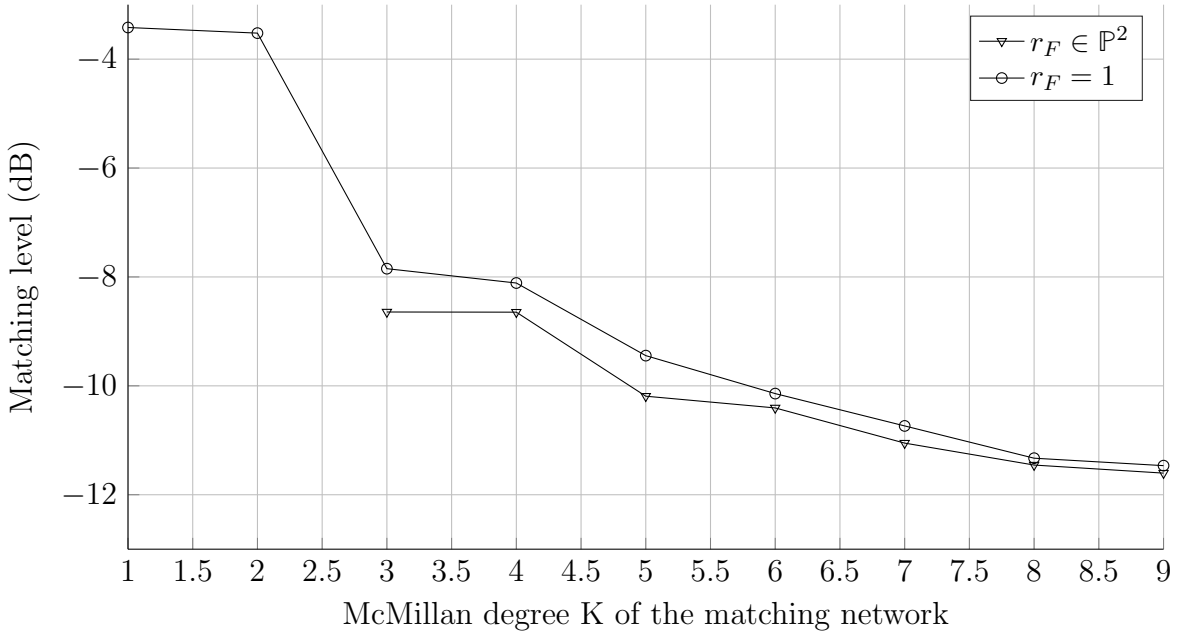


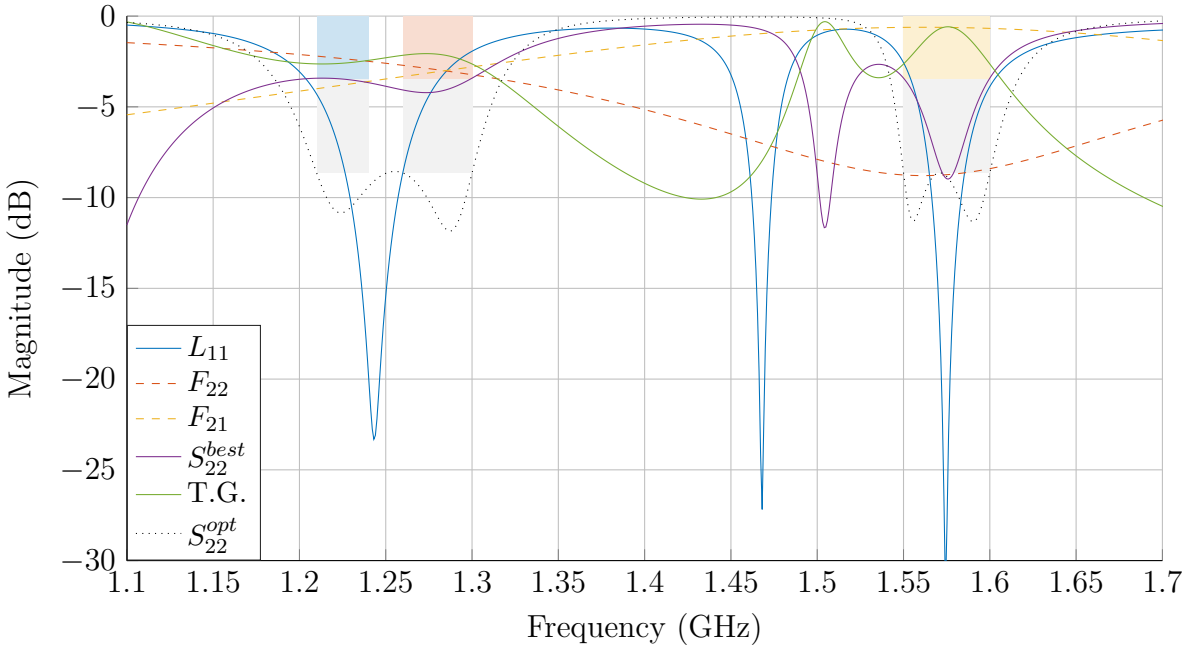
Figure 10.29: Matching level obtained without transmission zeros ($r_F = 1$) and with 2 finite transmission zeros ($r_F \in \mathbb{P}^2$).

the lower bound for degree $K = 1$. This bound, provided by the global system response in fig. 10.30b, is marked by the shaded squares. Similarly, in fig. 10.30a we see the best obtained matching filter of McMillan degree $K = 1$ (with dashed lines) along with the global response obtained after connecting this filter to the antenna (solid lines).

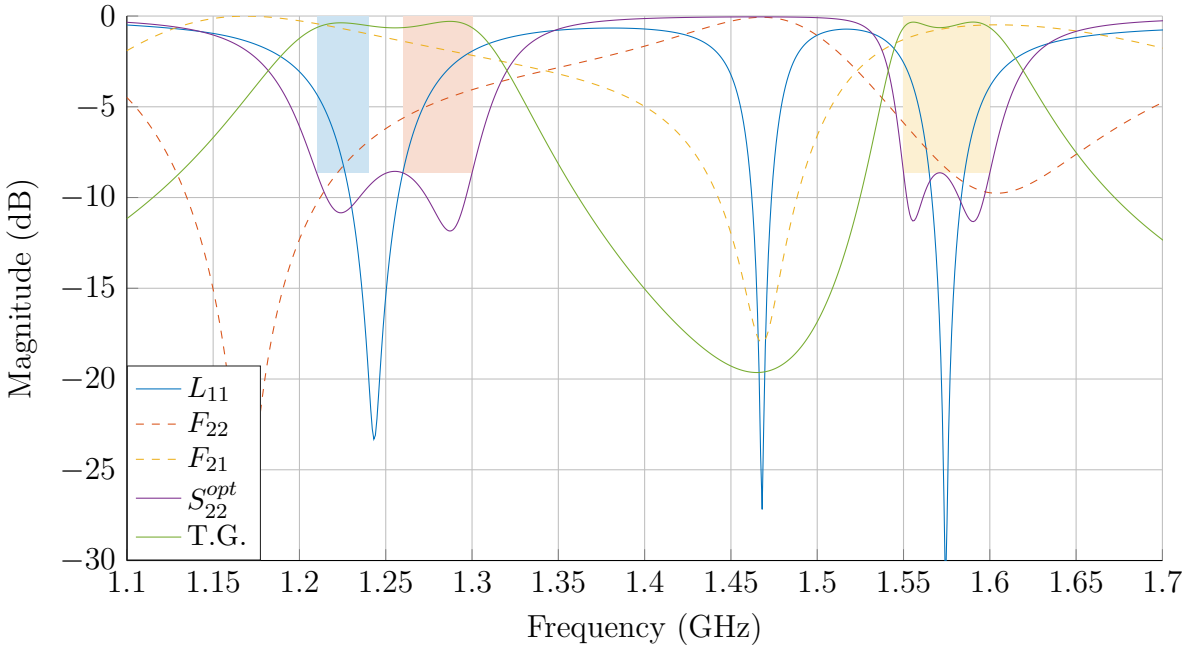
We also include, for comparison, the reflection of the optimal global system extracted from fig. 10.30b. By doing this we can visualize the huge optimality gap obtained in this case as the shaded gray area, namely the difference between the reflection level provided by the system reflection S_{22} (of degree $K = 1$) and the optimal reflection (of degree $K = 3$).

The analogous comparison is performed in fig. 10.31 for degree $K = 3$ and in fig. 10.32 for degree $K = 9$. In both cases (in figs. 10.31b and 10.32b) we can see the matching filter of degree $K + 2$ (5 and 11 respectively) together with the global system response which attain the lower bound for the matching level using a filter of degree K. Similarly in figs. 10.31a and 10.32a we show the sub-optimal filters of McMillan 3 and 9 respectively and the scattering parameters obtained in each case.

Additionally, as in the previous example, the reflection of the global system in fig. 10.31b has been plotted again with dotted lines in figs. 10.31a and 10.32a to compare with the reflection S_{22} of degree K. It can be note how the optimality gap, namely the gray square in figs. 10.31a and 10.32a decreases as the McMillan degree K increases.

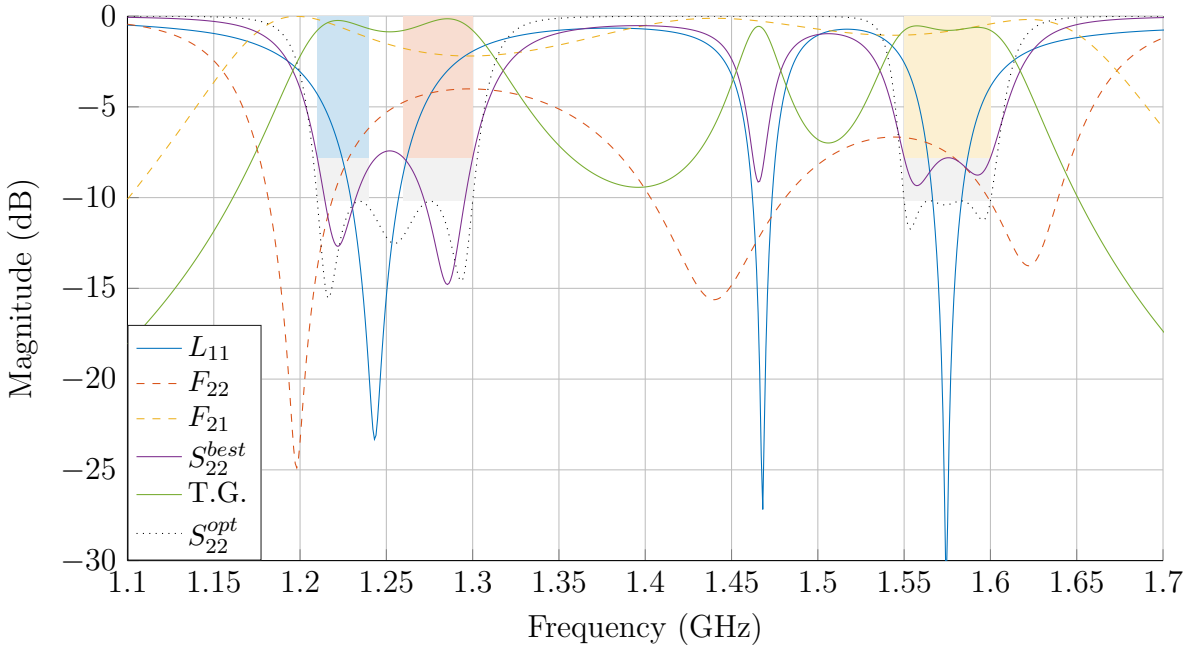


(a) Matching network of McMillan degree $K = 1$

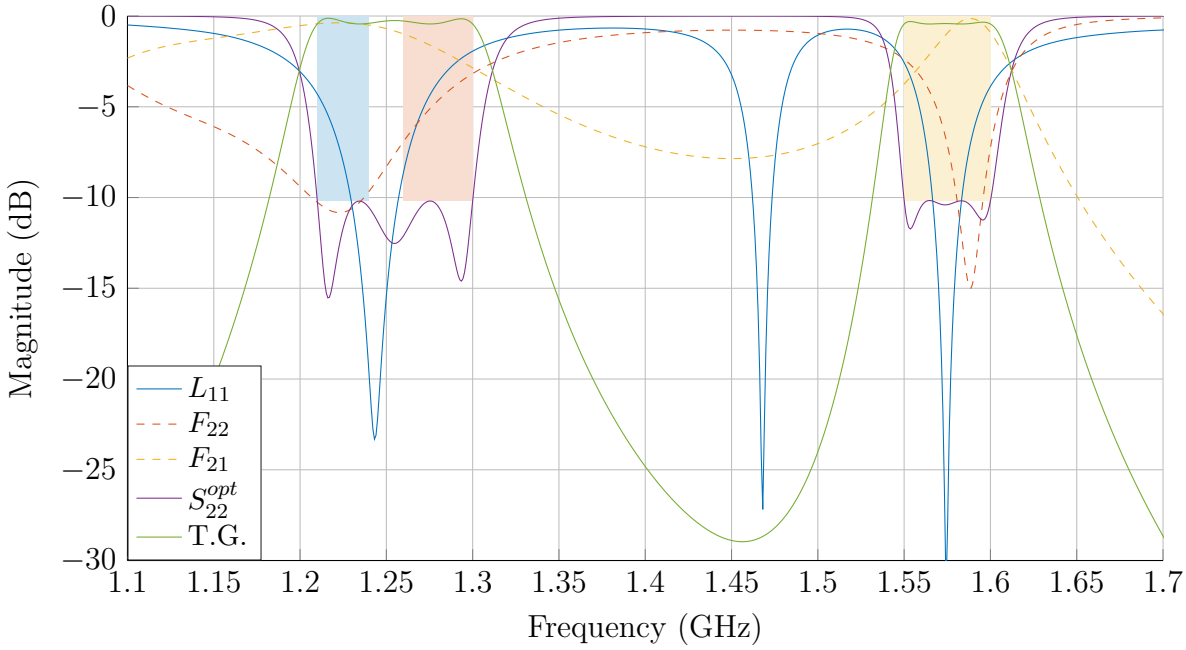


(b) Matching bound for McMillan degree $K = 1$

Figure 10.30: Result of matching a dual-band antenna

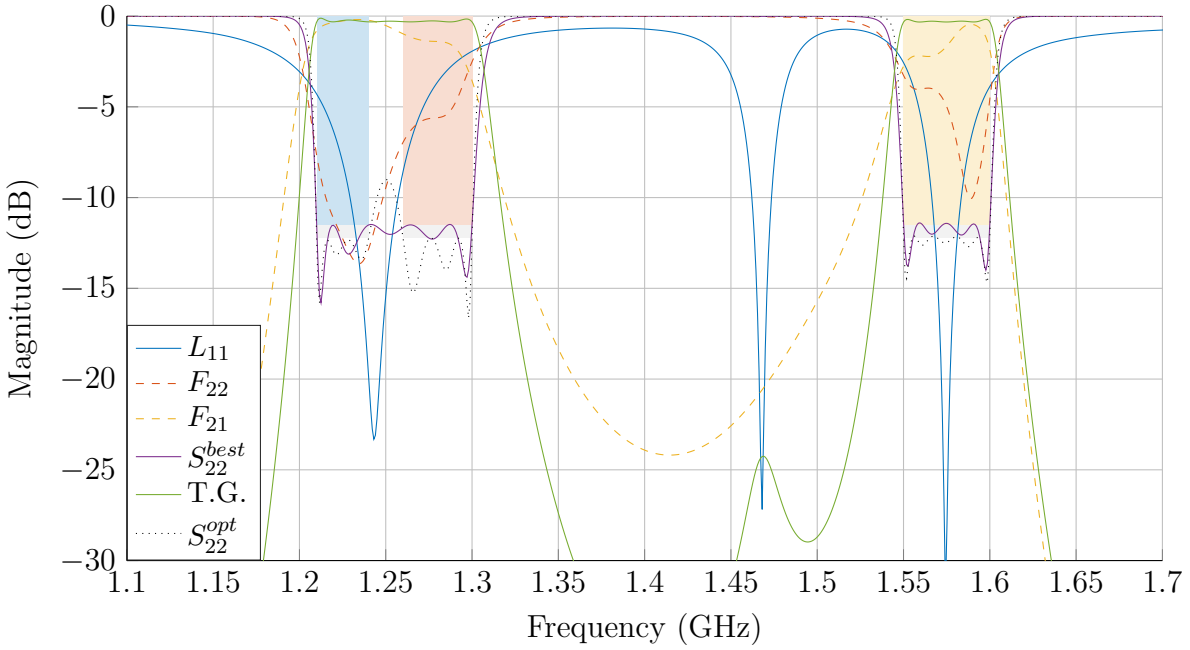


(a) Matching network of McMillan degree $K = 3$

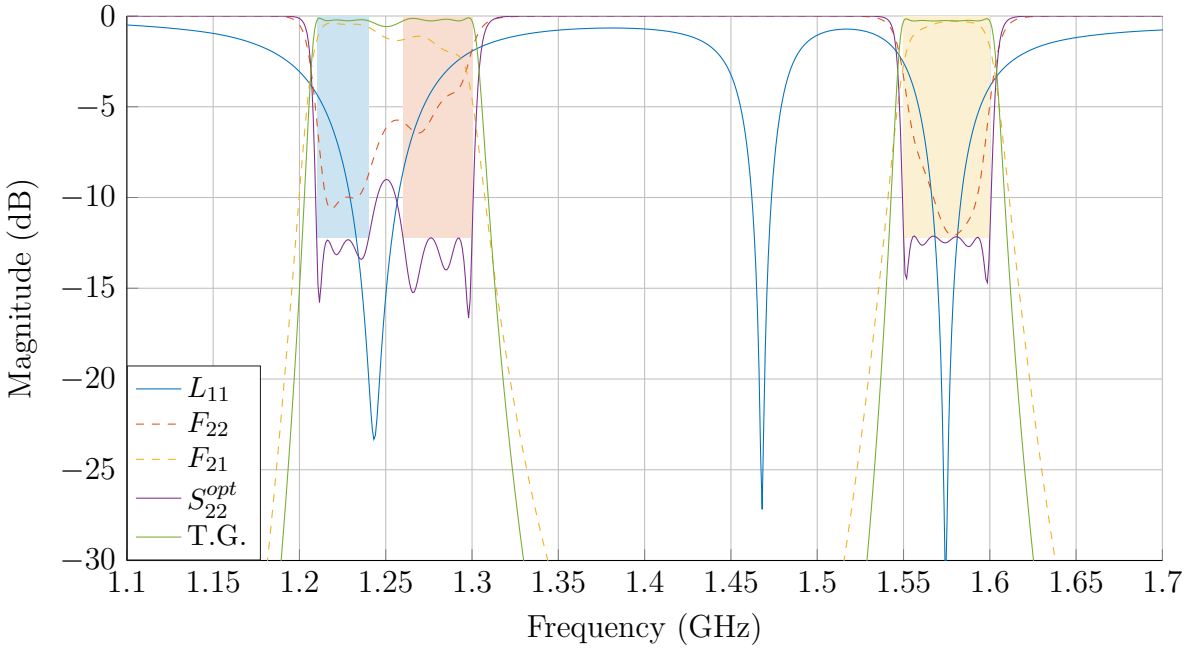


(b) Matching bound for McMillan degree $K = 3$

Figure 10.31: Result of matching a dual-band antenna



(a) Matching network of McMillan degree $K = 9$



(b) Matching bound for McMillan degree $K = 9$

Figure 10.32: Result of matching a dual-band antenna

10.5 Concluding remarks

Thanks to the examples studied in this chapter, it is possible to highlight, on the one hand, the dependence with the K value of the lower bound ψ_{opt}. We verify that there is a substantial difference between the value of ψ_{opt} corresponding to small values of K and the limit $\lim_{K\to\infty} \psi_{opt}$. This confirms that the limits in the literature for $K = \infty$, namely the results obtained in [51] are imprecise for finite values of K.

On the other hand we can also remark that the form of the response corresponding to the optimal matching filter, namely the parameters F_{22} and F_{21} also vary substantially with the value of K. We can compare the parameters F_{22}^{opt} and F_{21}^{opt} as well as F_{22}^{best} and F_{21}^{best} shown in figs. 10.30 to 10.32 which corresponds to $K = 1$, $K = 3$ and $K = 9$ respectively. Note for instance that the response shown in fig. 10.30 does not approach the result for $K = 9$ plotted in fig. 10.32. This fact highlights the interest of the presented algorithm, allowing for the calculation of filter responses with finite degree K as opposed to other procedures available in the literature which provides a matching filter of infinite degree.

References

[47] H. J. Carlin and P. P. Civalleri, *Wideband circuit design*, ser. Electronic engineering systems series. Boca Raton, Fla. CRC Press, 1998.

[48] L. Baratchart, M. Olivi, and F. Seyfert, "Boundary Nevanlinna-Pick interpolation with prescribed peak points. Application to impedance matching," *SIAM Journal on Mathematical Analysis*, 2017.

[49] B. L. Jonsson, S. Shi, L. Wang, F. Ferrero, and L. Lizzi, "On Methods to Determine Bounds on the Q-Factor for a Given Directivity," *IEEE Transactions on Antennas and Propagation*, 2017.

[50] F. Fezai, A. A. Nour, J. Sence, T. Monediere, F. Torres, R. Chantalat, S. Bila, and B. Jarry, "Low-profile dual-band circularly polarized microstrip antenna for GNSS applications," in *2015 9th European Conference on Antennas and Propagation (EuCAP)*, may 2015, pp. 1–4.

[51] J. W. Helton, "Broadbanding:Gain equalization directly from data," *Circuits and Systems, IEEE Transactions on*, vol. 28, no. 12, pp. 1125–1137, dec 1981.

Chapter 11:

Radiation efficiency and dissipation

In the first part of the thesis we have provided a lot of theory around the problem of matching and the numerical implementation of this problem. However, in all the theory presented so far nothing has been said about the effect of dissipation losses in the system. Indeed, in the problem of synthesis, as it happens with the traditional problem of synthesis of transfer functions for the design of filters, the devices are considered lossless at first.

However, it is also important to consider the real case with losses in the system. Note that in this case minimizing the reflection of the system is not equivalent to maximizing the transmission. In fact, if losses due to dissipation in the system are considered, the problem of matching loses its meaning since it is possible to minimize the reflection of the system with a matching filter that dissipates all the received power without transmitting or reflecting anything. This also occurs in the classical synthesis of filters when devices with losses are considered. Indeed, the classic pre-distortion filter design techniques treat this problem by calculating an optimal transfer function in the sense of maximizing transmission in the presence of losses. We can see, for example, in Darlington's contributions in [52] that the mentioned techniques are almost as old as the synthesis of filters.

In the synthesis of matching filters, the limits ψ_{opt} obtained previously for the reflection level also represent limits for the transmission if the quantity $\sqrt{1-\psi_{opt}|^2}$ is considered. Therefore, it is possible to apply the same criterion, namely the maximization of the transmission in the presence of losses, using as initial point, for example the optimum matching filter in the lossless case.

In this chapter we discuss a practical implementation of a matching filter in presence of losses. Additionally we apply the theory to an array of antennas where the radiation efficiency, namely the transmission, is to be maximised. Therefore, before addressing this practical implementation we introduce the concept of radiation efficiency of an antenna which can be considered as the equivalent to the transmission coefficient of a 2-port device.

11.1 Radiated efficiency

The radiation efficiency corresponds to the percentage of the energy delivered to the antenna that is effectively radiated into space and not dissipated in the physical structure of the antenna because of the finite conductivity of metals and losses in the dielectric materials. This amount of radiated energy is easily calculated thanks to Poynting's theorem [53].

Theorem 11.1.1 (Poynting)**.** *As part of the Poynting theorem, given the electric and magnetic field vectors E, H and the current density function J, we find the result stating that the variation of the energy density stored inside a volume V can be computed as the volume integral*

$$\frac{\partial}{\partial t}\int_V \mathcal{E}\mathrm{d}V = -\int_V \mathrm{div}(\Pi)\mathrm{d}V - \int_V E\cdot J\mathrm{d}V,$$

where $\int_V \mathcal{E}dV$ corresponds to the stored energy in the volume V and Π denotes the

Poynting vector $\Pi = E \times H$. *Additionally by the divergence theorem we have*

$$\frac{\partial}{\partial t}\int_V \mathcal{E}\mathrm{d}V = -\int_{\partial V}\Pi \mathrm{d}A - \int_V E\cdot J\mathrm{d}V,$$

with $\mathrm{d}A$ *the area differential element. Therefore, the variation of energy inside the volume equals the integral of the Poynting vector in a surface covering the given volume.*

We show next the motivation behind this theorem and the application to the problem that concerns us. Let us first remember Maxwell equations, since it is the basis for every electromagnetism based calculation, in particular we are interested about the curl equations.

$$\mathrm{rot}E = -\mu\frac{\partial H}{\partial t}, \tag{11.1}$$

$$\mathrm{rot}H = J + \varepsilon\frac{\partial E}{\partial t}, \tag{11.2}$$

where μ and ε denotes the magnetic and electric permittivity respectively while the rot notation denotes the *curl* operator. Similarly we indicate the *divergence* by the div notation. Developing now the expression $\mathrm{div}(E\times H)$ by the fundamental vector identity and introducing eqs. (11.1) and (11.2) we have

$$\begin{aligned}\mathrm{div}(E\times H) &= (\mathrm{rot}E)\cdot H - E\cdot(\mathrm{rot}H)\\ &= H\cdot\left(-\mu\frac{\partial H}{\partial t}\right) - E\cdot\left(J+\varepsilon\frac{\partial E}{\partial t}\right).\end{aligned}$$

Using now the product rule we have

$$\begin{aligned}E\frac{\partial E}{\partial t} &= \frac{1}{2}\frac{\partial}{\partial t}(E\cdot E),\\ H\frac{\partial H}{\partial t} &= \frac{1}{2}\frac{\partial}{\partial t}(H\cdot H).\end{aligned}$$

Therefore

$$\begin{aligned}-\mathrm{div}\Pi &= \frac{\mu}{2}\frac{\partial}{\partial t}(H\cdot H) + \frac{\varepsilon}{2}\frac{\partial}{\partial t}(E\cdot E) + E\cdot J\\ &= \frac{\partial}{\partial t}\frac{1}{2}\left(\mu H\cdot H + \varepsilon E\cdot E\right) + E\cdot J.\end{aligned}$$

Finally taking the integral on the volume V and using the dominated convergence theorem to exchange the derivative and integral operators we have

$$-\frac{\partial}{\partial t}\frac{1}{2}\int_V\left(\mu H\cdot H + \varepsilon E\cdot E\right)\mathrm{d}V = \int_V \mathrm{div}(\Pi)\mathrm{d}V + \int_V E\cdot J\mathrm{d}V. \tag{11.3}$$

Note that the term $\frac{1}{2}\int_V\left(\mu H\cdot H+\varepsilon E\cdot E\right)\mathrm{dV}$ corresponds to the magnetic and electric energy stored by the fields H and E inside the volume V. Furthermore the integral $\int_V E\cdot J\mathrm{d}V$ represents the power loss due to dissipation inside the volume V because of the existence of given conductivity density J. We obtain then that the negative variation

of the stored energy in the volume equals the integral of $\text{div}(\Pi) + E \cdot J$ as stated by the theorem.

In eq. (11.3) the left-hand terms represents the derivative of the electromagnetic energy stored in the volume V and therefore by the energy conservation theorem, the right-hand side indicates the power balance, namely the power dissipation inside the volume V and the power flow through the surface of the volume ∂V. Therefore as an immediate consequence of the Poynting's theorem, we conclude that the integral of the Poynting vector in eq. (11.3) corresponds to the electromagnetic power flux through the surface of the volume

$$P = \int_{\partial V} \Pi \ \mathrm{d}A = \int_{\partial V} E \times H \ \mathrm{d}A.$$

Note that the differential $\mathrm{d}A$ is a vector with direction normal to the surface ∂V, therefore the scalar product with $\mathrm{d}A$ has direction normal to ∂V. Denoting by E_T and H_T the component tangent to ∂V of the electric and magnetic field respectively we have

$$P = \int_{\partial V} |E_T||H_T| \ \mathrm{d}A. \tag{11.4}$$

Finally, for far field and free space propagation, we can use the free space impedance $\mu_0 c_0$ to express eq. (11.4) in terms of the component $|E_T|$ only

$$P = \int_{\partial V} \frac{|E_T|^2}{\mu_0 c_0} \mathrm{d}A.$$

We consider now a spherical volume V containing the radiating element shown in fig. 10.23a and section 10.4 and define the perpendicular directions Θ and Φ tangent to the surface of the volume ∂V as indicated in fig. 11.1. With this notation we can decompose $E_T = E_\Phi + E_\Theta$.

If we denote by a the incident wave and by b the reflected wave as shown in fig. 11.2, assuming that the reference impedance equals the generator impedance ($Z_0 = R_g$) we can define the radiation efficiency as the ratio between the total radiated power P and the incident power to the antenna which is given by the magnitude of a

$$\eta = \frac{P}{|a|^2}. \tag{11.5}$$

Let us now consider a single direction of radiation instead of the entire sphere. In this case, the punctual value of the radiated power in the direction (θ, ϕ) is given by the expression

$$P_{\theta,\phi} = \frac{|E_T(\theta,\phi)|^2}{\mu_0 c_0}.$$

Similarly we can define the radiation efficiency in the direction defined by (θ, ϕ) as

$$\eta_{\theta,\phi} = \frac{1}{\mu_0 c_0} \left| \frac{E_T(\theta,\phi)}{a} \right|^2.$$

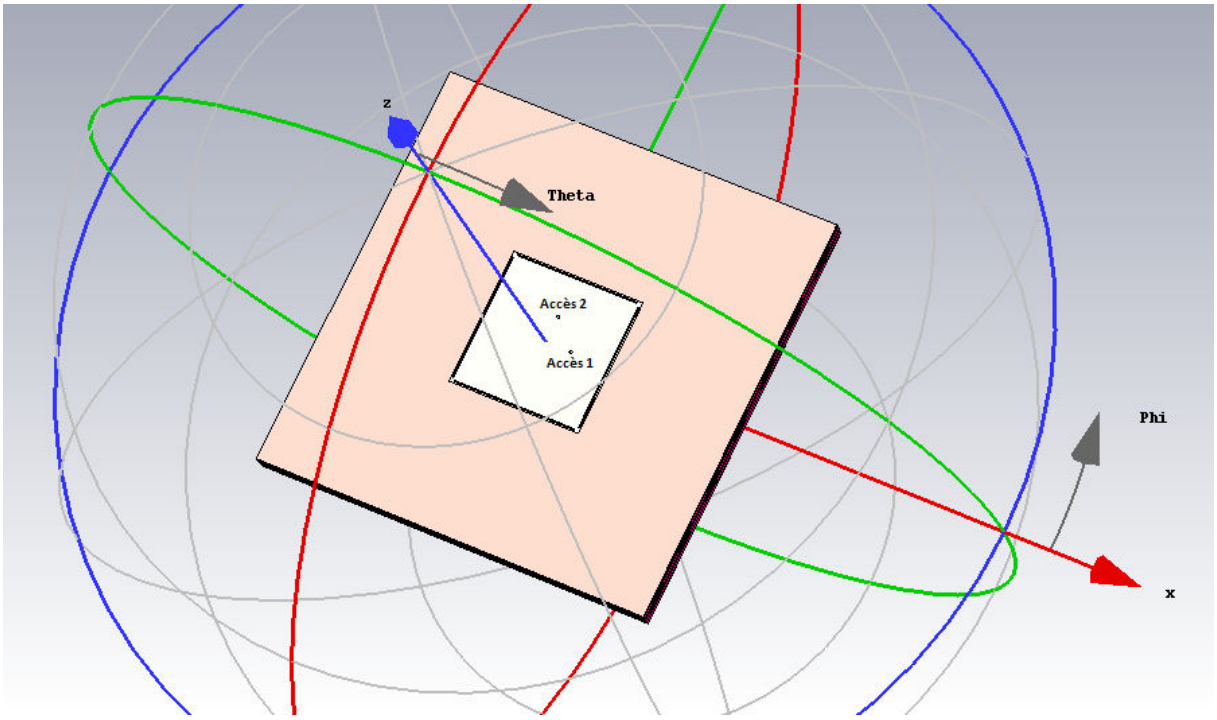


Figure 11.1: Definition of Θ and Φ directions

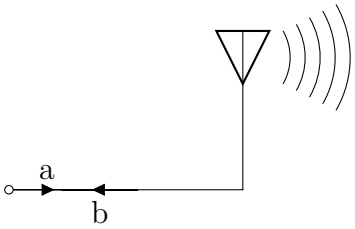


Figure 11.2: Incident and reflected waves

11.2 Extended scattering matrix

The expression of the radiation efficiency given by eq. (11.5) is similar to the transmission coefficient defined in eq. (2.10) as it provides the amount of power that is transmitted with respect to the incident wave a. Therefore we could now think of defining an equivalent matrix of scattering parameters where the transmission to free space represents an additional output port in order to calculate the power radiated by the antenna in terms of the incident waves only. While this is quite simple for a mono port antenna, if we consider an array of antennas or just an antenna with a double input port like in the previous case, things become more complicated.

Our objective now is to calculate an equivalent scattering matrix that allows us to calculate the power radiated by the antenna as a function of the excitation of each port, even in the case of the association of an arbitrary number of antennas. Thus let us consider now a N-ports device with scattering matrix S, we can compute the output wave at port 1 when several ports are excited by means of the superposition principle

$$b_1 = a_2 S_{1,2} + a_3 S_{1,3} + \cdots + a_N S_{1,N}. \tag{11.6}$$

Nevertheless note that the efficiency $\eta \in [0, 1]$ is a positive quantity with no phase information. This missing phase information prevent us from predicting how the radiating waves interact upon the excitation of several input ports simultaneously.

Definition 11.2.1 (Effective transmission)**.** *We define as effective transmission in the direction ϕ, θ, denoted by $S^E_{\phi,\theta}$, the ratio between the component of the radiated E field tangent to the surface ∂V and the incident wave a.*

$$S^E_{\theta,\phi} = \frac{E_T(\theta, \phi)}{a\sqrt{\mu_0 c_0}}.$$

Additionally in the direction (θ, ϕ) we have

$$\eta_{\theta,\phi} = |S^E_{\theta,\phi}|^2.$$

Consider now the schematic in fig. 11.3 representing the circular polarisation antenna introduced before. We denote by a_1, a_2 and b_1, b_2 the incident and reflected waves in ports 1 and 2 respectively. Additionally we denote by $S^E_{\theta,\phi,1}$ and $S^E_{\theta,\phi,2}$ the effective transmission associated to ports 1 and 2 in the direction (θ, ϕ), which are computed by exiting only one port of the antenna at a time. Therefore, by only exciting the i-th port of the antenna and calculating the electric field vector $E(\theta, \phi)$ we have

$$S^E_{\theta,\phi,i} = \frac{E_T(\theta, \phi, i)}{a_i\sqrt{\mu_0 c_0}},$$

with $E_T(\theta, \phi, i)$ represents the radiated E field tangent to the surface ∂V in the direction (θ, ϕ) when only the input port i is excited. It is important to note here that once divided by the value of a_i, assuming a linear behaviour for the antenna, the parameters $S^E_{\theta,\phi,i}$ obtained do not depend on the excitation a_i. Therefore we can define now the radiated wave $b_{\theta,\phi}$ in the direction (θ, ϕ) as

$$b_{\theta,\phi} = a_1 S^E_{\theta,\phi,1} + a_2 S^E_{\theta,\phi,2}. \tag{11.7}$$

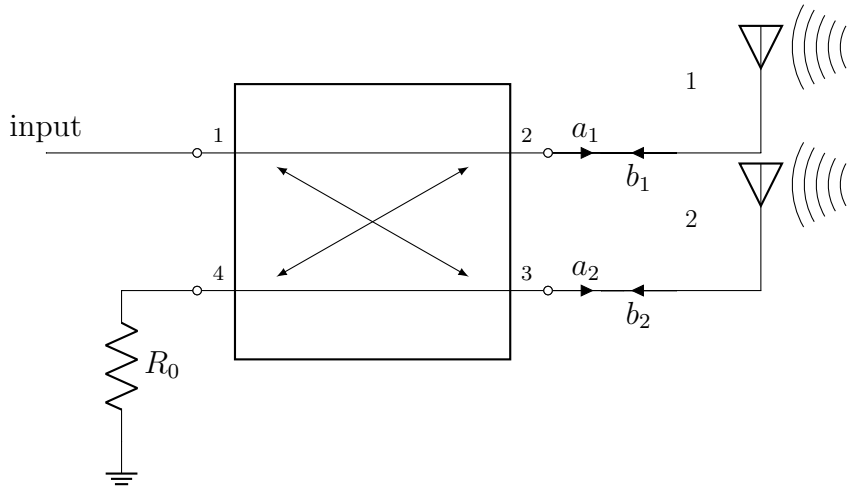


Figure 11.3: Incident and reflected waves in each input port of the RHCP antenna.

Note the similarity between eq. (11.7) and eq. (11.6). Indeed, these parameters $S^E_{\theta,\phi}$ allow us to model the antenna using a traditional scattering parameter matrix, obtaining a multi-port device where each direction (θ, ϕ) in the free space represents an output port. In the general case of an structure with N input ports we have

$$b_{\theta,\phi} = \sum_{i=1}^{N} a_i S^E_{\theta,\phi,i}. \tag{11.8}$$

We define now an extended scattering matrix of size $(N+1) \times (N+1)$ where an additional port corresponding to the transmission to the free space in the direction (θ, ϕ) has been added. Filling the gaps in the aforementioned scattering matrix extended with the parameters $S^E_{\theta,\phi,i}$ we obtain

$$A^E = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,N} & S^E_{\theta,\phi,1} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,N} & S^E_{\theta,\phi,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_{N,1} & A_{N,2} & \cdots & A_{N,N} & S^E_{\theta,\phi,N} \\ S^E_{\theta,\phi,1} & S^E_{\theta,\phi,2} & \cdots & S^E_{\theta,\phi,N} & - \end{pmatrix}$$

. Note that the last element in the previous matrix is missing. However, if we assume that the only input ports in the matrix A^E are the ports from 1 to N then the element $N+1, N+1$ is not necessary. This assumption is equivalent to considering that only the original ports of the antenna, from 1 to N, are excited. Therefore we can calculate the reflection in each of these ports and the coupling between them, by the scattering parameters provided by the original scattering matrix A , as well as the transmission to free space from each one of the input ports.

It is important to note that by reciprocity, the antenna behaves in the same way in transmission as in reception, so the missing parameters could also be defined in a similar way. Nevertheless, the theory developed below is made, without loss of generality, from the point of view of a transmitting antenna, thus only the N first columns of the matrix

A^E are necessary. Therefore we define the non-square $(N+1) \times N$ matrix A^e as the sub-matrix containing the first N columns of A^E

$$A^e = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,N} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N,1} & A_{N,2} & \cdots & A_{N,N} \\ S^E_{\theta,\phi,1} & S^E_{\theta,\phi,2} & \cdots & S^E_{\theta,\phi,N} \end{pmatrix}$$

.

Remark 11.2.1. *Note that this extended scattering matrix can be calculated from the electric field vector radiated by the antenna and allows to obtain the transmitted power for each set of excitations $a_1, a_2, \cdots, a_N$ under assumptions made previously and without the necessity to perform a new EM simulation to recalculate the radiated fields.*

11.3 Optimisation of the array efficiency

Let us now compute the radiated power in the direction (θ, ϕ). This power is given by

$$P_{\theta,\phi} = |b_{\theta,\phi}|^2.$$

Moreover we can again compute the radiation efficiency in the given direction as the ratio between the radiated power and the incident which is given by $|a_1|^2 + |a_2|^2 + \cdots + |a_N|^2$. We have

$$\eta_{\theta,\phi} = \frac{P_{\theta,\phi}}{\sum_{i=1}^{N} |a_i|^2} = \frac{|b_{\theta,\phi}|^2}{\sum_{i=1}^{N} |a_i|^2}. \tag{11.9}$$

Next we consider in a similar way the radiation in every direction (θ, ϕ). Note that considering the full surface covering the volume V, we have an infinite amount of directions in which power is transmitted. We are therefore facing a device with N input ports and an infinite number of output ports.

We introduce expression of $b_{\theta,\phi}$ given by eq. (11.8) in eq. (11.9) and integrate over the surface covering a volume V which contains the radiating structure. In this way we obtain an expression for the radiation efficiency of the array under consideration from the knowledge of the parameters $S^E_{\theta,\phi}$ and the waves a_i incident at each input port.

$$\eta = \iint_{\theta,\phi} \eta_{\theta,\phi} \, \mathrm{d}\theta \mathrm{d}\phi = \frac{1}{\sum_{i=1}^{N} |a_i|^2} \iint_{\theta,\phi} \left| \sum_{i=1}^{N} a_i S^E_{\theta,\phi,i} \right|^2 \, \mathrm{d}\theta \mathrm{d}\phi. \tag{11.10}$$

In eq. (11.10) we calculate the radiation efficiency from the integral of the pointing vector which is obtained as the vector sum of the field contributions corresponding to each of the input excitations to the antenna.

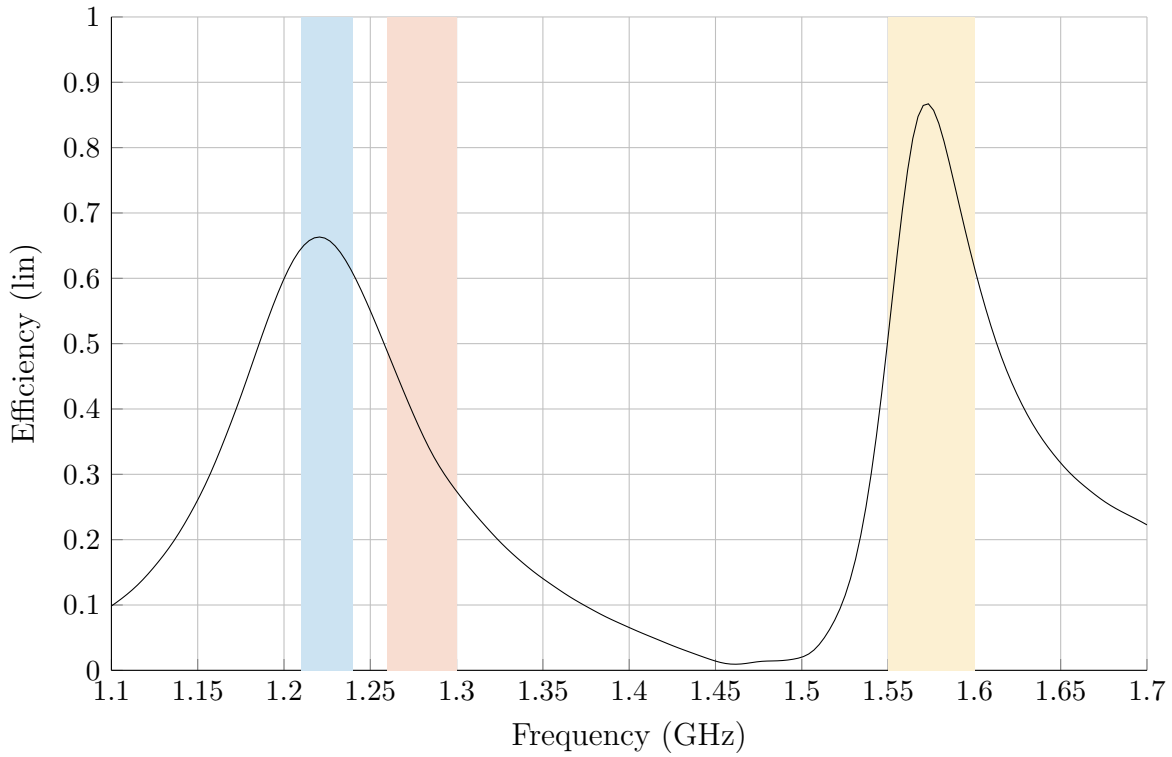


Figure 11.4: Efficiency of the dual-band antenna

11.4 Example: Four elements array

Next we return to the problem of matching within the bands listed in table 10.3. However, this time we consider an array of 4 radiating elements as shown in fig. 10.26, each with two input ports that are fed with phase and quadrature signals respectively. This array gives us a 3-dimensional control of the pointing direction in the space of the main radiation lobe.

Each of the radiating elements is fed throughout a 90-degrees hybrid coupler as illustrated in fig. 10.25. However, due to the mismatch of the antenna, a significant part of the energy is reflected and passes through the coupler in the opposite direction being dissipated in the 50Ω charge. Consequently, the total efficiency is degraded in the passband edges as shown in fig. 11.4.

Once again, in order to overcome the matching issue, a matching filter is introduced in each of the input ports to compensate for the mismatch of each radiating element as indicated in fig. 10.26. Nevertheless in this case it is necessary to match each of the 8 input ports to the 4-elements array. Figure 11.5 shows a diagram of the aforementioned structure, together with a part of the feeding network that provides the signals with the appropriate phase shift at the input ports of each radiant element. We have also included the matching filters preceding each feeding port of the antennas. In total we need a set of 8 matching filters, which are considered equal due to the symmetry of the structure.

Figure 11.6 also shows a top view of the built structure where you can see the four radiating elements along with the excitations of each of them. The matching filters

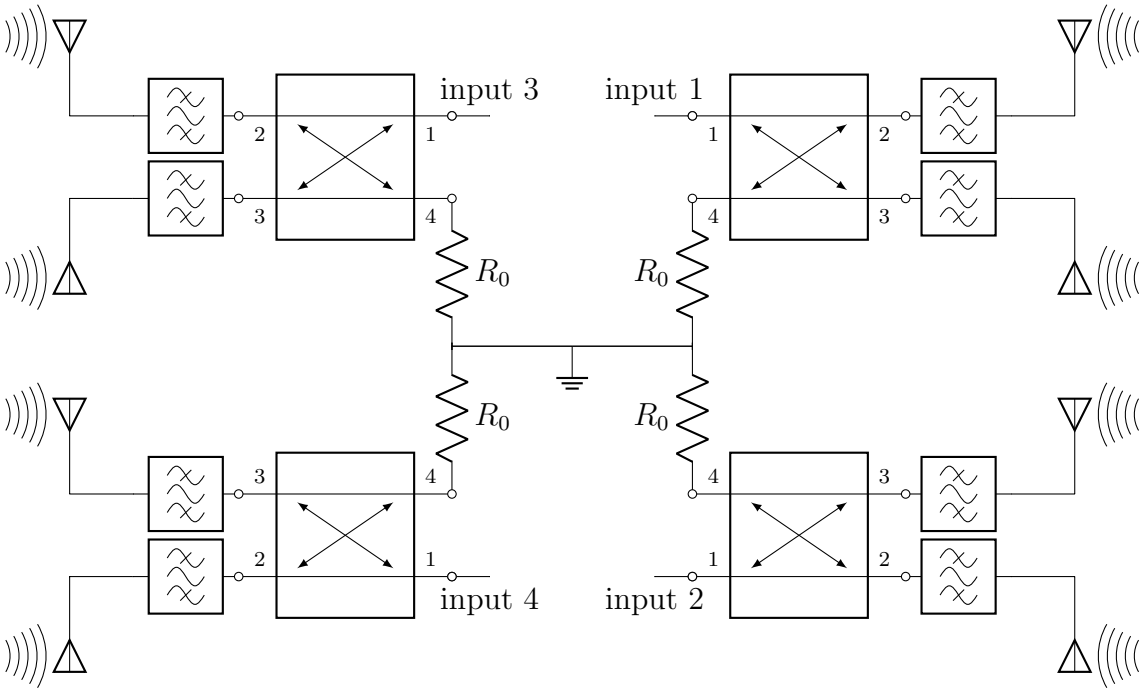


Figure 11.5: Array of radiating elements with matching filters.

together with the hybrid couplers are integrated in the same substrate on the back of each of the antennas.

The structure shown in fig. 11.6 can be parametrised by a scattering matrix A of size (8×8). This matrix is obtained directly from a scattering parameters analysis. In addition, for each of the 8 input ports, we perform an additional analysis using the software CST to calculate, in the far field, the electric field vector E at each position of the space. This allows us to also obtain the elements $S^E_{\theta,\phi,i}$ for all $\theta, \phi \in [-\pi, \pi]$ and with $i \in [1, 8]$. Note, as it has been mentioned previously, that there is an infinity amount of directions parametrised by the values θ, ϕ. However, in practice we sample the interval $[-\pi, \pi]$ selecting a finite number of values for θ and ϕ within that interval. This allows us to obtain a numerical approximation of the efficiency by transforming the integral provided in eq. (11.10) onto a finite sum. Once both parameters are available, on the one hand the matrix A of the array and on the other hand the parameters of equivalent transmission, we calculate the extended matrix A^e. This is a matrix with 8 columns, since the structure has 8 input ports, and a number of rows equal to the number of directions considered in space.

11.4.1 Filter model and optimisation

To implement the matching filters, we select the option corresponding to the result of degree 3 in table 10.4 with 2 transmission zeros. This filter gives us a matching level of -8.6338dB. Note in fig. 10.29 that this is the best possible option with a low degree since to improve the level of matching it would be necessary for a filter of degree 5, which is excessive for this practical implementation. The scattering parameters of this filter are shown in fig. 10.30b with dashed lines. As we have discussed previously, the presented

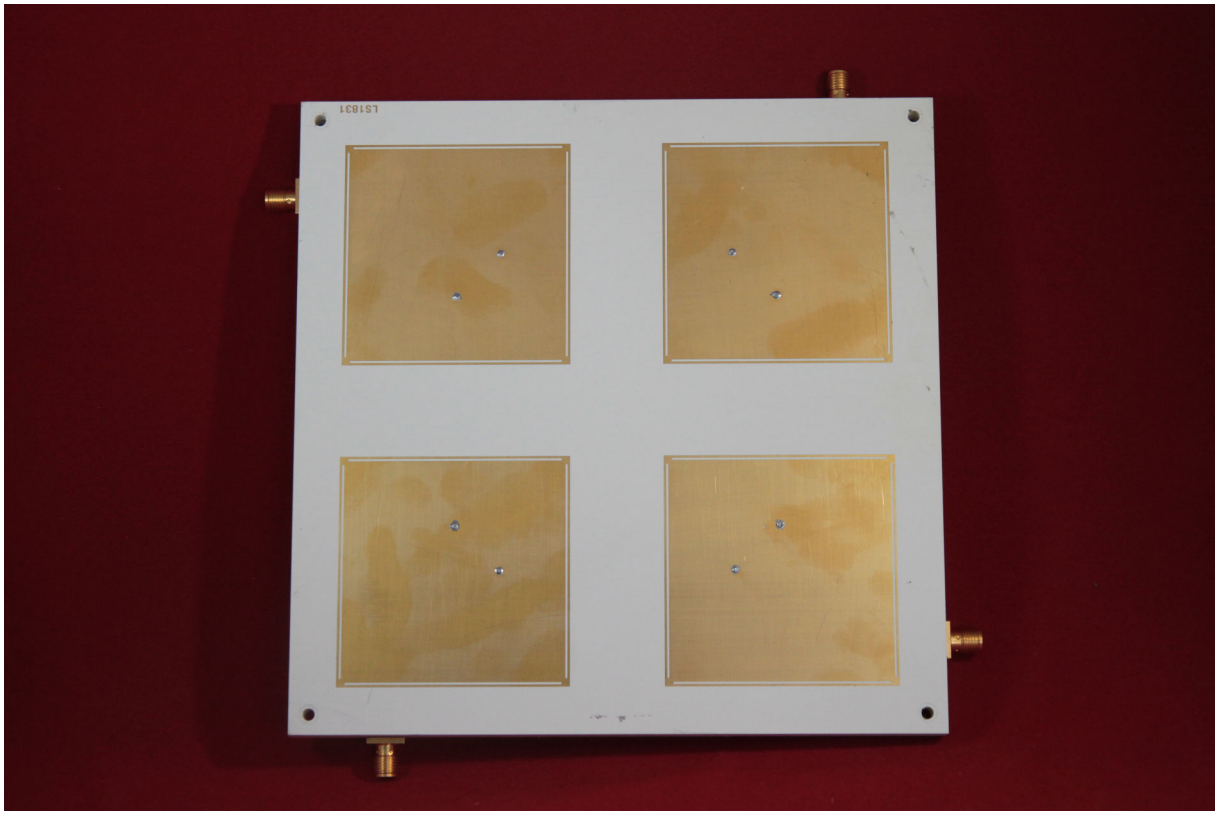


Figure 11.6: Dualband antenna: 4-elements array. Top view.

algorithm can provide us with information on where to place the transmission zeros. In this case it is important to note, as we can see in fig. 10.30b the presentation of one of these transmission zeros extremely close to the frequency axis around 1.47 GHz.

Next we perform a local optimization of the matching filters shown in fig. 11.5. Once again we make the assumption that they are all the same. To carry out this optimization, the filter previously obtained when only one of the elements was considered provides us with a good starting point. Note that in the case where the coupling between the radiating elements is null, namely the matrix A is diagonal, the solution obtained in the previous section represents the optimum solution in the case of an array since each of the radiating elements it can be considered separately.

11.4.2 Optimisation parameters

The optimization of matching filters F is done again by means of a rational model parametrized by two polynomials $p \in \mathbb{P}^3, r \in \mathbb{P}^2$ in the Belevitch form

$$F = \frac{1}{q} \begin{pmatrix} p^\star & -r^\star \\ r & p \end{pmatrix},$$

where q is the stable polynomial such that $qq^\star = pp^\star + rr^\star$. In this model, both polynomials p, r are taken as parameters what allows for both the reflection and transmission zeros to be optimised.

11.4.3 Optimisation criterium

In the optimization of matching filters, the objective is to maximize the efficiency calculated by eq. (11.10). This expression is equivalent to the vectorial sum of the effective transmissions in each direction of the space.

In this case we have considered the transmission in all directions to obtain the real efficiency of the array. However, if the objective is the maximization of directivity, or the pointing of the main lobe in a certain direction, we can consider a limited interval for the angles θ, ϕ, so that only a portion of the surface ∂V is taken into account.

The motivation to use the initial point shown in fig. 10.30b comes from the fact that for an antenna without dissipation losses, the solution shown in fig. 10.30b also maximizes the efficiency of the antenna, since the totality of the energy that it is not reflected it is transmitted. If we then consider a structure with weak dielectric losses, the solution to the efficiency problem is found in the vicinity to the optimal solution for the matching problem.

Remark 11.4.1. *Note that the efficiency is obtained as the transmission of the global multi-port system whose scattering matrix is computed by the chaining of the matrices corresponding to the matching filters with the extended scattering matrix A^e of the array.This extended matrix takes into account the transmission of each of the radiating elements as well as the reflections and coupling between the said elements.*

Moreover, the criterion of maximizing transmission requires that the reflection coefficient F_{22} of the filters must be matched to the reflection of each port of the antenna. However, in this case the transmission criterion also implies an even stronger condition, namely that the signals from the i-th matching filter, coupled between the radiating elements i and k and finally reflected by the k-th filter must be added constructively with the signal coming directly from the k-th input port.

This phenomenon is implicitly imposed in the process of maximizing the transmission of a multi-port device, since a destructive interference, on the contrary, would produce a lower efficiency and therefore a worse criterion in the optimization.

In fig. 11.7 the response of the matching filter resulting from the optimization is shown together with the initial point in fig. 10.30b. This response shares some characteristics with the initial response as the transmission zero around 1.47 GHz. However, both devices differ fundamentally in the McMillan degree. While the initial response obtained by solving the matching problem with a single radiant element is of degree 3, the filter response optimized considering the array of antennas shown in the figure is only of degree 1.

Note as we have already highlighted several times previously, the response of the optimal matching filter may not be of full degree, especially in the multi-band case. Moreover, in this case we have also moved from a structure with a single radiant element to an array in which 8 filters intervene, such as the one shown in fig. 11.7. Because of the coupling between the different radiating elements, each filter contributes to the

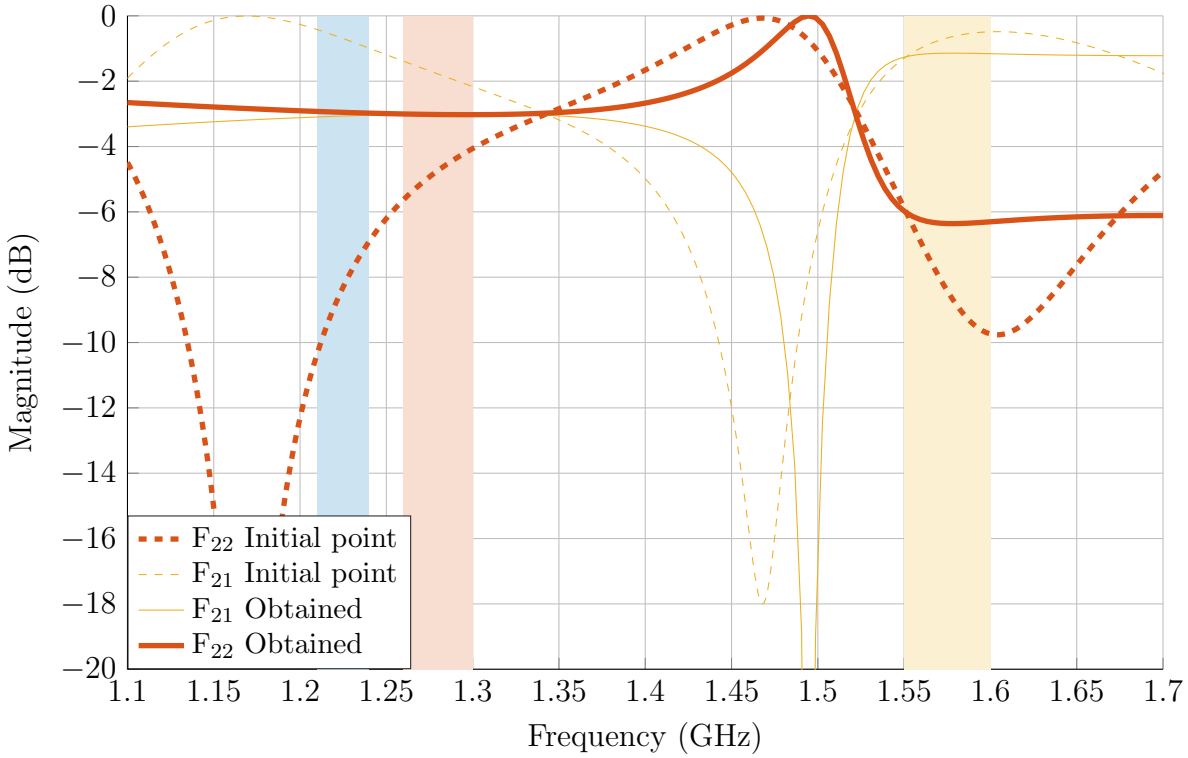


Figure 11.7: Result of efficiency optimisation

optimization of overall efficiency, since, as has already been pointed out, the efficiency criterium from the point of view of each matching filter takes into account the signals coming from the other filters, which must be reflected to the antenna with the appropriate phase. Therefore, the decrease in degree is only apparent since we start from a filter of degree 3 to obtained a network of 8 filters of degree 1 each.

11.4.4 Design of the matching filter

Given the result obtained in fig. 11.7 the objective is to design a filter that can be integrated with the antenna and which approximates as close as possible the response shown in fig. 11.7. From the objective response obtained above, we can extract the pi network shown in fig. 11.8a composed of lumped elements, namely capacitors and coils only, together with a transmission line at the input which interconnects with the antenna feeding port, implementing a phase shift of 1.87 radians at the frequency of 1.4 GHz.

The first step taken for the practical implementation of the structure shown in fig. 11.8a has been to neglect the 64.1 nH coil since due to its great value, the influence on the response is minimum. Note that a coil of a sufficiently large value in parallel can be approximated by an open circuit. Then the series resonator has been implemented by means of a resonator in microstrip technology while the 3.7 nH coil has been conserved in the form of a lumped component.

The matching matching filter has a transmission zero and a zero of reflection at finite frequencies. This implies that a direct coupling between the input and the output is

necessary to implement said zeros with a degree 1 response. In fig. 11.8, the final filter designed is shown. We can observe the direct input-output coupling, namely between the ports $P1$ and $P2$. We can also appreciate the $\lambda/4$ resonator together with transmission line which implements a $\lambda/4$ impedance inverter to introduce the transmission zero present in the target response. Finally we can see the ports $P3$ and $P4$ destined to the connection of the coil.

The structure shown in fig. 11.8 has been optimized to closely implement the target response, in the pass bands. In addition, to select the optimum value of the inductance shown in fig. 11.8a (of 3.8nH in the ideal circuit) a sweep has been made on the different commercial values of said component around the nominal value. A set of 8 copies of this filter have been integrated, along with the antenna in a single component, each filter connected to one of the input ports of the antenna as in the schematic shown in fig. 11.5.

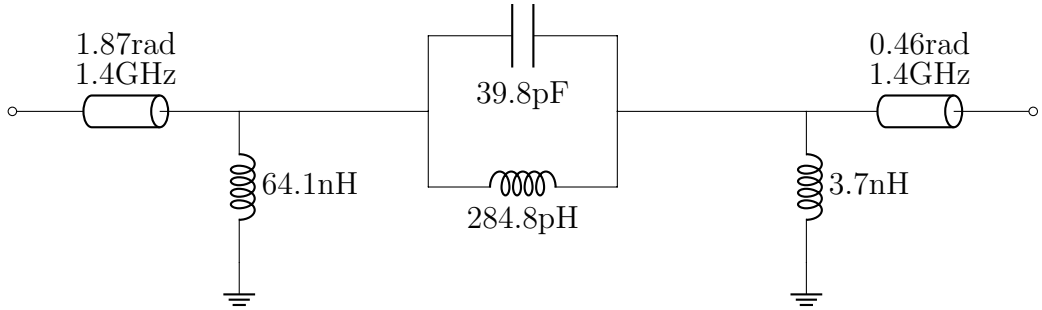
Figure 11.10a shows the back side of the antenna array illustrated in fig. 11.6 where we can see the implementation of the network from fig. 11.5. Additionally we provide in fig. 11.10b a close view of the feeding circuit of each radiating element. In this figure we can distinguish the 90 degrees hybrid coupler and the two matching filters whose layout is provided in fig. 11.8.

In fig. 11.9 the response obtained as a result of the filter optimization in fig. 11.8 is shown in both module and in phase. The optimization has been carried out within the bands shown in the figure in which the criterion is defined in this application, namely the maximization of efficiency.

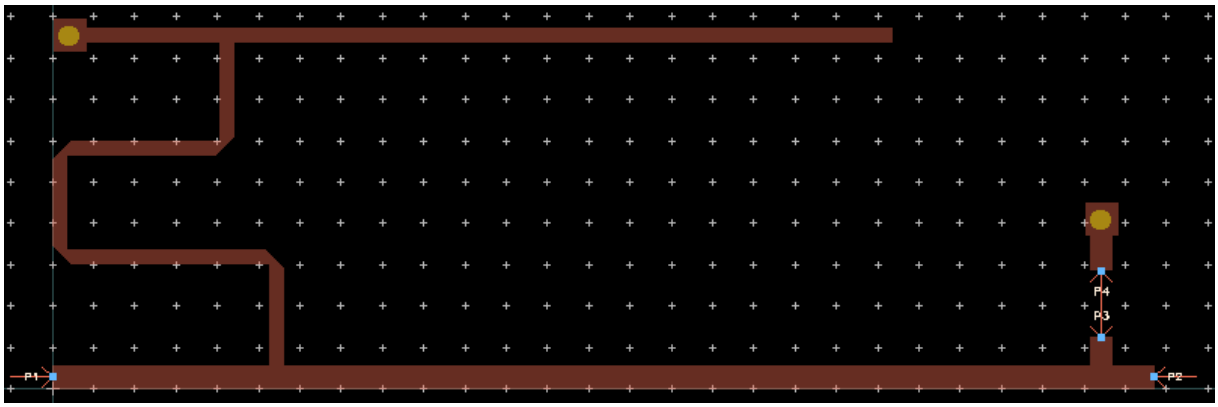
11.4.5 Global efficiency result

Using the matching structure presented in this chapter we can calculate the improvement in terms of radiation efficiency achieved by using the network of matching filters obtained. In fig. 11.4 we show a comparison between the efficiency of the array without matching elements (in blue), the efficiency obtained by the ideal matching filters (lossless) synthesized shown in yellow and finally in red the efficiency obtained through the matching network shown in fig. 11.10.

In fig. 11.11 it is possible to appreciate the efficiency improvement by adding the matching network, which is more pronounced in the GALILEO E6 band. Additionally, the decrease in the minimum efficiency between the result issue of the lossless synthesis and by the actual filters due to the dissipation. This dissipation causes the decrease from -1.1dB reached by the ideal curve at the frequency of 1.3GHz to -2.36 dB obtained at the frequency of 1.24 GHz with the matching network implemented. However, it is important to note that in this case, no selectivity requirements have been considered. Therefore, it is necessary to obtain a compromise between the gain in efficiency provided by the matching filters and the losses due to dissipation introduced by the filters themselves. This trade off is reached by choosing the McMillan degree of the matching filters. If this degree is too high, the losses introduced by the filters will exceed the obtained gain in efficiency and therefore deteriorating the overall performance of the system.



(a) Matching network extracted from the ideal response



(b) Implemented solution.

Figure 11.8: Design of the matching filter

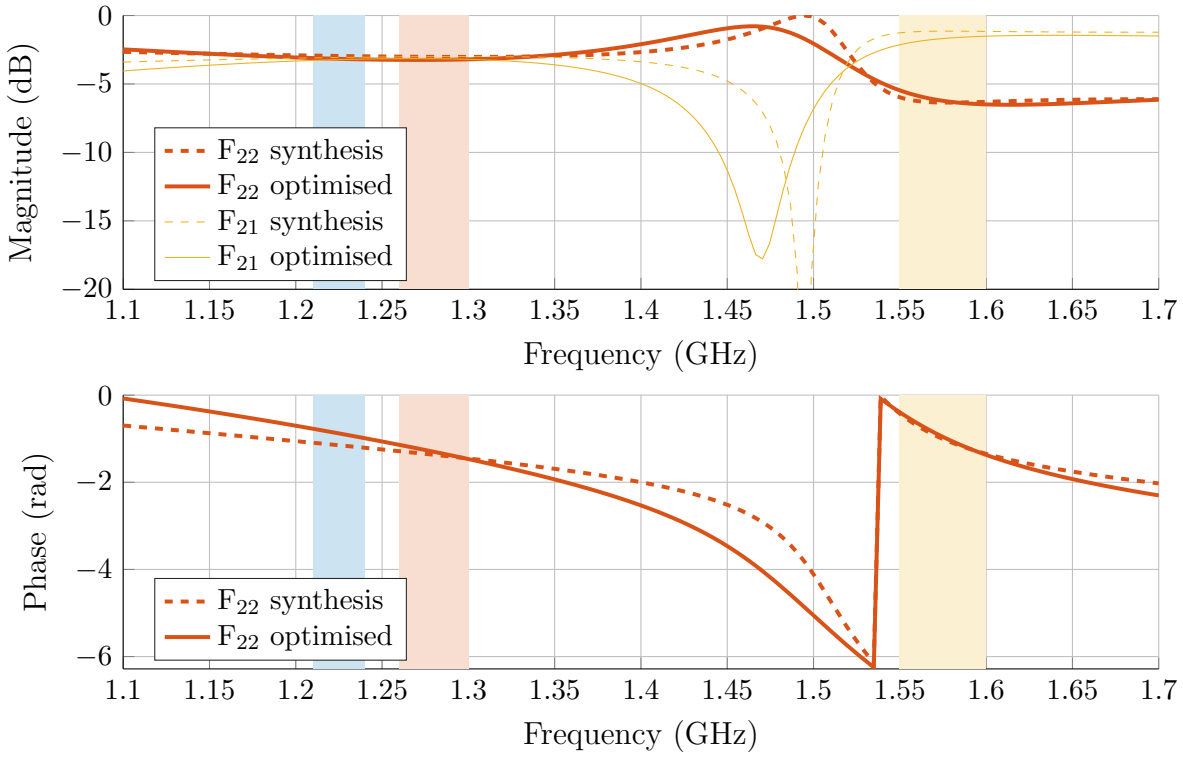
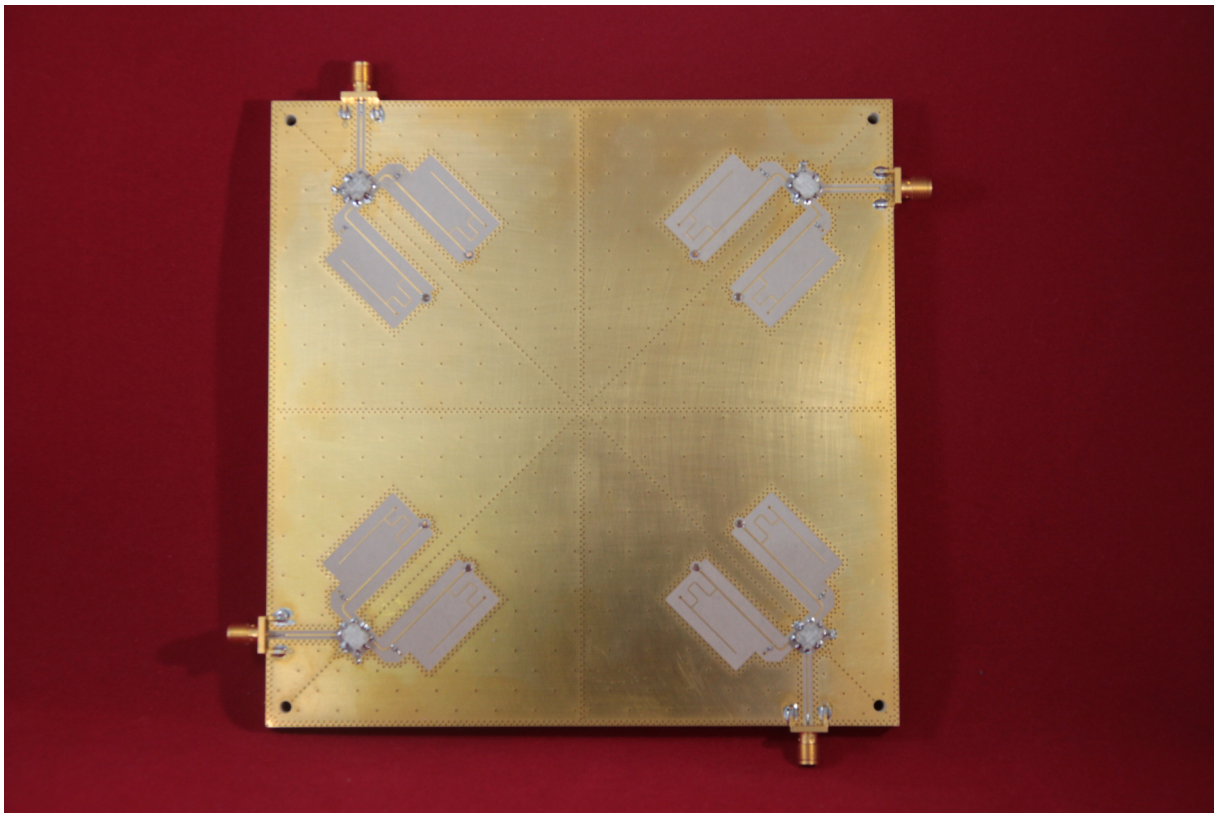
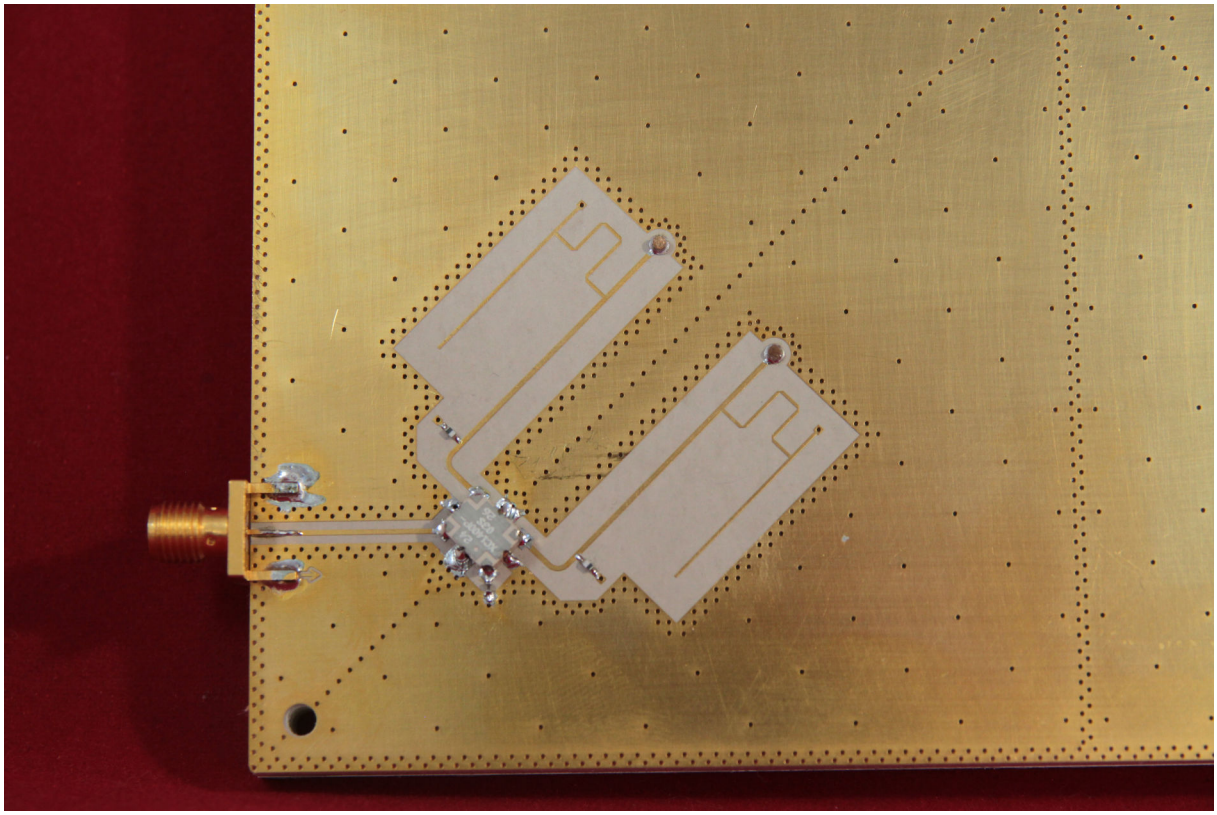


Figure 11.9: Result of filter optimisation



(a) Top view



(b) Close view

Figure 11.10: Matching filter array

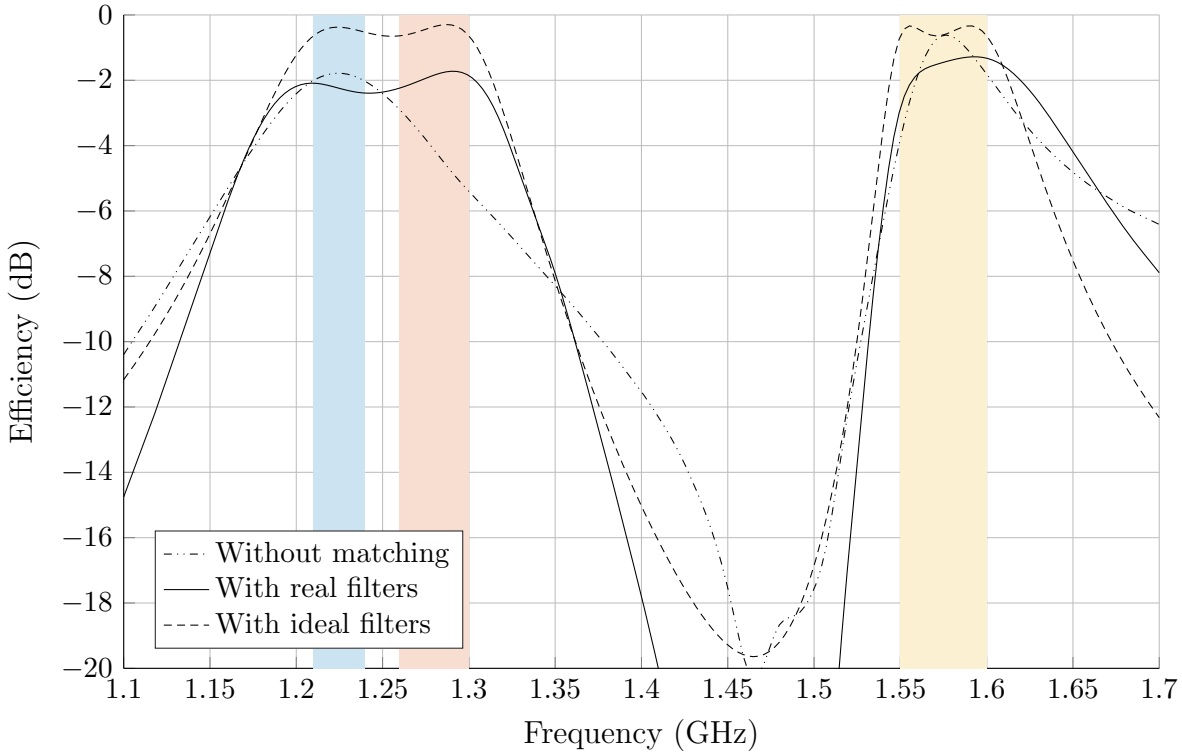


Figure 11.11: Final efficiency of the array.

Conversely, it should also be noted that when selectivity specification are imposed and the received signal need to be filtered whatsoever, no drawback is added by the use of matching filters since additional losses will be introduced by the filters in any case.

Finally, we present the final result of this study in fig. 11.12, namely the original efficiency of the array compared to the efficiency obtained through the use of matching filters, this time in a linear scale. We can see that the efficiency of the antenna has been improved from a minimum value of about 0.28% at the frecuency of 1.3 GHz to a value of 0.58 % attained by the final result at the frequency of 1.24 GHz as it is summarised in table 11.1.

Freq. (GHz)	Eff. without matching (%)	Eff. with matching (%)
1.21	0.64	0.62
1.24	0.61	0.58
1.26	0.49	0.61
1.30	0.28	0.63
1.55	0.50	0.60
1.60	0.61	0.72
min	0.28 (1.3GHz)	0.58 (1.24GHz)

Table 11.1: Summary of the obtained efficiency at each frequency.

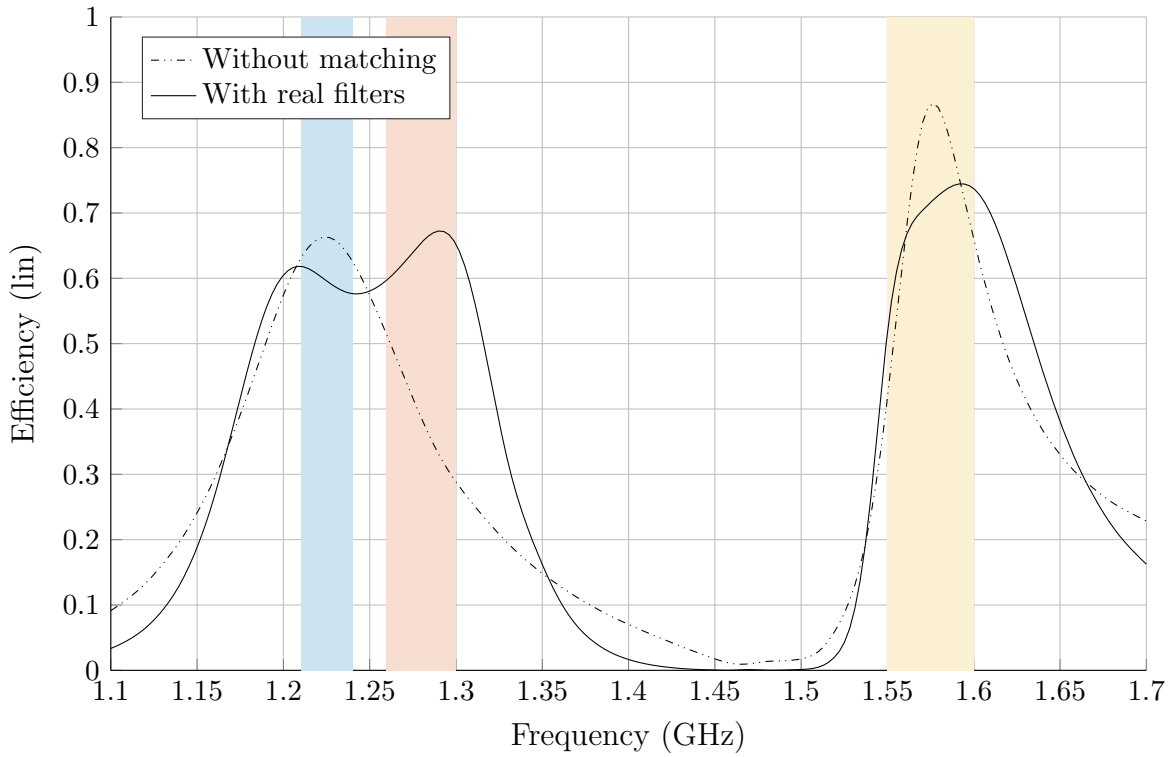


Figure 11.12: Final array efficiency (linear scale).

References

[52] S. Darlington, "Synthesis of Reactance 4-Poles Which Produce Prescribed Insertion Loss Characteristics: Including Special Applications To Filter Design," *Journal of Mathematics and Physics*, 1939.

[53] J. H. Poynting, "n the Transfer of Energy in the Electromagnetic Field," *Philosophical Transactions of the Royal Society of London (1776-1886)*, 1884.

Part V

Synthesis of Multiplexers

Chapter 12:

Introduction to multiplexer synthesis and state of art

The high power amplifiers used in satellite payloads require a narrowband signal to avoid the effects introduced by the non-linearity of these devices and operate with high efficiency. As a result, multiplexers are needed to separate a broadband signal in a given number of narrowband channels. Multiplexers are also used to combine a group of narrow-band channels into a broadband signal transmitted through a conventional antenna.

Figure 12.1 shows the simplified block diagram of a satellite payload system composed of a receiving antenna (uplink), a low noise amplifier (LNA) and an input multiplexer to separate the signal into narrower-band channels. The amplification chain composes high power amplifiers to amplify each channel separately and an output multiplexer to recombine the amplified channels into a single signal suitable to be transmitted. However, it is possible to distinguish between two categories of multiplexing applications.

On the one hand, we have input/output multiplexers. Input multiplexers divide a received broadband signal into a given number of narrowband channels. This task is performed after the broadband amplification of the signal by an LNA. Conversely, output multiplexers are used to recombine the frequency channels after the high power amplification. In this application, shallow insertion losses are a critical condition, since it has a direct impact on the transmitted signal.

On the other hand, we can speak about the transmitter-receiver diplexers. A duplexer is a two-channel multiplexer that is often used to share a single antenna between the transmitter and the receiver allowing to separate the uplink and downlink bands.

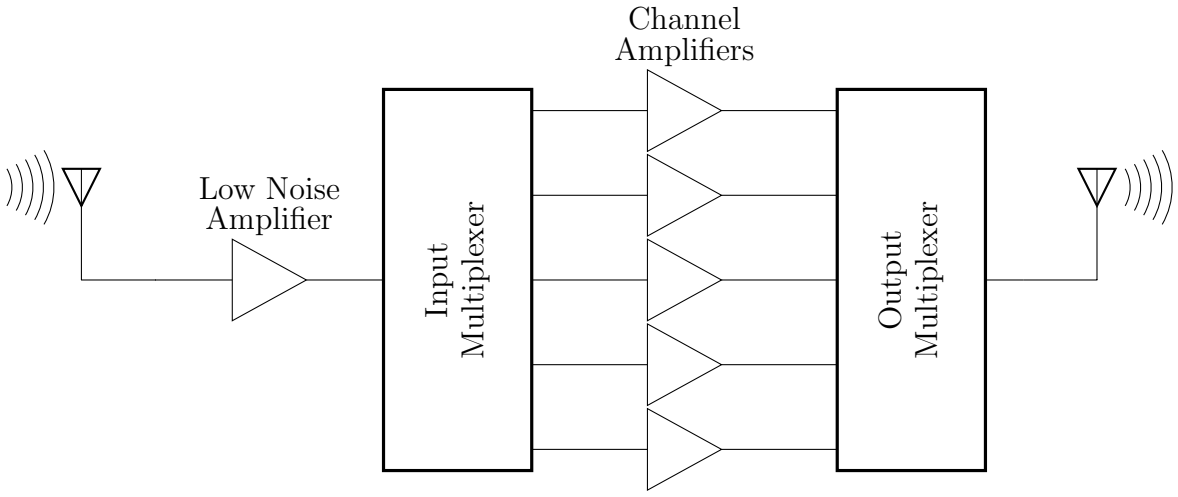


Figure 12.1: Simplified satellite payloads diagram.

12.1 Multiplexing techniques

The design of multiplexers is a recurring problem in the field of passive microwave devices which takes the synthesis of transfer functions a step further by considering a set of filters coupled together. In addition, the type of coupling between the different channels

can simplify or further complicate the design process. Indeed, depending on the type of coupling between channels we can distinguish different types of multiplexers, which do not differ simply in the topology, but also in the features and in the design techniques used to perform the dimensioning of the structure.

Below is a list of the main types of multiplexers according to the type of junction used, which allows us to review the main characteristics of each one as well as the design procedure necessary in each case. Nevertheless note that we can find a more exhaustive list in [54].

12.1.1 Modular multiplexer configurations

Below we present two types of structures characterized by having a modular design. This is possible since there is no interaction between the different channels, so it is possible to eliminate or add a new channel at any time without disturbing the operation of the other channels. These structures present mainly the disadvantage of the high level of losses and the large volume occupied.

12.1.1.1 Hybrid multiplexers

Multiplexers coupled by 90-degree hybrids are based on the direct properties of this hybrid. We can see an example in fig. 12.2. The 90 degree hybrid splits the input signals from ports 1 onto the port 2 and 3 meanwhile the input signals at ports 2 and 3 get combined to the port 4. Therefore in the structure shown in fig. 12.2 the broadband input signal passes through a directional coupler that divides the signal between two identical filters in the first channel. At the output, another hybrid combines these signals. Similarly, the reflected signals of both filters go back to the first 90-degree hybrid that recombines these signals on the direction to the next channel and cancels them on the return way. This structure made of the 90-degree hybrids together with the recombining and destructive path composes the manifold of the multiplexer in this case.

Through this structure, we manage to separate the signals transmitted and reflected by the filters, so that the transmitted signal, already filtered, is recovered at the output of the multiplexer, while the reflected signal, corresponding to the rest of the channels, is redirected to the next set of filters, without interfering with the input signal to the first channel.

As a negative aspect of this structure it is important to note that the signal that crosses the coupler at the entrance of each channel is divided into two separate branches, so two identical copies of the same filter are necessary. In addition a second coupler is necessary in each channel to recombine the signals coming from these two filters. As a result the structure corresponding to each channel represents a considerable volume needing to locate two filters and two couplers. It is also necessary to consider the increase in the level of losses introduced into the system due to the presence of cascaded couplers.

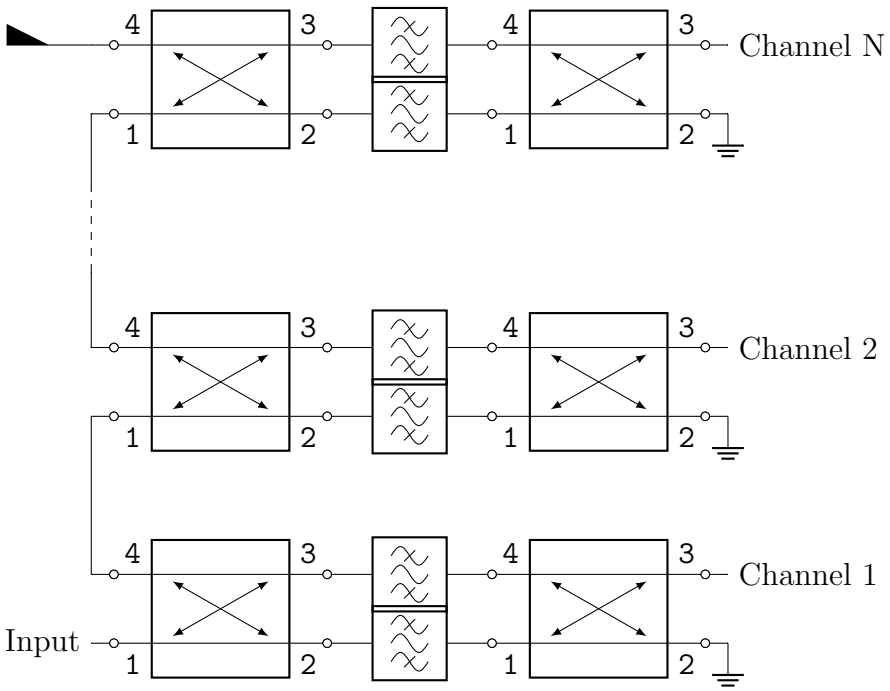


Figure 12.2: Hybrid coupled multiplexer.

12.1.1.2 Circulator-coupled multiplexers

Multiplexers based on couplings by circulators follow exactly the same principle as those based on 90-degree hybrids introduced in the previous section. The only difference is that the function performed by the hybrids, namely the direction of the signals reflected in one channel at the entrance of the next, is performed by a circulator. Therefore the advantages provided by the multiplexers with directional couplers related to the modular design are maintained. It is possible to design each channel independently of the structure of the multiplexer due to the unidirectional property of the circulator, as well as to add new channels at any time. Besides, this structure overcomes the disadvantage present in the multiplexer with directional filters in terms of the need for two filters in each channel. This fact leads to a reduction in size since each channel is composed only of a circulator and the corresponding filter.

We still have, however, the problem of cumulative losses since the signal at the entrance of the second channel must first pass through the first circulator, the input signal to channel 3 must pass through circulators 1 and two, and so on successively.

12.1.2 Non modular multiplexer configurations

Unlike modular structures, these structures work by coupling all channels directly to the input port. In this case the design is a simultaneous matching problem since each of the filters is matched, within its respective band, to the set of all the other filters, which are also matched in the corresponding band simultaneously. As a consequence, it is not possible to modify the channel configuration once the device has been designed since the elimination or addition of a new channel modifies the operation of the rest of the channels. However, as an advantage of these non-modular devices, we obtain a lower footprint due

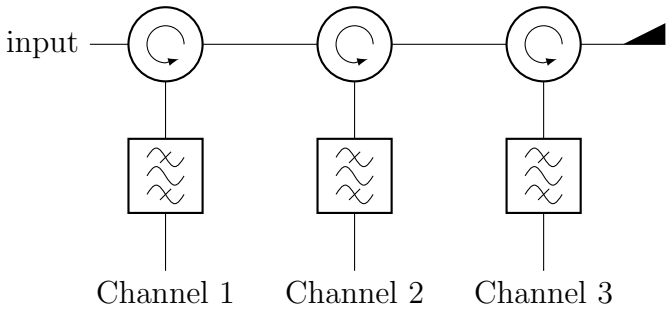


Figure 12.3: Circulator coupled multiplexer.

to the simplicity of the structure, which also leads to a lower level of losses.

12.1.2.1 Star junction multiplexers

Star-junction multiplexers consist of a certain number of filters, depending on the number of channels, directly coupled to a common input port as in fig. 12.4. This type of multiplexers represents, among the non-modular multiplexers, the type of multiplexers that is easier to synthesize. In fact, if we assume that the junction is small enough to be considered non-dispersive in frequency, the design of this type of device becomes a problem of simultaneous matching.

If we also suppose that the passband of the channels are sufficiently separated in frequency so that the impedance shown by each channel within the band of the other channels is constant, the synthesis of the multiplexer is equivalent to the classical synthesis of filters. The synthesis of star-junction multiplexers with channels spaced in frequency is already well know, and classical papers can be found in the literature. We can find for instance [55] where the method of synthesis of traditional filter prototypes, prior to the coupling matrix, based on impedance inverters, is used for the synthesis of multiplexers. Additionally, we can also find modern studies where these structures are synthesized by an equivalent circuit based on coupled resonators. For instance in [56, 57, 58].

However, this structure does not allow the inclusion of a large number of channels due to the topological restrictions that prevent the connection of more than a certain number of channels to the same input node, without including transmission lines or other devices. Therefore Star junction multiplexers are used when only a small number of channels is required, such as diplexers or triplexers [59, 60, 61, 62].

12.1.2.2 Manifold-coupled multiplexers

Finally, manifold multiplexers are the approach in which we have focused this work. These multiplexers are composed of a bandpass filter for each channel interconnected by transmission lines as shown in fig. 12.5. With a manifold configuration, the channels are not isolated from each other, and the reflected signal of a given filter influences the response of the whole device.

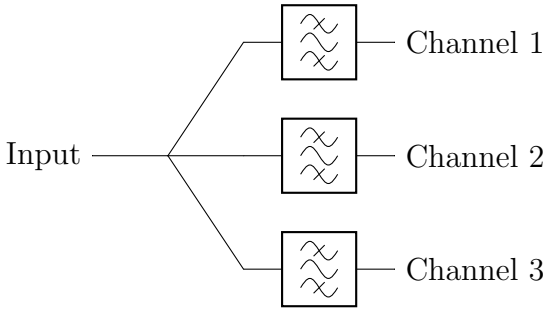


Figure 12.4: Star multiplexer.

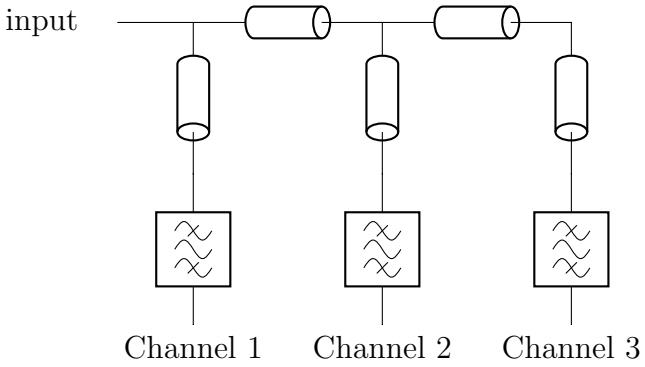


Figure 12.5: Manifold multiplexer.

This configuration is widely implemented for spatial applications since this approach offers the best result in terms of structure miniaturization and total insertion losses in each channel. Also, the inability to perform a modular design is usually not an inconvenience in the space sector where no physical access to the designed and operating devices is possible ([54, 63]).

In this context, some authors have developed optimization methods to achieve the desired behaviour [64, 65, 66, 67] and also synthesis techniques for the whole multiplexer [68]. This is the most relevant type of multiplexers for this thesis since it presents the greatest challenge in terms of synthesis of the channel filters and design of the structure. Therefore in this chapter multiplexers of the manifold type receive special attention. In the next section we recapitulate the existing design techniques as they have been presented over the years. In the same way, the most important contributions will be highlighted as they are of greater interest for the work done in this part.

12.2 State of art and techniques for manifold-type multiplexer synthesis

The literature on the design of multiplexers of manifold type is vast since many authors have tried over the years to obtain an analytical synthesis procedure similar to the synthesis of filters by means of the model of coupled resonators. However, the synthesis of filters benefits from the fact that the responses obtained by a structure formed by coupled resonators can be approximated in a relatively narrow band to a rational function. This

does not occur in the case of multiplexers of the manifold type. The reason is that the manifold, constituted mainly by transmission lines is by definition, not rational. Indeed, a moderately accurate manifold model must contain functions of exponential type.

12.2.1 Rhodes and Levy's theory. Classical multiplexer synthesis

Among the first contributions in manifold design we find in [69] the works of J. Rhodes and R. Levy who introduced the synthesis of this type of devices is an extension of the synthesis of double ended filters. The synthesis of double terminated filters supposes a constant impedance in each one of the ports.

It is interesting to note that this type of synthesis can easily be applied to the design of multiplexers as long as the impedance shown by each channel filter in the adjacent bands is constant. Therefore in [69] it is considered that the channels are sufficiently spaced to validate the previous assumption. This assumption is even more legitimate in the case of narrow-band channels where the variation of the impedance seen by each of the channel filters, namely the impedance shown by the rest of the multiplexer, is small.

In addition, the assumption is also made that the manifold response is independent of frequency. This means that the transmission lines that make up the manifold introduce a constant phase shift in frequency instead of a disperse phase shift. With these assumptions we are again in the case of the synthesis of a star-coupled multiplexer. Indeed, the structure obtained is a star multiplexer where a series of frequency-independent reactances, which we can consider as part of the star-junction, have been introduced between the different channels.

The theory presented here used elegant approaches to formulate the problem of the multiplexer synthesis in a rigorous way, as the synthesis of filters was formulated at the time, by means of a equivalent low-pass prototype computed by reactive elements, to reduce the number of condensers or coils, and impedance or admittance inverters. However, this effort to formulate a synthesis problem in an analytic way has progressively disappeared, and the synthesis of multiplexers has not followed the same evolution as that of microwave filters using coupled resonators. The reason lies undoubtedly in the difficulty to relax the assumption made about the absence of frequency-dispersion in the behaviour of the transmission lines.

12.2.2 Transition to optimisation-based synthesis

In the following years, while the synthesis of filters evolved from the newly introduced concept of the coupling matrix, the synthesis of multiplexers remained stagnant. The concept of the coupling matrix allowed to represent the interactions in the complete structure of a filter by simply using a matrix. In fact, we can find several contributions (i.e. [70]) where this concept applies equally to the design of multiplexers obtaining a much more tractable parametrisation than that obtained by Rhodes and Levy. This simplified parametrisation makes it possible to directly apply classical optimization

techniques to the whole structure.

Since then, multiplexer design has evolved along with non-linear optimization algorithms, relying increasingly on the increasing computational efficiency to solve the problem.

12.2.3 Fix-point algorithm. An iterating matching procedure

One of the recent design techniques that is worth noting is the fixed-point algorithm used in [71, 72]. This method seeks to formulate the problem of simultaneous matching as a problem of standard matching with a variable impedance in frequency similar to the problem dealt with in this thesis. To do this, it considers the synthesis of a single channel filter in each iteration and poses a matching problem assuming that the rest of the filters are fixed. However, to solve this problem, non-linear optimization techniques are used, so the optimality of the solution or even the convergence of the algorithm is not guaranteed. Additionally, it should be noted that the case in which the load shown by the other filters is relatively constant in frequency constitutes one of the rare cases in the aforementioned matching problem can be optimally solved. This is the case of multiplexers with sufficiently spaced channels.

Moreover, as the pass band of the channels becomes increasingly wider and the separation between them narrower, another problem has appears in the design of this type of devices, already complicated by itself. This is the problem of the manifold peaks.

12.2.4 Modern manifold synthesis. Dealing with the problem of manifold peaks.

The problem of manifold peaks consists mainly of the appearance of transmission zeros within the pass-band of one of the channels due to the phase recombination produced within the manifold. Remember that the design of these multiplexers of manifold type is equivalent to a problem of simultaneous matching where each of the filters matches the impedance shown by the rest of the multiplexer. However, if said impedance is 0 namely, a short-circuit, matching is not possible.

Although the case of finding a null impedance may seem strange, consider that each of the filters presents a practically uni-modular reflection within the adjacent bands. In addition, due to the phase shift produced by the transmission lines that make up the manifold, it is possible to introduce a virtual short circuit in any of the channels. This is illustrated in fig. 12.6, where the filter in channel 3 shows a short circuit at a frequency f_0 within the band of channel 2. If the transmission line L indicated in the figure takes then the value $\lambda_G/4$ at any frequency belonging either to the band of channel 1 or 3, a short circuit is introduced at the input of the multiplexer at that frequency. Therefore a manifold peaks occurs at the same frequency within the band of the respective channel.

This is an important disadvantage in the synthesis of multiplexers, so much so that it is even difficult to detect until the device has been physically constructed. The reason comes from the fact that, when using transmission lines of too long a

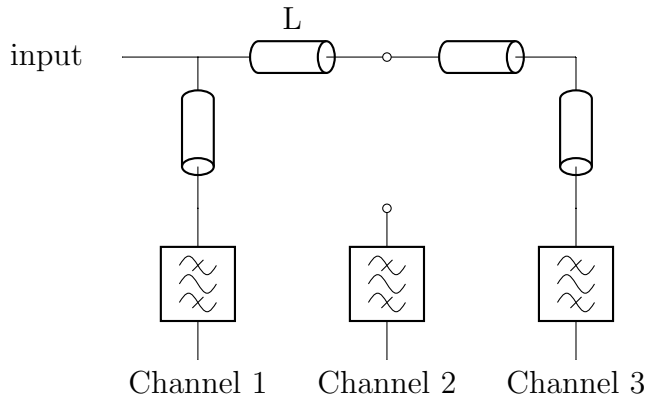


Figure 12.6: Occurrence of a virtual short circuit, namely a manifold peak.

length, the variation speed of the reflection seen by each of the channel filters increases. This causes the said manifold peaks to appear at a very precise frequency, making it difficult to detect them in a simulation if the sampling of the frequency axis is not too large. This is usually the case in practice since in order to perform the simulation, a refined sampling of the frequency axis would slow down the whole design process.

The typical solution to the problem of manifold peaks is to perform a first optimization where these peaks are ignored, then through an EM simulation with sufficient resolution the presence of the manifold manifold peaks in the response is revealed. Once these are localized, the contributions in this regard found in the literature (for instance in [73]) propose the modification of the physical structure to displace the frequency at which said peaks occur and perform a new iteration which consists of a successive optimization. Note that when modifying for example, the coupling topology of one of the filters changes the level of dispersion present in the structure and therefore also the position of said peaks.

References

[54] R. J. Cameron, R. Mansour, and C. M. Kudsia, *Microwave Filters for Communication Systems: Fundamentals, Design and Applications*. Wiley, 2007.

[55] J. D. Rhodes and R. Levy, "A Generalized Multiplexer Theory," *IEEE Transactions on Microwave Theory and Techniques*, 1979.

[56] G. Macchiarella and S. Tamiazzo, "Novel Approach to the Synthesis of Microwave Diplexers," *IEEE MTT-S International Microwave Symposium Digest*, vol. 54, no. 12, pp. 4281–4290, 2006.

[57] ——, "Synthesis of Star-Junction Multiplexers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 12, pp. 3732–3741, 2010.

[58] G. Macchiarella, "Synthesis of Star-Junction Multiplexers," *IEEE Microwave Magazine*, vol. 12, no. 6, pp. 101–109, 2011.

[59] T. Gál, J. Ladvánszky, and F. Lénárt, "Improvement of Waveguide Diplexer Components," in *Asia Pacific Microwave Conference*, Seoul, 2013, pp. 28–30.

[60] J. Li, H. Huang, Z. Zhang, W. Song, H. Shao, C. Chen, and and Wenhua Huang, "A Novel X-Band Diplexer Based on Overmoded Circular Waveguides for High-Power Microwaves," *IEEE Transactions on plasma science*, vol. 41, no. 10, pp. 2724–2728, 2013.

[61] F. Cheng, X. Lin, K. Song, Y. Jiang, and and Yong Fan, "Compact Diplexer With High Isolation Using the Dual-Mode Substrate Integrated Waveguide Resonator," *IEEE Microwave and Wireless Components Letters*, vol. 23, no. 9, pp. 459–461, 2013.

[62] H. Ezzeddine, P. Mazet, S. Bila, and S. Verdeyme, "Design of a Compact Dual-Band Diplexer with Dual-Mode Cavities," in *EuMC, European Microwave Conference*, vol. 42. Amsterdam, The Netherlands: EuMC, 2012, pp. 455–458.

[63] H. Hu, K.-L. Wu, and R. J. Cameron, "Stepped Circular Waveguide Dual-Mode Filters for Broadband Contiguous Multiplexers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 1, pp. 139–145, 2013.

[64] A. Morini, T. Rozzi, and M. Morelli, "New formulae for the initial design in the optimization of T-junction manifold multiplexers," in *IEEE International Microwave Symposium Digest*, Denver, CO, USA, 1997, pp. 1025–1028.

[65] L. Accatino and M. Mongiardo, "Hybrid Circuit-Full-Wave Computer-Aided Design of a Manifold Multiplexers Without Tuning Elements," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 9, pp. 2044–2047, 2002.

[66] M. S. Uhm, J. Lee, J. H. Park, and J. P. Kim, "An Efficient Optimization Design of a Manifold Multiplexer Using an Accurate Equivalent Circuit Model of Coupling Irises of Channel Filters," in *IEEE International Microwave Symposium Digest*, 2005, pp. 1263–1266.

[67] J. R. Montejo-Garai, J. A. Ruiz-Cruz, and J. M. Rebollar, "Full-Wave Design of H-Plane Contiguous Manifold Output Multiplexers Using the Fictitious Reactive Load Concept," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 8, pp. 2625–2632, 2005.

[68] G. Macchiarella and S. Tamiazzo, "Polynomial design of manifold multiplexers," in *IEEE International Microwave Symposium Digest*, Baltimore, MD, 2011, pp. 1–4.

[69] J. D. Rhodes and R. Levy, "Design of General Manifold Multiplexers," *IEEE Transactions on Microwave Theory and Techniques*, 1979.

[70] G. Tanne, S. Toutain, J. Favennec, P. Jarry, and C. Boschet, "Optimal design of contiguous-band output multiplexers (COMUX)," *Electronics Letters*, 2007.

[71] R. J. Cameron and M. Yu, "Design of manifold-coupled multiplexers," *IEEE Microwave Magazine*, 2007.

[72] M. Brumos, S. Cogollos, M. M. Mendoza, P. Soto, V. E. Boria, and M. Guglielmi, "Design of Waveguide Manifold Multiplexers with Dual-Mode Filters Using Distributed Models," in *IEEE International Microwave Symposium Digest*, Tampa Bay, Florida, 2014.

[73] D. Bariant, S. Bila, D. Baillargeat, S. Verdeyme, P. Guillon, D. Pacaud, and J. J. Herren, "Method of spurious mode compensation applied to manifold multiplexer design," *IEEE MTT-S International Microwave Symposium Digest*, 2002.

Chapter 13:

Manifold model and manifold peaks

In this chapter we introduce the first part of the design technique presented in part V of this thesis. The main objective is the synthesis and subsequent design of a multiplexer of the manifold type. As mentioned in the previous chapter, manifold multiplexers offer the greatest advantage in terms of compactness and insertion losses. However, they present the important inconvenience that constitutes the need to perform a very complex and computationally expensive synthesis.

In previous chapters, if the problem of matching comes into play, the decomposition of the global system into two devices, namely a matching filter and a load, has facilitated the task with regards to matching. This holds in both the use of traditional techniques of matching and in the procedure developed in this work, where a series of necessary conditions on the global system are determined solely by one of those two sub-devices. Indeed, one of the main reasons for the manifold multiplexer design complexity comes from the fact that the device to be designed is essentially multi-port, which can not be decomposed into two-port sub-devices only.

When it comes to synthesizing a device with more than two ports, the different paths that the input signal can follow inside the device before recombining at the output represent a quite complex combinatorial problem. Because of the said multipath effect, when this multi-port device is inserted in a larger block, interconnected with other devices, transmission zeros may appear between two of its ports, which were not present in the isolated device. This fact differs fundamentally from the behaviour obtained when two-port are connected in cascade, since in this case the transmission zeros obtained in the global system are necessarily present in at least one of the sub-devices.

In general, multiplexer synthesis involves dealing with multiple spikes that appear within the pass bands. These are transmission zeros at an extremely precise frequency that appear due to phase recombinations within the manifold. In the design of multiplexers, we often face the problem of avoiding these peaks. However, due to the nature of the design techniques traditionally used in the synthesis of multiplexers, the presence of said peaks is revealed thanks to an EM simulation of the complete structure once the design of the multiplexer has been completed. This implies that to solve the problem it is necessary to go back, modifying or restarting the synthesis of the device.

13.1 Multi-port scattering matrices

At this point, some of the concepts introduced in chapter 2 with respect to scattering matrices need to be extended or particularized for the case of multi-port devices.

It is important to note that the theory developed in chapter 2 was particularized for the case of 2×2 scattering matrices without loss of generality. Therefore, the mentioned theory can be extended to the general case of $N \times N$ scattering matrices. However, the concepts introduced in chapter 2 need to be adapted.

The main difference in the synthesis of multi-port scattering matrices with respect to the traditional synthesis of 2×2 matrices is the absence of a parametrization equivalent

to the Belevitch form. Indeed when it comes to synthesizing a multi-port device, no Belevitch form exists. The absence of a Belevitch form entails the absence of a Darlington equivalent that allows the parametrization of this $N \times N$ scattering matrix by means of a smaller number of rational Schur functions. This is one of the main problems we are going to face in this part of the thesis.

Finally we find the concept of transmission zeros that, although it does not directly apply as defined in chapter 2, can be extended to the case of scattering matrices of size $N \times N$.

13.1.1 Transmission zeros

Next, we provide an extension of the concept of transmission zeros for the case of multi-port devices. In this case the concept of transmission zero needs to be associated with a pair of input-output ports to the device.

In order to simplify the subsequent development, we assume that all multi-port scattering matrices S appearing in this chapter are reciprocal. This implies from definition 2.5.2 that

$$S(\lambda) = S(\lambda)^T \qquad \forall \lambda \in \mathbb{C}.$$

Therefore if the parameter $S_{i,j}$ vanishes at a point $\lambda \in \mathbb{C}$, then the parameter $S_{j,i}$ also vanishes at that point. Note in this case the reciprocity assumption is completely legit since if a non-reciprocal junction is allowed, then we can just avoid the problem of the interaction between channels by means of such non-reciprocal junction. This is indeed the case of the circulator-coupled multiplexer as the circulator structure represents a non-reciprocal manifold.

Remark 13.1.1. *It is important to note the numbering used for multi-port devices in this chapter. Unlike the classic numbering from* 1 *to* N *for a device with* N *ports, we have decided to use a 0-based numbering. Therefore, the rows and columns of a scattering array of size* $N \times N$ *are numbered from 0 to* $N-1$*. This numbering allows us to associate the port* i *of the junction or manifold to the* i*-th channel while port 0 corresponds to the common port of the multiplexer.*

We state the following definition for transmission zeros.

Definition 13.1.1 (Transmission zeros of multi-port devices)**.** *Let denote by transmission zeros between the ports* i, j *of the scattering matrix* S *the set of points* $\mathbb{O}_i^S(\mathbb{K})$ *with* $\Omega \subset \overline{\mathbb{C}^-}$ *defined as*

$$\mathbb{O}_i^S(\mathbb{K}) = \{\lambda \in \mathbb{K} \mid S_{0,i}(\lambda) S_{i,0}(\lambda) = 0\}.$$

The approach presented in this chapters is developed around the transmission zeros as in the second part of this work. With definition 13.1.1 we generalise the definition of transmission zeros provided in chapter 2 such that definition 13.1.1 with $i = 1$ and $\overline{\mathbb{C}^-} \subset \mathbb{K}$ corresponds to the original definition. Nevertheless, although the definition is general, we are only interested in this chapter on the transmission zeros occurring on the real

line. Additionally it should be noted in this case the concept of multiplicity is not required.

In the next section, we carry out a study of the necessary and sufficient conditions for the apparition of a transmission zero $\omega \in \mathbb{O}_i^L(\mathbb{K})$ with $\mathbb{K} \subset \overline{\mathbb{C}^-}$ in a sub-device L when it is inserted into a larger block. This study is necessary for the correct understanding of the notion of transmission zeros associated with multi-port devices. In order to get a first perspective of the problem, let us start with the simplest case, namely a two-channel multiplexer.

13.2 Manifold peaks in duplexer synthesis

We address now the problem of manifold peaks in a three port-device, our objective is to understand why such peaks appear, allowing us to develop an analytic strategy to avoid them. Consider then the duplexer schematic shown in fig. 13.1. In the figure we can see the two channel filters (Filter 1 and Filter 2) interconnected by a three port junction. The junction has been represented by the T-shaped sub-block in fig. 13.1.

Next we will apply a similar approach to that developed in chapter 4 in the study of matching between a matching filter and a two-port load. The objective therefore is to obtain a decomposition of the duplexer in different two-port sub-devices. However, as previously mentioned, it is not possible to decompose a multi-port device into two-port sub-devices only. This decomposition will be possible, however, if we authorize the use of three-port devices as well. Therefore, we will use the grouping indicated in fig. 13.1 where the Filter 1 of the duplexer has been connected to port 1 of a three-port device, composed of the junction together with the second filter. This decomposition serves us in the following section to determine in what positions of the frequency axis, and under what conditions we can find a transmission zero $\omega \in \mathbb{O}_i^L(\mathbb{K})$ with $\mathbb{K} \subset \overline{\mathbb{C}^-}$.

13.2.1 Channel filters and passbands

We consider the duplexer with two channel filters which are shown in fig. 13.1. We denote by $F^{(i)}$ the scattering matrix of Filter i, with $i \in [1, 2]$. Similarly we denote by $\mathbb{I}_1$ the passband of filter 1, and by $\mathbb{I}_2$ the passband of the second filter. Both $\mathbb{I}_1$ and $\mathbb{I}_2$ are composed of a finite union of compact real intervals $\mathbb{I}_1, \mathbb{I}_2 \subset \mathbb{R}$ as it was the case of the matching filter passband in previous chapters. Additionally both passbands do not intersect, namely $\mathbb{I}_1 \cap \mathbb{I}_2 = \emptyset$. We denote further by S the scattering matrix of the global duplexer in fig. 13.1. Note that with the provided definitions, the manifold peaks in the duplexer consists on the transmission zeros $\mathbb{O}_1^S(\mathbb{I}_1)$ and $\mathbb{O}_2^S(\mathbb{I}_2)$.

The aim of this chapter is not to synthesize the mentioned channel filters, but to determine the possible manifold peaks (transmission zeros) that could appear in the global duplexer when all the subsystems, namely the junction and the channel filters are assembled. For this reason, the channel filters are assumed to be fixed for the time being and the global system obtained from the chaining expression of the three sub-devices will be studied.

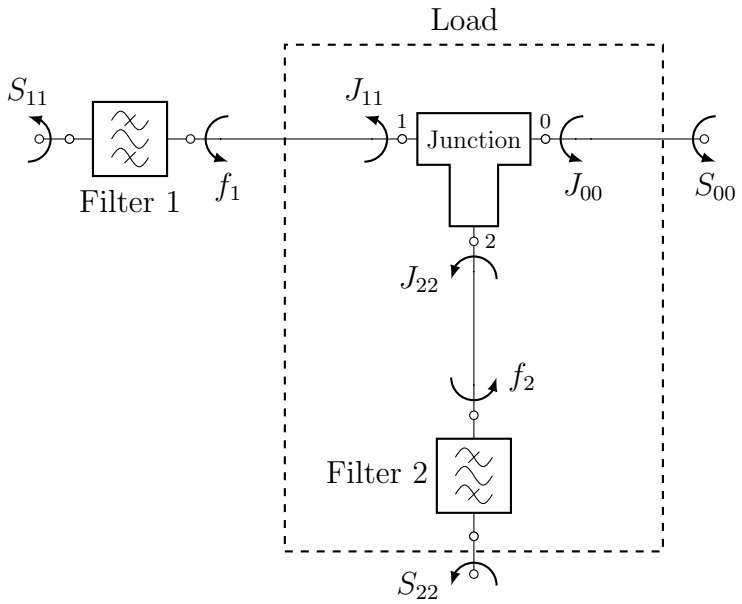


Figure 13.1: Diplexer schematic

We consider that the Filter i has no transmission zeros inside the interval $\mathbb{I}_i$ as otherwise a transmission zero $\alpha \in \mathbb{O}_i^S(\mathbb{I}_i)$ would be trivially present in the global system at the frequency $\alpha \in \mathbb{I}_i$ where the filter has a transmission zero. We have

$$|F_{22}^{(i)}(\omega)| = |F_{11}^{(i)}(\omega)| \leq 1 \qquad \forall \omega \in \mathbb{I}_i,$$
$$|F_{12}^{(i)}(\omega)| = |F_{21}^{(i)}(\omega)| \neq 0 \qquad \forall \omega \in \mathbb{I}_i.$$

Additionally we assume that both filters are reciprocal, namely

$$F_{21}^{(i)}(\omega) = F_{12}^{(i)}(\omega) \qquad \forall \omega \in \mathbb{R}.$$

13.2.2 Two-port chaining

We determine next the expression of the scattering matrix of the global device in fig. 13.1 as a function of the scattering matrix of Filter 1 and the scattering matrix of the Load. Note that the chaining expression introduced in chapter 2 does not apply to this case since the Load is a 3-port device. Therefore let us suppose now that Filter 2 is connected to a matched load as in fig. 13.2. We obtain in this way the cascade of filter with a load, both two-port devices. We denote by L the 2×2 scattering matrix of the load in fig. 13.2. We are looking for the transmission zeros of the system $S = F^{(1)} \circ L$ namely $\mathbb{O}_1^S(\mathbb{I}_1)$. Using now eq. (3.3) we obtain the 2×2 scattering matrix resulting of the cascade operation $F^{(1)} \circ L$.

$$S = F^{(1)} \circ L = \begin{pmatrix} F_{11}^{(1)} + \dfrac{F_{12}^{(1)} L_{11} F_{21}^{(1)}}{1 - F_{22}^{(1)} L_{11}} & \dfrac{F_{12}^{(1)} L_{12}}{1 - F_{22}^{(1)} L_{11}} \\ \dfrac{L_{21} F_{21}^{(1)}}{1 - F_{22}^{(1)} L_{11}} & L_{22} + \dfrac{L_{21} F_{22}^{(1)} L_{12}}{1 - F_{22}^{(1)} L_{11}} \end{pmatrix},$$

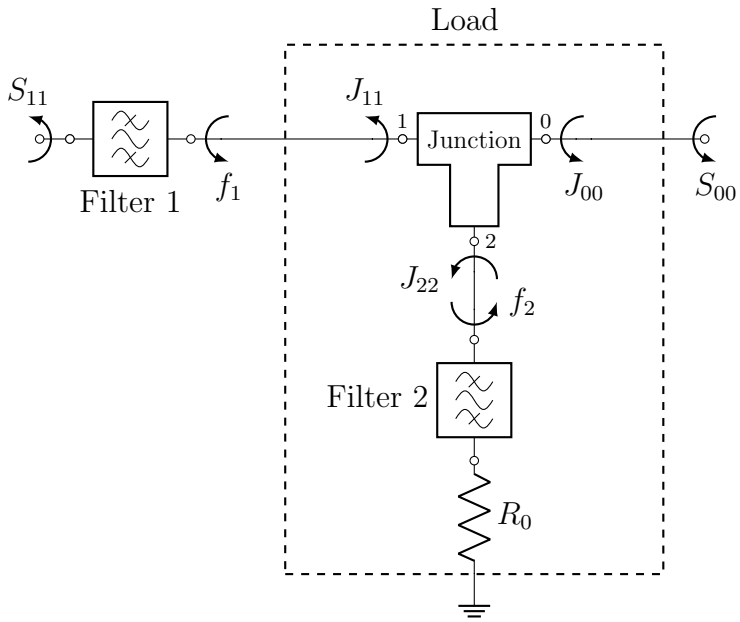


Figure 13.2: Diplexer seen as the chaining of two-port sub-devices

where

$$S_{12} = S_{21} = \frac{L_{21}F_{21}^{(1)}}{1 - F_{22}^{(1)}L_{11}}. \tag{13.1}$$

Note that $L_{11} \in \mathbb{\Sigma}$ then $|L_{11}(\omega)| \leq 1$ for all $\omega \in \mathbb{R}$. Therefore the denominator of eq. (13.1) cannot vanish in $\mathbb{I}_1$. Hence since $F_{21}^{(1)}(\omega)$ has no zeros in the interval $\mathbb{I}_1$, we conclude that $\mathbb{O}_1^S(\mathbb{I}_1)$ contains only the zeros $\mathbb{O}_1^L(\mathbb{I}_1)$ of the load.

$$\mathbb{O}_1^S(\mathbb{I}_1) \subset \mathbb{O}_1^L(\mathbb{I}_1).$$

Next we have to determine when the set $\mathbb{O}_1^L(\mathbb{I}_1)$ is not empty. This time the load L can not be decomposed in two sub-devices of two-ports since the filter 2 has to be connected to the port 2 of the junction and since we need to determine the zeros of transmission from port 0 to 1 of L, we can not close any of these ports 0 or 1. Thus we need to introduce now a generalised formula for the chaining of multi-port devices.

13.2.3 Chaining of multiport scattering matrices

In this section we consider the $N+1$-ports device with $(N+1) \times (N+1)$ scattering matrix J and the one port reflections $[f_n, f_{n+1}, \cdots, f_N] \in \mathbb{\Sigma}$. Each of these reflection represents the reflection used to close the respective port of the junction as indicated in fig. 13.3. Now consider that ports $[n, N]$ of the junction are closed by the reflections $[f_n, f_{n+1}, \cdots, f_N]$. As a result a $n \times n$ scattering matrix L is obtained. This operation is equivalent to the scalar chaining defined in eq. (3.2) when the load is a multi-port device. We provide then a generalised definition of scalar chaining.

Definition 13.2.1 (Scalar chaining onto a multi-port)**.** *Define* $T^{(n)}$ *as the diagonal matrix*

$$T^{(n)} = \begin{bmatrix} f_n & & \\ & \ddots & \\ & & f_N \end{bmatrix} \qquad n \leq N,$$

where $f_i \in \mathbb{\Sigma}$ *for all* $i \in [n, N]$*. Additionally partition the matrix* J*, which corresponds to the junction, in 4 sub-matrices as*

$$J = \begin{bmatrix} J_{1,1}^{(n)} & J_{1,2}^{(n)} \\ J_{2,1}^{(n)} & J_{2,2}^{(n)} \end{bmatrix} = \left[\begin{array}{ccc|ccc} J_{0,0} & \cdots & J_{0,n} & \cdots & J_{0,N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline J_{n,0} & \cdots & J_{n,n} & \cdots & J_{n,N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ J_{N,0} & \cdots & J_{N,n} & \cdots & J_{N,N} \end{array}\right].$$

We denote by $L = T^{(n)} \circ J$ *the* $n \times n$ *matrix resulting of closing the port* i *of* J *with the reflection* f_i *with* $i \in [n, N]$ *as shown in fig. 13.3. This matrix* L *takes the expression*

$$L = T^{(n)} \circ J = J_{1,1}^{(n)} + J_{1,2}^{(n)} \left(I^{(n)} - T^{(n)} J_{2,2}^{(n)}\right)^{-1} T^{(n)} J_{2,1}^{(n)}, \tag{13.2}$$

where $I^{(n)}$ *stands for the* $(N - n + 1) \times (N - n + 1)$ *identity matrix.*

Note the similarity between eq. (3.2) and eq. (13.2). Indeed if J is a 2×2 matrix and $T^{(2)} \in \mathbb{\Sigma}$ is a scalar function, then eq. (3.2) is obtained. It is important to note that, as in the case of scalar in which the chaining formula can degenerate if the two devices have a common transmission zero, in this case, zero-pole simplifications can also occur. Particularly the denominator of eq. (13.2) vanish if and only if the matrix $I^{(n)} - T^{(n)} J_{2,2}^{(n)}$ is singular. Nevertheless we overcome this issue as it will become clear later on, with the assumption that the $J_{2,2}^{(n)}$ is strictly passive, namely

$$J_{2,2}^{(n)\star}(\omega) J_{2,2}^{(n)}(\omega) \prec I^{(n)} \qquad \forall \omega \in \mathbb{R} \tag{13.3}$$

It is important to remark that this assumption does not imply that the junction J is lossy such as $J \prec I$. However it does imply $|J_{i,i}(\omega)| \leq 1$ with $n \leq i \leq N$.

13.2.4 Transmission Zeros in 3-Port Devices

We use now eq. (13.2) to determine how the manifold peaks, namely transmission zeros from port 0 to 1 of the load in fig. 13.2, are produced. To do so it is important to understand when the 3-port device J presents a transmission zero $\alpha \in \mathbb{O}_1^J(\mathbb{I}_1)$ when terminal 2 is closed by a reflection $f_2 \in \mathbb{\Sigma}$.

Let $J(\lambda)$ be the scattering matrix of the 3-port junction in fig. 13.2. If port 2 of J is closed by the reflection coefficient $f_2(\lambda) \in \mathbb{\Sigma}$ applying eq. (13.2) to compute $T^{(2)} \circ J$ with $T^{(2)} = f_2$, the following 2×2 matrix is obtained

$$\begin{aligned} L = f_2 \circ J &= \begin{bmatrix} J_{00} & J_{01} \\ J_{10} & J_{11} \end{bmatrix} + \begin{bmatrix} J_{0,1} & J_{0,2} \end{bmatrix} \frac{f_2}{1 - f_2 J_{22}} \begin{bmatrix} J_{1,0} \\ J_{2,0} \end{bmatrix} \\ &= \begin{bmatrix} J_{00} & J_{01} \\ J_{10} & J_{11} \end{bmatrix} + \begin{bmatrix} J_{02}J_{20} & J_{02}J_{21} \\ J_{12}J_{20} & J_{12}J_{21} \end{bmatrix} \frac{f_2}{1 - f_2 J_{22}}. \end{aligned} \tag{13.4}$$

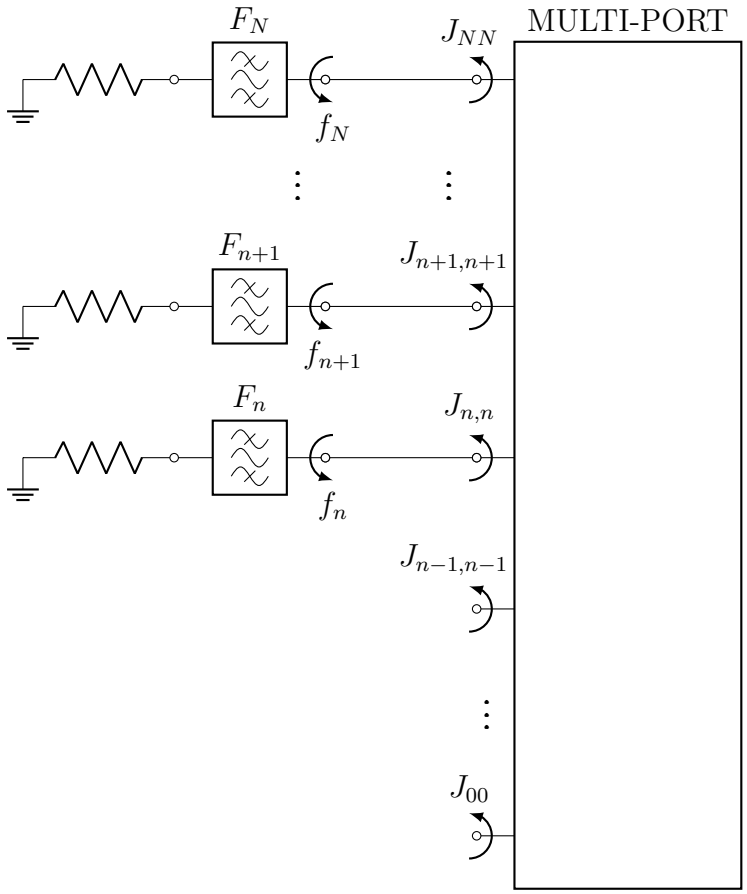


Figure 13.3: Schematic of channels filters connected to the junction.

From eq. (13.3) we have $|J_{22}(\omega)| \leq 1$ for all $\omega \in \mathbb{R}$. Hence the denominator of eq. (13.4) can not vanish on the real line, in particular on the interval $\mathbb{I}_1$. Then we compute the value of f_2 such that the previous matrix has a transmission zero at $\omega \in \mathbb{O}_1^L(\mathbb{I}_1)$. Assuming the matrix J is reciprocal, we seek the transmission zero by equating the element $L_{01}(\omega)$ to zero

$$\begin{aligned} J_{01}(\omega) + \frac{J_{02}(\omega)J_{21}(\omega)f_2(\omega)}{1 - f_2(\omega)J_{22}(\omega)} = 0 \qquad & \omega \in \mathbb{I}_1, \\ J_{01}(\omega) + f_2(\omega)(J_{02}(\omega)J_{21}(\omega) - J_{01}(\omega)J_{22}(\omega)) = 0 \qquad & \omega \in \mathbb{I}_1, \\ \frac{-J_{01}(\omega)}{J_{02}(\omega)J_{21}(\omega) - J_{01}(\omega)J_{22}(\omega)} = f_2(\omega) \qquad & \omega \in \mathbb{I}_1. \end{aligned}$$

Note that the previous denominator is the $(1,0)$ element of the cofactor matrix of J. The cofactor matrix takes the expression

$$C = \begin{bmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,N} \\ C_{2,1} & C_{2,2} & \cdots & C_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N,1} & C_{N,2} & \cdots & C_{N,N} \end{bmatrix}$$

where $C_{i,k} = (-1)^{i+k} \det\left(\mathrm{sub}_{i,k}(J)\right)$ and $\mathrm{sub}_{i,k}(J)$ is the sub-matrix of S where the i-th row and the k-th column are removed. Thus

$$f_2(\omega) = \frac{-J_{01}(\omega)}{-\det\begin{pmatrix} J_{01}(\omega) & J_{02}(\omega) \\ J_{21}(\omega) & J_{22}(\omega) \end{pmatrix}} = \frac{-J_{01}(\omega)}{C_{10}(\omega)} \qquad \forall \omega \in \mathbb{I}_1.$$

If we express now the inverse matrix of J by means of the cofactor matrix we have

$$J^{-1} = \frac{C^T}{\det J}$$

We have $C_{10} = [J^{-1}]_{01} \det(J)$, therefore

$$f_2(\omega) = \frac{-J_{01}(\omega)}{[J^{-1}(\omega)]_{01} \det(J(\omega))} \qquad \omega \in \mathbb{I}_1. \tag{13.5}$$

Finally, if we consider the matrix J to be lossless, this implies $J^{-1} = J^\star$ and together with the reprocity we have for all $\omega \in \mathbb{R}$, $J^{-1}(\omega) = \overline{J(\omega)}$. Then

$$f_2(\omega) = \frac{-J_{01}(\omega)}{\overline{J_{10}(\omega)} \det(J(\omega))} \qquad \forall \omega \in \mathbb{I}_1.$$

Note the previous expression is well defined if $J_{01}(\omega) \neq 0$. Conversely if $J_{01}(\omega) = 0$ at any frequency $\omega \in \mathbb{I}$ we have

$$\begin{aligned} L_{01}(\omega) = \frac{J_{02}(\omega)J_{21}(\omega)}{1 - f_2(\omega)J_{22}(\omega)} = 0 \qquad & \forall \omega \in \mathbb{I}_1, \\ J_{02}(\omega)J_{21}(\omega)f_2(\omega) = 0 \qquad & \forall \omega \in \mathbb{I}_1, \end{aligned}$$

what, assuming $J_{02}(\omega), J_{21}(\omega)$ do not vanish at the same point as $J_{01}(\omega)$, provides us the result $f_2(\omega) = 0$. Let consider the case where $J_{01}(\omega) \neq 0$ for all $\omega \in \mathbb{I}_1$ as it would be the general case for an arbitrary junction. Then we state the following theorem whose proof has been already done in this section.

Theorem 13.2.1 (Transmission zeros of 3-port devices)**.** *A 3-port device with ports* $\{0, i, k\}$ *and* 3×3 *scattering matrix* J*, present a transmission zero* $\alpha \in \mathbb{O}_i^J(\mathbb{I})$ *if and only if port* k *of* J *is closed by the impedance with reflection coefficient* $f_k(\omega)$ *where*

$$f_k(\alpha) = \frac{-J_{0,i}(\alpha)}{\overline{J_{0,i}(\alpha)} \det J(\alpha)} \qquad \alpha \in \mathbb{I}. \tag{13.6}$$

Remark 13.2.1. *Note that we made the assumption that* J *is a lossless scattering matrix. This assumption does not influence the nature of the obtained result, as if the load is not lossless, then eq.* (13.5) *can be used instead of eq.* (13.6)*. However eq.* (13.6) *indicates that if the junction is lossless, then port 2 of* J *has to be closed necessarily by a lossless impedance in order for a transmission zero* $\alpha \in \mathbb{O}_1^J(\mathbb{I}_1)$ *to appear.*

From a physical point of view, reflection f_2 introduces a virtual short-circuit in the 3-port device. Additionally, note that the fact that the junction was assumed lossless, implies the channel filters must be lossless as well for a manifold peak to occurs and the reflection in eq. (13.6) is uni-modular

$$f_k(\alpha) = \frac{e^{j\,(\pi + 2\arg J_{0,i}(\alpha))}}{\det J(\alpha)} \qquad \alpha \in \mathbb{I}_k.$$

Additionally a transmission zero of the k-th filter should be present at the point α. This phenomenon, however, never happens in an exact way in theory unless the user places a transmission zero of the k-th filter on a particular frequency of the band $\mathbb{I}_i$ with $i \neq k$. This is one of the reasons why it is so difficult to detect the dreaded manifold peaks during the circuital optimization stage since in ideal conditions the appearance of the mentioned peaks is very limited. However, in practice, when we try to address the implementation of these devices, we can find a situation extremely close to the aforementioned scenario.

For example, in the case where the k-th filter presents a reflection f_k whose zeros are sufficiently far from the band $\mathbb{I}_i$ which also verifies $\lim_{\omega\to\infty} |f_k(\omega)| = 1$. In this case the reflection $f_k(\omega)$ in the band $\omega \in \mathbb{I}_i$ will present an absolute value arbitrarily close to the unit (depending on the distance between the band $\mathbb{I}_i$ and the closest point in frequency where f_k is cancelled). In this scenario, it only remains that the phase condition on f_k indicated in 12.4 is satisfied at a frequency $\alpha \in \mathbb{I}_1$ to cause this manifold peak at the frequency α. Nevertheless, if the k-th filter is connected to the junction or manifold by a dispersive transmission line, the phase of $f_k(\omega)$ will present a monotonous variation in frequency with a periodicity in frequency equivalent to a wavelength. Therefore there will be a frequency $\alpha \in \mathbb{R}$ at which the manifold peak will occur, repeating itself at each wavelength. In the extreme case in which the band $\mathbb{I}_i$ covers a bandwidth corresponding to at least one wavelength, a manifold peak will inevitably appear at at least one frequency $\alpha \in \mathbb{I}_i$.

13.2.5 Transmission lines

We come back now to the duplexer in fig. 13.1. To gain control over the location of the points $\omega \in \mathbb{O}_i^S(\mathbb{I})$, where $k \in [1, 2]$ it is required to play with the phase reflection

coefficients $f_k(\omega)$ on the interval $\mathbb{I}$. In order to obtain some control over this phase, two transmission lines are added in the interconnection between each of the filters and the junction. In this way we obtain the scheme shown in fig. 13.4.

In this chapter we assume that transmission lines have a transmission whose phase is exponentially decaying. We assume further that this transmission is uni-modular. Additionally let us suppose that filter k presents an uni-modular reflection in every band $\mathbb{I}_i$ with $i \neq k$ respecting the condition of uni-modularity over the reflection f_k seen from each of the terminals of the junction.

$$|F_{22}^{(k)}(\omega)| = 1 \qquad \forall\omega \in \mathbb{I}_i \qquad i \neq k.$$

It is important to emphasize that these assumptions are only made with the objective of obtaining a clearer explanation. In practice, it is possible to use any transmission line model, as long as power losses are not introduced into the system, which would imply a reflection f_i of a modulus smaller than 1. Let $\Phi^{(1)}(\omega)$ and $\Phi^{(2)}(\omega)$ represent the scattering matrices of the transmission lines 1 and 2 respectively. We have

$$\Phi^{(1)}(\omega) = \begin{pmatrix} 0 & e^{-j\beta_1\omega} \\ e^{-j\beta_1\omega} & 0 \end{pmatrix}$$
$$\Phi^{(2)}(\omega) = \begin{pmatrix} 0 & e^{-j\beta_2\omega} \\ e^{-j\beta_2\omega} & 0 \end{pmatrix}.$$

Now we apply again the previous reasoning to study the necessary conditions for the appearance of a transmission zero $\alpha \in \mathbb{O}_1^S(\mathbb{I}_1)$. The transmission line 1 essentially modifies the phase of the signal that passes through it and can not therefore introduce a transmission zero in the path between the filter 1 and the junction. Then line 1 does not influence the occurrence of transmission zeros $\omega \in \mathbb{O}_1^S(\mathbb{I}_1)$ which can only be introduced by the load. However, line 2 directly influences the position of the transmission zeros $\alpha \in \mathbb{O}_1^L(\mathbb{I}_1)$ since it modifies the phase of reflection seen from terminal 2 of the junction.

Therefore, transmission lines 1 and 2 can be used to control the position of zeros $\alpha \in \mathbb{O}_1^S(\mathbb{I}_1)$ and $\alpha \in \mathbb{O}_2^S(\mathbb{I}_2)$ independently since the phase of reflections f_1 and f_2 depends directly on these transmission lines as

$$f_1(\beta_1) = F_{22}^{(1)}(\omega)e^{-2j\beta_1\omega} \qquad \forall\omega \in \mathbb{R},$$
$$f_2(\beta_2) = F_{22}^{(2)}(\omega)e^{-2j\beta_2\omega} \qquad \forall\omega \in \mathbb{R}.$$

Particularly, the procedure to be carried out consists of two stages

- Find β_2 corresponding to the transmission line 2 so that

$$F_{22}^{(2)}(\omega)e^{-2j\beta_2\omega} \neq \frac{-J_{0,1}(\omega)}{\overline{J_{0,1}(\omega)}\det J(\omega)} \qquad \forall\omega \in \mathbb{I}_1,$$
$$e^{2j\beta_2\omega} \neq \frac{\overline{J_{0,1}(\omega)}\det J(\omega)F_{22}^{(2)}(\omega)}{-J_{0,1}(\omega)} \qquad \forall\omega \in \mathbb{I}_1.$$

  Therefore if we take for instance the principal determination for the logarithm we have the following constraint on the value β_2

$$j\beta_2 \neq \frac{1}{2\omega}\log\left(\frac{\overline{J_{0,1}(\omega)}\det J(\omega)F_{22}^{(2)}(\omega)}{-J_{0,1}(\omega)}\right) \qquad \forall\omega \in \mathbb{I}_1.$$

- Find β_1 corresponding to the transmission line 1 so that

$$F_{22}^{(1)} e^{-2j\beta_1\omega} \neq \frac{-J_{0,2}(\omega)}{\overline{J_{0,2}(\omega)} \det J(\omega)} \qquad \forall \omega \in \mathbb{I}_2$$

$$e^{2j\beta_1\omega} \neq \frac{\overline{J_{0,2}(\omega)} \det J(\omega) F_{22}^{(1)}(\omega)}{-J_{0,2}(\omega)} \qquad \forall \omega \in \mathbb{I}_2.$$

Similarly to the previous case we have now

$$j\beta_1 \neq \frac{1}{2\omega} \log \left(\frac{\overline{J_{0,2}(\omega)} \det J(\omega) F_{22}^{(1)}(\omega)}{-J_{0,2}(\omega)} \right) \qquad \forall \omega \in \mathbb{I}_2.$$

Manifold peaks are therefore avoided as long as we pick the values of β_1, β_2 such that previous equation are satisfied. We state this result with the following theorem.

Theorem 13.2.2 (Transmission lines value)**.** *Given the structure shown in fig. 13.4 and let J represent the scattering matrix of the junction with port numbering $\{0, i, k\}$. Denote by $\mathbb{I}_i$ the passband of the i-th channel and by $\Phi^{(k)}$ the scattering matrix of the k-th transmission line*

$$\Phi^{(k)}(\omega) = \begin{pmatrix} 0 & e^{-j\beta_k\omega} \\ e^{-j\beta_k\omega} & 0 \end{pmatrix}.$$

Manifold peaks are avoided in channel i if and only if $e^{j\beta_k} \in \mathbb{T} \setminus \Psi_k$ where $\Psi_k \in \mathbb{T}$ is the image set of the application $G_k : \mathbb{I}_k \longrightarrow \Psi_k$ defined as

$$G_k(\omega) = \left(\frac{\overline{J_{0,i}(\omega)} \det J(\omega) F_{22}^{(k)}(\omega)}{-J_{0,i}(\omega)} \right)^{\frac{1}{2\omega}} \qquad \forall \omega \in \mathbb{I}_i.$$

Remark 13.2.2. *It is important to note that $e^{j\beta_k}$ is a constant function in the range $\mathbb{I}_i$. So if G_k is a surjective application from $\mathbb{I}_i$ onto Ψ_k, such that $\mathbb{T} \subset \Psi_k$ then $\mathbb{T} \setminus \Psi_k$ is the empty set. This means that there is no value of β_k with which the appearance of manifold peaks in the i-th band is avoided.*

This values β_k implicitly depends on the length of the k-th transmission line allowing for the determination of the transmission lines lengths according to the technology used to implements such lines. Additionally, we supposed in this section that $\beta_k(l_k)$ is not frequency dependent. However if the transmission line model involves a frequency dependency $\beta_k(l_k, \omega)$, then the condition to avoid manifold peaks in channel i becomes

$$e^{j\beta(\omega)} \neq G_k(\omega) \qquad \forall \omega \in \mathbb{I}_i.$$

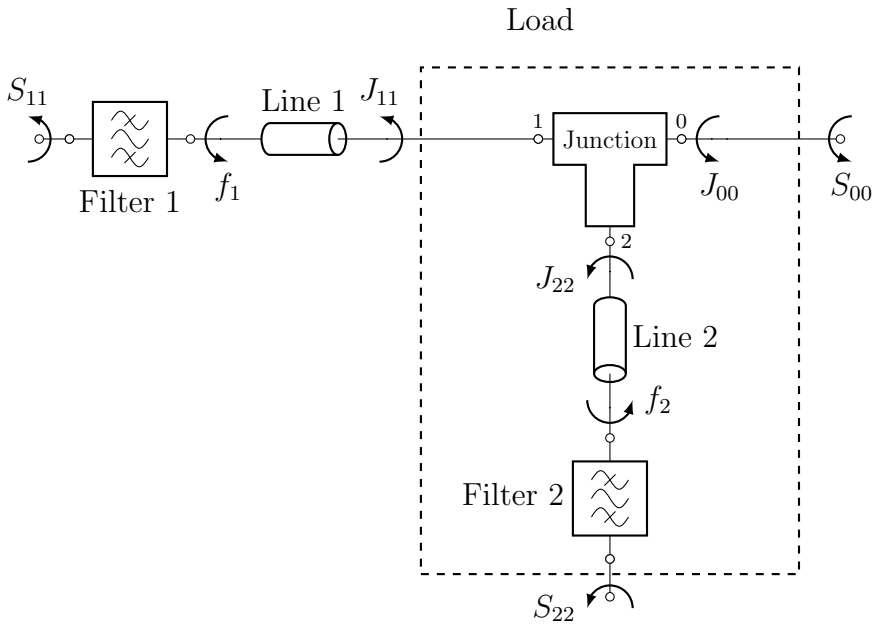


Figure 13.4: Diplexer schematic with transmission lines

13.3 Manifold model

After the intense study carried out in the case of a duplexer on how to avoid manifold peaks by appropriately choosing the values of the transmission lines, we must apply this study to the general case of a junction (manifold) of $(N+1)$-ports.

While in the case of a duplexer the task of determining the lengths of all the transmission lines in such a way that no manifold peak appears in any of the pass bands is relatively simple, in the general case of a junction of N-ports, this task becomes an extremely complicated combinatorial problem. In this case, to avoid the manifold peaks in the i-th channel it is not enough to adjust one of the transmission lines, but it would be necessary to determine all the transmission lines of each of the j-th channels with $j \neq i$.

To overcome such problem we use a *divide and conquer* strategy. This strategy consists on dividing the manifold in a given number of sub-devices each of them equivalent to the duplexer shown in fig. 13.4. With the mentioned sub-division, we obtain the model shown in fig. 13.5 representing a N-channel multiplexer. This model is based of several 3-port devices, denoted hereinafter by elements J_i, interconnected by transmission lines.

Note that, similarly to the duplexer case, transmission zeros from channel ports to the common port can not be introduced by the transmission lines since they only modify the phase of the signal. Therefore transmission zeros happen inside the J elements.

The schematic model in fig. 13.5 is composed of $N-1$ sections equivalent to the structure in fig. 13.4. Each of these sections contains a 3-port junction whose scattering matrix is denoted by $J^{(k)}$ with $2 \leq k \leq N$. Ports 1 and 2 of each junction $J^{(k)}$ is connected to a transmission line denoted by $\Phi_1^{(k)}$ and $\Phi_2^{(k)}$ respectively. The scattering matrices of

the transmission lines $\Phi_i^{(k)}$ with $i \in \{1,2\}$ are defined as

$$\Phi_i^{(k)} = \begin{pmatrix} 0 & e^{j\beta_i^{(k)}\omega} \\ e^{j\beta_i^{(k)}\omega} & 0 \end{pmatrix} \qquad i \in \{1,2\} \qquad k \in [2,N].$$

We denote by $L_i^{(k)} \in \mathbb{\Sigma}$ the respective reflection shown by the load at the opposite end of each transmission line $\Phi_i^{(k)}$ and by $f_i^{(k)} \in \mathbb{\Sigma}$ the reflection shown at ports i of the k-th junction. Reflection coefficients $L_i^{(k)}(\omega)$ and $f_i^{(k)}(\omega)$ are related as

$$f_i^{(k)}(\omega) = L_i^{(k)}(\omega) \cdot e^{2j\beta_i^{(k)}\omega} \qquad i \in \{1,2\} \qquad k \in [2,N]$$

Finally we denote by S the scattering matrix of the complete multiplexer.

As in the case of the duplexer, we will use the $\mathbb{I}_i$ notation to refer to the passband of the i-th channel. This sub-band $\mathbb{I}_i$ is constituted by a finite union of compact intervals of the real line. In addition, we assume that the intersection between the bands $\mathbb{I}_i \cap \mathbb{I}_k$ is empty for all $i,k \in [1,N]$ and with $i \neq k$.

Remark 13.3.1. *It is important to note that the ordering of the channels in the figure is important and determines the result in terms of the necessary transmission lines. In this example we have chosen to order the channels from left to right from channel 1 to channel N.*

With a different arrangement (for example, a right-to-left arrangement), the result obtained would be different. However, since there is only a finite number of combinations between N channels, it is possible to test all of them and choose among the different options, the one that provides the most satisfactory result.

With the decomposition performed in fig. 13.5 we can apply the same analysis performed in the case of the duplexer in the previous section. However, in this case it is necessary to differentiate between two types of channels depending on the number of interconnections between each channel port to the common port to the multiplexer. In particular, a distinction must be made between main branches and secondary branches. This classification determines the dependency between the optimal lengths for the transmission lines. The values of the transmission lines in the secondary branches depend on the already fixed values of the lines in the main branches. Next, each of these types of branches is analysed, introducing the concept of priority when determining the transmission lines that constitute the manifold so that the manifold peaks are avoided.

13.3.1 Main branch

The main branches are constituted by the channels whose path from the channel filter to the common port of the multiplexer passes through a maximum number of elements $J^{(k)}$. Furthermore, depending on the topology obtained in the multiplexer schematic, there could be 1 or several different main branches. If the obtained scheme is sufficiently complex, then there could even be secondary branches that in turn are sub-divided into one or more main and secondary branches.

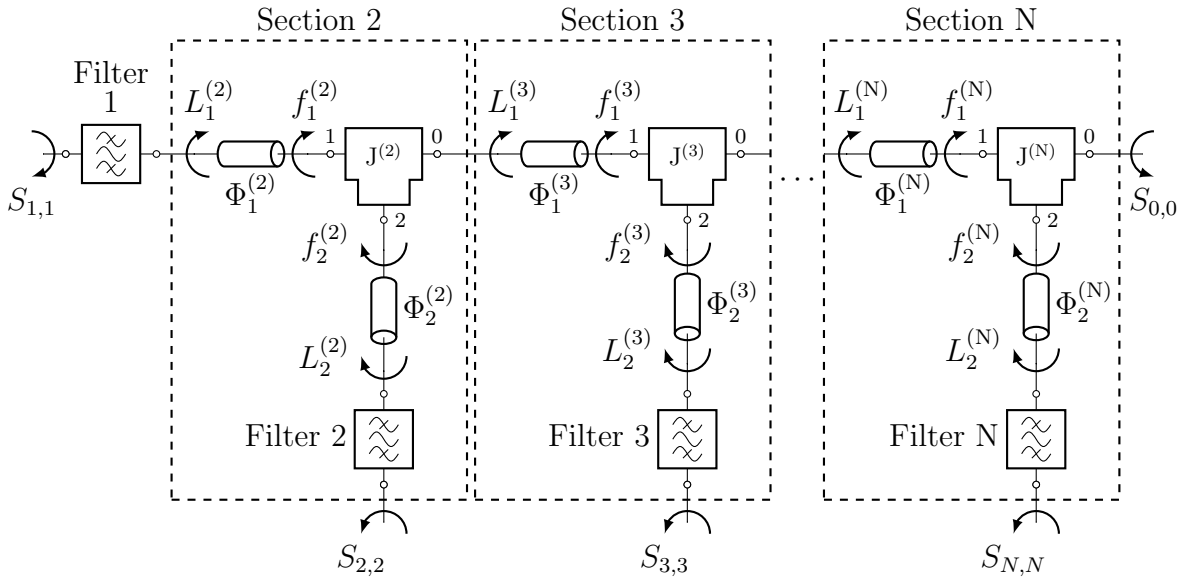


Figure 13.5: Multiplexer schematic

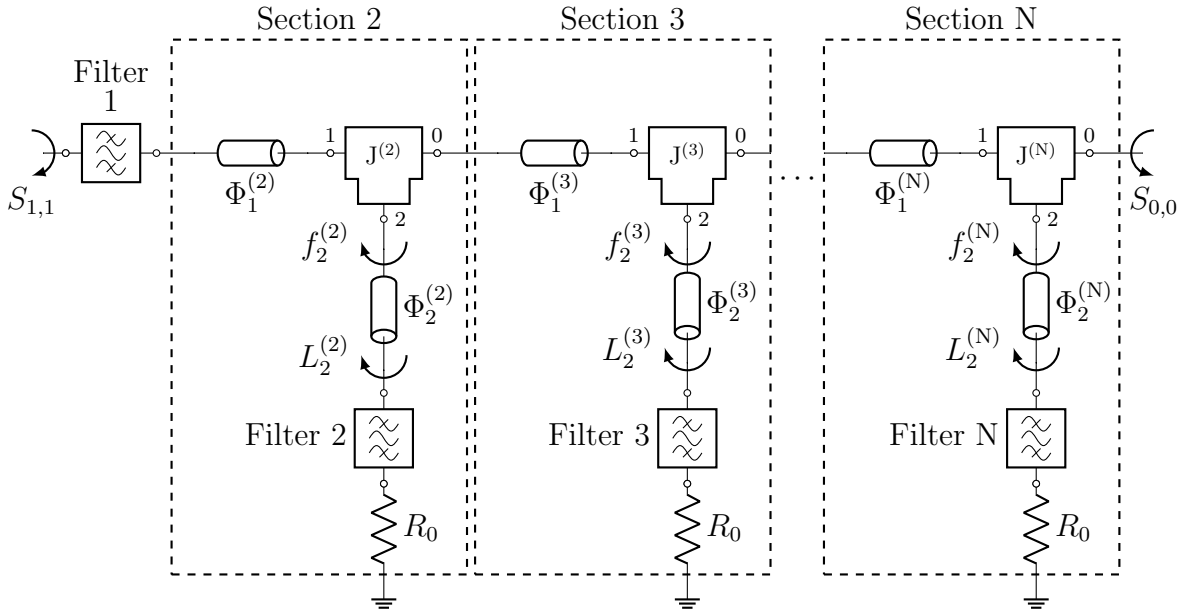


Figure 13.6: Multiplexer schematic

Consider again the structure shown in fig. 13.5. We can see that the road from port 1 to port 0 necessarily crosses all junctions. Therefore channel 1 belongs to a main branch. It is necessary to note that the channel 2 also belong to a main branch since the path from channel 2 to the common port also crosses all the J-junctions. To illustrate the proposed procedure, we choose the main branch as the one corresponding to the first channel. Working with this branch we can not only determine the position of the manifold peaks $\alpha \in \mathbb{O}_1^S(\mathbb{I}_1)$ unequivocally, but also obtain certain necessary conditions (not sufficient) to avoid the appearance of manifold peaks in each one of the remaining channels.

To study the path from terminal 1 to terminal 0, we consider that all other channels are terminated in a matched load, thus obtaining the 2-port device shown in fig. 13.6 constituted by the chaining of N 2-port device. Let us treat first the manifold peaks in channel 1. In fig. 13.6 the global scattering matrix, denoted by S_1 is obtained as the cascade of the matrix of filter 1, denoted here by F_1, and a sequence of two port scattering matrices corresponding to each one of the sections. Let us then denote by Σ_1 the scattering matrix of section k with $2 \leq k \leq N$. We have

$$S_1 = F_1 \circ \Sigma_1 \circ \cdots \circ \Sigma_1.$$

We are now looking for the manifold peaks $\alpha \in \mathbb{O}_1^S(\mathbb{I}_1)$, namely the transmission zeros of the system S_1. As a cascade of two ports devices, the transmission zeros of the system S_1 must be present in at least one of the sub-devices. Thus we have

$$\mathbb{O}_1^S(\mathbb{I}_1) = \mathbb{O}_1^{F_1}(\mathbb{I}_1) \cup \mathbb{O}_1^{\Sigma_1}(\mathbb{I}_1) \cup \cdots \cup \mathbb{O}_1^{\Sigma_N}(\mathbb{I}_1).$$

Note the filter F_1 cannot have transmission zeros inside its own passband $\alpha \in \mathbb{I}_1$. Furthermore, as discussed before, the transmission line $\Phi_1^{(k)}$ can not introduce neither a transmission zero inside the k-th section as it has an unimoduluar transmission coefficient. Hence it only remains to ensure that each J-element does not introduce a transmission zero in the main branch. From theorem 13.2.2 and with the notation defined in this section we have

$$e^{j\beta_2^{(k)}} \in \mathbb{T} \setminus \Psi_2^{(k)}(\mathbb{I}_1),$$

where $\Psi_2^{(k)}(\mathbb{I}_1) \subset \mathbb{T}$ is the image set of the application $G_2^{(k)} : \mathbb{I}_1 \longrightarrow \Psi_2^{(k)}(\mathbb{I}_1)$ defined as

$$G_2^{(k)}(\omega) = \left(\frac{\overline{J_{0,1}^{(k)}(\omega)} \det J^{(k)}(\omega) L_2^{(k)}(\omega)}{-J_{0,1}^{(k)}(\omega)} \right)^{\frac{1}{2\omega}} \qquad \omega \in \mathbb{I}_1.$$

Nevertheless we can now also consider the path from filter 2 to the common port shown in fig. 13.7. It must be noted here that the transmission zeros in band $\mathbb{I}_2$ in sections from 3 to N are are also present in $\mathbb{O}_2^S(\mathbb{I}_2)$ as the only path to the filter 2 goes through sections Σ_k with $3 \leq k \leq N$. We obtain therefore the following necessary condition over $e^{j\beta_2^{(k)}}$

$$e^{j\beta_2^{(k)}} \in \mathbb{T} \setminus \Psi_2^{(k)}(\mathbb{I}_2) \qquad \forall k \in [3, N],$$

with $\Psi_2^{(k)}(\mathbb{I}_2) \subset \mathbb{T}$ the image of the set I_2 under the application $G_2^{(k)} : \mathbb{I}_2 \longrightarrow \Psi_2^{(k)}(\mathbb{I}_2)$ defined as before. Applying now the same argument for the transmission zeros $\mathbb{O}_k^S \mathbb{I}_k$ in

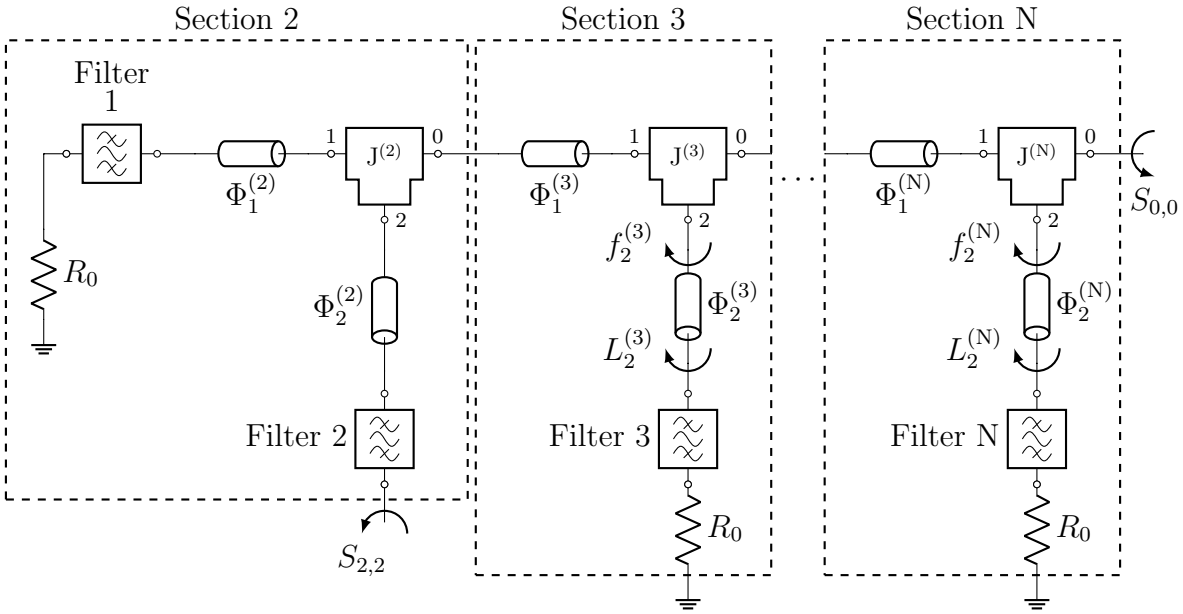


Figure 13.7: Multiplexer schematic

each channel with $2 \leq k \leq N$ we reach a generalised condition over the values $e^{j\beta_2^{(k)}}$ that is necessary to avoid manifold peaks in the k-th channel with k from 2 to N. In the most general case, this condition would be necessary to avoid manifolds peaks in every channel that belong to the main branch under study apart from the channel used to obtain the mentioned branch.

Theorem 13.3.1 (Transmission zeros in a main branch)**.** *Given the schematic in fig. 13.5. The following condition is necessary to avoid manifold peaks in $\mathbb{O}_k^S \mathbb{I}_k$ in every channel with $2 \leq k \leq N$*

$$e^{j\beta_2^{(k)}} \in \mathbb{T} \setminus \Psi_2^{(k)}(\mathbb{K}_k) \qquad \forall k \in [2, N],$$

where $\mathbb{K}_k$ is the union of the passbands of every channel whose path to common port goes through the section k, namely

$$\mathbb{K}_k = \bigcup_{i=1}^{k-1} \mathbb{I}_i$$

and $\Psi_2^{(k)}(\mathbb{K}_k)$ the image set of $\mathbb{K}_k$ under the application $G_2^{(k)}(\omega)$ defined as

$$G_2^{(k)}(\omega) = \left(\frac{\overline{J_{0,1}^{(k)}(\omega)} \det J^{(k)}(\omega) L_2^{(k)}(\omega)}{-J_{0,1}^{(k)}(\omega)} \right)^{\frac{1}{2\omega}} \qquad \forall \omega \in \mathbb{K}_k.$$

Here concludes the analysis of the main branch with the aim of avoiding the manifold peaks. Defining this main branch has given us a set of necessary conditions that must be met by each of the transmission lines which, even without belonging to the main branch in question, are connected to it.

It is important to note that we have not yet said anything about the transmission lines that constitute the main branch under study. In effect, the necessary conditions obtained are independent of the transmission lines belonging to the main branch in question. However, it is necessary that the transmission lines already treated are fixed before moving on to the next stage. This next stage consists in the determination of the remaining transmission lines, which are part of the main branch.

13.3.2 Secondary branches

The last step in the dimensioning of the manifold is the adjustment of the remaining transmission lines, which belong to one of the main branches. To carry out this task it is necessary to define now the concept of secondary branch.

Given the scheme of a multiplexer composed of 3-port sub-devices interconnected by transmission lines similar to the structure shown in fig. 13.5. Now considering the main branch defined in the previous section, formed by the path from filter 1 to the common port.

The secondary branches associated with the aforementioned main branch consist of each of the branches contained entirely in the main branch with the exception of one single transmission line which connect the channel filter to the main branch. Additionally, when secondary branches are considered, the notion of branch length is necessary as secondary branches are sorted in function of its length.

Definition 13.3.1 (Length of a branch)**.** *We define the length of a secondary branch as the minimum number of 3-port elements in the path from the corresponding channel filter to the common port (number 0) of the multiplexer.*

Example 13.3.1 (Secondary branches)**.** *Consider the main branch in fig. 13.5 from filter 1 to the common port. The paths from filter k with $2 \leq k \leq N$ to the common port are secondary branches of length $N - k + 1$.*

The length of each secondary branch determines the ordering in which they have to be considered. Let us now continue with the example treated in this chapter, namely the multiplexer in fig. 13.5. We must consider the k-th secondary branches from $k = 2$ to $k = N$ in that particular order.

1. *Secondary branch 2.* The first to be treated in this section is the path from filter 2 to the common port. We consider now the schematic in fig. 13.8. Note that we have determined in the previous section the transmission lines $\Phi_2^{(k)}$ with $3 \leq k \leq N$ to ensure that the transmission through sections 3 to N is not vanishing in the band of channel 2, namely $\mathbb{I}_2$. We fix now these transmission lines $\Phi_2^{(k)}$. Therefore manifold peaks $\alpha \in \mathbb{O}_2^S(\mathbb{I}_2)$ can only be introduced now by $J^{(2)}$.

$$\mathbb{O}_2^S(\mathbb{I}_2) \subset \mathbb{O}_2^{J^{(2)}}(\mathbb{I}_2).$$

According to theorem 13.2.2, the set of manifold peaks $\mathbb{O}_2^{J^{(2)}}(\mathbb{I}_2)$ is empty if and only if we have $e^{j\beta_1^{(2)}} \in \mathbb{T} \setminus \Psi_1^{(2)}(\mathbb{I}_2)$ being $\Psi_1^{(2)}(\mathbb{I}_2)$ the image of the set $\mathbb{I}_2$ under the

application $G_1^{(2)}$ defined as

$$G_1^{(2)}(\omega) = \left(\frac{\overline{J_{0,2}^{(2)}(\omega)} \det J^{(2)}(\omega) L_1^{(2)}(\omega)}{-J_{0,2}^{(2)}(\omega)} \right)^{\frac{1}{2\omega}} \qquad \forall \omega \in \mathbb{I}_2.$$

2. *Secondary branch 3.* Let us now fix the transmission line $\Phi_1^{(2)}$ to an admissible value according to the previous criterium. We consider the path from filter 3 to port number 0 (the common port) in fig. 13.9. Using the same argument as in the previous case, transmission zeros $\alpha \in \mathbb{O}_3^S(\mathbb{I}_3)$ can only be introduced by the 3-port junction $J^{(3)}$.

   Our objective now is to determine the admissible values for the transmission line $\Phi_1^{(3)}$ connected at port 1 of $J^{(3)}$. It should be noted that the opposite end of the line $\Phi_1^{(3)}$ is now loaded by a fix load with reflection coefficient $L_1^{(3)}$ since every transmission line at the left of $\Phi_1^{(3)}$ has already been fixed. Therefore we have a manifold peak in the band $\mathbb{I}_3$ if and only if $e^{j\beta_1^{(3)}} \in \Psi_1^{(3)}(\mathbb{I}_3)$. This set $\Psi_1^{(3)}(\mathbb{I}_3)$ represents as before the image set of the application $G_1^{(3)} : \mathbb{I}_3 \longrightarrow \Psi_1^{(3)}(\mathbb{I}_3)$ with

$$G_1^{(3)}(\omega) = \left(\frac{\overline{J_{0,2}^{(3)}(\omega)} \det J^{(3)}(\omega) L_1^{(3)}(\omega)}{-J_{0,2}^{(3)}(\omega)} \right)^{\frac{1}{2\omega}} \qquad \forall \omega \in \mathbb{I}_3.$$

3. *Secondary branch n.* Finally, the described procedure is repeated with each of the secondary branches, ordered from highest to lowest length, determining space of admissible values for the corresponding transmission line $\Phi_1^{(n)}$ until reaching the last channel, namely the channel with the shortest path to the common port. In the above discussion here, the last mentioned step consists of determining the transmission line $\Phi_1^{(N)}$ to avoid the transmission zeros $\alpha \in \mathbb{O}_2^{J^{(N)}}(\mathbb{I}_N)$ in the path shown in fig. 13.10 from channel N to the common port through the junction $J^{(N)}$. In the general case, the transmission line $\Phi_1^{(n)}$ is selected according the the following theorem.

   Theorem 13.3.2 (Transmission zeros in secondary branches)**.** *Consider now the schematic in fig. 13.10 where every transmission line $\Phi_2^{(k)}$ with $2 \leq k \leq N$ has been fixed to ensure no transmission zeros appear in the main branch.*

   Assume further that the transmission lines $\Phi_1^{(i)}$ with $2 \leq i < n$ are also fixed. We consider now the section n with $1 \leq n \leq N$. We denote by L_1^n the reflection presented by the load seen from the transmission line $\Phi_1^{(n)}$ at the terminal opposite from $J^{(N)}$. Then manifold peaks $\alpha \in \mathbb{O}_2^S(\mathbb{I}_n)$ are avoided if and only if

$$e^{j\beta_1^{(n)}} \in \mathbb{T} \setminus \Psi_1^{(n)}(\mathbb{I}_n),$$

   where $\Psi_1^{(n)}(\mathbb{I}_n) \subset \mathbb{T}$ represents the image of the passband $\mathbb{I}_n$ under the application

$$G_1^{(n)}(\omega) = \left(\frac{\overline{J_{0,2}^{(n)}(\omega)} \det J^{(n)}(\omega) L_1^{(n)}(\omega)}{-J_{0,2}^{(n)}(\omega)} \right)^{\frac{1}{2\omega}} \qquad \forall \omega \in \mathbb{I}_n.$$

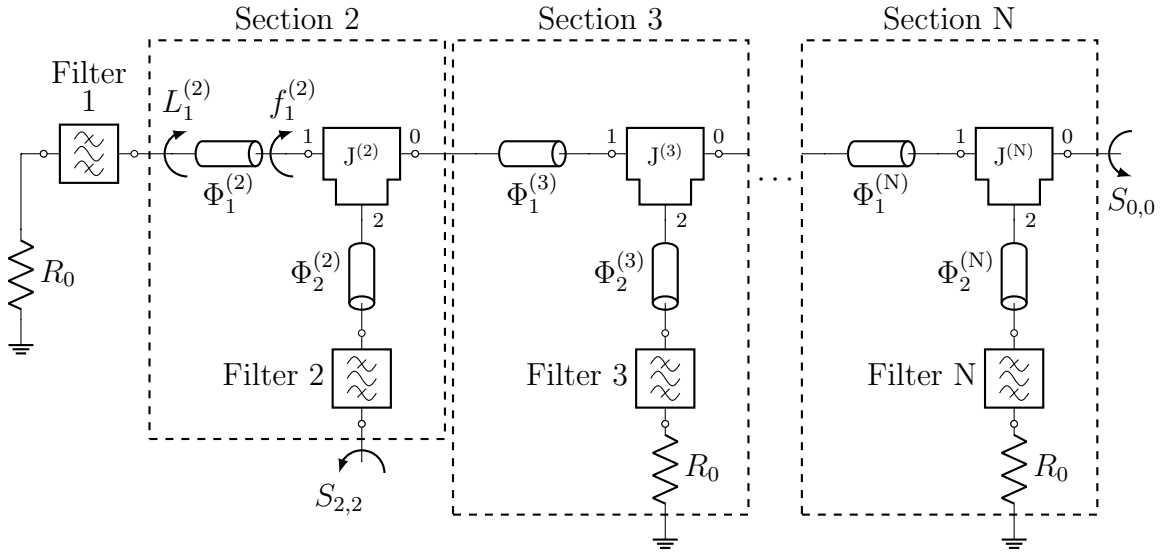


Figure 13.8: Multiplexer schematic

13.4 Concluding remarks

With the analysis of the manifold peaks which can appear in the secondary branches we finish the present chapter in which a systematic procedure has been presented to optimally size the manifold in order to minimize the risk of occurrence of manifold peaks inside of the bands of any of the channels. Note that for the synthesis process presented here to be well determined, it is necessary to define the criterion by which to select the lengths of each of the transmission lines that constitute the manifold among all the admissible lines. This criterion can be, for example, the minimum lengths for which all the manifold peaks present are at a certain safety distance of the pass bands. However, in this work we have considered as a criterion the maximization of the minimum distance of all manifold peaks to the passbands.

Note that due to the periodic character of the manifold peaks in the presence of transmission lines with a frequency dispersive response, the maximization of the minimum distance between the peaks and the bands implies that each of the bands will be centred between two peaks of manifold. Therefore, in the case where the bandwidth of the channels increases, assuming that a solution without manifold peaks within the bands is still possible, we will find peaks at the lower and upper edges of the bands.

Before finishing the present chapter it is interesting to note that the analysis of the manifold presented here is done once the topology of said manifold has been determined. Therefore it is possible to combine this analysis with the numerous existing techniques in the literature to handle the manifold peaks as the use of certain structures that influence the level of dispersion induced by the manifold.

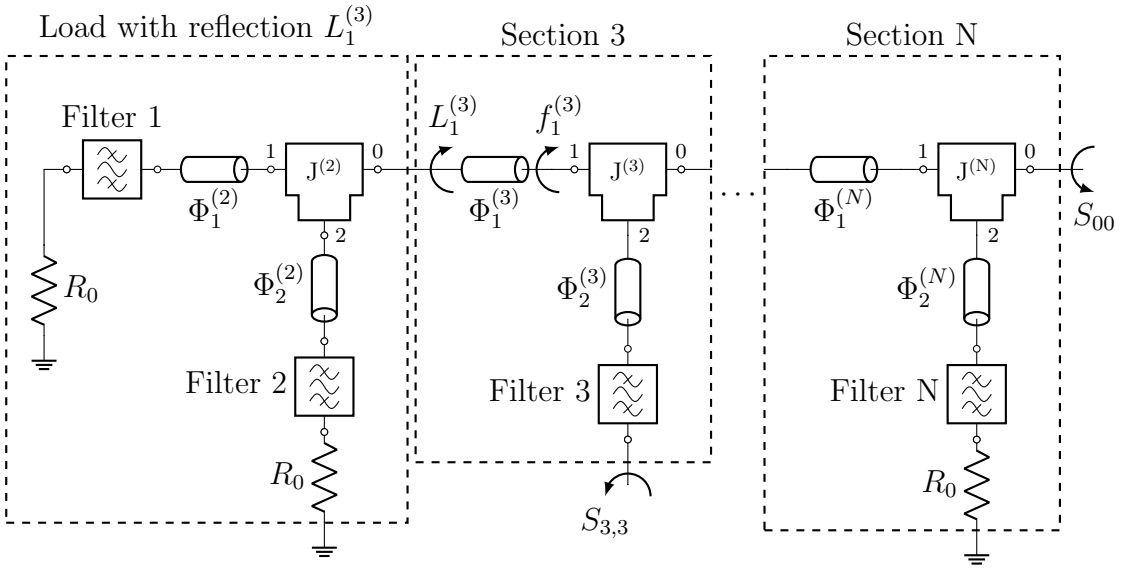


Figure 13.9: Multiplexer schematic

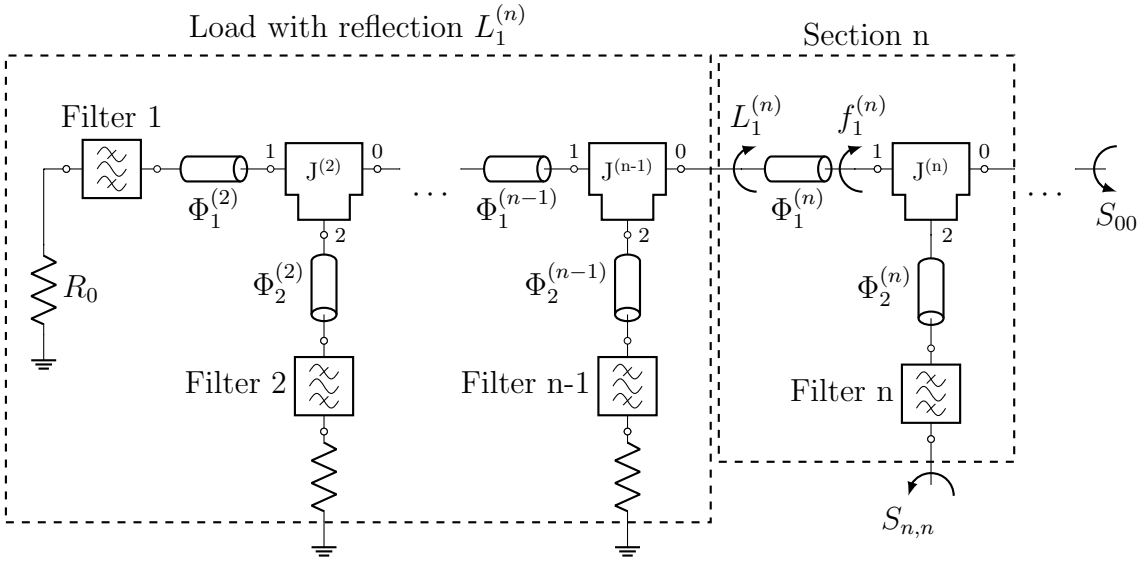


Figure 13.10: Multiplexer schematic

Chapter 14:

Multiport point-wise matching. Application to multiplexer synthesis

The algorithm introduced in chapter 13 provides an equivalent model of the manifold once the channel filters have been fixed. The aforementioned model is built from transmission lines and basic 3-port junctions and ensures the absence of manifold peaks in any of the channels. This result is achieved by means of a systematic analysis of the obtained responses, through the interconnection of the sub-blocks constituting the manifold, with respect to the scattering parameters of each of block separately.

In this chapter we present the second part of the design technique for multiplexers. This second part consists of the procedure complementary to the theory developed in the previous chapter. Particularly, this time we assume that the manifold has already been calculated is fixed with an optimum structure. Then we calculate the matching filters that are matched simultaneously to the load seen from each one of the ports of the manifold, which includes the filters corresponding to others channels, to provide the desired level of reflection in each of the pass bands. Unlike the philosophy used in the previous chapter, where a divide and conquer strategy was used, this time the synthesis of the channel filters is not performed by decomposition into smaller parts, on the contrary we consider all of them as a single multi-port block. The formulation of the problem therefore consists in determining, simultaneously, the set of matching filters that minimize the reflection corresponding to each of the channels when the multiplexer is assembled.

In general, the design method presented in part V of this thesis consists in the division of the problem into two well differentiated parts. On the one hand the determination of the transmission lines that form the manifold, and on the other hand the synthesis of the matching filters in the form of a multi-port device. Furthermore, the simultaneous filter synthesis entails a high computational efficiency, compared with the tradition synthesis procedures for multiplexers.

14.1 Framework and notation

In the previous chapter, a procedure for the dimensioning of a multi-port structure has been developed. This structure was subject to a series of restrictions, such as the fact that it can be built through an arbitrary number of 3-port connections connected by transmission lines. In addition, a certain hierarchy between the different branches, namely the different paths from the common port to each of the channels, was required. These conditions restrict the topology of the dimensioned structures to those of the manifold type with certain exceptions or variations.

In this chapter, however, we consider a generic multi-port junction which is not constrained to be of manifold-type. Additionally we denote by N the number of channels and by J the $(N+1)\times(N+1)$ scattering matrix of the junction. Note here that we adopt, similarly as in the previous chapter, a 0-based numbering for the rows and column of the matrix J such that the index 0 corresponds to the common port meanwhile the index i with $i \in [1, N]$ denotes the port associated to the $i-th$ channel. Additionally, as in previous chapter, we denote by $\mathbb{I}_i$ with $i \in [1, N]$ the passband of the i-th channel. We assume that the passbands $\mathbb{I}_i$ have empty intersection

14.1.1 Belevitch model of channel filters

Along with the junction described above, we also have a set of N filters that complete the multiplexer. Each of those filters is constrained to have a McMillan degree M_i with $i \in [1, N]$ and a prescribed transmission polynomial $R_i = r_i r_i^\star \in \mathbb{P}_+^{2N}$. The parametrisation used for each of the channel filters which ensures the previous requirements consists of a rational model with the Belevitch structure as shown in chapter 2. With this model, the scattering matrix of the filter corresponding to the channel i, denoted here as $F_{22}^{(i)}$ is written as

$$F_{22}^{(i)} = \frac{1}{q(p_i)} \begin{pmatrix} p_i^\star & -r_i^\star \\ r_i & p_i \end{pmatrix},$$

where $p_i \in \mathbb{P}^{M_i}$ and $q(p_i) \in \mathbb{P}^{M_i}$ is the stable polynomial such that

$$q(p_i)q(p_i)^\star - p_i p_i^\star = R_i.$$

Each of the channel filters is then parametrised in terms of the polynomial $p_i \in \mathbb{P}^N$ only. This Belevitch parametrisation is recurrent in the field of filter synthesis and has also been applied to multiplexer design in [74].

14.2 Algorithm to synthesize the channel filters

Now we tackle the synthesis of the channel filters as a matching problem in which the i-th filter is matched to the impedance shown by the rest of the multiplexer in the i-th band. At the same time, the aforementioned filter i should provide a rejection as strong as possible in the bands of the other channels in order to avoid transmission to the port i of the multiplexer within the frequency band corresponding to the channel j with $i \neq j$.

For this purpose we use the point-wise-matching technique introduced in [75] (see section 3.4). To illustrate the procedure, consider as a simple example the case of channel 1 we have the load with reflection $L_{1,1}^{(1)}$ which is composed of the $N+1$-ports junction with the i-th filter with $i \in [2, N]$ connected to the i-th port and terminated at the opposite end by the reference impedance as in fig. 14.1. According to theorem 3.4.1 the load $L_{1,1}^{(1)}$ can be perfectly matched in a set $M_1 + 1$ points $\xi_m^{(1)}$ with a filter $F^{(1)}$ of McMillan degree M_1 meanwhile a perfect rejection in a different set of points $\zeta_m^{(1)}$, with $m \in [1, M_1]$ some of them possible at infinity. Nevertheless the unique network $F^{(1)}$ providing perfect matching at the points $\xi_m^{(1)}$ with $i \in [1, M_1]$ presents a reflection $F_{2,2}^{(1)}$ such that the phase a high frequencies given by

$$\lim_{\omega \to \infty} \arg F_{2,2}^{(1)}(\omega) = \phi$$

takes an arbitrary value $\phi \in [-\pi, \pi]$. This phase ϕ is usually not a drawback as it is implemented by adjusting the transmission line $\Phi^{(1)}$ connected at the right port of Filter 1 fig. 14.1.

Nevertheless in this case the transmission line $\Phi^{(1)}$ is an essential part of the junction and it has already been adjusted to deal with the problem of manifold peaks in the band

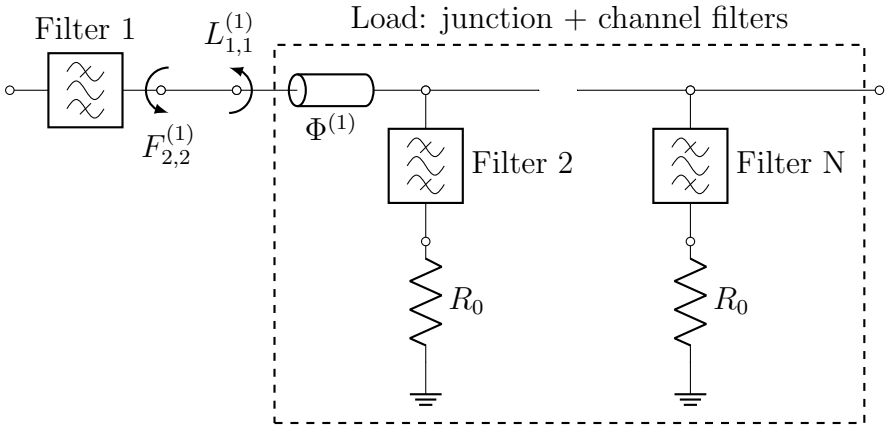


Figure 14.1: Synthesis of filter N to match the rest of the multiplexer.

of filter 1. Therefore a further modification of the line $\Phi^{(1)}$ is not allowed. To overcome this issue, we fix a particular interpolation condition at infinity such that

$$\lim_{\omega\to\infty} F_{2,2}^{(1)}(\omega) = 1 \tag{14.1}$$

The price to be paid for having this additional interpolation condition is losing one of the matching points $\xi_m^{(1)}$. It is interesting to note that the condition imposed at the infinite frequency over the reflection $F_{2,2}$ is equivalent to a perfect matching point where the load to be matched is an open circuit. The synthesis problem concerning the filter of channel 1 is stated as finding the function $F_{2,2}^{(1)}(p_1)$ such that M_1 frequencies of perfect matching $\xi_m^{(1)}$ and M_1 different frequencies $\xi_m^{(1)}$ of perfect rejection are prescribed

$$\begin{aligned}[F_{2,2}^{(1)}(p_1)](\xi_m^{(1)}) &= L_{1,1}^{(1)}(\xi_m^{(1)}),\\ |[F_{2,2}^{(1)}(p_1)](\zeta_m^{(1)})| &= 1,\end{aligned} \tag{14.2}$$

for all $m \in [1, M_1]$ and where $[F_{2,2}^{(1)}(p_1)](\omega)$ tends to an open circuit as $\omega \to \infty$.

14.2.1 Simultaneous computation of matching filters

In this section, instead of dealing with each of the N filters separately, we consider that all of them are contained in a single multi-port device with a total of $2N$ ports as illustrated in fig. 14.2 where each of the ports from 1 to N are connected to the ports from 1 to N of the junction. For each channel, we fix the polynomial $R_i \in \mathbb{P}_+^{2M_i-2}$ having roots at the left and right of the i-th passband in order to contribute to the selectivity of the i-th filter. Additionally, a set of matching points $\xi_m^{(i)}$ is distributed within each passband with $i \in [1, N]$ and $m \in [1, M_i]$. The obtained multi-port device, composed by the union of N different filters, which are isolated from each other, can be parametrized by the set of reflection coefficients $f_i \in \mathbb{S}_{R_i}^{M_i}$ such that the reflection of the multiplexer at the i-th port vanishes at the points $\xi_m^{(i)}$, namely

$$S_{i,i}(\xi_m^{(i)}) = 0 \qquad \forall m \in [1, M_i] \qquad \forall i \in [1, N].$$

This implies $f_i(\xi_m^{(i)}) = \overline{L_i(\xi_m^{(i)})}$ where L_i is the reflection of the load seen from the port 2 of the filter i. In this case the load includes as well other filters as it was illustrated in fig. 14.1. The problem to solve is then stated as:

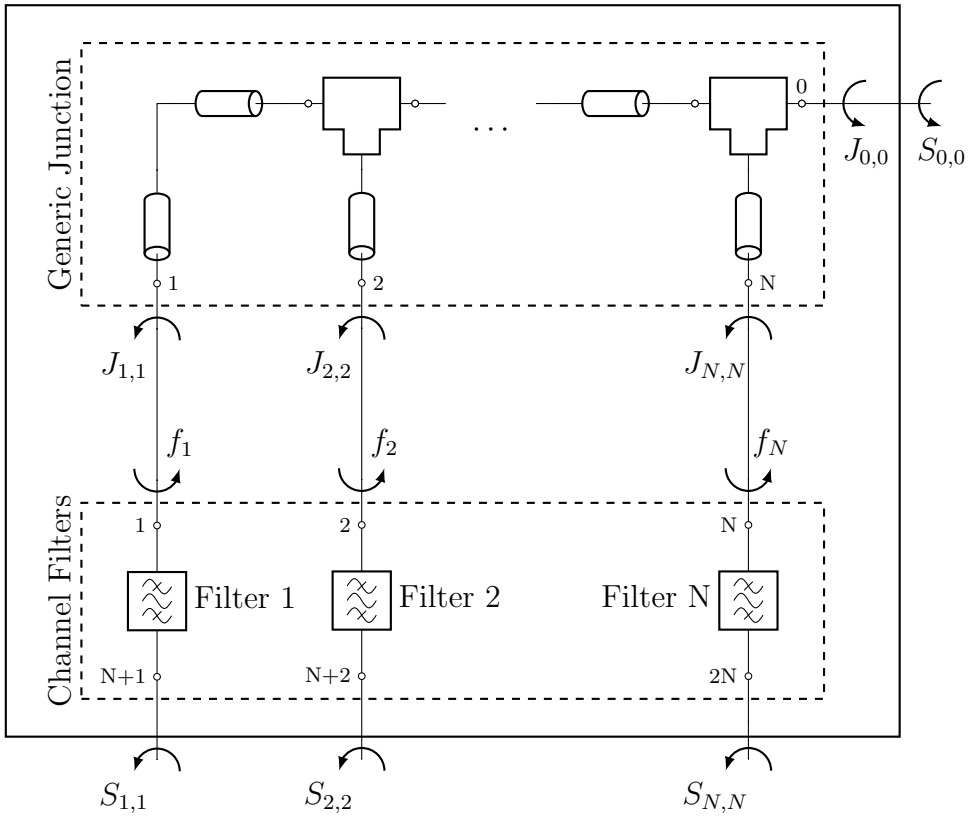


Figure 14.2: Multiplexer structure

Problem 14.2.1 (Simultaneous matching)**.** *Find* $f_i \in \mathbb{S}_{R_i}^{M_i}$*, such that:*

$$f_i[\xi_m^{(i)}] - \overline{L_i[\xi_m^{(i)}]} = 0 \qquad i \in [1, N] \qquad m \in [1, M_i], \tag{14.3}$$

where L_i *is the load seen by filter* i *when the other filters are connected to the manifold.*

Remark 14.2.1. *Note that the output reflection* f_i *allows us to recover the* 2×2 *scattering matrix* $F^{(i)}$ *of the filter* i *such that* $F_{22}^{(i)} = f_i$*. This matrix is uniquely determined from* f_i *up to an uni-modular constant. Nevertheless since only the parameter* F_{22} *of each filter plays a role in the design, we consider only the set of Schur functions* $[f_1, \cdots, f_N]$ *while the extension of* f_i *from a scalar function to a* 2×2 *matrix is performed trivially afterwards.*

14.2.2 Multi-port load

We have stated a simultaneous matching problem where each filter is perfectly matched in a set of M_i points to a load which depends on the other filters. Thus we need now an expression for this load with respect of each sub-device in the system, namely the junction and the channel filters. This expression can be found in definition 13.2.1, indeed with the appropriate ordering of rows/columns of the scattering matrix J, eq. (13.2) can be used to derive an expression for the load L_i with respect to the reflection coefficients $f_i(\xi_m^{(i)})$.

Les us now partition the matrix J such that the i-th column and row are extracted

$$J = \left[\begin{array}{ccc|c|ccc} J_{1,1} & \cdots & J_{1,i-1} & J_{1,i} & J_{1,i+1} & \cdots & J_{1,N} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ J_{i-1,1} & \cdots & J_{i-1,i-1} & J_{i-1,i} & J_{i-1,i+1} & \cdots & J_{i-1,N} \\ \hline J_{i,1} & \cdots & J_{i,i-1} & J_{i,i} & J_{i,i+1} & \cdots & J_{i,N} \\ \hline J_{i+1,1} & \cdots & J_{i+1,i-1} & J_{i+1,i} & J_{i+1,i+1} & \cdots & J_{i+1,N} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ J_{N,1} & \cdots & J_{N,i-1} & J_{N,i} & J_{N,i+1} & \cdots & J_{N,N} \end{array}\right].$$

We denote by V_i the column vector $[J_{k,i}]$ with $k \in [1, N]$ and $k \neq i$ (note that by reciprocity we have $V_i = V_i^T$)

$$V_i = \left[\begin{array}{ccc||ccc} J_{1,i} & \cdots & J_{i-1,i} & J_{i+1,i} & \cdots & J_{N,i} \end{array}\right]^T$$

and by W_i the sub-matrix of J where the rows and columns with index 0 and i are removed

$$W_i = \left[\begin{array}{ccc||ccc} J_{1,1} & \cdots & J_{1,i-1} & J_{1,i+1} & \cdots & J_{1,N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ J_{i-1,1} & \cdots & J_{i-1,i-1} & J_{i-1,i+1} & \cdots & J_{i-1,N} \\ \hline\hline J_{i+1,1} & \cdots & J_{i+1,i-1} & J_{i+1,i+1} & \cdots & J_{i+1,N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ J_{N,1} & \cdots & J_{N,i-1} & J_{N,i+1} & \cdots & J_{N,N} \end{array}\right].$$

Additionally we define F_i as the diagonal matrix which have the functions f_k in the diagonal with $k \neq i$

$$F_i = \left[\begin{array}{ccc||ccc} f_1 & & & & & \\ & \ddots & & & & \\ & & f_{i-1} & & & \\ \hline\hline & & & f_{i+1} & & \\ & & & & \ddots & \\ & & & & & f_N \end{array}\right].$$

Definition 14.2.1 (Load of channel i)**.** *The expression of the reflection L_i seen by Filter i when all other filters are connected to the junction is given by*

$$L_i = J_{i,i} + V_i^T \left(I - F_i W_i\right)^{-1} F_i V_i. \tag{14.4}$$

This expression represents the multi-port scalar chaining where all ports apart from the port i of the junction J are closed by the scalar reflection coefficients f_k with $k \neq i$, providing then the scalar reflection L_1 to be matched by the filter i. Note that the row and column of J with index 0 does not intervene in the computation of the reflection L_i.

14.2.3 A multi-port continuation algorithm

The goal now is to obtain the set of functions f_i satisfying eq. (14.3). Equation (14.3) consists of a set of non-linear equalities, and the solution to problem 14.2.1 satisfying those equalities can be obtained by different approaches. In this chapter we propose the resolution of the aforementioned problem by means of a homotopy similar to the one introduced in section 3.4. With this purpose we formulate an extended problem which depends on a parameter λ and whose solution is trivial for $\lambda = 0$. Furthermore, as λ tends to 1, the solution to this extended problem converges towards the solution of problem 14.2.1.

To perform the continuation, we add a complex parameter $\lambda_m^{(i)}$ for each $\xi_m^{(i)}$ in the expression of $L_i(\xi_m^{(i)})$ obtaining

$$\Lambda_{i,m} = J_{i,i} + \lambda_m^{(i)} V_i^T \left[I - F_i W_i\right]^{-1} F_i(P) V_i \tag{14.5}$$

evaluated at $\xi_m^{(i)}$. Therefore problem 14.2.1 becomes

Problem 14.2.2 (Matching conditions as a function of λ)**.** *Find $f_i \in \mathbb{S}_{R_i}^{M_i}$, such that:*

$$f_i(\xi_m^{(i)}) - \overline{\Lambda_{i,m}} = 0 \qquad i \in [1, N] \qquad m \in [1, M_i]. \tag{14.6}$$

Note that for $\lambda_m^{(i)} = 0$ we have $\Lambda_{i,m} = J_{i,i}(\xi_m^{(i)})$ for which the solution to the problem is trivial. The objective then is to continuate this initial solution, varying these parameters from $\lambda_m^{(i)} = 0$ to $\lambda_m^{(i)} = 1$ for all i, m in small increments $\Delta\lambda$. At each step, the increment to the polynomials p_i (denoted by Δp) is computed by

$$\Delta p = \left[\mathbf{J}\lambda\right]^{-1} \Delta\lambda,$$

where $\mathbf{J}\lambda$ is the Jacobian of the vector of parameters $\lambda_m^{(i)}$ where $i \in [1, N]$ and for all $m \in [1, M_i]$ with respect to the polynomials p_i. Note here that when $\lambda_m^{(i)} = 1$, the filters with output reflections $[f_1, \cdots, f_N]$ are perfectly matched to the manifold simultaneously at the points $\xi_m^{(i)}$.

The efficiency of the algorithm proposed here depends fundamentally on the good conditioning of the Jacobian matrix $\mathbf{J}\lambda$. In particular, the fact that said matrix is non-singular, for a vector λ determined, implies that, given the variation $\Delta\lambda$, there exists a unique direction Δp that indicates the variation to be applied to the parameters of the problem so that eq. (14.6) is still satisfied.

Additionally, it should be noted that the path from $\lambda_m^{(i)} = 0$ to $\lambda_m^{(i)} = 1$ is of relevance here since accidents can happen during the continuation (points where the Jacobian matrix is singular). In this case, we can modify the trajectory to follow in order to avoid such problematic points. For each of the parameters $\lambda_m^{(i)}$ we can choose a path in the complex plane from the starting point 0 to 1.

14.3 Numerical implementation

In this section we provide a detailed development of the numerical procedure used to determine the solution to problem 14.2.1. In addition, as part of this development, we

derive the analytical formulas necessary for the efficient calculation of the aforementioned solution.

In order to ensure eq. (14.2) we impose that the polynomials p_i and $q(p_i)$ are monic. Therefore, similarly as introduced in chapter 7, the polynomial p_i is parametrised by the vector $\Theta_i \in \mathbb{R}^{2M_i}$ containing the $2M_i$ real coefficients $[\theta_1, \cdots, \theta_{2M_i}]$ with respect to the basis of Tchebyshev polynomials such that

$$p_i(\omega) = \mathit{T}_{M_i}(\omega) + \sum_{k=1}^{M_i} (\theta_k + j\theta_{M_i+k})\ \mathit{T}_{M_i-k}(\omega),$$

where T_k is the Tchebyshev polynomial of degree k. Therefore if we denote the basis vector $\mathit{B}_k(\omega)$ as

$$\mathit{B}_k(\omega) = [\,\mathit{T}_k(\omega), \cdots,\ \mathit{T}_0(\omega), j\,\mathit{T}_k(\omega), \cdots, j\,\mathit{T}_0(\omega)]^{\,T},$$

then we have

$$p_i(\omega) = \mathit{T}_{M_i} + \mathit{B}_{M_i-1}(\omega)^T\Theta_i.$$

Consider now the map $\mathcal{Q}$ introduced in eq. (7.21) that associates to each polynomial $p_i \in \mathbb{P}^N$ the coefficients vector T_i of the positive polynomial $p_i p_i^\star$. Note that since the polynomial p_i is monic, no additional normalisation is needed. Then we have

$$\mathcal{Q}_i : \Theta_i \in \mathbb{R}^{2M_i} \longrightarrow T_i \in \mathbb{R}^{2M_i}.$$

Additionally denote by $\frac{1}{2}\mathbf{\Xi}_i(\Theta_i)$ the Jacobian matrix of $\mathcal{Q}_i$ at the point Θ_i. Let us now compute the Jacobian of the evaluation $[q(p_i)](\xi_m^{(i)})$ for each $i \in [1, M_i]$ with respect to the coefficients of p_i. We have $q(p_i)q(p_i)^\star = p_i p_i^\star + R_i$, derivating with respect to p_i as it was done in chapter 7 we obtain

$$\begin{aligned} \mathrm{D}[q(p_i)](\xi_m^{(i)}) &= \mathrm{D}p_i(\xi_m^{(i)}), \\ \mathbf{\Xi}_i(\Phi_i)\mathbf{J}_q(\Theta_i) &= \mathbf{\Xi}_i(\Theta_i). \end{aligned}$$

Therefore

$$\mathbf{J}_q(\Theta_i) = \mathbf{\Xi}_i(\Phi_i)^{-1} \cdot \mathbf{\Xi}_i(\Theta_i).$$

14.3.1 Derivation of analytic formulas providing the derivative of the load

Unlike the case of matching 2-port devices where the load is fixed and the matching problem arises in terms of the matching filter only, this time the load also depends on the matching filters. This fact motivates the calculation of the derivative of the reflection coefficient L_i at the points $\xi_m^{(i)}$ with respect to $f_k(\xi_m^{(i)})$ with $k \neq i$ and $m \in [1, M_i]$. Therefore we consider now the evaluation of the load L_i in eq. (14.4) at the points $\xi_m^{(i)}$. Similarly consider the $N \times N$ matrix $J(\xi_i)$ where again the row and column with index 0

is removed and the matrix $F_i(\xi_m^{(i)}$ evaluated at $\xi_m^{(i)}$. The derivative of $L_i(\xi_m^{(i)})$ with respect to $f_k(\xi_m^{(i)})$ with $i \in [1, N]$ and $k \neq i$ takes the expression

$$\mathrm{D}_{f_k} L_i = \left({F_{i,k}}^T \left(I - F_i W_i \right)^{-1} F_i V_i \right)^2$$

evaluated at $(\xi_m^{(i)})$ where $F_{i,k}$ is defined as the row of F_i containing the element f_k

$$F_{i,k} = \begin{cases} \left[\begin{array}{ccc} {0_{k-1}}^T & f_k & {0_{N-(k+1)}}^T \end{array} \right]^T & \text{if} \quad k < i \\ \left[\begin{array}{ccc} {0_{k-2}}^T & f_k & {0_{N-k}}^T \end{array} \right]^T & \text{if} \quad k > i \end{cases},$$

where 0_k is the zero column vector in $\mathbb{R}^k$.

Once we have the derivative of the function $L_i(\xi_m^{(i)})$ with respect to $f_i(\xi_m^{(i)})$, we only need to calculate the derivative of the function $f_i(\xi_m^{(i)})$

$$f_i(\xi_m^{(i)}) = \frac{p_i(\xi_m^{(i)})}{[q(p_i)](\xi_m^{(i)})},$$

with respect to the parameters of the problem, namely the coefficients of the polynomial p_i. Note that we have already computed the Jacobian matrix of the coefficients of the polynomial $q(p_i)$ with respect to the vector Θ_i with the coefficients of p_i in eq. (7.25). Consider now the vector of matching points $\xi_m^{(i)}$ for all $i \in [1, N]$ and for all $m \in [1, M_i]$

$$X = [\xi_1^{(1)}, \cdots, \xi_{M_1}^{(1)}, \xi_1^{(2)}, \cdots, \xi_{M_2}^{(2)}, \cdots\cdots, \xi_{M_N}^{(N)}]^T.$$

We have $X \in \mathbb{R}^M$ with $M = \sum_{i=1}^{N} M_i$. The evaluation of at the points $X = [x_1, x_2, \cdots, x_M]$ can now be obtained by means of the basis matrix $\mathbf{Б}_i$ defined as

$$\mathbf{Б}_i = \left[Б_{M_i-1}(x_1), Б_{M_i-1}(x_2), \cdots, Б_{M_i-1}(x_M) \right].$$

Since the coefficients of polynomials $p_i, q(p_i)$ were decomposed in real and imaginary parts, we also consider both parts here, hence

$$\mathbf{J}_{q(p_i)}(X) = \left[\begin{array}{cc} {\mathbf{Б}_i}^T & {\mathbf{Б}_i}^T \end{array} \right] \mathbf{\Xi}_i(\Phi_i)^{-1} \mathbf{\Xi}_i(\Theta_i).$$

Let us consider now the application $\mathcal{F}_i : \mathbb{R}^{2M_i} \longrightarrow \mathbb{C}^M$ that associates to the vector Θ_i the evaluation of the function f_i at the points x_m for all $m \in [1, M]$.

$$\mathcal{F}_i : \Theta_i \longrightarrow [f_i(x_1), f_i(x_2), \cdots, f_i(x_M)]^T.$$

Additionally note that the Jacobian matrix of $\mathcal{F}_i$ with respect to the vector Θ_i is given as in eq. (9.14) by the expression

$$\mathbf{J}\mathcal{F}_i = \begin{bmatrix} [q(p_i)](x_1) & & \\ & \ddots & \\ & & [q(p_i)](x_M) \end{bmatrix}^{-1} - \begin{bmatrix} f_i(x_1) & & \\ & \ddots & \\ & & f_i(x_M) \end{bmatrix} \mathbf{J}_{q(p_i)}(X).$$

Each Jacobian matrix $\mathbf{J}\mathcal{F}_i$ has therefore the size $M \times 2M_i$. If we consider finally the application $\mathcal{F} : \mathbb{R}^{2M} \longrightarrow \mathbb{C}^{N \cdot M}$ that associates to the set of vectors $\{\Theta_1, \Theta_2, \cdots, \Theta_N\}$ the evaluation of all functions f_i with $i \in [1, N]$ at every point x_m where $m \in [1, M]$

$$\mathcal{F} : \begin{bmatrix} \Theta_1 \\ \Theta_2 \\ \vdots \\ \Theta_N \end{bmatrix} \longrightarrow \begin{bmatrix} \mathcal{F}_1(\Theta_1) \\ \mathcal{F}_2(\Theta_2) \\ \vdots \\ \mathcal{F}_N(\Theta_N) \end{bmatrix},$$

With the previous formulation we can now easily construct the Jacobian matrix of $\mathcal{F}$ as the block-diagonal matrix

$$\mathbf{J}\mathcal{F} = \begin{bmatrix} \mathbf{J}\mathcal{F}_1 & & \\ & \ddots & \\ & & \mathbf{J}\mathcal{F}_N \end{bmatrix}.$$

It is important to note that the present section up to this point we have developed an expression for the derivative of the reflection coefficient L_i evaluated in the perfect matching frequencies $\xi_m^{(i)}$ with respect to the vector of f_i with $i \in [1, N]$, namely $[f_1(\xi_m^{(i)}), f_2(\xi_m^{(i)}), \cdots, f_N(\xi_m^{(i)})]$. Also by means of the matrix $\mathbf{J}\mathcal{F}$ we can obtain the derivative of the load $L_i(\xi_m^{(i)})$ with respect to the coefficients of the polynomials p_i.

14.3.2 Derivative of the Chaining Expression

Concluding the implementation of the fixed-point algorithm presented in this chapter, it is necessary to express the derivative of the vector of parameters $\lambda_m^{(i)}$ introduced in eq. (14.5) with respect to $f_i(\xi_m^{(i)})$ for all $i \in [1, N]$ and with $m \in [1, M_i]$. Introducing eq. (14.5) into eq. (14.6) we obtain

$$\overline{f_i(\xi_m^{(i)})} - J_{i,i}(\xi_m^{(i)}) - \lambda_m^{(i)} V_i^T \left(I - F_i W_i\right)^{-1} F_i V_i = 0. \tag{14.7}$$

At each point $\xi_m^{(i)}$ there is a value of $\lambda_m^{(i)}$ such that eq. (14.7) holds

$$\lambda_m^{(i)} = \frac{\overline{f_i(\xi_m^{(i)})} - J_{i,i}(\xi_m^{(i)})}{V_i^T \left(I - F_i W_i\right)^{-1} F_i V_i}. \tag{14.8}$$

Finally compute the derivative of $\lambda_m^{(i)}$ with respect to f_k

$$\mathrm{D}_{f_k} \lambda_m^{(i)} = \begin{cases} \left(\overline{V_i^T \left(I - F_i W_i\right)^{-1} F_i V_i}\right)^{-1} & k = i \\ \left(J_{i,i}(\xi_m^{(i)}) - \overline{f_i(\xi_m^{(i)})}\right) \dfrac{F_{i,k} \left(I - F_i W_i\right)^{-1} F_i V_i}{V_i^T \left(I - F_i W_i\right)^{-1} F_i V_i} & k \neq i \end{cases}.$$

It is important to note that for any arbitrary value of $f_i(\xi_m^{(i)})$, this analytical expression allows us to directly calculate the value of $\lambda_m^{(i)}$ that satisfies the matching condition in eq. (14.7) under the condition that the denominator in eq. (14.8) is non-zero and the matrix $(I - F_i W_i)$ is non-singular. Furthermore note that the matrix $(I - F_i W_i)$ is singular if and only if the reflection L_i is lossless. Nevertheless since the port 0 has been removed from the matrix J, a lossless L_i implies no transmission toward the common port. This is not possible if we suppose that no manifold peak appears at the frequency $\xi_m^{(i)}$.

References

[74] G. Macchiarella and S. Tamiazzo, "Synthesis of Star-Junction Multiplexers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 12, pp. 3732–3741, 2010.

[75] L. Baratchart, M. Olivi, and F. Seyfert, "Boundary Nevanlinna-Pick interpolation with prescribed peak points. Application to impedance matching," *SIAM Journal on Mathematical Analysis*, 2017.

Chapter 15:

Multiplexer design and results

The theory developed in the preceding chapters at the beginning of this part may seem quite abstract a priori without a practical example that serves as a support for the correct understanding of it. Especially when certain somewhat unusual definitions such as the concept of main and secondary branches is introduced. This absence of practical applications and real examples is mitigated to a certain extent by the inclusion of several examples of application of the theory introduced above. Specifically in this chapter we detail the design of an X-band triplexer for the space sector. This triplexer is implemented in waveguide technology with a prototype made by additive manufacturing techniques.

Therefore, this chapter, and at the same time this last part, are of a more applied character compared to part 2. In fact, in this chapter we discuss more technical aspects of the implementation and manufacture of the devices designed such as technology chosen for the implementation, the level of dissipation, or the adequate tuning of the final structure.

15.1 Application and target specifications

In order to exemplify the theory developed in part V, we considered the design of an X-band triplexer with the specifications listed in table 15.1. These specifications correspond to the following passband $\mathbb{I}_1, \mathbb{I}_2, \mathbb{I}_3$ in GHz

$$\begin{aligned}\mathbb{I}_1 &= [11.932, 12.168],\\ \mathbb{I}_2 &= [11.677, 11.913],\\ \mathbb{I}_3 &= [11.145, 11.645].\end{aligned}$$

In what the transmission concerns, we have a criterion in terms of the flatness of the transmission coefficient from each channel to the common port. As it is indicated in table 15.1, the maximum variation of insertion losses within any of the passbands is of 1.1 dB. This requirement could be met through the use of predistorted filtering functions achieving an equalization of the transmission along the pass band. Nevertheless, the aforementioned predistortion technique deviates from the objectives of this thesis. As an alternative, in this chapter we consider filters implemented in waveguide technology which provides us with a high quality factor.

15.2 Reference filters

Before approaching the design of the multiplexer according to the theory developed in part V, we made 3 filters in waveguide technology, each of which implements a Tchebyshev type 6 response with 2 transmission zeros, one on each side of the band. Through these reference filters, we can validate the feasibility of the proposed technology in order to meet the specifications required for each channel. It is important to remember that by using channel filters of degree M_i it is possible to set M_i matching points in each passband. Therefore in the response of the multiplexer in each channel, namely the parameters $S_{i,i}$ and $S_{i,0}$ of the final structure, we have the same amount of perfect matching points and transmission zeros .

This implies that by using the reference filters we obtain a response which, although the points of perfect matching and perfect rejection are not located exactly on the same frequency, is of the same type as that provided by the final structure. Each of these filters satisfies the specifications required for the corresponding channel in terms of selectivity as can be seen in fig. 15.2. With respect to the insertion losses, we obtain the minimum quality factor $Q_0 = 2000$ required per resonator such that the cited response with 6 poles and 2 transmission zeros satisfies the specifications listed in table 15.1. A view of the channel transmission showing that the insertion loss requirement is verified with a value $Q_0 = 2000$ is provided in fig. 15.3.

		Channel 1	*Channel 2*	*Channel 3*
General	Centre frequency	12.050 GHz	11.795 GHz	11.395
	Bandwidth	236 MHz	236 MHz	500 MHz
	Return loss	21 dB	21 dB	21 dB
	Insertion loss	10 dB	10 dB	10 dB
Rejection	$Fc \pm 144$ Mhz	20 dB	20 dB	
	$Fc \pm 300$ MHz			20 dB
	$Fc \pm 380$ MHz	40 dB	40 dB	
	$Fc \pm 800$ MHz			40 dB
IL variation	$Fc \pm 70$ Mhz	0.25 dB	0.25 dB	
	$Fc \pm 100$ MHz	0.4 dB	0.4 dB	
	$Fc \pm 118$ MHz	1.1 dB	1.1 dB	
	$Fc \pm 150$ MHz			0.25 dB
	$Fc \pm 200$ MHz			0.4 dB
	$Fc \pm 250$ MHz			1.1 dB

Table 15.1: Required specifications for each channel filter

15.2.1 Coupling topology and waveguide implementation

We shall now decide the technology and the coupling topology used to implement the filters. This step is important since it determines the out of band behaviour of the channel filters. Note that in chapter 13 we assumed that the channel filters are fixed and used the out-of-band reflection of those filters to determine the transmission lines which compose the manifold. In this way we manage to avoid the manifold peaks inside the bands.

In addition, in chapter 14, the aforementioned reflection outside the band of the filters is normalized to a certain value. Particularly in eq. (14.1) we impose a normalization of $F_{22}^{(i)}$ at infinity. This ensures that the out-of-band trend of $F_{22}^{(i)}$ is known. Alternatively a different normalization is possible, for example at a certain frequency in the adjacent passband. This normalisation allows us to make the assumption that the reflection of the matching filters (or more specifically its phase) obtained as a result of the simultaneous synthesis algorithm is not drastically modified with respect to the reflection of the reference filters.

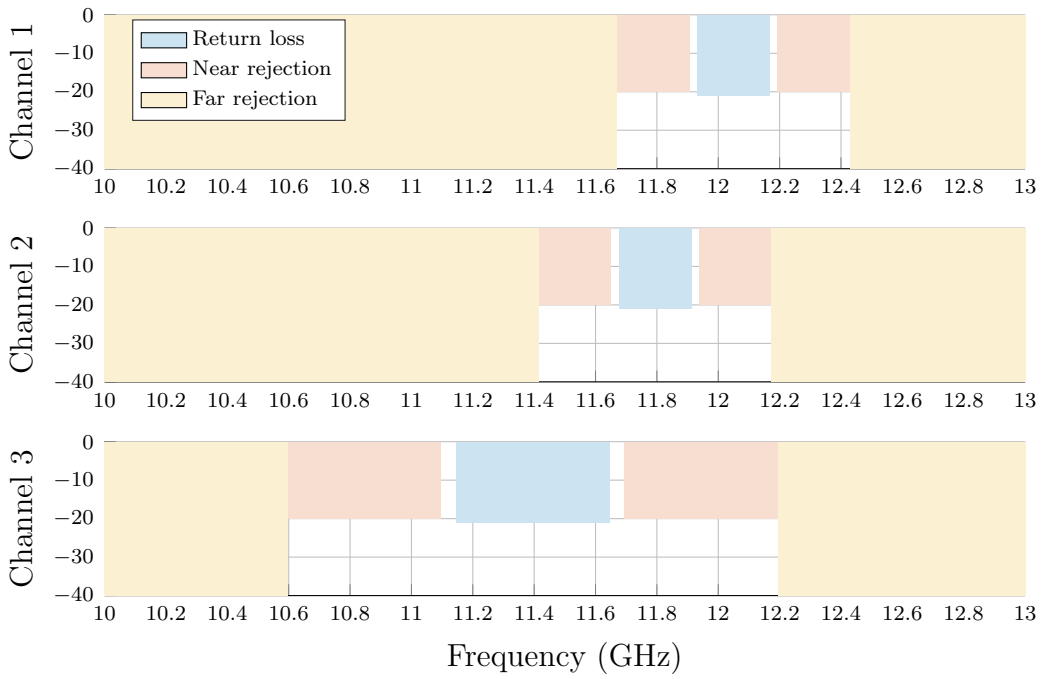


Figure 15.1: Rejection and return loss levels (in dB) indicated in table 15.1.

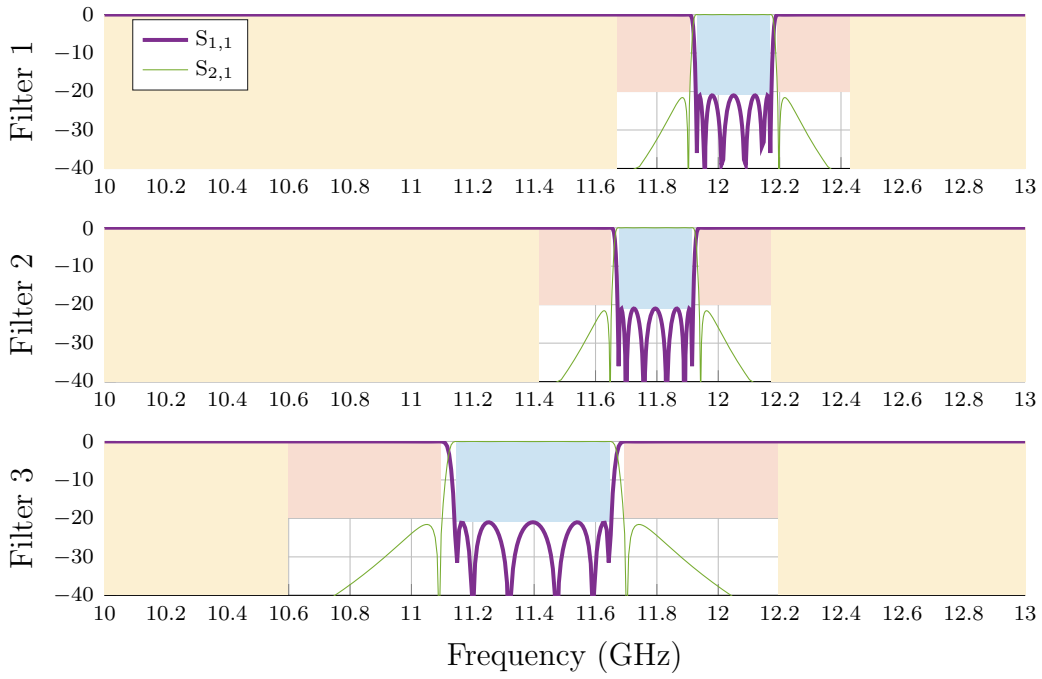


Figure 15.2: Scattering parameters of reference response.

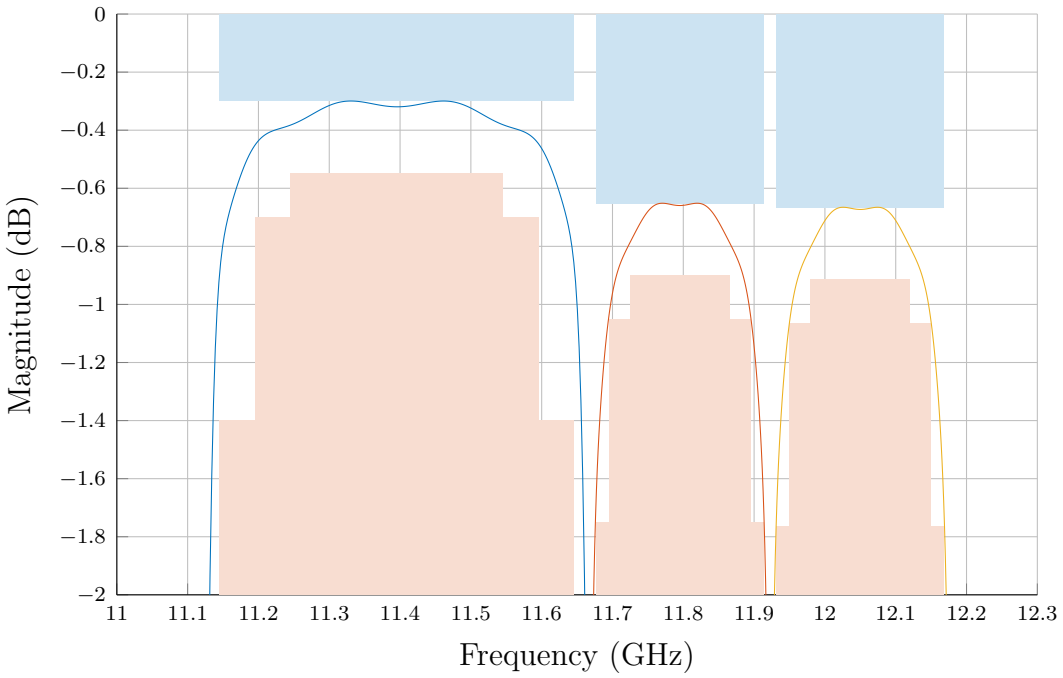


Figure 15.3: Insertion loss of the reference filters (in dB) with a quality factor $Q_0 = 2000$ together with the specifications indicated in table 15.1.

We choose for this example the cross-coupled topology shown in fig. 15.4 which allows for 2 transmission zeros and 6 reflection zeros. The three filtering functions shown in figs. 15.2 and 15.3 are represented by the same coupling matrix after the normalisation with respect to the respective bandwidth and centre frequency. For all of them we have the coupling matrix

$$M = \begin{pmatrix} 0 & 1.01 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1.01 & 0 & -0.35 & 0.76 & 0 & 0 & 0 & 0 \\ 0 & -0.35 & -0.94 & 0 & -0.07 & 0.19 & 0 & 0 \\ 0 & 0.76 & 0 & 0.20 & -0.19 & 0.51 & 0 & 0 \\ 0 & 0 & -0.07 & -0.19 & 0.94 & 0 & 0.35 & 0 \\ 0 & 0 & 0.19 & 0.51 & 0 & -0.20 & 0.76 & 0 \\ 0 & 0 & 0 & 0 & 0.35 & 0.76 & 0 & 1.01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.01 & 0 \end{pmatrix}.$$

These filters are implemented in waveguide technology through rectangular resonant cavities with an fully inductive structure. The structure of each filter consists of four in-line cavities, where the resonant mode TE_{101} is used for the input and output cavities meanwhile the two remaining cavities make use of the higher order modes TE_{102} and TE_{201}. The filter structure can be seen in fig. 15.5. This type of all inductive dual-mode structures were introduced in [76] and have been widely studied in the recent years. Additionally, a more detailed work on inductive dual mode filters, including the structure shown in fig. 15.5, can be found in [77]. This choice of topology allows us to reach, with a standard manufacturing technique, the required quality factor while obtaining a reasonably result in terms of volume and footprint of the structure.

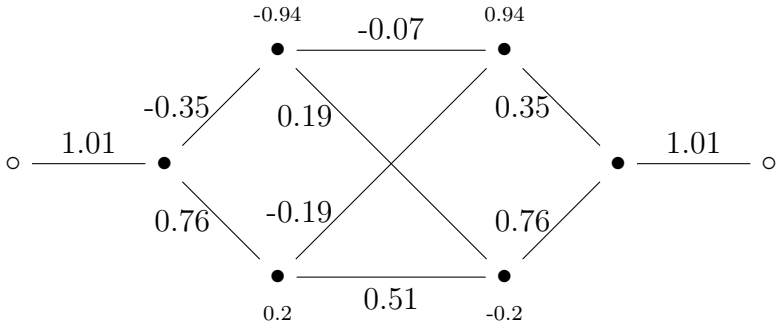


Figure 15.4: Coupling topology

15.2.2 Manufactured prototype

Concluding the study of the reference filters a prototype of the reference filter has been realised by additive manufacturing. This prototype, which is shown in appendix E provides, after being implemented by the structure shown in fig. 15.5, the response shown in fig. 15.6.

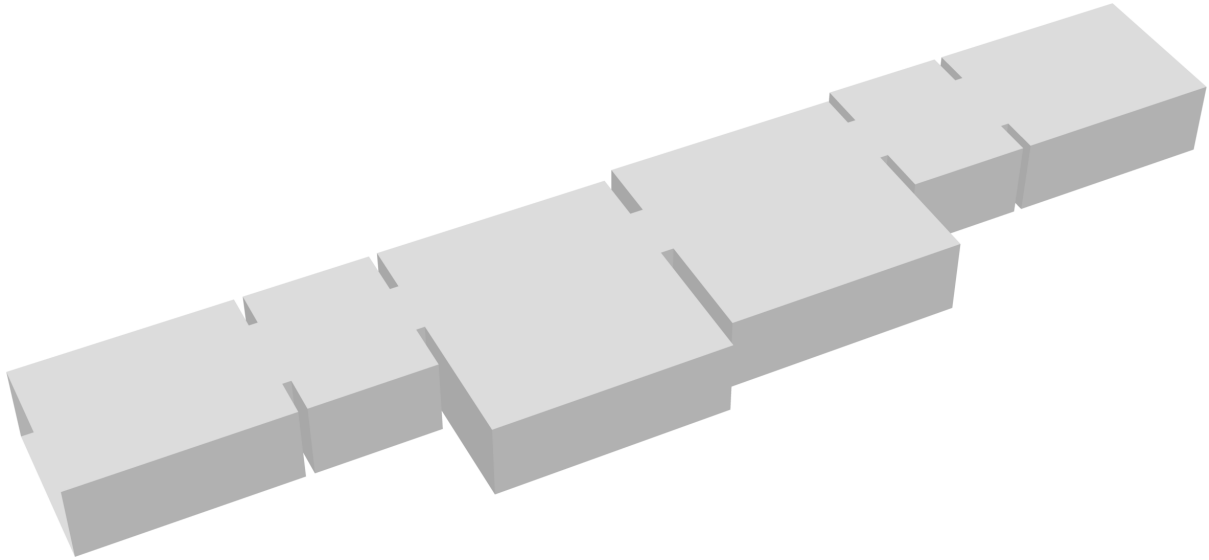


Figure 15.5: 3D view of the structure used to implement the reference filters.

The manufactured filters shows an unloaded quality factor Q_0 per resonator of about $Q_0 \approx 1200$. Note that this result does not reach the value of $Q_0 = 2000$ required to meet the specifications, however considering the fact that it corresponds to a plastic prototype the obtained result is exceptional. This results also indicates that the required value $Q_0 = 2000$ would be attained easily by means of a standard high quality manufacturing technique such that metal milling and silver plating.

The synthesized reference filters provide us with an estimate of the out-of-band phase of each of the channel filters. We show in fig. 15.7 the out-of-band phase of each reference filter.

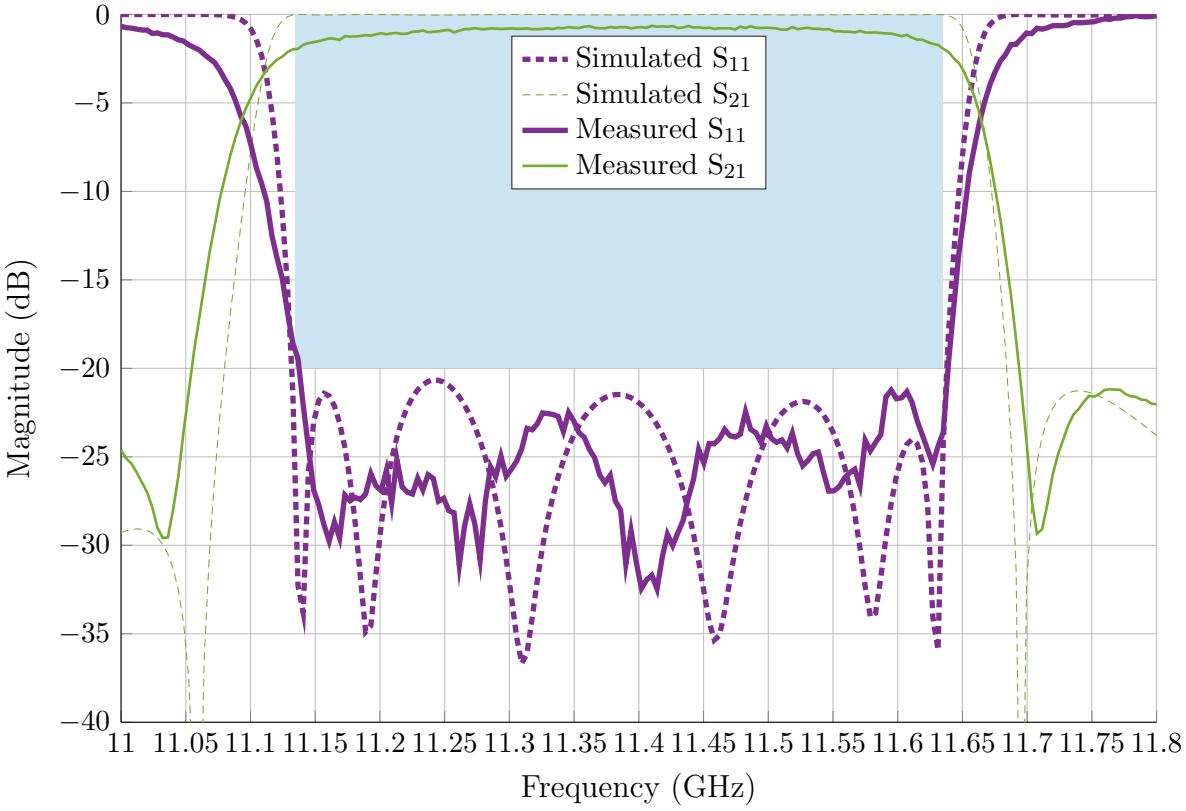


Figure 15.6: Response of the reference filter after tuning (the manufactured device is shown in appendix E) compared to the EM simulated (lossless) response.

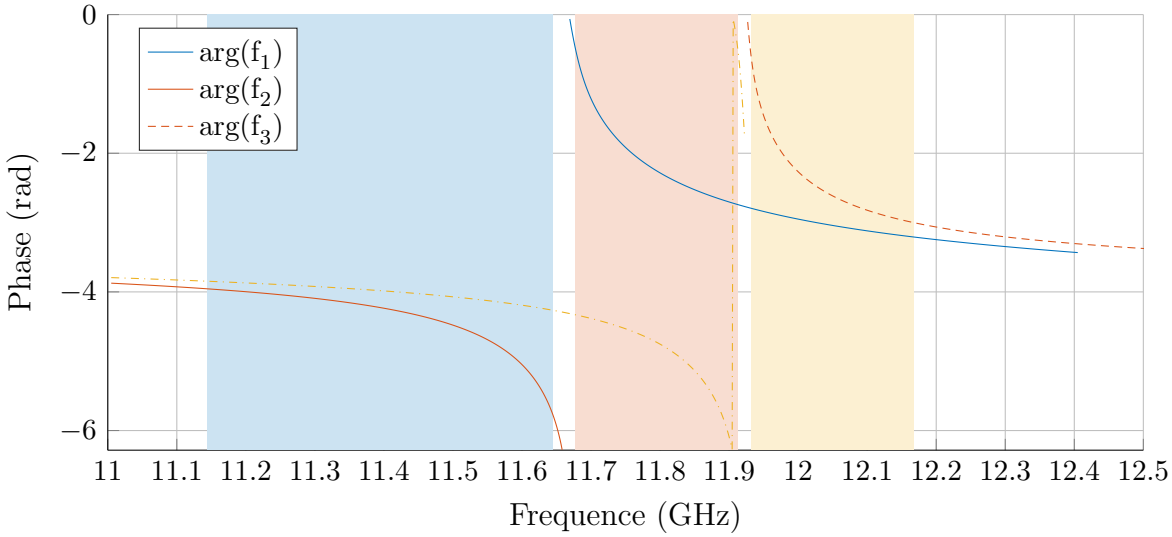
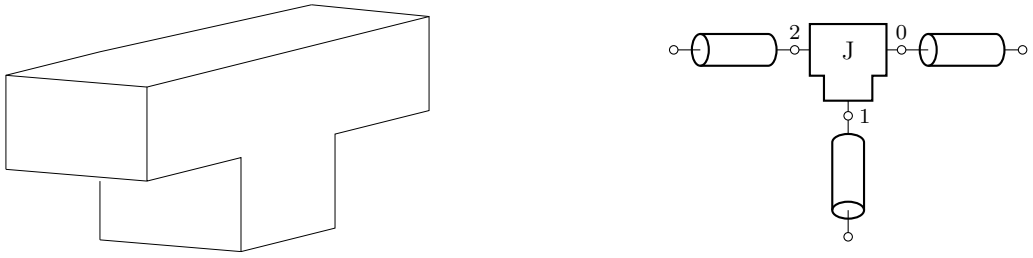


Figure 15.7: Out-of-band phase of each reference filter

15.3 Manifold synthesis

To carry out the implementation of the triplexer to which this chapter is dedicated, we selected a fishbone structure with three channels coupled by a manifold. With the fishbone configuration, the channels are arranged alternating to the right and to the left of the manifold. This configuration can be seen in fig. 15.9. The manifold is composed by two T-junctions interconnected by the transmission line $\Phi^{(0)}$. The model taken for the T-junctions is a 3-port EM-simulation of the waveguide T shown in fig. 15.8a. The simulated structured is modelised by the circuit provided in fig. 15.8b. We de-embed the transmission lines at each access of the 3-port obtaining the core of the T-junction J where a minimal line length per terminal is considered. Additionally the i-th channel is connected to the manifold by means of the transmission line $\Phi^{(i)}$.



(a) Simulated structure

(b) Circuit model

Figure 15.8: T-junction model.

Note that both T-junctions appearing in fig. 15.9 are identical. We denote then their scattering matrix by J. We compute now the uni-modular reflection $P_1(\omega)$ given by eq. (13.6) that produces a transmission zero in $J_{2,0}$ at the frequency ω when it is used to close the port 1 of the junction J. Similarly we compute $P_2(\omega)$ such that $J_{1,0}(\omega) = 0$ with $P_2(\omega)$ closing the port 2. We have

$$P_2(\omega) = \frac{-J_{02}(\omega)}{\overline{J_{20}(\omega)} \det J(\omega)},$$
$$P_1(\omega) = \frac{-J_{21}(\omega)}{\overline{J_{12}(\omega)} \det J(\omega)}.$$

Using the previously designed reference filters, along with the theory developed in the previous chapters, we can determine the lengths of the transmission lines in fig. 15.9 that minimise the risk of encountering a manifold peak in any of the channel passbands. First it is necessary to identify the different branches that constitute the manifold. In this structure we can distinguish a main branch, namely the horizontal path from port 3 to the common port and two secondary branches. As it has been explained in the previous chapter, the design will consist of two stages. The first stage aims to avoid the manifold peaks in the main branch within the passband of channel 3. For this we will adjust the transmission lines connected to the main branch, which are arranged vertically in fig. 15.9. The second phase consists of guaranteeing the absence of manifold peaks in each of the secondary branches, namely the path from the common port to channel 1 and to channel 2 through the corresponding adjustment of the length of the transmission lines $\Phi^{(3)}$ and $\Phi^{(0)}$.

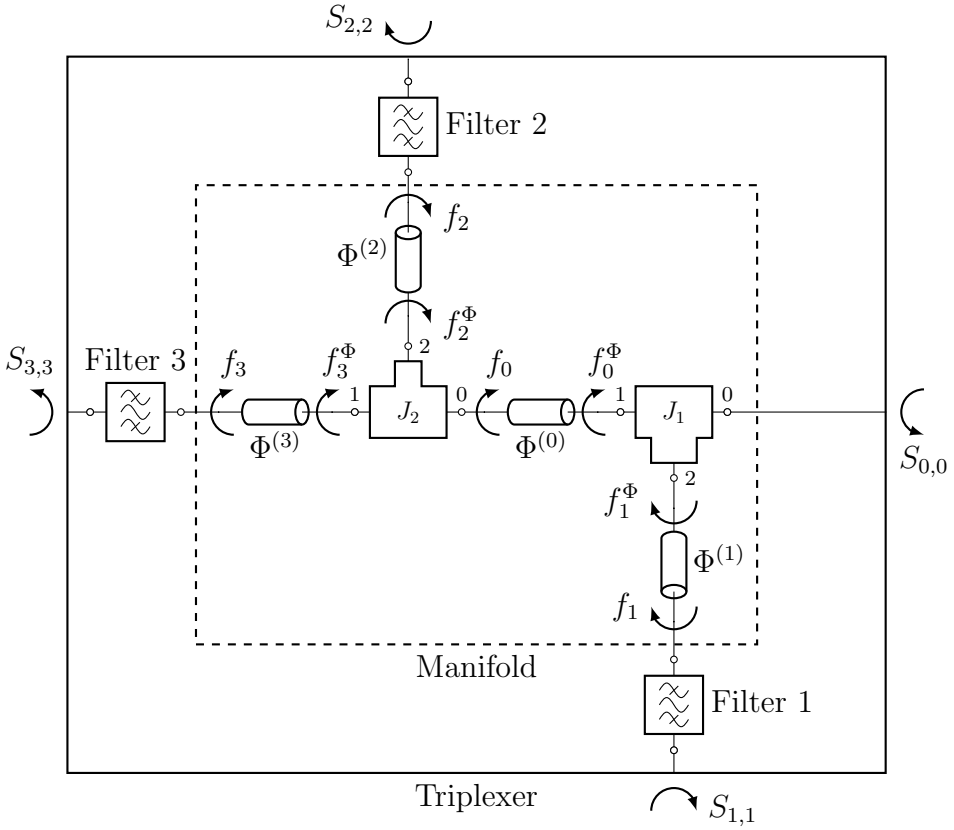


Figure 15.9: Multiplexer schematic

15.3.1 Main branch

We compute in this section the length of the lines $\Phi^{(1)}$ and $\Phi^{(2)}$ (vertical transmission lines in fig. 15.9). We consider the transmission along the main (horizontal) branch of the manifold. The junction $J^{(1)}$ must allow the transmission from terminal 2 to 1 in the bands of channels 2 and 3, meanwhile transmission from terminal 2 to 1 of $J^{(2)}$ in band 3 must be possible. Transmission lines $\Phi^{(1)}$ and $\Phi^{(2)}$ are used to shift the phase of the reflections f_1 and f_2 such that the phase showed at terminal 1 of T_1 does not coincide with $P_1(\omega)$ within the bands 2 and 3. Equivalently the phase brought to terminal 1 of T_2 must not coincide with $P_1(\omega)$ in the third band. We can see in fig. 15.10 that the phase of $f_1(\omega)$ and $f_2(\omega)$ does not coincide with $P_1(\omega)$ within the bands of interest. Nevertheless we choose minimal length of $1mm$ is chosen for $\Phi^{(1)}$ and $\Phi^{(2)}$ to ensure that manifold peaks does not occurs at the edges of passbands $\mathbb{I}_2$ and $\mathbb{I}_3$.

15.3.2 Secondary branches

Next, after $\Phi^{(1)}$ and $\Phi^{(2)}$ are selected and reference filters 1 and 2 connected, we compute the lines $\Phi^{(3)}$ and $\Phi^{(0)}$ (horizontal transmission lines in fig. 15.9). We can see in fig. 15.11 that $f_3(\omega)$ coincides with $P_2(\omega)$ in the band of channel 2. Therefore J_2 introduces a transmission zero from terminal 3 to 1 and a manifold peak will appear in channel 2. To avoid this problem, reflection f_3 is shifted by means of a line $\Phi^{(3)}$ of $7mm$ obtaining the reflection f_3^Φ (shown in fig. 15.11) which does not intersect $P_2(\omega)$ within the band of

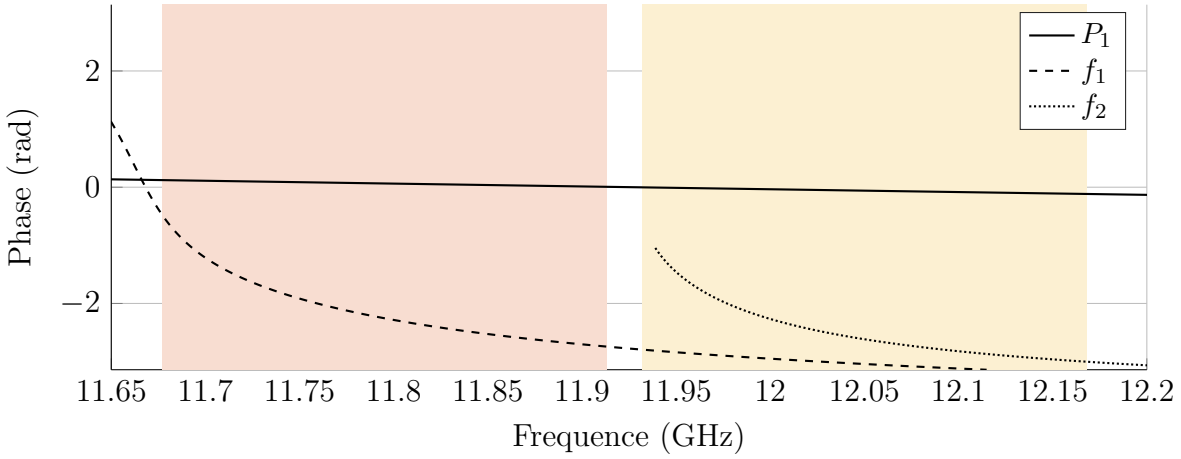


Figure 15.10: Relevant phases to adjust $\Phi^{(1)}$ and $\Phi^{(2)}$

interest. Finally we verify that $f_0(\omega)$ (the reflection of the load seen from port 2 of T_1) does not coincide with $P_2(\omega)$ within the band I_1 as we can see in fig. 15.11. Finally a minimal length is taken for $\Phi^{(0)}$ as in the case of $\Phi^{(1)}$ and $\Phi^{(2)}$.As a result of the previous procedure we obtain the lengths listed in table 15.2 for each transmission line.

Transmission line	$\Phi^{(0)}$	$\Phi^{(1)}$	$\Phi^{(2)}$	$\Phi^{(3)}$
Length (mm)	2	7	1	1

Table 15.2: Selected transmission line length

As it can be seen in fig. 15.11, the result in this second stage is much more tight due to the dispersion introduced by the transmission lines $\Phi^{(1)}$ and $\Phi^{(2)}$, which accumulates throughout the entire process. Furthermore, this effect is increased by the fact that the channels have a considerable relative bandwidth (the relative bandwidth of the band $\mathbb{I}_3$ is 4.4 %) together with the narrow channel spacing (note that the variation of the out-of-band phase is faster the greater the proximity to the band in question, as can be seen in fig. 15.7). However, generally, as long as the bandwidth per channel and the gap between channels allow it, this algorithm can be continued to deal with an arbitrary number of channels.

In this example, the length of the transmission lines has been selected to maximize the distance of the manifold peaks to the corresponding passband in both stages of the process, both in the main branch and in the secondary branches. It is important to note that a compromise must be reached between moving the manifold peaks away from the band in the main branch and in the secondary branches. Indeed if the length selected for $\Phi^{(1)}$ and $\Phi^{(2)}$ is excessive, we would be introducing a high level of frequency dispersion, which would complicate the process of adjusting $\Phi^{(3)}$ and $\Phi^{(0)}$ in the next step. In order to provide a precise criterium to determine the optimal length of the transmission lines we chose the values of $\Phi^{(0)}$, $\Phi^{(1)}$, $\Phi^{(2)}$ and $\Phi^{(3)}$ such that the minimum distance of all manifold peaks to their correspondent passband is maximised.

As long as dispersive transmission lines are used, due to the periodicity in frequency of the response of these elements, manifold peaks are necessarily repeated periodically.

Therefore the best location for the mentioned peaks is that in which the passband is centred in the middle of two occurrences of said peaks. In the same way, if two manifold peaks occur within the same passband originated by the same element J_i, then the appearance of said peaks within the band can not be avoided without reducing the bandwidth thereof.

Note that due to the variation in frequency of the phase introduced by the transmission lines, with the presented algorithm we will obtain manifold peaks to the right and to the left of the passband, at a greater or lesser distance. If we now consider the reflection L_i of the load seen by each of the matching filters, this reflection will be uni-modular in the points where the manifold peaks occur. Similarly, the closer to a manifold peak ω is, the closer $|L_i(\omega)|$ to 1. Hence after the matching filters are synthesised we expect to obtain for each channel i a load with reflection L_i of small modulus in the middle of the passband meanwhile $|L_i(\omega)|$ grows as ω approaches the edges of the passband.

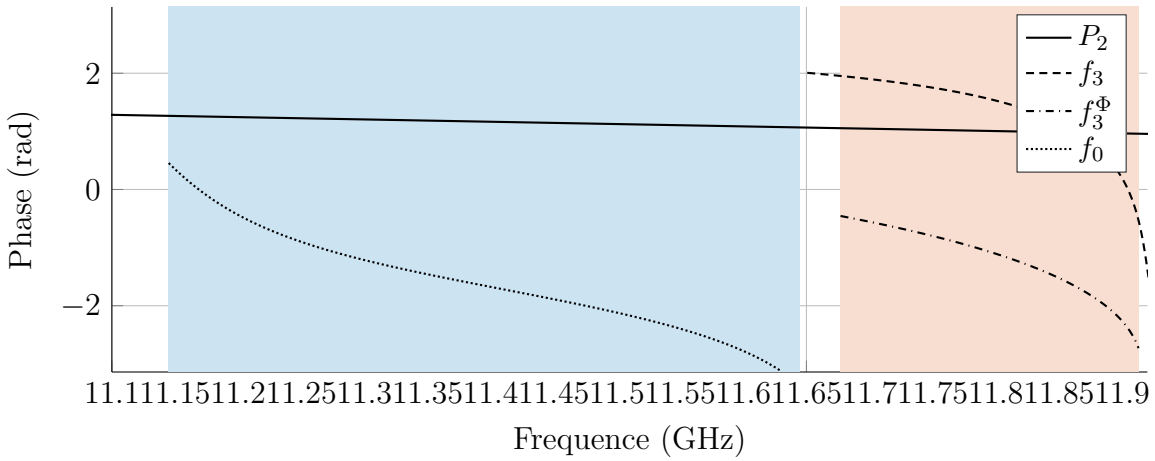


Figure 15.11: Relevant phases to adjust $\Phi^{(3)}$ and $\Phi^{(0)}$

15.4 Matching filters

Once the structure of the manifold is determined and the transmission lines that compose it are determined, the synthesis of the channel filters can be addressed. Each of these filters has been modelled in the previous chapter by means of a rational scattering matrix with the Belevitch structure. Once the structure of the manifold is determined and the transmission lines that compose it are determined, the synthesis of the channel filters can be addressed. Each of these filters has been modelled in the previous chapter by means of a rational scattering matrix with the Belevitch structure. However, once the channel filter technology, namely waveguide filters with inductive character in our case, and the coupling topology used, the topology with cross coupling shown above, it is possible to further refine the model used for the matching filters in such a way that the result obtained in the synthesis stage is much closer to the actual response provided by the filters once manufactured. The aforementioned model is obtained by the rational approximation of the response provided by each of the reference filters.

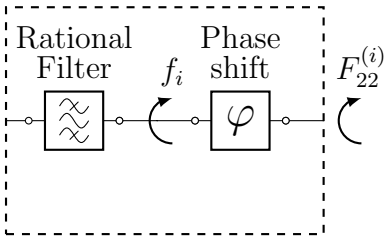


Figure 15.12: Equivalent model for the channel filters

15.4.1 Extraction of rational model

The extraction of a rational model for the reference filters is a crucial task for the subsequent procedure of dimensioning the manifold. In this chapter we use a rational model with the Belevitch form to model the channel filters. However, the reference filters implemented in the previous section show, as a general rule, a non-rational response. The main factor behind this problem is the set of time delays that occur within the structure. Considering these delays, the reflection function of the reference filters $F_{22}(i)$ can be approximated within the passband by

$$F_{22}^{(i)} = \frac{p_i(\omega)}{q_i(\omega)} e^{A\omega+B} = f_i(\omega) e^{A\omega+B} \qquad \omega \in \mathbb{I}_i,$$

with q_i the stable polynomial such that $q_i^\star q_i = p_i^\star p_i + R_i$ and f_i denotes the rational factor of $F_{22}^{(i)}$. Note that the product by $e^{A\omega+B}$ corresponds to the cascade operation with an all-pass device $S^{(i)}$ as shown in fig. 15.12. The all-pass device has the scattering matrix

$$S^{(i)}(\omega) = \begin{pmatrix} 0 & e^{\frac{1}{2}(A\omega+B)} \\ e^{\frac{1}{2}(A\omega+B)} & 0 \end{pmatrix}.$$

During the optimisation, this element $S^{(i)}$ is fixed for each channel and only the rational factor is synthesised allowing us to apply the Belevitch form for the modelling of the rational factor.

15.4.2 Synthesis algorithm

The synthesis of the matching filters is carried out as explained in chapter 14, that is, solving eq. (14.3) through the algorithm introduced previously. In this case we have $N = 3$ and $M_1 = M_2 = M_3 = 6$. To begin with we distribute the points $\xi_m^{(i)}$ within the band of each channel i with $m \in [1, 6]$. Additionally we fix the transmission polynomial for each channel having roots at the previous points. Therefore we are looking for the optimal set of functions $\left\{f_1^{opt}, f_2^{opt}, f_3^{opt}\right\} \in \mathbb{\Sigma}_{R_i}^6$ for $i \in [1, 3]$ such that

$$f_i^{opt}[\xi_m^{(i)}] = \overline{L_i^{opt}(\xi_m^{(i)})} \qquad \forall i \in [1, 3] \qquad \forall m \in [1, 6].$$

The algorithm introduced in chapter 14 is based on the parametrisation of eq. (14.3) by means of a vector of parameters $\lambda_{i,m}$ for all i, m so that for $\lambda_{i,m} = 0$ the solution is trivial and for $\lambda_{i,m} = 1$ we obtain the solution to eq. (14.3). The presented algorithm is based on the continuation of an initial solution obtained for $\lambda_{i,m} = 0$ by varying the parameter

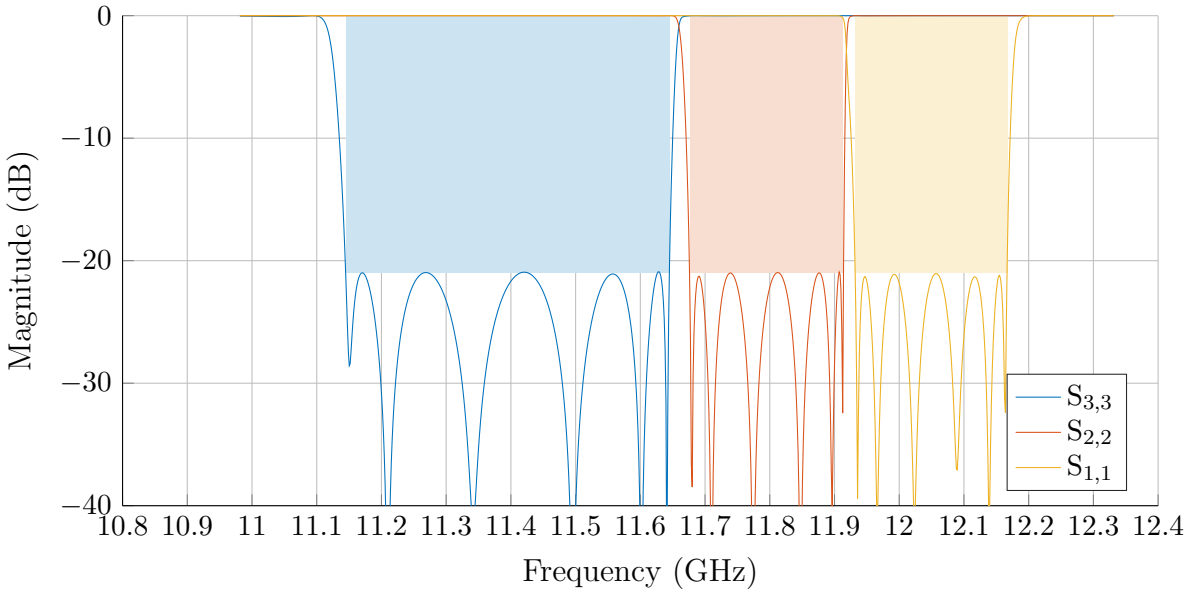
vector $\nu_{i,m}$ from its initial value to $\nu_{i,m} = 1$.

The circuit model of each filter is obtained using the presented method. The computation takes an average time of one second in a laptop with CPU *i7-6600U* under the environment *Matlab R2019a* [78]. After the computation is concluded, the global response shown in figs. 15.13a and 15.13b is obtained.

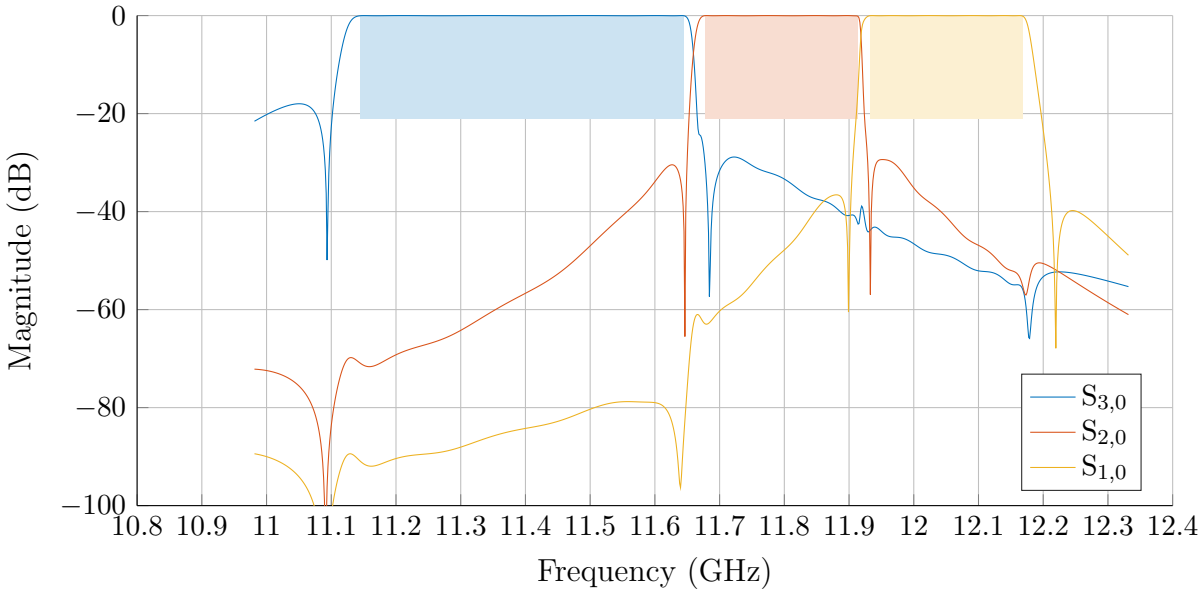
In fig. 15.13a we can see the 6 matching points set in the passband of each channel. The position of these matching points has been optimized to obtain a constant oscillation level in all the bands. It is important to note that the position of these matching points does not correspond exactly to the position of the roots of the Tchebyshev polynomial of degree 6 in the corresponding interval, unless the load seen by the matching filter is of degree 1. Additionally , the reflection level has been adjusted to $-21dB$ modifying the dominant coefficient of each of the transmission polynomials R_i.

In fig. 15.13b the transmission of each channel to the common port is shown. We can also distinguish the position set for each transmission zero, two of them per channel, on each side of the passbands. Finally, what is more important, is the presence of manifold peaks on the left and right of the bands.

In particular we can distinguish for example the peak in the frequency of 11.66 GHz. These manifold peaks were already preceded in the previous section where it was proved that said peaks would necessarily appear on the sides of the bands due to the bandwidth of the channels. To properly visualize the position of said manifold peaks, it is interesting to study the reflection of the load $L_{1,1}^{(i)}$ seen by each of the matching filters. This reflection has been graphed in fig. 15.14 where we can appreciate its magnitude in each of the bands. In particular we can see how this reflection $L_{1,1}^{(i)}(\omega)$ is lower in the centre of the band i while growing toward $0dB$ when (ω) approaches the edge of the band. In particular, in fig. 15.14a we can see the manifold peak mentioned above to the right of channel 3 (the lowest channel in frequency represented in blue) around 11.66 GHz.

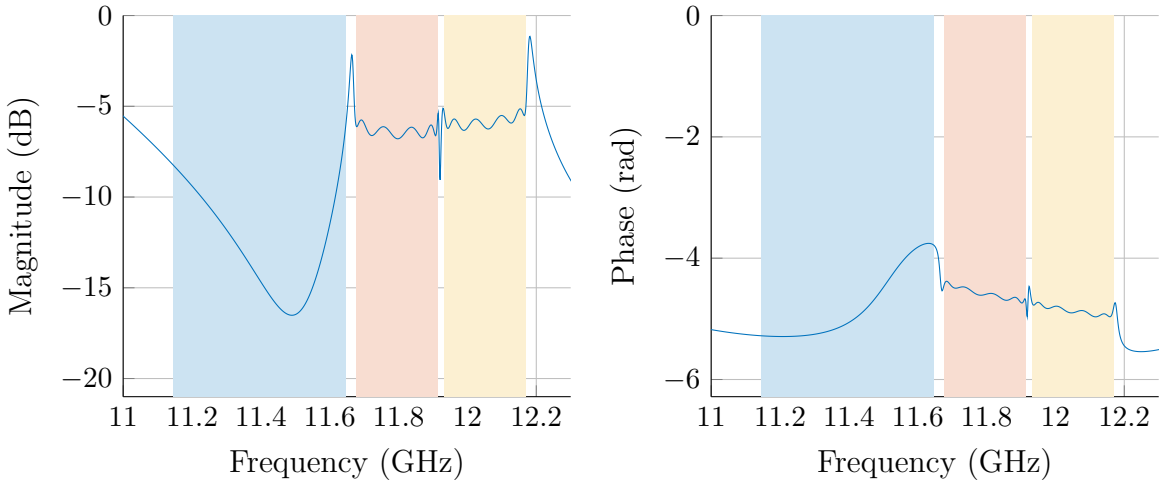


(a) Reflection $S_{i,i}$

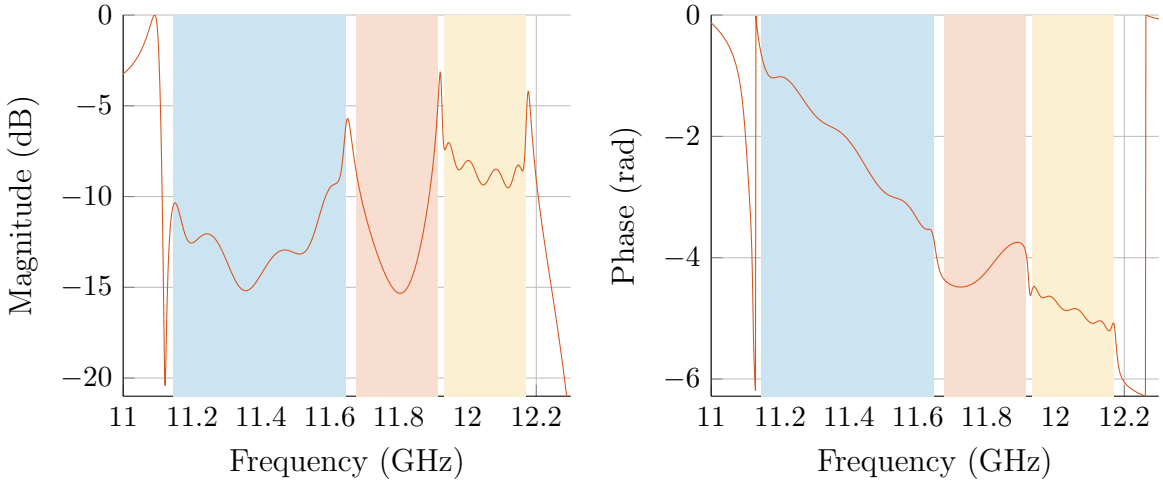


(b) Transmission toward the common port

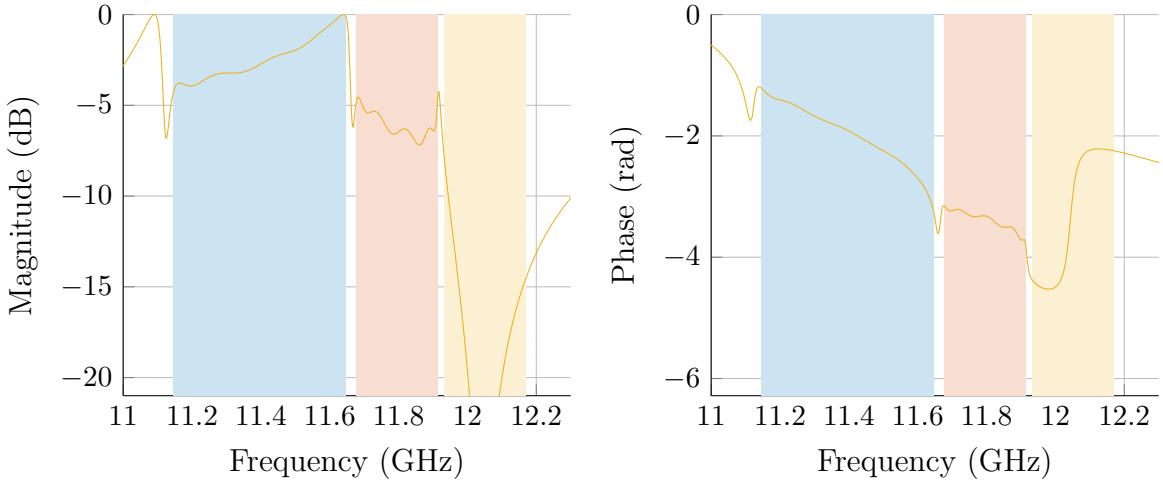
Figure 15.13: Target scattering parameters of the global structure.



(a) Channel 1: $L_{1,1}^{(1)}$



(b) Channel 2: $L_{1,1}^{(2)}$



(c) Channel 3: $L_{1,1}^{(3)}$

Figure 15.14: Load reflection $L_{1,1}^{(i)}$ to be matched by each filter.

15.5 Results

The circuit model of each filter is obtained using the presented method. Note that the manifold was dimensioned at the beginning and the matching filters have been synthesized with the manifold already fixed. Furthermore, the synthesis of the filters carried out previously provides us with the circuit model of each filter, which we represent in terms of coupling matrices, so that all the filters are simultaneously matched to the manifold. By using the obtained circuit for the EM-design of the waveguide filters, this procedure allows us to obtain the objective response for the global structure shown in figs. 15.13a and 15.13b only by optimizing the channel filters separately. Thus we avoid the need for a global optimization of the entire structure, which can become excessively slow in the case of a generic multiplexer.

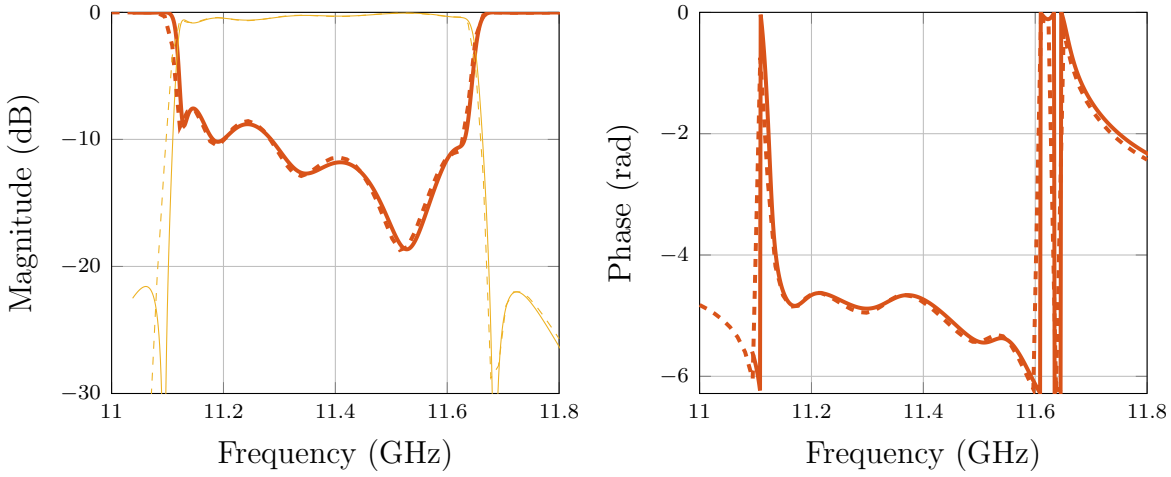
15.5.1 Channel filters optimisation

We show in fig. 15.15 the result of each filter optimisation compared to the goal provided by the presented procedure. These optimisations are carried out with the help of the EM-simulation software *Microwave Wizard* [79] and *Ansys Electronic Desktop* [80] together with the segmentation technique presented in [81, 82]. It should be highlighted here that a good fit of the target response both in magnitude and phase is required here in order to recover afterwards the target global response. This fit in phase of the target response is particularly important since a channel filter providing the wrong phase, even with the proper magnitude, modifies the matching condition between the filters and the manifold preventing us from obtaining the global response shown in figs. 15.13a and 15.13b. Note that the optimization in fig. 15.15 is not perfect since small errors are always made between the EM result and the objective result. However, these errors are less than the errors due to tolerance in the manufacturing process and will be corrected by the use of tuning screws. We include already in the design of each channel filters a set of tuning screws in each coupling and each cavity. One in the centre of the single mode cavities, and four in each dual-mode cavity, all of them positioned at the locations of maximum electric field. These screws are inserted to allow for a compensation of a 50 μm error in the dimensions, namely the attainable precision in the manufacturing process.

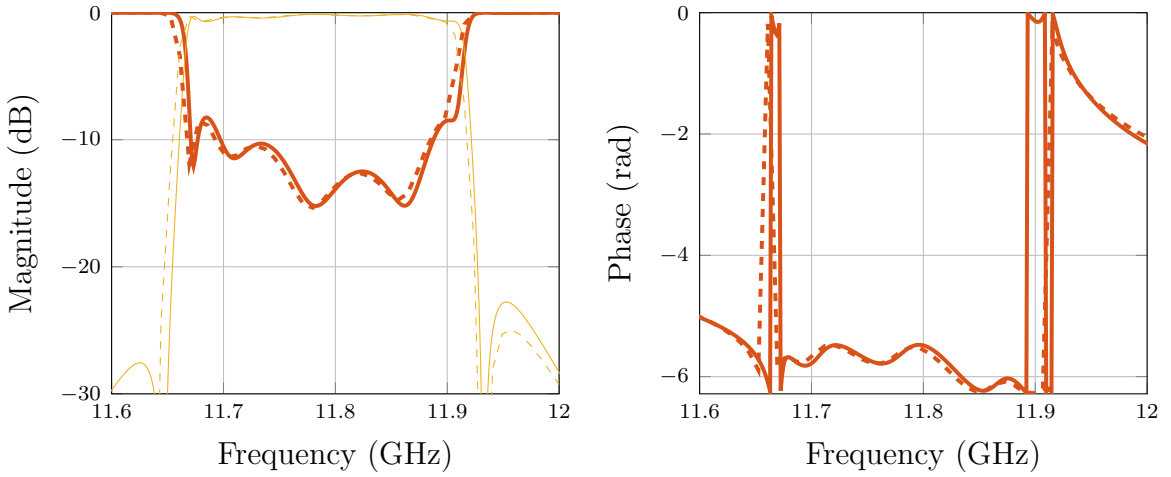
It is now important to compare the out-of-band phase obtained by the EM result of each filter shown in fig. 15.15. Remember that for the dimensioning of the manifold we made the assumption that the out-of-band phase of the matching filters is not strongly modified with respect to the out-of-band phase of the reference filters. This comparison is provided in fig. 15.16 where we can see that the maximum discrepancy between the phase of f_i^{ref} and f_i^{final} occurs near the edge of the passband I_i while both functions tends to the same value as ω tends to infinity as it is imposed in the filter model. Therefore this result validates the assumption made at the manifold design stage.

15.5.2 Global response

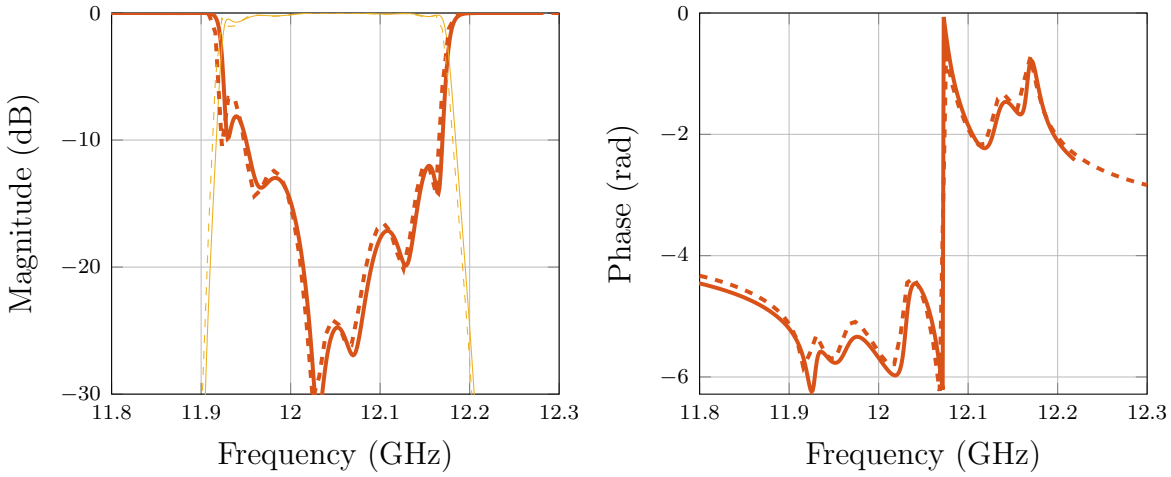
Once the three filters are optimised with the fit shown in fig. 15.15, we assemble the global structure of the triplexer using the manifold previously designed. The obtaining global structure composed of the three channel filters connected to the manifold is shown



(a) Channel 1



(b) Channel 2



(c) Channel 3

Figure 15.15: Filter response. $\boldsymbol{S_{22}}$: thick; $\boldsymbol{S_{21}}$: thin. Circuit response (solid line) vs EM simulation (dashed line).

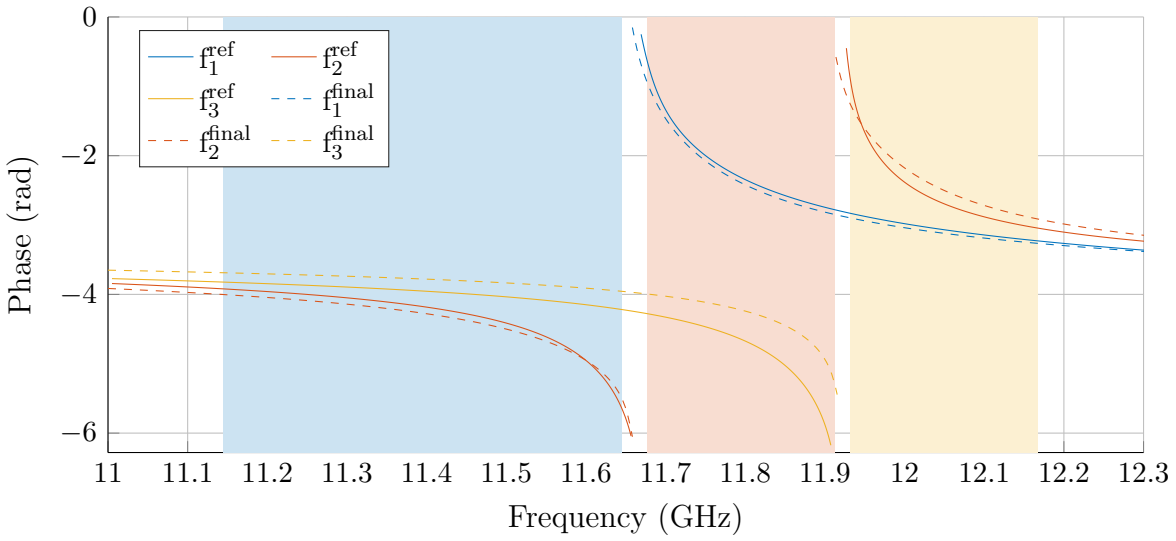


Figure 15.16: Out-of-band phase. f_i^{ref}: reference filters. f_i^{final}: final filters

in fig. 15.17. At this point, without need for a global optimisation, we verify the response of the global structure and compare it to the target result. This comparison can be seen in fig. 15.19. We can verify that the result is good enough without performing any additional optimization. The differences between the triplexer EM response and the objective response are due to the small errors in the optimization separately from each of the channel filters, which is why we consider that the device is ready for manufacturing, correcting possible errors in the tuning stage later on.

The per-channel reflection and transmission parameters measured after tuning the structure are shown in fig. 15.20. As it can be noticed, the measured response shows a frequency shift toward lower frequencies of about 10% the total bandwidth. This fact is due to a manufacturing error higher than expected that could not be compensated by means of the tuning screws. It can also be noted that the complete fishbone structure shown in fig. 15.17 allows the insertion of the tuning screws in each of the cavities and couplings of the channel filters. This condition was imposed as a design restriction, limiting the compactness of the structure which presents filters oriented towards alternating directions.

15.6 Concluding remarks

It is possible to notice at the end of this chapter that in this last part of the thesis a greater effort has been made in the realization of prototypes that validate the theory presented. However, we have not deepened the study of the manufacturing techniques used. We believe that the said study is already widely covered in the literature while the proposed synthesis technique represents a mayor contribution to the field of multiplexers design. For this reason, we have decided to offer a detailed study of the proposed algorithm which, although it is still at a rather preliminary stage, represents one of the main lines of future research issue of the work done.

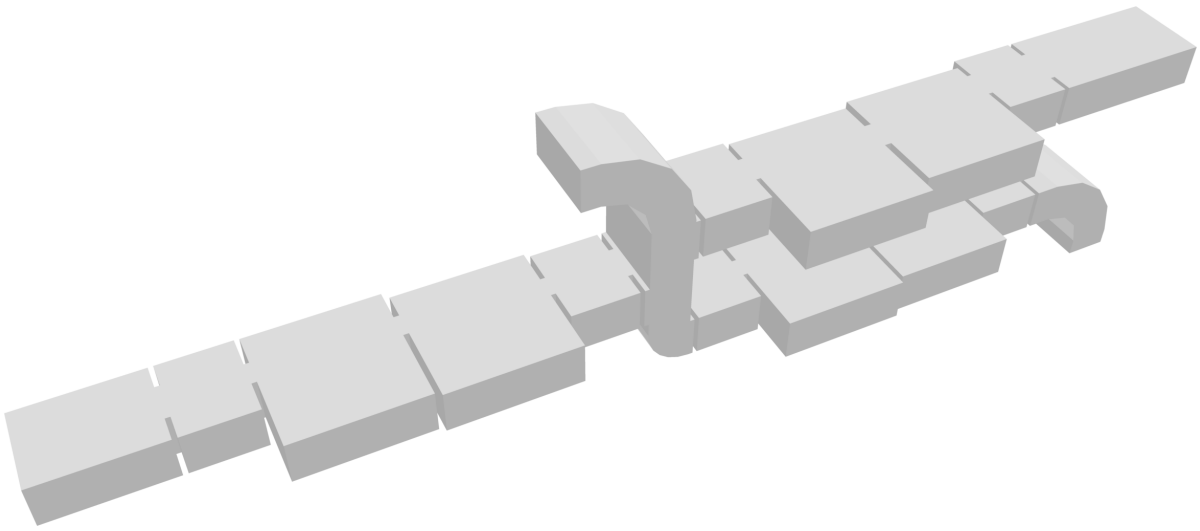


Figure 15.17: 3D view of the triplexer waveguide structure.

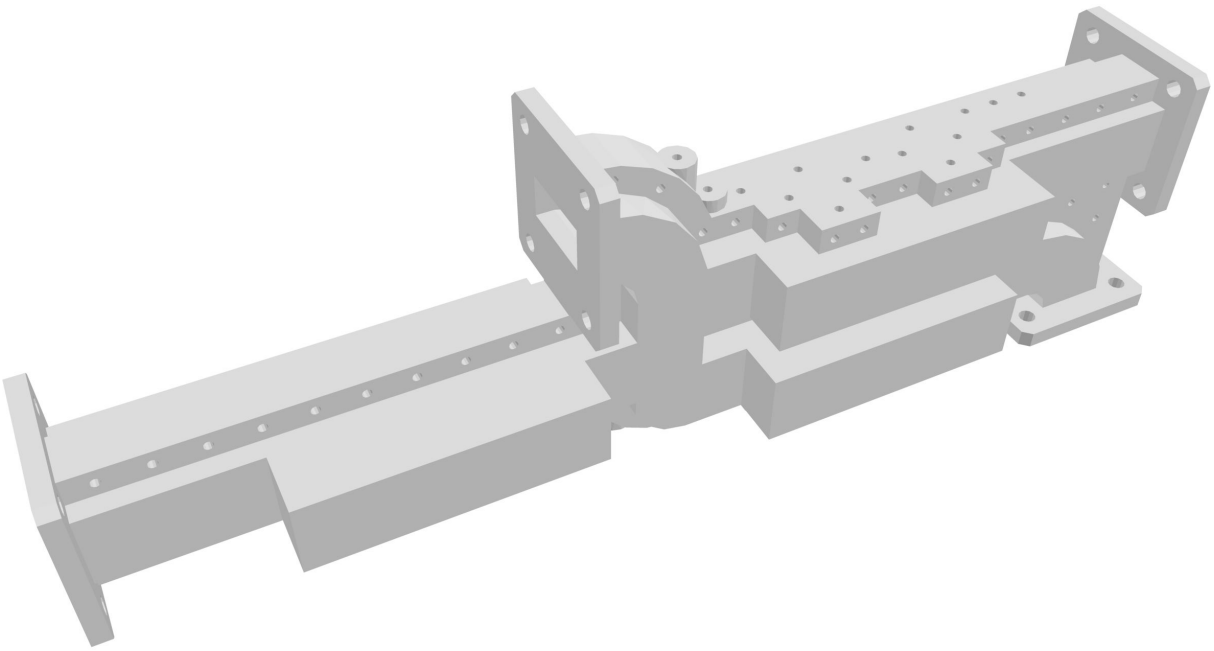
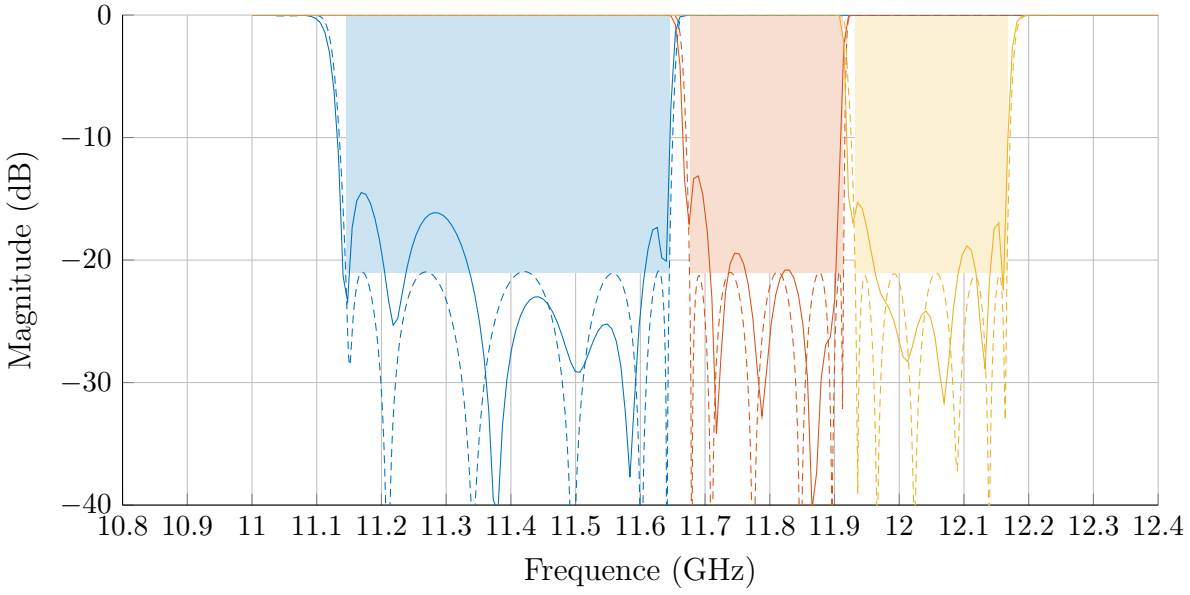
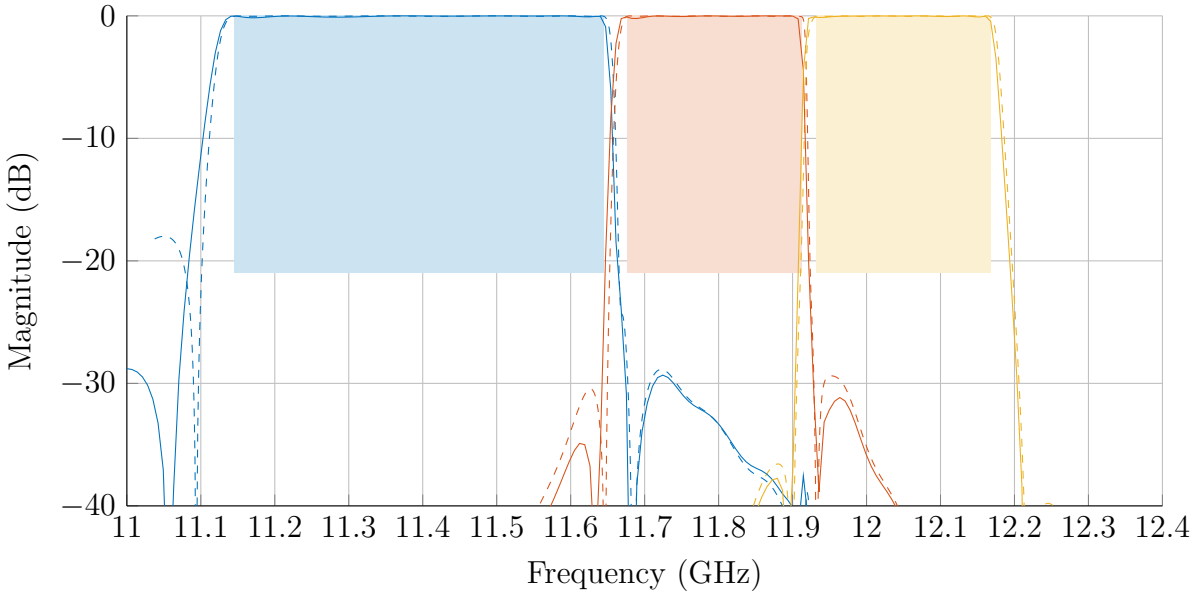


Figure 15.18: 3D view of the triplexer structure.

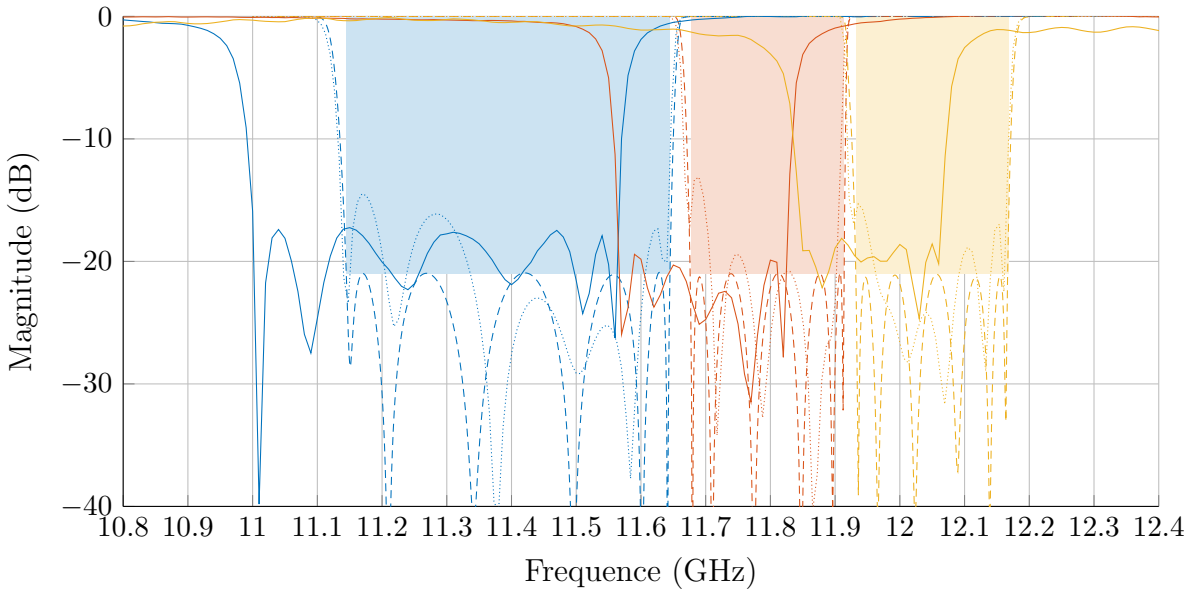


(a) Reflection $S_{i,i}$

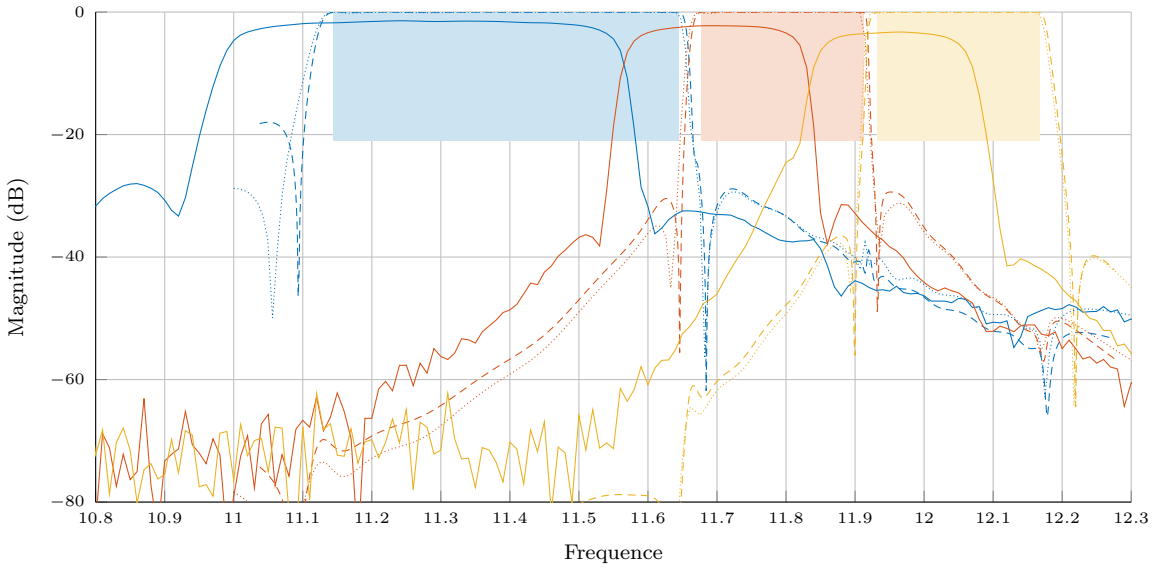


(b) Transmission toward the common port

Figure 15.19: Triplexer response in dB: circuit (dashed line) vs EM simulation (solid line) without need of a global optimisation.



(a) Reflection $S_{i,i}$



(b) Transmission toward the common port

Figure 15.20: Triplexer response in dB: circuit (dashed line) vs EM simulation (dotted line) vs measurements (solid line).

References

[76] M. Guglielmi, P. Jarry, E. Kerherve, O. Roquebrun, and D. Schmitt, "A New Family of All-Inductive Dual-Mode Filters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 10, pp. 1764–1769, 2001.

[77] D. Martínez Martínez, "Band-pass waveguide filters and multiplexers design by the structure segmentation technique," Ph.D. dissertation, Universidad Politécnica de Cartagena, sep 2014. [Online]. Available: http://hdl.handle.net/10317/5310

[78] Mathworks, "MATLAB." [Online]. Available: https://fr.mathworks.com/products/matlab.html

[79] MiCIAN, "MicroWave Wizard." [Online]. Available: https://www.mician.com/

[80] Ansoft, "HFSS 12.0," www.ansoft.com, 2010.

[81] D. Martinez Martinez, A. Pons Abenza, A. Romera Perez, J. Hinojosa, F. Quesada-Pereira, A. Alvarez-Melcon, and M. Guglielmi, "Advanced filter design technique based on equivalent circuits and coupling matrix segmentation," *International Journal of Circuit Theory and Applications*, 2018.

[82] A. R. Perez, A. P. Abenza, D. M. Martinez, A. A. Melcon, and F. D. Quesada Pereira, "Filter design for folded canonical topologies based on equivalent circuit segmentation," *AEU - International Journal of Electronics and Communications*, 2019.

Part VI

Conclusion and perspectives

Chapter 16:

Concluding discussion

In this thesis, a considerable amount of work has been done on the problem of matching complex impedances and their applications. This is one of the beautiful problems in engineering that, although having been introduced in the first half of the last century and studied by numerous authors since then, both from the point of view of its practical applications as well as from an analytical perspective, still remains present. The main reason is the inherent complexity of this problem, due to both the wide variety of loads that can be faced, and the high non-linearity of the equations involved in the chaining operations of these loads with the respective matching networks

During the course of this document, we have been able to corroborate, as it has been noted previously by the different authors who faced the problem of impedance matching in the past, the complexity of the mentioned problem. Particularly in the case of matching networks of fixed degree, this complexity appears in two different ways, either in terms of a rational approximation problem where the criterion to be minimized is expressed in terms of the pseudo-hyperbolic distance to a prescribed function, or in the form of a complex interpolation problem where the interpolating function is expressed as a rational function, with a number of interpolation conditions located in the complex plane. In both cases, if the degree of the aforesaid rational function is bounded, the problem obtained is clearly non-linear. This non-linearity prevents us, for example, from directly applying modern optimization techniques unless the load under consideration is relatively simple.

In order to deal with the problems mentioned, we have formulated a convex problem within the framework of the original matching theory presented by Fano and Youla in the past century. In this way we overcome one of the main disadvantages of this theory as is the rigidity of its practical applications since a prescribed model, either Tchebyshev or Butterworth, was required for the global response. We also provide a generalized algorithm where no model of the network under consideration is required.

16.1 The problem

The convex formulation of the matching problem is achieved by means of a relaxation of the set of functions among which the response of the optimal matching network is sought. This relaxation provides us with fundamental limits that dictate the best possible result in terms of matching for a given load, whose complex impedance varies in frequency. The provided results were unknown until now for a matching network of bounded degree. These limits allow us to determine, for example, that a certain level of matching can not be reached once the load is set, even making a series of strong relaxations on the matching network, such as the absence of power dissipation. The only assumptions made in the matching network are passivity, causality, and stability.

However, these results are not obtained for free, as always, there is a price to pay. The price is given in terms of the sharpness of the obtained bounds, and depends largely on the degree of the load under consideration. Indeed, for a load of degree 1, the computed limitations are sharp without any sacrifice been made. However, when the load is of higher degree, an optimality gap appears between the lower bound and the best result

achieved by a matching filter of fixed degree. This optimality gap is considerably small for a low degree loads and increases as the load's degree increases.

The presented optimization problem is accompanied by a series of interesting properties such as the existence and uniqueness of the solution which indicate that a correct numerical implementation is possible, with the guaranty that the only existing optimal result is obtained in a finite time.

As a result, we obtain a semi-defined non-linear program (NLSDP) which, although being a convex program, does not fit into any of the more typical formulations which can be solved optimally by the solvers available today. However, it is still possible to find some publication in the literature where the NLSDP programs are treated, although without an available implementation capable of correctly handling the problem presented here. We are therefore facing one of the most complex problems in optimization among which the optimality of the solution can be guaranteed. In this thesis we have decided to implement a solver, based on the algorithms available in the literature of NLSDP programs in recent years, adapted to the matching problem we face. The presented algorithm is complemented with a series of heuristic procedures destined to calculate, once the problem of convex optimization has been solved providing the fundamental limits of matching for the load under consideration, a matching network of the desired degree which reaches a result in matching terms as close as possible to those bounds. In addition, we also present a characterization of said optimal solution which, although not allowing direct calculation of it, can be used to certify that a certain result obtained is not incorrect, which would imply, among others, a problem in the implementation of the algorithm.

Next, we also made an interesting discussion about the two applications considered in the present study to illustrate the benefits of the theory developed. These applications concern the problem of matching with two types of loads of a very different nature. The first type is the impedance provided by different antennas, which is represented by a complex function variable in frequency. The second type of loads consists of the impedance seen from the output of each of the channel filters that compose the multiplexer.

16.2 Application to antenna matching

The problem of antenna matching is a classic in modern communication, particularly in the mobile word where the miniaturisation of communication devices and batteries has made of power efficiency a crucial aspect. To achieve the sought energy efficiency, two fundamental aspects are essential in the conception of the global device. These aspects are, on the one hand, the minimization of the dissipation losses in the dielectric materials that make up two physical devices and on the other hand the minimization of the losses caused by the power reflected at the input of the different subsystems.

Particularly when we speak about an antenna, it is implicitly said that the antenna does not show a constant impedance at its input terminal. Usually an important part of the design of the said antenna consists of obtaining a determined impedance, in most cases 50Ω at a specific frequency. However, in the field of radiating devices there is no equivalent

model established as it is the model of coupled resonators in the synthesis of microwave filters. In addition, the design of antennas is conditioned by numerous external factors, such as the interaction between the radiating structure and the rest of the system, namely batteries, metal housing, and other conductive parts which modify the behaviour of the antenna. These effects make the design of the radiating structures extremely complicated.

Therefore, the normalization of the input impedance shown by the antenna to a reference impedance, say 50Ω, imposed only for reasons of standardization supposes an additional restriction in the design of antennas, in addition to the usual requirements in terms of of directivity, efficiency, polarisation, main to secondary lobe level, etc. This requirement is however artificial since it is not related to the fundamental functioning of the antenna itself. Through the algorithms provided in this thesis, the matching condition is transferred to the microwave filter commonly used in the reception/transmission chain to eliminate unwanted signals.

The synthesis of filters, on the other hand, has a more advanced level of maturation and thanks to tools such as the coupling matrix or filter optimization assisted by rational approximation techniques it is possible to design a filter of a certain degree which handles the same time the filtering and matching requirements in a similar way to the design of traditional filters with filtering criteria only.

From the point of view of the design of antennas, by relaxing the condition of matching, there is greater freedom to optimize other criteria that are more relevant to radiation, such as those mentioned above. It is important to note that an antenna does not cease to be a device for the transmission/reception of signals, so it is in this aspect that the design must be focused.

In addition, with the study of antenna arrays, we take this transfer of the matching criterion from the antenna to the matching filter even further. In the case of an array of antennas, the impedance shown at the input of each of the elements is conditioned by the presence of other radiating elements, whether passive or active. This fact is transparent from the point of view of the matching filter since the objective of the matching problem is to adapt a generic load variable in frequency. However, when matching this impedance, from the point of view of the array, the matching filter also fulfils the task of reflecting the signals coming from the rest of the radiating elements that reach this filter due to the coupling between the different elements. These signals are reflected by the matching filter and retransmitted with the appropriate phase so that, through constructive interference with the rest of the radiated signals, they contribute to the overall radiation of the structure.

16.3 Application to the synthesis of multiplexers

Finally, concluding the work done in this thesis, we find the problem of multiplexer synthesis. In this problem the frequency variable impedance which must be matched is obtained as a result of another matching problem which in turn depends on the result of the first. In this case we are facing a simultaneous matching problem where all the

channel filters must be matched to the impedance shown by the rest of the multiplexer. This impedance is not like the case of antennas, a fixed impedance set depends on other matching filters which are also variable in the problem. These particularities make necessary an important adaptation of the classic matching theory to allow the handling of this problem of simultaneous matching.

In addition, in this case the nature of the problem and the application to real devices provide an even more complex problem than the aforementioned simultaneous matching. In fact, we are faced with a problem of filter synthesis where the mastery of filter techniques is not enough, even considering that the synthesis of ideal filters is possible. This is due to the manifold influence. In the problem of matching presented above, even having claimed that the optimal matching of a generic impedance is possible, there is a case where the problem is poorly conditioned and no transmission is possible between the matching filter and the load. This is the case in which the impedance shown by the load is a short circuit. In the synthesis of multiplexers, especially multiplexers of the manifold type, this case can be found recurrently. The reason comes from the fact that the channel filters exhibit a reflection almost unimodular in the bands of the adjacent channels and the manifold, basically composed of transmission lines, provides effect of shifting the phase of this reflection to the point where a short circuit might appear in the impedance seen from any of the channels. This short circuit is commonly known as a manifold peak. If this occurs, the synthesis of the channel filters is not possible with any synthesis technique.

The first of the innovative contributions introduced in the field of multiplexer design consists of a preliminary analysis of the manifold in order to predict the position of the manifold peaks allowing. This allows us to design the manifold accordingly to avoid such peaks.

The second contribution presented consists of the simultaneous synthesis of the channel filters once the manifold has been designed. This synthesis is carried out through a point-wise matching algorithm based on the continuation from an arbitrary initial solution to the final response of the matching filters, ensuring the perfect matching of each of the filters in a set of points fixed on the frequency axis.

Note that the design of multiplexers usually involves the optimization of the multiplexer structure using direct optimization techniques. This process is extremely slow due to the complexity of the global structure and the numerous amount of local minima found during the optimization. By means of the presented algorithm we also achieve a dissociation between the synthesis of the manifold and the synthesis of the channel filters, which represents a major advance in the design of this type of devices.

Finally, by designing and manufacturing one of these devices, namely a triplexer for satellite applications with extremely tight frequency specifications, we were able to validate the presented theory by obtaining the synthesis of the channel filters simultaneously in a matter of seconds while a much longer time would have been necessary through traditional synthesis techniques.

Chapter 17:

Perspectives and open questions

At the end of the study conducted in this work, having obtained several results of considerable importance, especially in relation to the fundamental limits in the problem of matching with bounded degree filters, we also have an important part of prospective results. These results have been outlined during this work, indicating possible lines of investigation to pursue. However, they have not been further developed due to the limited time available.

In particular, we consider that the most relevant results in the context of the matching problem are still to come in the form of a definitive solution to the matching problem by eliminating the optimality gaps shown in chapter 9 as a final implementation of an algorithm of synthesis or transfer functions for multiplexers with guaranteed optimality.

These are, however, the most ambitious results to the problems dealt with in this thesis, which could, as far as our current understanding goes, not have a unique optimal solution in all cases. Furthermore, even if such an optimal solution exists, there may not be a method to obtain it in a guaranteed manner.

However, there are also other less ambitious and more immediate research lines which seek to respond to some of the most interesting open questions about the work done. Following the outline of the previous chapter, where we have made a summary of the results obtained, we provide a quick list of all these open questions to serve as a guide for possible future work on the problem of matching in its different forms . This list constitutes a summary of the missing results, which would have a greater importance in our study by completing the obtained results.

17.1 Optimal synthesis of transfer function for matching synthesis

Even with the results obtained in part II of this thesis, we can not consider that the matching problem is over. Especially when the most significant results have been obtained in terms of lower fundamental limits to the solution of said problem, without obtaining the optimal solution guaranteed in the majority of cases. It is important to note that in chapter 9 a vague connection has been obtained between the solution of the relaxed problem and the solution of the original problem which seeks to obtain the best filter of degree K in terms of matching with a load L which is fixed. We next make a hypothesis about the number of zeros in the analyticity domain of the optimal function for problem 4.1.1.

Conjecture 17.1.1. *Consider the function $S_{22}^{best} \in \mathbb{S}_R^K$ solution to problem 4.1.1 with a load L of degree M and $R = R_F R_L$ with R_F a positive polynomial of degree K and $R_L \in \mathbb{P}_+^{2M}$ the transmission polynomial of the load. We state that the function S_{22}^{best} can be expressed as*

$$S_{22}^{best} = S_{22}^{O} \cdot S_{22}^{B} \tag{17.1}$$

where S_{22}^{O} is a minimum phase function and S_{22}^{B} is a Blaschke product of degree exactly $M-1$.

The motivation to formulate this conjecture comes from the necessary conditions in the form of integrals that must be satisfied by the global system S once the load L is fixed.

We already know that it is possible to construct a load of degree M, namely a reducible load, for which the optimal solution to problem 4.1.1 can be expressed in the form given by eq. (17.1) with a function S_{22}^B of a degree strictly less than $M-1$. However we can still ask if the function S_{22}^B is in the general case of degree $M-1$ and even more, if S_{22}^B is never of degree greater or equal to M. Note that in the relaxed problem the Blaschke product obtained as part of the solution to the relaxed problem is a maximum of M. This is one of the results obtained in the relaxed formulation of the problem that is uncertain when considering problem 4.1.1.

In addition, further results can be investigated with respect to problem 4.1.1. Note that problem 4.1.1 is not convex, therefore we can not guarantee for the moment the uniqueness of the optimal solution and little can be said about it. We can investigate if the aforementioned solution S_{22}^{best}, assuming that this function presents exactly $M-1$ zeros in $\mathbb{C}^-$, can be calculated from problem 9.2.2 if the positions of such zeros are known in advance. This leads to the following conjecture

Conjecture 17.1.2. *Consider an arbitrary load L of McMillan degree M with the transmission polynomial R_L. Let $S_{22}^{best} \in \mathbb{\Sigma}_R^N$ be the optimal solution to problem 4.1.1 over an interval $\mathbb{I}$, where $N \geq M$ and $R = R_F R_L$ with R_F a fixed polynomial. Assume now that the function S_{22}^{best} can be written as in eq.* (17.1) *with*

$$S_{22}^B = \prod_{k=1}^{M-1} \frac{\lambda - \xi_k}{\lambda - \overline{\xi_k}} \qquad \xi_k \in \mathbb{C}^- \qquad \forall k \in [1, 2...M-1]$$

Now take $\Xi = \prod_{k=1}^{M-1} (\lambda - \xi_k)\big(\lambda - \overline{\xi_k}\big)$ and solve problem 9.2.2 with the same load and the same values of $\mathbb{I}, N, R$. After solving problem 9.2.2 we obtain an outer function $u_{\Xi P_{opt}}$ and a Blaschke product b_{opt} such that $S_{22} = u_{\Xi P_{opt}} b_{opt} \in \mathbb{F}^{N+M-1}$. We have $b_{opt} = S_{22}^B$ and therefore $S_{22} = S_{22}^{best}$, namely the solution to problem 4.1.1 is computed by means problem 9.2.2, which is a convex problem.

Therefore another interesting question arises to determine if those $M-1$ zeros of S_{22}^{best} inside $\mathbb{C}^-$ can be estimated a priory by some procedure.

Finally, on the way to the solution of problem 4.1.1 we could ask ourselves what other additional information gives us the relaxed problem about the optimality of the solution obtained for problem 4.1.1. For example, we know that if this solution is outer, then it is also the optimal solution to problem 4.1.1. Yet other details are unknown to our eyes, like the information provided by the fix point algorithm proposed in chapter 9. Suppose for instance that from the relaxed problem we obtain a function $S_{22} \in \mathbb{F}_R^N$ issue of the fixed-point algorithm presented, can we say anything about the optimality of this function for problem 4.1.1?

For instance we could assume that the fix point algorithm stated in chapter 9 converges and study the optimality of the provided solution for problem 4.1.1.

Conjecture 17.1.3. *Consider a load L of degree M along with the values of $\mathbb{I}, N, R$ as before. Take now the function $S_{22} = u_{\Xi P_{opt}} b_{opt}$ computed from the solution to problem 9.2.2 and assume that b_{opt} is of degree $M-1$ and S_{22} of degree N. Therefore $M-1$ simplifications occurs between $u_{\Xi P_{opt}}$ and b_{opt}. The computed function S_{22} is optimal for problem 4.1.1.*

17.2 Efficiency optimisation

It is also important to note that in parallel to the synthesis of transfer functions for matching filters, a study of the radiation efficiency provided by a given antenna has also been carried out.

17.2.1 Efficiency optimization

In the context of the maximization of efficiency, namely the transmission of the system, the first missing result is the optimality of the solution obtained in the presence of losses due to dissipation in the system. In addition, on the way to the aforementioned uniqueness we can study the following aspects

- First, the existence and uniqueness of the optimal solution to the matching problem, namely problem 4.1.1 with non-lossless devices. The uniqueness implies, in the case where the losses are sufficiently low, that the optimal solution to the problem with losses is relatively close to the lossless problem. Therefore we can consider the calculation of this optimal solution by solving the lossless problem presented in this thesis followed by a local optimization from the obtained solution and considering the corresponding level of losses in order to maximize the efficiency of the system.

- The next question that arises is whether the optimal solution in the case of a system with a high level of dissipation can be obtained by means of a continuation algorithm, which starts with calculating the solution to the system considering lossless devices, then increasing the level of losses in small increments so that the solution in each iteration is obtained simply by a deformation of the solution obtained in the previous iteration. Along with the uniqueness of the solution in each iteration, an algorithm of this type could provide the optimal solution to the problem of transmission with lossy devices.

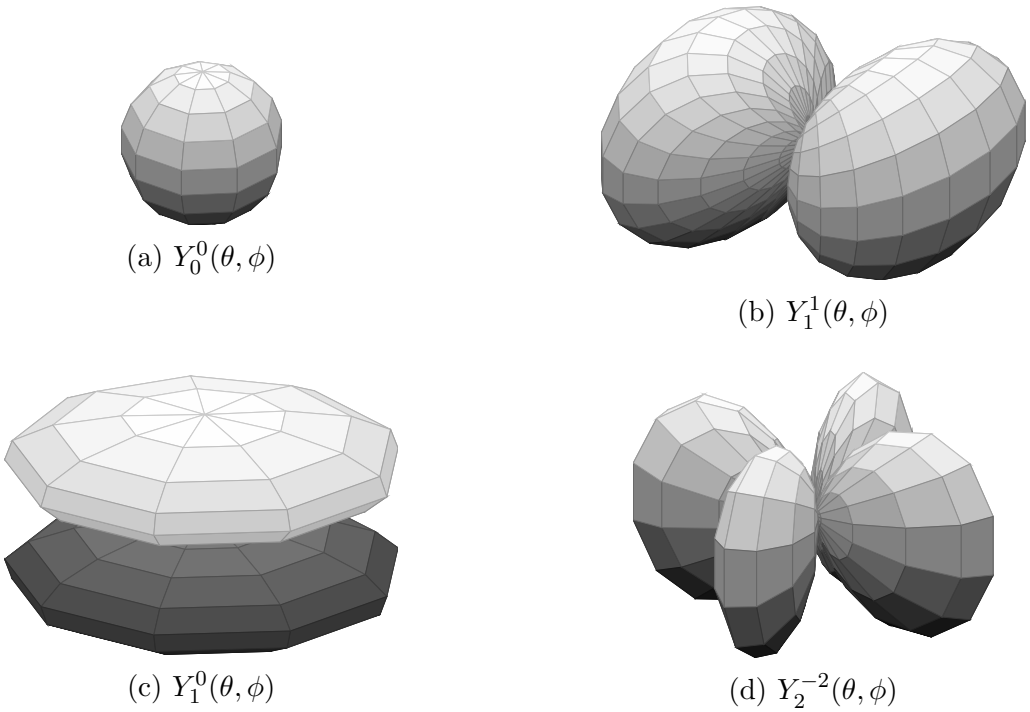
17.2.2 Parametrisation of the antenna array

In chapter 11 the radiation efficiency of an antenna array has been expressed in terms of a matrix of equivalent scattering parameters. However, the definition of these equivalent scattering parameters is still very immature and needs to be improved.

In particular, we have considered as an equivalent output port of the antenna each of the spatial directions parametrised by two angles (θ, ϕ). This parametrisation introduces a new transmission parameter for each of those directions. However, there are infinite directions in the free space and therefore the obtained scattering matrix is theoretically of infinite size. This matrix is reduced by a sampling of the angles (θ, ϕ), nevertheless a

better model is possible.

It is important to note that a function defined in the sphere, such as, for example, the radiation diagram of an antenna can be decomposed into a series of orthogonal functions in the sphere. A possible set of the mentioned orthogonal functions is the set of spherical harmonics which is usually used for this purpose. With an illustrating purpose, we show in fig. 17.1 some of the lowest order spherical harmonics.



(a) $Y_0^0(\theta, \phi)$

(b) $Y_1^1(\theta, \phi)$

(c) $Y_1^0(\theta, \phi)$

(d) $Y_2^{-2}(\theta, \phi)$

Figure 17.1: Spherical harmonics (magnitude of the real part)

These spherical harmonics can perform the same function as the different resonance modes within a resonance cavity. Therefore, in the same way as in the traditional filter design where the coupling matrix defines the coupling between the different resonant modes present in the structure, it is possible to define an equivalent model for the antennas. In this equivalent model the output ports are represented by each of these spherical harmonics, thus parametrising the effective transmission of the antenna to each of these harmonics. Note that in the case of simple antennas, the radiation pattern in space can be decomposed into a small number of spherical harmonics, namely lower order harmonics.

17.3 Multiplexer synthesis

As with the synthesis of matching filters, we can find a large number of open questions throughout part V of this thesis, dedicated to the design of multiplexers. Indeed, the theory provided in part V is largely prospective since a significant number of results are still absent. These results can be classified into two categories, each corresponding to one of the parts of the presented algorithm, namely the dimensioning of the manifold in the first instance as described in chapter 13 and the synthesis of the channel filters afterwards following the algorithm provided in chapter 14.

17.3.1 Manifold synthesis

With respect to the dimensioning of the manifold, the presented algorithm allows to obtain, for a set of determined channel filters, the dimensions of said manifold so that the manifold peaks are as far as possible from the passbands. In this study, a rational model for channel filters is assumed, allowing for a normalization of the phase of the reflection parameter of each of these filters is imposed at an infinite frequency. This condition has the objective of ensuring that the response of the channel filters does not vary too much in the adjacent bands.

However, it is still uncertain whether the aforementioned normalization at infinity is the optimal choice. Indeed, we could think that, at least in the case of a multiplexer with passbands extremely close to each other, a normalization of the reflection of each filter in the adjacent passband is a better choice.

17.3.2 Synthesis of channel filters

In the design of channel filters we find the main prospective aspect in the synthesis of multiplexers. In chapter 14 the synthesis begins by computing these channel filters when the channels are completely isolated. In this case the solution is trivial since the multiplexer synthesis problem is equivalent to a classic matching problem for each channel. Next, the presented algorithm is based on the continuation of this trivial solution from the initial point where the channels are isolated, to the final state of the multiplexer, varying a set of isolation parameters from 0 to 1. However it could happen that at some point of the trajectory from 0 to 1 the Jacobian matrix of the application that relates the isolation parameters with the coefficients of the problem is singular. In these cases we say that an accident occurs. These accidents can be caused, for example by the appearance of a manifold peak in the passband in the case that the preceding stage of manifold design has not been satisfactorily performed.

In this context a large number of interesting open questions appear. Among these questions we can find the following two.

- Are there other kinds of accidents in the presented continuation algorithm apart from the manifold peaks?
- In the event that accidents are encountered, can they always be *contoured* by a deformation of the trajectory from 0 to 1 in the complex plane, or can there be an accident barrier around the optimum point in the problem that absolutely prevents us from reaching that optimal solution?

Appendices

Appendix A:

Angular derivatives

The angular derivatives are defined as the derivative of the phase of a Schur function $S_{22} \in \mathbb{\Sigma}$ at a point $\alpha \in \mathbb{R}$ where $|S_{22}(\alpha)| = 1$. This occurs in the transmission zeros of the global system S. At these points, due to the condition of losslessness and the definition of transmission zeros we obtain $|S_{22}(\alpha)| = 1$. Therefore, only the phase of the function S_{22} is variable in the transmission zeros α which converts the interpolation condition $S_{22}(\lambda) = K_1$ where $|K_1| = 1$ in a degenerate condition. In other words, we are facing a medium interpolation condition since the aforementioned condition only carries on a real parameter.

One possible manner to overcome this problem is to add an additional condition on the phase, in particular the derivative of the phase of S_{22} at the point α, value called the angular derivative of S_{22}. In this way, the problem of interpolation with conditions on the real axis can be formulated in an equivalent way to that obtained by considering transmission zeros inside the complex plane.

Note that the concept of angular derivative which is widely available in the literature. In this context we must highlight [83] where a rigorous definition of angular derivatives is made. In the same context, this appendix is devoted to the study of the main properties of the angular derivatives which we summarise below.

A.1 Definition

Definition A.1.1 (Angular derivatives)**.** *We define the angular derivative of S_{22} as the derivative of the function $j\log(S_{22}(\lambda))$ at the transmission zeros $\alpha_i \in \mathbb{R}$.*

$$\operatorname{ang}_k S_{22}(\alpha_i) = j\left(\mathrm{D}^k(\log S_{22})\right)[\alpha_i] \qquad |S_{22}(\alpha_i)| = 1, \tag{A.1}$$

where $\left(\mathrm{D}^k(\log S_{22})\right)[\alpha_i]$ represents the k-th derivative of the function S_{22} evaluated at α_i.

Note that being α_i being a transmission zero of the system S, implies that the function S_{22} in eq. (A.1) does not vanish in a neighbourhood of α_i. Therefore the function $\operatorname{ang} S_{22}(\lambda) = j\log(S_{22}(\lambda))$ is analytic in a neighbourhood of each transmission zero $\alpha_i \in \mathbb{R}$. Additionally note that all possible determination of the complex logarithm only differ by a constant. Thus the angular derivative does not depends on the selected determination of the log.

A.2 Properties

We first prove some elementary properties about angular derivatives.

Theorem A.2.1 (Properties of angular derivatives)**.** *Suppose S_{22} is not constant. The angular derivative is real and strictly positive:*

$$\operatorname{ang} S_{22}(\alpha_i) > 0.$$

If m_i is the multiplicity of the transmission zero α_i then

$$\forall k \in 1 \dots 2m_i - 1, \;\; \Im(\operatorname{ang}_k(S_{22}(\alpha_i))) = 0$$

and

$$\Im(\mathrm{ang}_{2m_i}(S_{22}(\alpha_i))) < 0.$$

Proof. Around a transmission zero $\alpha_i \in \mathbb{R}$ we have,

$$\begin{aligned}\log(S_{22}(\alpha_i+\lambda)) &= \log(S_{22}(\alpha_i)) + \mathrm{D}(\log(S_{22}(\alpha_i)))\lambda + o(\lambda)\\ &= j\arg(\log(S_{22}(\alpha_i))) + \mathrm{D}(\log(S_{22}(\alpha_i)))\lambda + o(\lambda).\end{aligned}$$

The real part of $\log(S)$ is locally maximal on $\mathbb{R}$ in λ: this indicates that $\mathrm{Dlog}((S_{22}(\alpha_i))$ is pure imaginary. Moreover by the maximum principle the modulus of S_{22} around the point α_i cannot increase while moving inwards $\mathbb{C}^-$, that is

$$\forall \alpha \in \mathbb{C}^-,\ \ \mathrm{D}(\log(S_{22}(\alpha_i)))\alpha \in \mathbb{C}^-.$$

This imposes $\Im(\mathrm{Dlog}((S_{22}(\alpha_i))) \leq 0$.

Suppose now that $\mathrm{D}(\log(S_{22}(\alpha_i))) = 0$. As S_{22} is not a constant, one of its derivative is non zero say the k^{th}, $(k > 1)$ and again we need

$$\mathrm{D}(\log(S_{22}(\alpha_i)))\alpha^k \in \mathbb{C}^-$$

for all $\alpha \in \mathbb{C}^-$. But this is impossible because z^k take all possible arguments when the direction α varies in $\mathbb{C}^-$, and therefore $\Im(\mathrm{Dlog}((S_{22}(\alpha_i))) < 0$.

Eventually note that on the real line $2\log(|S_{22}|(\lambda)) = \log(1 - S_{12}(\lambda)S_{12}^*(\lambda))$, where the numerator of the fraction $S_{12}S_{12}^*$ has a zero of order $2m_i$, as the fraction itself. We therefore have

$$\forall k \in 1 \ldots 2m_i - 1,\ \ \Re(\mathrm{D}^k(\log(S_{22})(\alpha_i))) = 0$$

and

$$\Re(\mathrm{D}^{2m_i}(\log(S_{22})(\alpha_i))) \neq 0.$$

Noting again that the modulus of S_{22} can not increase in the real vicinity of α_i imposes that $\Re(\mathrm{D}^{2m_i}(\log(S_{22})(\alpha_i))) < 0$. □

Next we introduce the first of the main theorems in this appendix which provides us with an alternative expression for the angular derivative.

Theorem A.2.2 (Integral expression of the angular derivatives)**.** *Let $S_{22} \in \mathbb{\Sigma}_R^N$ with $N \in \mathbb{N}$ and $R \in \mathbb{P}_+^{2N}$ such that $R(\alpha_i) = 0$ at a point $\alpha_i \in \mathbb{R}$. Denote by $\beta_i \in \mathbb{C}^-$ with $i \in [1, n]$ and $n \leq N$ the zeros of S_{22} in the lower half plane. The angular derivatives of S_{22} at α_i are expressed as*

$$\mathrm{ang}S_{22}(\alpha_i) = \frac{-1}{2\pi}\int_{\mathbb{R}} \frac{\log|S_{22}^O(\tau)|^2}{(\tau-\alpha_i)^2}d\tau + 2\sum_{n=1}^{N}\Im\left(\frac{1}{\beta_n-\alpha_i}\right). \tag{A.2}$$

To perform the proof, we first decompose the function S_{22} as the product of a minimum phase function S_{22}^O times a Blaschke product S_{22}^B. Then we have

$$\begin{aligned}S_{22}(\alpha_i) &= S_{22}^O(\alpha_i)\cdot S_{22}^B(\alpha_i),\\ \log S_{22}(\alpha_i) &= \log S_{22}^O(\alpha_i) + \log S_{22}^B(\alpha_i),\\ \mathrm{ang}S_{22}(\alpha_i) &= \mathrm{ang}S_{22}^O(\alpha_i) + \mathrm{ang}S_{22}^B(\alpha_i).\end{aligned}$$

Next we threat both terms S_{22}^O and S_{22}^B separately. We give first special attention to the angular derivatives of a Blaschke product. Those derivatives can be expressed in function of the zeros of the Blaschke product only as stated in the following lemma.

Lemma A.2.1 (Angular derivatives of a Blaschke product)**.** *Consider the Blaschke product S_{22}^B of degree $N > 0$*

$$S_{22}^B = \prod_{n=1}^{N} \frac{\lambda - \beta_n}{\lambda - \overline{\beta_n}} \qquad \beta_n \in \mathbb{C}^-,\ \forall n \in [1, N].$$

The angular derivatives of the function S_{22}^B are computed as

$$\mathrm{ang}_k S_{22}^B(\alpha_i) = 2 \sum_{n=1}^{N} \Im \left(\frac{1}{(\beta_n - \alpha_i)^k} \right) \qquad k \geq 1.$$

Proof. We follow now Fano's procedure, first divide by $S_{22}^B(\alpha_i)$

$$S_{22}^B(\lambda) = S_{22}^B(\alpha_i) \frac{S_{22}(\lambda)}{S_{22}^B(\alpha_i)} = \prod_{n=1}^{N} \frac{\lambda - \beta_n}{\lambda - \overline{\beta_n}} \frac{\alpha_i - \overline{\beta_n}}{\alpha_i - \beta_n} = \prod_{n=1}^{N} \frac{1 - \frac{\lambda - \alpha_i}{\beta_n - \alpha_i}}{1 - \frac{\lambda - \alpha_i}{\overline{\beta_n} - \alpha_i}},$$

$$\log S_{22}^B(\lambda) = \log S_{22}^B(\alpha_i) + \sum_{n=1}^{N} \log \left(1 - \frac{\lambda - \alpha_i}{\beta_n - \alpha_i} \right) - \sum_{n=1}^{N} \log \left(1 - \frac{\lambda - \alpha_i}{\overline{\beta_n} - \alpha_i} \right).$$

Each factor $\log \left(1 - \frac{\lambda - \alpha_i}{\beta_n - \alpha_i} \right)$ can be developed around 1 when $\lambda \to \alpha_i$:

$$- \log \left(1 - \frac{\lambda - \alpha_i}{\beta_n - \alpha_i} \right) \Big|_{\lambda \to \alpha_i} = - \log S_{22}^B(\alpha_i) + \sum_{k=1}^{\infty} \frac{1}{k} \frac{(\lambda - \alpha_i)^k}{(\beta_n - \alpha_i)^k},$$

therefore

$$\begin{aligned} \log(S_{22}^B(\lambda))\big|_{\lambda \to \alpha_i} &= \log(S_{22}^B(\alpha_i)) + \sum_{k=1}^{\infty} \left(\sum_{n=1}^{N} \frac{1}{(\overline{\beta_n} - \alpha_i)^k} - \frac{1}{(\beta_n - \alpha_i)^k} \right) (\lambda - \alpha_i)^k \\ &= \log(S_{22}^B(\alpha_i)) - \sum_{k=1}^{\infty} 2j \sum_{n=1}^{N} \Im \left(\frac{1}{(\beta_n - \alpha_i)^k} \right) (\lambda - \alpha_i)^k . \end{aligned} \tag{A.3}$$

The Taylor development of $\log(\lambda)$ around α_i

$$\log(S_{22}^B(\lambda))\big|_{\lambda \to \alpha_i} = \log(S_{22}^B(\alpha_i)) - j \sum_{k=1}^{\infty} \frac{\mathrm{ang}_k S_{22}^B(\alpha_i)}{k} (\lambda - \alpha_i)^k . \tag{A.4}$$

Now by inspection, comparing eqs. (A.3) and (A.4) we obtain the values $\mathrm{ang}_k S_{22}^B(\alpha_i)$

$$\mathrm{ang}_k S_{22}^B(\alpha_i) = 2 \sum_{n=1}^{N} \Im \left(\frac{1}{(\beta_n - \alpha_i)^k} \right) \qquad k \geq 1.$$

□

Next let us compute the derivative of the minimum phase factor S_{22}^O.

Lemma A.2.2 (Angular derivatives of a minimum phase function)**.** *Consider a minimum phase function S_{22}^O. We have*

$$\mathrm{ang}(\log(S_{22}^O))(\alpha_i) = \frac{-1}{2\pi}\int_{\mathbb{R}} \frac{\log|S_{22}^O(\tau)|^2}{(\tau-\alpha_i)^2} d\tau.$$

Proof. The evaluation of the logarithm of the outer function S_{22}^O at a point $\lambda \in \mathbb{C}^-$ is given in terms of its real part on the boundary by the classical Riesz-Herglotz representation:

$$\log(S_{22}^O)(\lambda) = \frac{1}{j\pi}\int_{\mathbb{R}} \log|S_{22}^O(\tau)| \left(\frac{1}{\lambda-\tau} + \frac{\tau}{1+\tau^2}\right) d\tau.$$

Later representation is unique up to an imaginary constant. Here the normalization $\Im(log(S_{22}(-i))) = 0$ has been taken. Differentiation with respect do $z \in \mathbb{C}^-$, yields

$$\mathrm{D}\log(S_{22}^O)(\lambda) = \frac{-1}{j\pi}\int_{\mathbb{R}} \frac{\log|S_{22}^O(\tau)|}{(\lambda-\tau)^2} d\tau.$$

Eventually taking the limit $\lambda \to \alpha_i$ and noting that $log(|S_{22}^O(\alpha)|)$ has a zero of order two in $\tau = \alpha_i$ allows to apply the Lebesgue's dominated convergence theorem and yields,

$$\mathrm{D}\log(S_{22}^O)(\alpha_i) = \frac{-1}{j\pi}\int_{\mathbb{R}} \frac{\log|S_{22}^O(\tau)|}{(\alpha_i-\tau)^2} d\tau,$$

which concludes the proof. □

A.3 Degenerate chaining

After introducing these properties, we use them to study the chaining operation, when trasmission zeros $\alpha_i \in \mathbb{R}$ are considered. Denote now by L a 2×2 lossless scattering matrix and define the Schur function $F_{22} \in \mathbb{\Sigma}$. We proved in lemma 3.5.1 that the function issue of the chaining operation $S_{22} = F_{22} \circ L$ is a Schur function. To do so we used the fact that the denominator of S_{22} can not vanish at a point $\alpha_i \in \mathbb{C}^-$ (inside the domain of analyticity). Nevertheless, if a transmission zero happens at the boundary ($\alpha_i \in \mathbb{R}$), a simplification in S_{22} may occurs. We provide then the following lemma

Lemma A.3.1 (Simplifications at transmission zeros)**.** *Given the 2×2 scattering matrix L and $F_{22} \in \mathbb{\Sigma}$, the denominator of the function $S_{22} \in \mathbb{\Sigma}$ computed as $S_{22} = F_{22} \circ L$ could have a zero simple (not multiple) at each transmission zero $\alpha_i \in \mathbb{R}$ that is common to L and F_{22}, namely $|L_{11}(\alpha_i)| = |L_{22}(\alpha_i)| = |F_{22}(\alpha_i)| = 1$. In this case a simplification occurs in S_{22}.*

Proof. Consider the expression of $S_{22} = F_{22} \circ L$ at $\alpha_i \in \mathbb{R}$

$$S_{22} = \frac{L_{22}}{\overline{L_{11}}} \frac{\overline{L_{11}} - F_{22}}{1 - L_{11}F_{22}}. \tag{A.5}$$

Note that the denominator vanish at $\alpha_i \in \mathbb{R}$ if and only if $F_{22}(\alpha_i)L_{11}(\alpha_i) = 1$, namely $F_{22}(\alpha_i) = \overline{L_{11}(\alpha_i)}$ and $|F_{22}(\alpha_i)| = |L_{11}(\alpha_i)| = 1$ (both L and F have a transmission zero at α_i). Suppose now that $F_{22}(\alpha_i)L_{11}(\alpha_i) = 1$ and compute the derivative

$$\mathrm{D}(1 - F_{22}L_{11}) = -F_{22}\mathrm{D}L_{11} - L_{11}\mathrm{D}F_{22} = -F_{22}L_{11}\left(\frac{\mathrm{D}F_{22}}{F_{22}} + \frac{\mathrm{D}L_{11}}{L_{11}}\right).$$

In the parenthesis we have the derivatives of $\log F_{22}$ and $\log L_{11}$, Therefore

$$\mathrm{D}(1 - F_{22}L_{11}) = jF_{22}L_{11}\left(j\mathrm{D}\log F_{22} + j\mathrm{D}\log L_{11}\right).$$

Evaluated at α_i we have $F_{22}(\alpha_i)L_{11}(\alpha_i) = 1$. Additionally note that by theorem A.2.1 we have $j\mathrm{D}\log F_{22}(\alpha_i) = \mathrm{ang}F_{22}(\alpha_i) > 0$ and $j\mathrm{D}\log L_{11}(\alpha_i) = \mathrm{ang}L_{11}(\alpha_i) > 0$. Thus the derivative of the denominator can not vanish at α_i.

$$\mathrm{D}(1 - F_{22}L_{11})[\alpha_i] = \mathrm{ang}F_{22}(\alpha_i) + \mathrm{ang}L_{11}(\alpha_i) > 0.$$

Finally if $F_{22}(\alpha_i) = \overline{L_{11}(\alpha_i)}$ and $|F_{22}(\alpha_i)| = |L_{11}(\alpha_i)| = 1$, namely the denominator of eq. (A.5) vanishes, then the numerator vanish as well producing the simplification. □

Corollary A.3.1. *Given a scattering matrix L of degree M with transmission zeros*

$$\alpha_i, \alpha_2, \cdots, \alpha_{Mr} \in \mathbb{R}$$

and

$$\alpha_{Mr+1}, \cdots, \alpha_M \in \mathbb{C}^-.$$

Let $F_{22} \in \mathbb{\Sigma}^K$. From the previous theorem a maximum of M_r simplification may occur after chaining the function F_{22} of degree K with the 2×2 matrix L of McMillan degree M. Therefore the function S_{22} resulting of the operation $S_{22} = F_{22} \circ L$ is of degree N with

$$K + M - Mr \leq N \leq K + M.$$

After showing that one simplification might occur at each transmission zero $\alpha_i \in \mathbb{R}$, we are interested to know how the angular derivatives are modified at those points upon the chaining operation.

Lemma A.3.2 (Degenerate chaining)**.** *Consider again the 2×2 scattering matrix L and the function $F_{22} \in \mathbb{\Sigma}$. The reflection of the global system $S_{22} = F_{22} \circ L$ satisfies at each transmission zero $\alpha_i \in \mathbb{R}$:*

$$\begin{aligned} S_{22}(\alpha_i) &= L_{22}(\alpha_i), \\ \mathrm{ang}S_{22}(\alpha_i) &= \mathrm{ang}L_{22}(\alpha_i) - \frac{|\mathrm{D}L_{21}(\alpha_i)|^2}{\mathrm{ang}L_{11}(\alpha_i) + \mathrm{ang}F_{22}(\alpha_i)}. \end{aligned} \tag{A.6a}$$

Proof. From eq. (3.3), the function $\log(S_{22})$ can be expressed as:

$$\log(S_{22}) = \log(L_{22}) + \log\left(1 + \varphi\right) \qquad ,$$

with

$$\varphi = \frac{L_{21}F_{22}L_{12}}{L_{22}(1 - L_{11}F_{22})}.$$

Note here that φ is not defined at α_i. Indeed, by lemma A.3.1, one pole-zero cancellation occurs in $\varphi(\lambda)$ at $\lambda = \alpha_i$. Let us compute the limit $\lim_{\lambda\to\alpha_i} \varphi(\lambda)$ by applying the *L'Hôpital* rule

$$\begin{aligned}\lim_{\lambda\to\alpha_i} \varphi(\lambda) &= \lim_{\lambda\to\alpha_i} \frac{\mathrm{D}(L_{21}F_{22}L_{12})[\lambda]}{L_{22}(\lambda)\mathrm{D}(1-L_{11}F_{22})[\lambda]} \\ &= \lim_{\lambda\to\alpha_i} \frac{-1}{L_{22}(\lambda)} \cdot \frac{\mathrm{D}(L_{21}F_{22}L_{12})[\lambda]}{L_{11}(\lambda)\mathrm{D}F_{22}(\lambda) + F_{22}(\lambda)\mathrm{D}L_{11}(\lambda)} \\ &= \frac{-j}{L_{22}(\alpha_i)F_{22}(\alpha_i)L_{11}(\alpha_i)} \cdot \frac{\mathrm{D}(L_{21}F_{22}L_{12})[\alpha_i]}{\mathrm{ang}F_{22}(\alpha_i) + \mathrm{ang}L_{11}(\alpha_i)} \end{aligned},$$

where $\mathrm{ang}F_{22}(\lambda) + \mathrm{ang}L_{11}(\lambda) > 0$.

Note that $L_{21}F_{22}L_{12} = 0$. Therefore the derivative of φ vanish and we obtain

$$S_{22}(\alpha_i) = L_{22}(\alpha_i)$$

Next we compute the limit $\lim_{\lambda\to\alpha_i} \mathrm{D}^{2m_i-1}\varphi(\lambda)$ at α_i, namely the value of the angular derivative of S_{22} after simplification. We have

$$\lim_{\lambda\to\alpha_i} \mathrm{D}^{2m_i-1}\phi(\lambda)[\alpha_i] = \frac{-j}{L_{22}(\alpha_i)F_{22}(\alpha_i)L_{11}(\alpha_i)} \cdot \frac{\mathrm{D}(L_{21})[\alpha_i]F_{22}(\alpha_i)\mathrm{D}(L_{12})[\alpha_i]}{\mathrm{ang}F_{22}(\alpha_i) + \mathrm{ang}L_{11}(\alpha_i)}. \tag{A.8}$$

Note from eq. (2.16) that

$$L_{11} \cdot \overline{\mathrm{D}L_{21}} = -\overline{L_{22}} \cdot \mathrm{D}L_{12} \qquad \forall\lambda \in \mathbb{R}. \tag{A.9}$$

From eqs. (A.8) and (A.9) and since $|L_{22}(\alpha_i)|^2 = 1$ we have

$$\lim_{\lambda\to\alpha_i} \mathrm{D}\phi(\lambda) = j\frac{|\mathrm{D}L_{21}[\alpha_i]|^2}{\mathrm{ang}F_{22}(\alpha_i) + \mathrm{ang}L_{11}(\alpha_i)}.$$

Finally eq. (A.6a) follows

$$\begin{aligned}\mathrm{ang}S_{22}(\alpha_i) &= \mathrm{ang}L_{22}(\alpha_i) + j\lim_{\lambda\to\alpha_i} \mathrm{D}\phi(\lambda) \\ &= \mathrm{ang}L_{22}(\alpha_i) - \frac{|\mathrm{D}L_{21}[\alpha_i]|^2}{\mathrm{ang}F_{22}(\alpha_i) + \mathrm{ang}L_{11}(\alpha_i)}.\end{aligned}$$ □

From this result we obtain a necessary condition that relates the angular derivatives of L_{22} and $S_{22} = F_{22} \circ L$ at a transmission zero $\alpha_i \in \mathbb{R}$.

Corollary A.3.2. *Last term in eq.* (A.6a) *is a real positive quantity. Thus given the* 2×2 *lossless scattering matrix* L *and the Schur function* f*, then the function* $S_{22} \in \mathbb{\Sigma}$ *obtained from the chaining operation* $S_{22} = f \circ L$ *at a transmission zero* $\alpha_i \in \mathbb{R}$ *satisfies*

$$\begin{aligned}S_{22}(\alpha_i) &= L_{22}(\alpha_i), \\ \mathrm{ang}S_{22}(\alpha_i) &\leq \mathrm{ang}L_{22}(\alpha_i).\end{aligned}$$

References

[83] J. A. Ball, I. Gohberg, and L. Rodman, *Interpolation of rational matrix functions*, ser. Operator theory, advances and applications. Birkhäuser Verlag, 1990. [Online]. Available: https://books.google.fr/books?id=iR{_}vAAAAMAAJ

Appendix B:

Nevanlinna-Pick interpolation and schur recursion with interpolation conditions inside the lower half plane

In chapter 4, a large part of the theory concerning the matching problem has been developed based on the Nevanlinna-Pick interpolation problem and the Schur recursal algorithm. In addition, numerous demonstrations known in the literature are referenced which can be found in [84].

In order to provide the reader with a quick review of the classical concepts in chapter 4 and at the same time without saturating too much chapter 4 with long demonstrations, we have decided to include in this appendix a part of the theory related to the mentioned interpolation problem.

We recall therefore Nevanlinna parametrisation of the solutions of a Schur interpolation problem. We will use that parametrisation later on to obtain an alternative characterisation of the functions satisfying Fano-Youla's interpolation conditions. The Schur recursion can be found in [84, chapter IV, section 6], however to fit the notation used in the rest of this thesis and for the reader convenience we particularise it for analytic functions in the lower half plane.

B.1 Schur recursion

We are concerned about the following interpolation problem

Problem B.1.1 (Nevanlinna-Pick interpolation problem)**.** *Consider the set of points* $\alpha_1, \alpha_2 \cdots \alpha_M \in \mathbb{C}^-$ *and* $\gamma_1, \gamma_2 \cdots \gamma_M \in \mathbb{D}$*. Now state the interpolation conditions*

$$f(\alpha_i) = \gamma_i \qquad \forall i \in [1, M]. \tag{B.1}$$

First define the set of Schur functions satisfying the first m interpolation conditions in eq. (B.1)

$$\mathbb{E}^m = \{f \in \mathbb{\Sigma} \mid f(\alpha_i) = \gamma_i;\ 1 \leq i \leq m\},$$

with $m \leq M$.

Note that

$$\mathbb{E}^m \subset \mathbb{E}^{m-1} \subset \cdots \subset \mathbb{E}^2 \subset \mathbb{E}^1.$$

Also note that the only condition for $\mathbb{E}^1$ not to be empty is $|\gamma_1| \leq 1$. Moreover if $|\gamma_1| = 1$, from the principle of maximum modulus, $\mathbb{E}^1$ contains only one function $f(x) = \gamma_1$. Additionally, we can obtain the following characterisation of $\mathbb{E}^1$ due originally to Nevanlinna and closely related to Fano-Youla's characterisation.

Theorem B.1.1 (Nevanlinna characterisation of $\mathbb{E}^1$)**.** *The function* $f_1 \in \mathbb{\Sigma}$ *verifies the simple interpolation condition* $f_1(\alpha_1) = \gamma_1$ *if and only if*

$$f_1(\lambda) = \frac{\gamma_1(\lambda - \overline{\alpha_1}) + (\lambda - \alpha_1) f_2(\lambda)}{(\lambda - \overline{\alpha_1}) + \overline{\gamma_1}(\lambda - \alpha_1) f_2(\lambda)}, \tag{B.2}$$

where $f_2 \in \mathbb{\Sigma}$.

Proof of sufficiency. First we show that all f_1 in the form eq. (B.2) is a Schur function.

$$f_1(\lambda) = \frac{\gamma_1(\lambda - \overline{\alpha_1}) + (\lambda - \alpha_1) f_2(\lambda)}{(\lambda - \overline{\alpha_1}) + \overline{\gamma_1}(\lambda - \alpha_1) f_2(\lambda)} = \frac{\gamma_1 + \frac{\lambda - \alpha_1}{\lambda - \overline{\alpha_1}} f_2(\lambda)}{1 + \overline{\gamma_1}\frac{\lambda - \alpha_1}{\lambda - \overline{\alpha_1}} f_2(\lambda)} = \frac{\gamma_1 - \hat{f}_2(\lambda)}{1 - \overline{\gamma_1}\hat{f}_2(\lambda)},$$

where $b(\lambda) = -\frac{\lambda-\alpha_1}{\lambda-\overline{\alpha_1}}$ is a Blaschke product. Note now that if $\hat{f}_2 = bf_2 \in \mathbb{S}$ and $\gamma_1 \in \mathbb{S}$ then f_1 is a Schur function. To conclude the proof we check that f_1 satisfy the interpolation condition

$$f_1(\alpha_1) = \frac{\gamma_1(\alpha_1 - \overline{\alpha_1})}{(\alpha_1 - \overline{\alpha_1})} = \gamma_1.$$

□

Proof of necessity. To proof necessity, we check that every $f_1 \in \mathbb{S}$ satisfying $f_1(\alpha_1) = \gamma_1$ can be express in the form eq. (B.2). Let $f_1 \in \mathbb{S}$, and compute f_2 by inverting eq. (B.2). Then we proof that $f_2 \in \mathbb{S}$.

$$f_2(\lambda) = \frac{f_1(\lambda - \overline{\alpha_1}) - \gamma_1(\lambda - \overline{\alpha_1})}{(\lambda - \alpha_1) - f_1\overline{\gamma_1}(\lambda - \alpha_1)} = \frac{f_1 - \gamma_1}{1 - f_1\overline{\gamma_1}} \frac{(\lambda - \overline{\alpha_1})}{(\lambda - \alpha_1)},$$

which corresponds to the product of the function $h(\lambda) = \frac{\gamma_1 - f_1}{1 - f_1\overline{\gamma_1}}$ and the inverse of the Blaschke product b. We can check that $|f_2(\lambda)| \leq 1$ for all $\lambda \in \mathbb{R}$ since $|h(\lambda)| = |\delta(\gamma_1, f_1(\lambda))| \leq 1$ and $|b(\lambda)| = 1$ for all $\lambda \in \mathbb{R}$. Furthermore the function b^{-1} has one single pole in $\mathbb{C}^-$ at $\lambda = \alpha_1$. This pole cancels out due to the fact that $f_1(\alpha_1) = \gamma_1$ which makes the numerator vanish as well at $\lambda = \alpha_1$. Thus $f_2 \in \mathbb{S}$. □

Using Fano-Youla's theory a physical interpretation to this characterisation is possible. We are characterising the Schur functions (i.e. passive stable reflection coefficients) that satisfy an interpolation condition $f_1(\alpha_1) = \gamma_1$.

The characterisation says $f_1 \in \mathbb{E}^1$ if and only if it can be obtained as the chaining of a function $f \in \mathbb{S}$ with a 2×2 scattering matrix, $S_{22} = f \circ L$, where

$$L(\lambda) = \frac{1}{\lambda - \overline{\alpha_1}} \begin{pmatrix} -\overline{\gamma_1}(\lambda - \alpha_1) & \sqrt{1 - |\gamma_1|^2}(\lambda - \overline{\alpha_1}) \\ \sqrt{1 - |\gamma_1|^2}(\lambda - \alpha_1) & \gamma_1(\lambda - \overline{\alpha_1}) \end{pmatrix}. \tag{B.3}$$

Note that this matrix satisfy

$$L_{21}(\alpha_1) = 0,$$
$$L_{22}(\alpha_1) = \gamma_1.$$

The main idea is illustrated in fig. B.1. The function L_{22} belongs to $\mathbb{E}^1$. Moreover, since the L_{22} has a tranmission zero at α_1, the value at α_1 is not modified after closing port one by chaining any Schur reflection. Additionally, if $f_1 \in \mathbb{E}^1$, then we can dechain the load L from f_1 obtaining the parametrisation $f_1 = f_2 \circ L$.

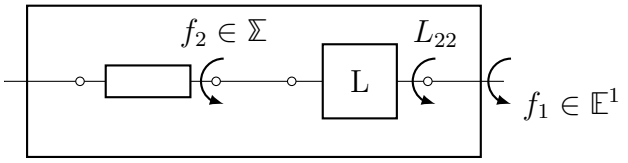


Figure B.1: Conceptual parametrisation of the functions $S_{22} \in \mathbb{E}^1$

Computing from eq. (3.6) the expression of f_1 we obtain

$$\begin{aligned} f_1(\lambda) = f_2(\lambda) \circ L(\lambda) &= \frac{L_{22}(\lambda)}{L_{11}^{\star}(\lambda)} \frac{L_{11}(\lambda)^{\star} - f_2(\lambda)}{1 - L_{11}(\lambda) f_2(\lambda)} \\ &= \frac{\lambda - \alpha_1}{\lambda - \overline{\alpha_1}} \frac{\gamma_1 \frac{\lambda - \overline{\alpha_1}}{\lambda - \alpha_1} + f_2(\lambda)}{1 + \overline{\gamma_1} \frac{\lambda - \alpha_1}{\lambda - \overline{\alpha_1}} f_2(\lambda)} \\ &= \frac{\gamma_1(\lambda - \overline{\alpha_1}) + (\lambda - \alpha_1) f_2(\lambda)}{(\lambda - \overline{\alpha_1}) + \overline{\gamma_1}(\lambda - \alpha_1) f_2(\lambda)}, \end{aligned}$$

which is equivalent to eq. (B.2).

Remark B.1.1. *Note that if $|\gamma_1| = 1$, by the maximum modulus principle $f_1 = \gamma_1$ is the unique interpolation function. Equivalently from the previous expression we obtain $f_1 = \gamma_1$ independently of the function f_2*

$$f_1(\lambda) = \frac{\gamma_1(\lambda - \overline{\alpha_1}) + (\lambda - \alpha_1) f_2(\lambda)}{(\lambda - \overline{\alpha_1}) + \overline{\gamma_1}(\lambda - \alpha_1) f_2(\lambda)} = \gamma_1 \frac{(\lambda - \overline{\alpha_1}) + \overline{\gamma_1}(\lambda - \alpha_1) f_2(\lambda)}{(\lambda - \overline{\alpha_1}) + \overline{\gamma_1}(\lambda - \alpha_1) f_2(\lambda)} = \gamma_1.$$

In this case $\mathbb{E}^1$ is a singleton containing only the function $f_1 = \gamma_1$. From the physical point of view, a unimodular interpolation condition such as $|\gamma_1| = 1$ implies $L_{11} = \gamma_1$ and therefore $L_{12} = 0$. As it is illustrated in fig. B.2, the interpolation condition is satisfied with a single phase shift with no transmission. The function f_2 has no influence in this case.

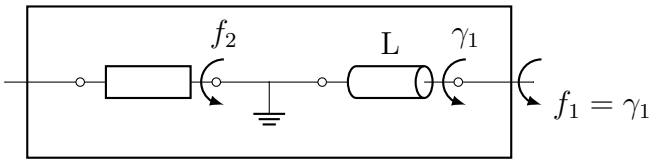


Figure B.2: Unique unimodular reflection satisfying $f_1(\alpha_1) = \gamma_1$

Suppose now $\mathbb{E}^1$ contains more than one function and $f_1 \in \mathbb{E}^1$, then

$$f_1(\lambda) = \frac{\gamma_1(\lambda - \overline{\alpha_1}) + (\lambda - \alpha_1) f_2(\lambda)}{(\lambda - \overline{\alpha_1}) + \overline{\gamma_1}(\lambda - \alpha_1) f_2(\lambda)},$$

with $f_2 \in \mathbb{\Sigma}$. Now since $f_1 \in \mathbb{E}^1$ for any arbitrary function $f_2 \in \mathbb{\Sigma}$, we can compute the required values of $f_2(\alpha_2)$ such that $f_1 \in \mathbb{E}^2$. That is, the values of $f_2(\alpha_2)$ such that the interpolation conditions also holds at α_2 ($f_1(\alpha_2) = \gamma_2$). We have

$$f_1(\alpha_2) = \gamma_2 = \frac{\gamma_1(\alpha_2 - \overline{\alpha_1}) + (\alpha_2 - \alpha_1) f_2(\alpha_2)}{(\alpha_2 - \overline{\alpha_1}) + \overline{\gamma_1}(\lambda_2 - \alpha_1) f_2(\alpha_2)}.$$

Then we obtain the value of $f_2(\alpha_2)$ as

$$f_2(\alpha_2) = \frac{\gamma_2 - \gamma_1}{1 - \overline{\gamma_1}\gamma_2} \frac{\alpha_2 - \alpha_1}{\alpha_2 - \overline{\alpha_1}} = \gamma_2^{(2)}.$$

The idea then is to parametrise again f_2 such that

$$f_2(\alpha_2) = \gamma_2^{(2)} \tag{B.4}$$

Therefore, as before, $\mathbb{E}^2$ is not empty iff $|\gamma_2^{(2)}| \leq 1$

$$\left| \frac{\gamma_1 - \gamma_2}{1 - \overline{\gamma_1}\gamma_2} \frac{\alpha_2 - \alpha_1}{\alpha_2 - \overline{\alpha_1}} \right| \leq 1.$$

Since $\Im(\alpha_i) \leq 0$ we have

$$\left| \frac{\alpha_2 - \alpha_1}{\alpha_2 - \overline{\alpha_1}} \right| \leq 1.$$

Thus we obtain a condition on the maximum pseudo-hyperbolic distance from γ_1 to γ_2

$$\delta(\gamma_2, \gamma_1) \leq \left| \frac{\alpha_2 - \alpha_1}{\alpha_2 - \overline{\alpha_1}} \right|. \tag{B.5}$$

Again, if $|\gamma_2^{(2)}| = 1$ then the only function satisfying eq. (B.4) is $f_2 = \gamma_2^{(2)}$ and therefore the set $\mathbb{E}^2$ is a singleton containing only the Blaschke product

$$f_1(\lambda) = \frac{\gamma_1(\lambda - \overline{\alpha_1}) + (\lambda - \alpha_1)\gamma_2^{(2)}}{(\lambda - \overline{\alpha_1}) + \overline{\gamma_1}(\lambda - \alpha_1)\gamma_2^{(2)}} = -\overline{\gamma_2^{(2)}} \frac{\gamma_1(\lambda - \overline{\alpha_1}) + (\lambda - \alpha_1)\gamma_2^{(2)}}{\overline{\gamma_1}(\lambda - \alpha_1) + (\lambda - \overline{\alpha_i})\overline{\gamma_2^{(2)}}}.$$

If we suppose $\mathbb{E}^2$ contains at least two different functions, namely eq. (B.5) holds with inequality, then the parametrisation we are seeking is

$$f_1 = \frac{\gamma_1(\lambda - \overline{\alpha_1}) + (\lambda - \alpha_1)f_2}{(\lambda - \overline{\alpha_1}) + \overline{\gamma_1}(\lambda - \alpha_1)f_2} \qquad f_2 = \frac{\gamma_2^{(2)}(\lambda - \overline{\alpha_2}) + (\lambda - \alpha_2)f_3}{(\lambda - \overline{\alpha_2}) + \overline{\gamma_2^{(2)}}(\lambda - \alpha_2)f_3}, \tag{B.6}$$

with $f_3 \in \mathbb{\Sigma}$. Now we compute the interpolation conditions over f_3 at α_3 such that $f_1 \in \mathbb{E}^3$. First we need to propagate that condition from f_1 to f_2

$$f_1(\alpha_3) = \gamma_3 = \frac{\gamma_1(\alpha_3 - \overline{\alpha_1}) + (\alpha_3 - \alpha_1)f_2(\alpha_3)}{(\alpha_3 - \overline{\alpha_1}) + \overline{\gamma_1}(\alpha_3 - \alpha_1)f_2(\alpha_3)} \quad \longrightarrow \quad f_2(\alpha_3) = \gamma_3^{(2)} = \frac{\gamma_3 - \gamma_1}{1 - \overline{\gamma_1}\gamma_3} \frac{\alpha_3 - \alpha_1}{\alpha_3 - \overline{\alpha_1}}.$$

Then we can write the interpolation value of $f_3(\alpha_3)$

$$f_3(\alpha_3) = \frac{\gamma_3^{(2)} - \gamma_2^{(2)}}{1 - \overline{\gamma_2^{(2)}}\gamma_3^{(2)}} \frac{\alpha_3 - \overline{\alpha_2}}{\alpha_3 - \alpha_2} = \gamma_3^{(3)}.$$

Once again, the set $\mathbb{E}^3$ is non-empty iff $|\gamma_3^{(3)}| \leq 1$

$$\delta(\gamma_3^{(2)}, \gamma_2^{(2)}) \leq \left| \frac{\alpha_3 - \alpha_2}{\alpha_3 - \overline{\alpha_2}} \right|.$$

Therefore problem B.1.1 is feasible iff at each step of the Schur recursion the interpolation value $\gamma_k^{(k)}$ imposed on $f_k(\alpha_i)$ satisfies $|\gamma_k^{(k)}| \leq 1$.

This recursion can be seen as a succession of elementary blocks in the form eq. (B.3). Each of the blocks L_i ensures the interpolation condition $f(\alpha_i) = \gamma_i$ introducing a transmission zero at α_i (fig. B.3). In addition, the output reflection of the block L_i at α_i is

computed such that $f_1(\alpha_i) = \gamma_i$

$$\begin{aligned} L_{1_{22}}(\alpha_1) &= \gamma_1, \\ L_{2_{22}}(\alpha_2) \circ L_1(\alpha_2) &= \gamma_2, \\ L_{3_{22}}(\alpha_3) \circ L_2(\alpha_3) \circ L_1(\alpha_3) &= \gamma_3, \\ &\vdots \\ f_{M+1}(\alpha_M) \circ L_M(\alpha_M) \circ \cdots \circ L_2(\alpha_M) \circ L_1(\alpha_M) &= \gamma_M. \end{aligned}$$

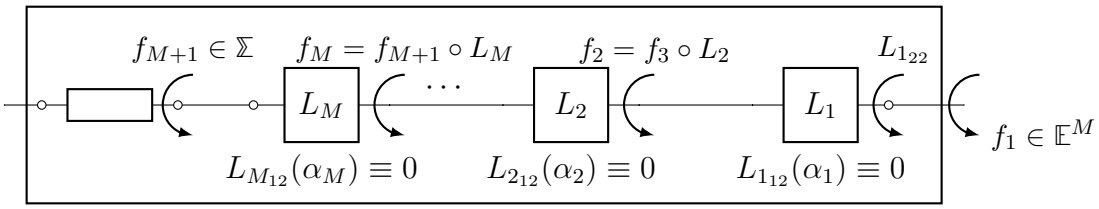


Figure B.3: Parametrisation of the interpolation problem $f(\alpha_i) = \gamma_i$

Note that if the m-th reflection $L_{m_{22}}(\alpha_i)$ becomes unimodular, the recursion stops obtaining the structure shown in fig. B.4. In this case the function f_1 is a unique Blaschke product of degree m.

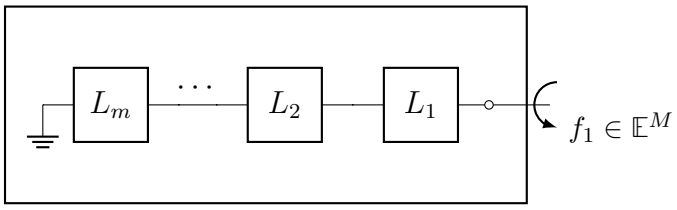


Figure B.4: Construction of the interpolant function when the solution to the interpolation problem is a unique Blaschke product of degree $m < M$.

Theorem B.1.2 (Cardinality of $\mathbb{E}^M$)**.** *The set $\mathbb{E}^M$ contains at least two functions, iff*

$$\begin{aligned} |\gamma_k| &< 1, \\ \delta(\gamma_{k+1}^{(k)}, \gamma_k^{(k)}) &< \left| \frac{\alpha_{k+1} - \alpha_k}{\alpha_{k+1} - \overline{\alpha_k}} \right| \qquad \forall k \in [1, M-1], \end{aligned}$$

where $\gamma_k^{(1)} = \gamma_k$ for all $k \in [1, M]$ and

$$\gamma_k^{(l+1)} = \frac{\gamma_k^{(l)} - \gamma_l^{(l)}}{1 - \overline{\gamma_l^{(l)}} \gamma_k^{(l)}} \frac{\alpha_k - \alpha_l}{\alpha_k - \overline{\alpha_l}} \qquad \forall l \in [1, M-1] \qquad \forall k \in [l+1, M].$$

If we suppose that $\mathbb{E}^3$ contains at least two functions, then it can be characterised by adding to eq. (B.6) the interpolation condition $f_3(\alpha_3) = \gamma_3^{(3)}$. Thus can obtain the characterisation of $\mathbb{E}^M$ proceeding by induction [84, Chapter IV,Lemma 6.1]

Theorem B.1.3 (Nevanlinna characterisation of $\mathbb{E}^M$)**.** *Given the set of points*

$$\alpha_1, \cdots, \alpha_M \in \mathbb{D},$$

and the values

$$\gamma_1, \cdots, \gamma_M \in \mathbb{C}^-.$$

If the set of Schur interpolant $\mathbb{E}^M$ is a singleton, it contains only a Blaschke product of degree $m < M$. Assuming $\mathbb{E}^M$ is not a singleton, then $S_{22}(\lambda) \in \mathbb{E}^M$ iff

$$S_{22}(\lambda) = \frac{A_M(\lambda) + B_M(\lambda) f(\lambda)}{C_M(\lambda) + D_M(\lambda) f(\lambda)}, \tag{B.7}$$

where $f \in \mathbb{\Sigma}$, $A_0 = 0$, $B_0 = 1$, $C_0 = 1$, $D_0 = 0$ and

$$\begin{aligned}
A_i(\lambda) &= (\lambda - \overline{\alpha_i}) \left(A_{i-1}(\lambda) + \gamma_k^{(k)} B_{i-1}(\lambda) \right), \\
B_i(\lambda) &= (\lambda - \alpha_i) \left(B_{i-1}(\lambda) + \overline{\gamma_k^{(k)}} A_{i-1}(\lambda) \right), \\
C_i(\lambda) &= (\lambda - \overline{\alpha_i}) \left(C_{i-1}(\lambda) + \gamma_k^{(k)} D_{i-1}(\lambda) \right), \\
D_i(\lambda) &= (\lambda - \alpha_i) \left(D_{i-1}(\lambda) + \overline{\gamma_k^{(k)}} C_{i-1}(\lambda) \right).
\end{aligned}$$

Remark B.1.2. *Note that this characterisation is equivalent to Fano-Youla's characterisation since chaining of the elements $L = L_M \circ \cdots \circ L_1$ in fig. B.3 is a matrix having transmission zeros at the point α_i and reflection $L(\alpha_i) = \gamma_i$ for all $i \in [1, M]$. Therefore the Schur function S_{22} satisfying $S_{22}(\alpha_i) = \gamma_i$ are parametrised as $f \circ L$ with $f \in \mathbb{\Sigma}$ (fig. B.5).*

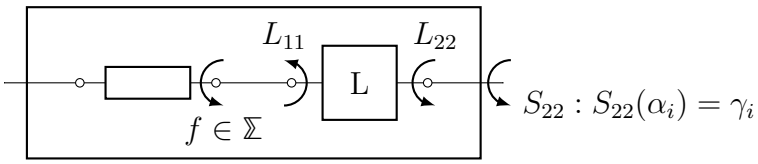


Figure B.5: Nevanlinna-Youla Parametrisation of the functions $S_{22} \in \mathbb{E}^M$

Remark B.1.3. *Also note that $A_0 = D_0^\star$ and $B_0 = C_0^\star$. Now assuming $A_{i-1} = D_{i-1}^\star$ and $B_{i-1} = C_{i-1}^\star$ we have*

$$\begin{aligned}
A_i &= (\lambda - \overline{\alpha_i}) \left(A_{i-1} + \gamma_k^{(k)} B_{i-1} \right) = (\lambda - \overline{\alpha_i}) \left(D_{i-1}^\star + \gamma_k^{(k)} C_{i-1}^\star \right) = D_i^\star, \\
B_i &= (\lambda - \alpha_i) \left(B_{i-1} + \overline{\gamma_k^{(k)}} A_{i-1} \right) = (\lambda - \alpha_i) \left(C_{i-1}^\star + \overline{\gamma_k^{(k)}} D_{i-1}^\star \right) = C_i^\star.
\end{aligned}$$

Therefore we obtain by induction $A_M = D_M^\star$ and $B_M = C_M^\star$.

Remark B.1.4. *Next assume $|C_i(\omega)| > |D_i(\omega)|$ for all $\omega \in \mathbb{R}$ and develop*

$$|C_i + \gamma_k^{(k)} D_i|^2 = |C_i|^2 + 2\Re\left(\gamma_k^{(k)} D_i \overline{C_i} \right) + |\gamma_k^{(k)} D_i|^2, \tag{B.8}$$

$$|D_i + \overline{\gamma_k^{(k)}} C_i|^2 = |D_i|^2 + 2\Re\left(\gamma_k^{(k)} D_i \overline{C_i} \right) + |\gamma_k^{(k)} C_i|^2. \tag{B.9}$$

From eqs. (B.8) *and* (B.9) *we have*

$$|C_i + \gamma_k^{(k)} D_i|^2 - |D_i + \overline{\gamma_k^{(k)}} C_i|^2 = (1 - |\gamma_k^{(k)}|^2)\left(|C_i|^2 - |D_i|^2\right) > 0.$$

Then if $|C_i| > |D_i|$ *we have* $|C_{i+1}| > |D_{i+1}|$ *and since* $|C_0| > |D_0|$ *we obtain*

$$|C_M(\omega)| > |D_M(\omega)| \qquad \forall \omega \in \mathbb{R}.$$

Note the same recursion holds for the leadings terms of the polynomials C_{i+1} *and* D_{i+1}*, namely* $ld(C_{i+1})$ *and* $ld(D_{i+1})$

$$ld(C_{i+1}) = ld(C_i) + \gamma_k^{(k)} ld(D_i),$$
$$ld(D_{i+1}) = ld(D_i) + \overline{\gamma_k^{(k)}} ld(C_i).$$

Therefore since $ld(C_0) > ld(D_0)$ *by induction we have* $ld(C_M) \geq ld(D_M)$*. If we now take the limits when* $\omega \to \infty$

$$\lim_{\omega \to \infty} |C_M(\omega)| - |D_M(\omega)| > 0.$$

Moreover if $C_{i-1}(\lambda) \neq 0$ *for all* $\lambda \in \overline{\mathbb{C}^-}$ *then by Rouche's theorem* $C_i(\lambda) = (\lambda - \overline{\alpha_i})\left(C_{i-1}(\lambda) + \gamma_i^{(i)} D_{i-1}(\lambda)\right)$ *has no zeros in the lower half plane. Thus* $C_M(\lambda) \neq 0$ *for all* $\lambda \in \overline{\mathbb{C}^-}$*. We conclude that*

$$L_{11}(\lambda) \equiv -\frac{D_M(\lambda)}{C_M(\lambda)} \in \mathbb{\Sigma}.$$

Now from eq. (B.7)

$$S_{22} = \frac{B_M}{C_M} \frac{\frac{A_M}{B_M} + f}{1 + \frac{D_M}{C_M} f} = \frac{C_M^\star}{C_M} \frac{\frac{D_M^\star}{C_M^\star} + f}{1 + \frac{D_M}{C_M} f} = \frac{L_{11}}{L_{22}^\star} \frac{f - L_{11}^\star}{1 - L_{11} f} = f \circ L,$$

with $L_{22}(\lambda) \equiv -\frac{D_M^\star(\lambda)}{C_M(\lambda)}$*. We obtain the expression of the chaining operation illustrated in fig. B.5 showing the relation between Nevanlinna characterisation and Fano-Youla's results. This result can be found in [84, Chapter IV, Lemma 6.2-6.3]*

References

[84] J. B. Garnett, *Bounded Analytic Functions*, ser. Pure and Applied Mathematics. Elsevier Science, 1981. [Online]. Available: https://books.google.fr/books?id=DVLO9gJ66{_}YC

Appendix C:

Schur recursion with simple interpolations conditions on the real line

In appendix B we reviewed the Nevanlinna-Pick interpolation problem which assumes a set of interpolation conditions inside the analyticity domain.

This interpolation problem appears upon writing the necessary and sufficient conditions that ensure the extraction of a fixed load from the port 2 of the global system. Those conditions are stated as a Schur interpolation conditions at the transmission zeros α_i of the load with the interpolation values provided by the reflection coefficient of the load at port 2.

Although the case of transmission zeros within the complex plane may seem the most general case, which in theory is true since in real life transmission zeros never occur exactly on the frequency axis, it is also quite common to find a load with a transmission zero sufficiently close to the real axis or far enough from the passband so that it can be considered at infinity. In these cases, Fano-Youla's parametrisation considering transmission zeros inside the complex plane provides very poorly conditioned interpolation conditions, close to an indetermination. To overcome this problem, in practice we assume that the aforementioned zeros are exactly on the real axis so they are treated in a special manner.

With this motivation we review, in the present appendix, an interpolation problem similar to that formulated by Nevanlinna-Pick where the interpolation occurs in the real axis and the interpolation values take a slightly more complex form introducing an inequality on the angular derivative of the interpolating function at the interpolation points. This problem was first studied in [85] and a detailed theory can be found in [86] or [87] where the problem of boundary interpolation is studied with the notion of angular derivatives. In this Appendix we provide, following Nevanlinna theory, the parametrisation of the feasible set $\mathbb{G}$ when the interpolation conditions occurs on the boundary, namely the load to be de-embedded presents transmission zeros on the frequency axis. This constitute the case analogous to the theory developed in appendix B for a load with transmission zeros inside the complex plane.

C.1 Elementary de-chaining matrix for a transmission zero on the boundary

Before dealing with the characterisation of $\mathbb{G}$, it is necessary to first introduce the concept of elementary section which allows the de-chaining of a simple transmission zero on the real axis. We consider now a function $S_{22} \in \mathbb{\Sigma}^N$ with a transmission zero $\alpha_i \in \mathbb{R}$, namely $|S_{22}(\alpha_i)| = 1$. Additionally we have

$$S_{22}(\alpha_i) = \beta_i, \tag{C.1}$$
$$\text{ang} S_{22}(\alpha_i) = \gamma_i. \tag{C.2}$$

The purpose of this section is to show that the function S_{22} can be expressed as the chaining $S_{22} = F_{22} \circ C$ where C is a 2×2 scattering matrix of McMillan degree 1 and $F_{22} \in \mathbb{\Sigma}^{N-1}$ with $|F_{\alpha_i}| \neq 1$. In other words the transmission zero at the point $\alpha_i \in \mathbb{R}$ can

be extracted by dechaining the matrix

$$C = \frac{1}{\lambda - (\alpha_i + j/\rho)} \begin{pmatrix} j\overline{\beta_i}/\rho & \lambda - \alpha_i \\ \lambda - \alpha_i & -j\beta_i/\rho \end{pmatrix}. \tag{C.3}$$

Let us therefore state the theorem first.

Theorem C.1.1 (De-chaining of simple transmission zero on the boundary)**.** *Let S_{22} be a non-constant Schur function satisfying eqs.* (C.1) *and* (C.2) *with $|\beta_i| = 1$ and $\alpha_i \in \mathbb{R}$. Then a matrix C in the form given by eq.* (C.3) *of McMillan degree 1 can be extracted from S_{22}. Namely S_{22} is written as $S_{22} = S_2 \circ C$ with $S_2 \in \mathbb{\Sigma}$ and*

Proof. For this proof we operate on the positive real function given by the admittance associated to S_{22} when the value $Y_0 = j\beta_i$ is taken as reference. This admittance is obtained by Cayley transform as

$$Y_1(\lambda) = \frac{Y_0 + S_{22}(\lambda)}{Y_0 - S_{22}(\lambda)} = \frac{\beta_i - jS_{22}(\lambda)}{\beta_i + jS_{22}(\lambda)}. \tag{C.4}$$

Note that $Y_1 \in PR(\overline{\mathbb{C}^-})$. Moreover by the interpolation condition in eq. (C.1) the function $Y_1(\lambda)$ in eq. (C.4) has a pole at α_i. Additionally, positive real functions have only simple poles and zeroes on the boundary of the analyticity domain (in this case the real axis) with strictly positive residues at each of those poles and strictly positive derivative at each zero on the boundary. Let us now the residue of $Y_i(\lambda)$ at the pole on the real axis at α_i. This residue takes the expression

$$\mathrm{res}_{\alpha_i}(Y_1) = \frac{\beta_i + S_{22}(\alpha_i)}{-(-j\mathrm{D}S_{22}(\alpha_i))},$$

where $-j\mathrm{D}S_{22}(\alpha_i)$ represents the derivative of S_{22} in the direction $-j$, namely corresponding to the lower half plane. Using now eq. (C.2) we have

$$\mathrm{D}S_{22}(\alpha_i) = -jS_{22}(\alpha_i)\mathrm{ang}S_{22}(\alpha_i) = -j\beta_i\gamma_i.$$

Therefore

$$\mathrm{res}_{\alpha_i}(Y_1) = \frac{2}{\gamma_i}.$$

Next we subtract from Y_1 an elementary positive real term Y_R with the expression

$$Y_R = \frac{2/\rho}{s - j\alpha_i},$$

with $\rho > 0$. Again from the elementary theory on positive real functions, the function Y_1 remains positive real after subtracting Y_R as long as $2/\rho \leq \mathrm{res}_{\alpha_i}(Y_1)$. Additionally if $2/\rho = \mathrm{res}_{\alpha_i}(Y_1)$ the pole of Y_1 at $j\alpha_i$ is removed completely and therefore the McMillan degree is decreased by one. After subtraction we have

$$Y_2 = Y_1 - Y_R = \frac{\beta_i + S_{22}}{\beta_i - S_{22}} - Y_R = \frac{\beta_i(1 - Y_R) + S_{22}(1 + Y_R)}{\beta_i - S_{22}}.$$

After subtracting the function Y_R we can now invert the Cayley transform to obtain the Schur function S_2 corresponding to the positive real function Y_2. Therefore

$$\begin{aligned} S_2 &= -\beta_i \frac{1-Y_2}{1+Y_2}, \\ &= -\beta_i \frac{\beta_i Y_R - S_{22}(2+Y_R)}{\beta_i(2-Y_R)+S_{22}Y_R}. \end{aligned}$$

Finally let us express the function S_{22} in terms of S_2. We have

$$\begin{aligned} S_2\beta_i(2-Y_R) + S_2S_{22}Y_R &= -\beta_i^2 Y_R - S_{22}\beta_i(2+Y_R), \\ S_{22}(S_2Y_R - \beta_i(2+Y_R)) &= -\beta_i^2 Y_R - S_2\beta_i(2-Y_R). \end{aligned}$$

Thus using the fact that $|\beta_i| = 1$ we obtain the expression for S_{22}

$$S_{22} = \frac{\beta_i Y_R + S_2(2-Y_R)}{(2+Y_R) - S_2\overline{\beta_i}Y_R}.$$

Introducing now the expression of $Y_R(\lambda)$ with the change of variable $s = j\lambda$, namely $Y_R(\lambda) = \frac{-j2/\rho}{\lambda-\alpha_i}$, we have

$$S_{22} = \frac{-j\beta_i/\rho + [(\lambda-\alpha_i) + j/\rho]S_2}{[(\lambda-\alpha_i) - j/\rho] - \overline{j\beta_i/\rho}S_2}. \tag{C.5}$$

Note that expression eq. (C.6) corresponds to the chaining operation $S_{22} = S_2 \circ C$ where C is the 2×2 elementary matrix expressed in the Belevitch form as

$$C = \frac{1}{q_C}\begin{pmatrix} -p_C^\star & r_C^\star \\ r_C & p_C \end{pmatrix},$$

where $q_C q_C^\star = p_C p_C^\star + r_C r_C^\star$. Let us the expression of the chaining operation $S_2 \circ C$

$$\begin{aligned} S_2 \circ C &= C_{22}\frac{C_{21}C_{21}S_2}{1-C_{11}S_2} \\ &= \frac{p_C}{q_C} + \frac{\frac{r_C^\star}{q_C}\frac{r_C}{q_C}S_2}{1+\frac{p_C^\star}{q_C}S_2} \\ &= \frac{p_C q_C + (p_C p_C^\star + r_C r_C^\star)S_2}{q_C(q_C + p_C^\star S_2)} \\ &= \frac{p_C + q_C^\star S_2}{q_C + p_C^\star S_2}. \end{aligned} \tag{C.6}$$

Identifying the terms in eqs. (C.5) and (C.6) we have

$$\begin{aligned} p_C &= -j\beta_i/\rho, \\ r_C &= \lambda - \alpha_i, \\ q_C &= \lambda - (\alpha_i + j/\rho). \end{aligned}$$

This matrix C corresponds to the matrix introduced in theorem C.1.1 concluding therefore the proof. □

C.2 Interpolation problem with boundary interpolation conditions

Let give now the characterisation of the problem analogous to problem B.1.1 when the interpolation conditions happens at the boundary ($\mathbb{R}$) of the analyticity domain. We consider in this case the interpolation problem where uni-modular values and the angular derivatives are specified at the points α_i. Let us denote now by f the Schur function S_{22} in the previous section.

Problem C.2.1 (Nevanlinna-Pick with boundary conditions)**.** *Let $\alpha_1, \alpha_2 \cdots \alpha_M$ be distinct points in $\mathbb{R}$, then consider the set of uni-modular values $\beta_1, \beta_2 \cdots \beta_M \in \mathbb{T}$ and positive real values $\gamma_1, \gamma_2 \cdots \gamma_M \in \mathbb{R} : \gamma_i \geq 0$. Now we consider the following interpolation conditions over the function $f \in \mathbb{\Sigma}$*

$$f(\alpha_i) = \beta_i \qquad \forall i \in [1, M], \tag{C.7a}$$
$$\text{ang} f(\alpha_i) \leq \gamma_i \qquad \forall i \in [1, M]. \tag{C.7b}$$

We define now a class of interpolant functions to problem C.2.1

$$\mathbb{B}^m = \{f \in \mathbb{\Sigma} \mid f(\alpha_i) = \beta_i;\ \text{ang} f(\alpha_i) \leq \gamma_k;\ 1 \leq i \leq m\},$$

with $m \leq M$. We have

$$\mathbb{B}^m \subset \mathbb{B}^{m-1} \subset \cdots \subset \mathbb{B}^2 \subset \mathbb{B}^1.$$

Note that if $\beta_1 \in \mathbb{T}$ and $\gamma_1 = 0$, $\mathbb{B}^1$ contains only the function $f(x) = \beta_1$. Additionally, we can obtain the following characterisation of $\mathbb{B}^1$ with an argument equivalent to the Schur recursion. We use the degenerate matrix introduced in theorem C.1.1 to obtain a function $f_1 \in \mathbb{B}^1$, namely satisfying $C_{22}(\alpha_1) = \beta_1$ and $\text{ang} C_{22}(\alpha_1) \leq \beta_1$. Then from eq. (3.6) we state the characterisation of $\mathbb{B}^1$

Theorem C.2.1 (Characterisation of $\mathbb{B}^1$ with boundary interpolation conditions)**.** *Given $\beta_1 \in \mathbb{T}$ and $\gamma_1 > 0$, the function $f_1 \in \mathbb{\Sigma}$ verifies $f_1(\alpha_1) = \beta_1$ and $\text{ang} f_1(\alpha_i) \leq \gamma_1$ if and only if*

$$f_1(\lambda) = \frac{\beta_1 + \overline{\Gamma}_1(\lambda) f_2(\lambda)}{\Gamma_1(\lambda) + \overline{\beta}_1 f_2(\lambda)}, \tag{C.8}$$

where $\Gamma_1(x) = 1 + j\gamma_1(\lambda - \alpha_i)$ and $f_2 \in \mathbb{\Sigma}$.

Proof of sufficiency. We verify that the function f_1 in eq. (C.8) verifies eqs. (C.7a) and (C.7b) for $i = 1$. Note that $\Gamma_1(\alpha_i) = 1$ therefore $f_1(\alpha_1) = \beta_1$ follows since

$$f_1(\lambda) = \frac{\beta_1 + f_2(\lambda)}{1 + \overline{\beta}_1 f_2(\lambda)} = \beta_1 \frac{1 + \overline{\beta}_1 f_2(\lambda)}{1 + \overline{\beta}_1 f_2(\lambda)} = \beta_1.$$

Additionally $\text{ang} f_1(\alpha_i) \leq \gamma_1$ follows from corollary A.3.2. □

Proof of necessity. To proof follows from theorem C.1.1. □

Function $f_2(\lambda)$ can be obtain by inverting eq. (C.8)

$$f_2(\lambda) = \frac{\beta_1 - \Gamma_1(\lambda) f_1(\lambda)}{\overline{\beta}_1 f_1(\lambda) - \overline{\Gamma}_1(\lambda)}.$$

We develop next the characterisation of the set $\mathbb{B}^m$ using an inductive argument as it was done in appendix B for the interpolation conditions inside the analyticity domain.

Remark C.2.1. *Note if $\gamma_1 = 0$ then $\Gamma_1(\lambda) = 1$ and we obtain $f_1 = \beta_1$ independently of the function f_2*

$$f_1(\lambda)|_{\Gamma=1} = \left.\frac{\beta_1 + \overline{\Gamma}_1(\lambda) f_2(\lambda)}{\Gamma_1(\lambda) + \overline{\beta}_1 f_2(\lambda)}\right|_{\Gamma_1(\lambda)=1} = \frac{\beta_1 + f_2(\lambda)}{1 + \overline{\beta}_1 f_2(\lambda)} = \beta_1 \frac{1 + \overline{\beta}_1 f_2(\lambda)}{1 + \overline{\beta}_1 f_2(\lambda)} = \beta_1.$$

In this case $\mathbb{B}^1$ is a singleton containing only the function $f_1 = \beta_1$.

Suppose $\mathbb{B}^1$ contains more than one function and $f_1 \in \mathbb{B}^1$, then

$$f_1(\lambda) = \frac{\beta_1 + \overline{\Gamma}_1(\lambda) f_2(\lambda)}{\Gamma_1(\lambda) + \overline{\beta}_1 f_2(\lambda)},$$

with $f_2 \in \mathbb{\Sigma}$. Now since $f_1 \in \mathbb{B}^1$ for any arbitrary function $f_2 \in \mathbb{\Sigma}$, we can compute the required values of $f_2(\alpha_2)$ and $\mathrm{ang} f_2(\alpha_2)$ such that $f_1 \in \mathbb{B}^2$. We have

$$f_1(\alpha_2) = \beta_2 = \frac{\beta_1 + \overline{\Gamma}_1(\alpha_2) f_2(\alpha_2)}{\Gamma_1(\alpha_2) + \overline{\beta}_1 f_2(\alpha_2)}.$$

Then we obtain the value of $f_2(\alpha_2)$ as

$$f_2(\alpha_2) = \beta_2 \frac{\beta_1 \overline{\beta}_2 - \Gamma_1(\alpha_2)}{\overline{\beta}_1 \beta_2 - \overline{\Gamma}_1(\alpha_2)} = \beta_2^{(2)}.$$

Note here that

$$\left| \frac{\beta_1 \overline{\beta}_2 - \Gamma_1(\alpha_2)}{\overline{\beta}_1 \beta_2 - \overline{\Gamma}_1(\alpha_2)} \right| = 1.$$

Therefore we have $|\beta_2^{(2)}| = |\beta_2| = 1$. Now compute $\mathrm{ang} f_1(\alpha_2) = j\mathrm{D}\log f_1(\lambda)|_{\lambda=\alpha_2}$

$$j\mathrm{D}\log f_1(\lambda) = j \frac{f_2(\lambda)\mathrm{D}\overline{\Gamma}_1(\lambda) + \overline{\Gamma}_1(\lambda)\mathrm{D}f_2(\lambda)}{\beta_1 + \overline{\Gamma}_1(\lambda) f_2(\lambda)} - j \frac{\mathrm{D}\Gamma_1(\lambda) + \overline{\beta}_1 \mathrm{D}f_2(\lambda)}{\Gamma_1(\lambda) + \overline{\beta}_1 f_2(\lambda)},$$

where $\mathrm{D}\Gamma_1(\lambda)) = j\gamma_1$. Evaluating at $\lambda = \alpha_2$ and using $f_2(\alpha_2) = \beta_2^{(2)}$

$$\begin{aligned}
\mathrm{ang} f_1(\alpha_2) &= \frac{\beta_2^{(2)}\gamma_1 + \overline{\Gamma}_1(\alpha_2) j\mathrm{D}f_2(\alpha_2)}{\beta_1 + \overline{\Gamma}_1(\alpha_2)\beta_2^{(2)}} + \frac{\gamma_1 - \overline{\beta}_1 j\mathrm{D}f_2(\alpha_2)}{\Gamma_1(\alpha_2) + \overline{\beta}_1 \beta_2^{(2)}} \\
&= \frac{\gamma_1 + \overline{\beta}_2^{(2)}\overline{\Gamma}_1(\alpha_2) j\mathrm{D}f_2(\alpha_2)}{\overline{\Gamma}_1(\alpha_2) + \beta_1 \overline{\beta}_2^{(2)}} + \frac{\gamma_1 - \overline{\beta}_1 j\mathrm{D}f_2(\alpha_2)}{\Gamma_1(\alpha_2) + \overline{\beta}_1 \beta_2^{(2)}} \\
&= \gamma_1 \Re\left(\frac{1}{\overline{\Gamma}_1(\alpha_2) + \overline{\beta}_1 \beta_2^{(2)}} \right) + \\
&+ \left(\frac{\overline{\beta}_2^{(2)} \overline{\Gamma}_1(\alpha_2)}{\overline{\Gamma}_1(\alpha_2) + \beta_1 \overline{\beta}_2^{(2)}} - \frac{\overline{\beta}_1}{\Gamma_1(\alpha_2) + \overline{\beta}_1 \beta_2^{(2)}} \right) j\mathrm{D}f_2(\alpha_2).
\end{aligned}$$

We have $\Re\left(\Gamma_1(\lambda)\right) = 1$, then

$$\begin{aligned}\operatorname{ang} f_1(\alpha_2) &= \gamma_1 \frac{1}{1+\Re\left(\overline{\beta}_1\beta_2^{(2)}\right)} + \\ &+ \frac{\overline{\beta}_2^{(2)}|\overline{\Gamma}_1(\alpha_2)|^2 + \overline{\beta}_1\overline{\Gamma}_1(\alpha_2) - \overline{\beta}_1\overline{\Gamma}_1(\alpha_2) - \overline{\beta}_2^{(2)}}{\left|\Gamma_1(\alpha_2)+\overline{\beta}_1\beta_2^{(2)}\right|^2} j\mathrm{D}f_2(\alpha_2) \\ &= \gamma_1 \frac{1}{1+\Re\left(\overline{\beta}_1\beta_2^{(2)}\right)} + \frac{|\Gamma_1(\alpha_2)|^2 - 1}{\left|\Gamma_1(\alpha_2)+\overline{\beta}_1\beta_2^{(2)}\right|^2} \cdot j\overline{\beta}_2^{(2)}\mathrm{D}f_2(\alpha_2).\end{aligned}$$

Note that $j\overline{\beta}_2^{(2)}\mathrm{D}f_2(\alpha_2)$ is the angular derivative $\operatorname{ang} f_2(\alpha_2)$. Thus

$$\begin{aligned}\operatorname{ang} f_1(\alpha_2) &= \gamma_1 \frac{1}{1+\Re\left(\overline{\beta}_1\beta_2^{(2)}\right)} + \frac{|\Gamma_1(\alpha_2)|^2 - 1}{\left|\Gamma_1(\alpha_2)-\overline{\beta}_1\beta_2^{(2)}\right|^2} \cdot \operatorname{ang} f_2(\alpha_2) \\ &= \frac{\gamma_1}{1+\Re\left(\overline{\beta}_1\beta_2^{(2)}\right)} + \frac{\gamma_1^2(\alpha_2-\alpha_1)^2}{\left|\Gamma_1(\alpha_2)+\overline{\beta}_1\beta_2^{(2)}\right|^2} \cdot \operatorname{ang} f_2(\alpha_2).\end{aligned}$$

We compute the value of $\operatorname{ang} f_2(\alpha_2)$ such that $\operatorname{ang} f_1(\alpha_2) = \gamma_2$

$$\operatorname{ang} f_2(\alpha_2) = \frac{\left|\Gamma_1(\alpha_2)+\overline{\beta}_1\beta_2^{(2)}\right|^2}{\gamma_1^2(\alpha_2-\alpha_1)^2}\left(\gamma_2 - \frac{\gamma_1}{1+\Re\left(\overline{\beta}_1\beta_2^{(2)}\right)}\right) = \gamma_2^{(2)}.$$

The question that arises now is whether $\gamma_2^{(2)}$ is positive. We have

$$\gamma_2^{(2)} \geq 0 \quad \Longleftrightarrow \quad \gamma_2 \geq \frac{\gamma_1}{1+\Re\left(\overline{\beta}_1\beta_2^{(2)}\right)}.$$

If $\gamma_2^{(2)} = 0$ we obtain $f_2 = \beta_2^{(2)}$. Then the class $\mathbb{B}^2$ is a singleton containing only a Blaschke product, namely the function

$$f_1(\lambda) = \frac{\beta_1 + \overline{\Gamma}_1(\lambda)\beta_2^{(2)}}{\Gamma_1(\lambda)+\overline{\beta}_1\beta_2^{(2)}} = \beta_2^{(2)}\frac{\beta_1\overline{\beta}_2^{(2)} + \overline{\Gamma}_1(\lambda)}{\Gamma_1(\lambda)+\overline{\beta}_1\beta_2^{(2)}}.$$

Conversely if $\gamma_2^{(2)} > 0$ we can parametrise the functions $f_1 \in \mathbb{B}^2$ as

$$f_1(\lambda) = \frac{\beta_1 + \overline{\Gamma}_1(\lambda)f_2(\lambda)}{\Gamma_1(\lambda)+\overline{\beta}_1 f_2(\lambda)} \qquad\qquad f_2(\lambda) = \frac{\beta_2^{(2)} + \overline{\Gamma}_2^{(2)}(\lambda)f_3(\lambda)}{\Gamma_2^{(2)}(\lambda)+\overline{\beta}_2^{(2)} f_3(\lambda)},$$

where $\Gamma_1(\lambda) = 1 + j\gamma_1(\lambda-\alpha_1)$ and $\Gamma_2^{(2)}(\lambda) = 1 + j\gamma_2^{(2)}(\lambda-\alpha_2)$.

Let us denote now by $\gamma_k^{(i)}$ the angular derivative imposed on f_i at the point α_k while β_k^i represents the value $f_i(\alpha_k)$ for all $l \in [1, M]$ and $k \in [l, M]$. Additionally we set $\beta_k^{(1)} = \beta_k$

and $\gamma_k^{(1)} = \gamma_k$. Defining now The values β_k^i, $\gamma_k^{(i)}$ for all $i \in [2, M]$ are obtained recursively as

$$\beta_k^{(i)} = \beta_k^{(i-1)} \frac{\beta_{i-1}^{(i-1)} \overline{\beta}_k^{(i-1)} - \Gamma_{i-1}^{(i-1)}(\alpha_k)}{\overline{\beta}_{i-1}^{(i-1)} \beta_k^{(i-1)} - \overline{\Gamma}_{i-1}^{(i-1)}(\alpha_k)}$$

$$\gamma_k^{(i)} = \frac{\left|\Gamma_{i-1}^{(i-1)}(\alpha_k) + \overline{\beta}_{i-1}^{(i-1)} \beta_k^{(i)}\right|^2}{\left(\gamma_{i-1}^{(i-1)}(\alpha_k - \alpha_{i-1})\right)^2} \left(\gamma_k^{(i-1)} - \frac{\gamma_{i-1}^{(i-1)}}{1 + \Re\left(\overline{\beta}_{i-1}^{(i-1)} \beta_k^{(i)}\right)}\right),$$

with $\Gamma_k^{(i-1)}(\lambda) = 1 + j\gamma_k^{(i-1)}(\lambda - \alpha_k)$.

We state now that problem C.2.1 is feasible if and only if at each step of the recursion the maximum angular derivative $\gamma_k^{(k)}$ imposed on $f_k(\alpha_k)$ satisfies $\gamma_k^{(k)} \geq 0$

Theorem C.2.2 (Cardinality of $\mathbb{B}^M$ with boundary interpolation conditions)**.** *The $\mathbb{B}^M$ contains at least two functions, if and only if*

$$\gamma_k^{(k)} \geq 0 \qquad \forall k \in [1, M].$$

Proceeding by induction we obtain the characterisation of $\mathbb{B}^M$ (assuming it contains at least two functions)

Theorem C.2.3 (Characterisation of $\mathbb{B}^M$ with boundary interpolation conditions)**.** *Consider the set of points $\alpha_1, \cdots, \alpha_M \in \mathbb{D}$ and $\gamma_1, \cdots, \gamma_M \in \mathbb{C}^-$, define the polynomials $A_i(\lambda)$, $B_i(\lambda)$, $C_i(\lambda)$ and $D_i(\lambda)$ for all $i \in [2, M]$ as*

$$\begin{aligned}
A_i(\lambda) &= A_{i-1}(\lambda)\Gamma_i^{(i)}(\lambda) + B_{i-1}\beta_i^{(i)}, \\
B_i(\lambda) &= B_{i-1}(\lambda)\overline{\Gamma}_i^{(i)}(\lambda) + A_{i-1}\overline{\beta}_i^{(i)}, \\
C_i(\lambda) &= C_{i-1}(\lambda)\Gamma_i^{(i)}(\lambda) + D_{i-1}\beta_i^{(i)}, \\
D_i(\lambda) &= D_{i-1}(\lambda)\overline{\Gamma}_i^{(i)}(\lambda) + C_{i-1}\overline{\beta}_i^{(i)},
\end{aligned}$$

with $\Gamma_k^{(i)} = 1 + j\gamma_k^{(i)}(\lambda - \alpha_k)$ while for $i = 0$ we have $A_0 = 0$, $B_0 = 1$, $C_0 = 1$, $D_0 = 0$. Then $S_{22}(\lambda) \in \mathbb{B}^M$ if and only if

$$S_{22}(\lambda) = \frac{A_M(\lambda) + B_M(\lambda) f(\lambda)}{C_M(\lambda) + D_M(\lambda) f(\lambda)},$$

where $f \in \mathbb{\Sigma}$.

C.3 Dealing with transmission zeros at infinity

In appendix A we provide the necessary and sufficient conditions to be able to express the function S_{22} as the chainig of a Schur function with the given load $S_{22} = F_{22} \circ L$. Those are given as a set of interpolation conditions at the transmission zeros of the load α_i. However we have not considered the case where the transmission zero happens at infinity ($\alpha_i = \infty$) yet. It can be shown that, if the load L has a transmission zero at infinity, the theory developed in appendix A can be reformulated equivalently by applying the change of variable $\lambda \to \frac{1}{\lambda}$.

References

[85] R. Gudipati and W. K. Chen, "Explicit formulas for the design of broadband matching bandpass equalizers with Chebyshev response," in *Circuits and Systems, 1995. ISCAS '95., 1995 IEEE International Symposium on*, vol. 3, apr 1995, pp. 1644—-1647 vol.3.

[86] D. Sarason, "Nevanlinna-Pick interpolation with boundary data," *Integral Equations and Operator Theory*, 1998.

[87] J. A. Ball, I. Gohberg, and L. Rodman, *Interpolation of rational matrix functions*, ser. Operator theory, advances and applications. Birkhäuser Verlag, 1990. [Online]. Available: https://books.google.fr/books?id=iR{_}vAAAAMAAJ

Appendix D:

Global system synthesis with optimal oscillating transfer functions. Proof of unicity.

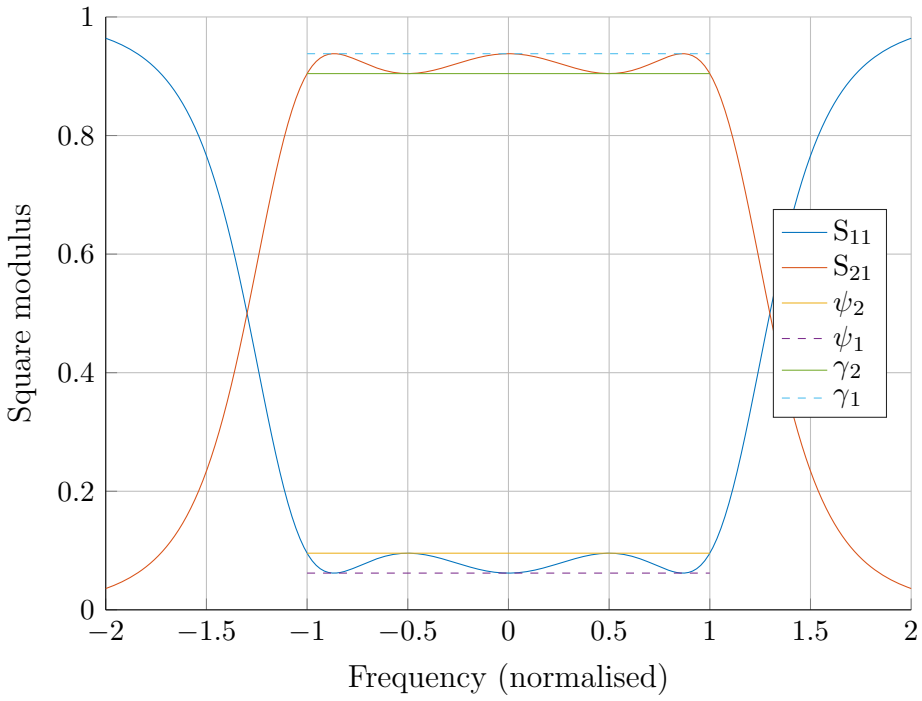


Figure D.1: Equioscillating response.

In Fano, the theory of matching was introduced for the first time from the point of view of the global system. This theory dictated in the simplest case, that a two-port system S of McMillan degree N can be represented as the chaining of a device of degree K and a fixed load L of degree M, assuming that both the system and the load has no finite transmission zeros, if and only if the following integral condition holds

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\log\left|\frac{1}{S_{22}(\lambda)}\right|d\lambda \leq \Im\left(\frac{\partial}{\partial\lambda}\log L_{22}\left(\frac{1}{\lambda}\right)\right)_{\lambda=0}. \tag{3.16}$$

Also with the objective of implementing a finite degree function N which approximates a constant reflection with magnitude as small as possible within the passband and equal to unity everywhere else in the frequency axis, an oscillating type of response, computed from the Tchebyshev class of polynomials, was considered. This type of response is shown in fig. D.1 where it can be seen that the reflection level oscillates in the passband between two levels ψ_1 and ψ_2. Similarly, as it can be seen in fig. D.1, the modulus squared of the global transmission S_{21} also oscillates between two values γ_1 and γ_2 with

$$\psi_1 = 1-\gamma_1,$$
$$\psi_2 = 1-\gamma_2.$$

In this appendix, we indicate how to obtain the response of the aforementioned type, which is optimal in terms of matching, namely with the lowest possible reflection level ψ_2. In addition, we also show that this optimal solution is unique.

D.1 Synthesis of oscillating Tchebyshev responses

Using the Belevitch form to parametrise the global system with a rational matrix of degree N, it can easily be shown that the modulus squared of the polynomials appearing in eq. (2.38) and providing the response in fig. D.1 is expressed as

$$\begin{aligned} rr^\star &= \beta^2 - \alpha^2 && \beta > \alpha, \\ pp^\star &= T_N^2 + \alpha^2, \\ qq^\star &= T_N^2 + \beta^2, \end{aligned}$$

with T_N the Tchebyshev polynomial of degree N and

$$\begin{aligned} \alpha^2 &= \left(\frac{\gamma_2(1-\gamma_1)}{\gamma_1 - \gamma_2} \right), \\ \beta^2 &= \left(\frac{\gamma_2}{\gamma_1 - \gamma_2} \right). \end{aligned}$$

Additionally, the spectral factorisation of $|p|^2$ and $|q|^2$ is uniquely determined by imposing the minimum phase character of the polynomial p and the stability of q.

In order to allow for the dechaining of the load afterwards, function S_{22} must verify eq. (3.16). It must be remarked that assuming a rational form for S_{22}

$$S_{22}(\lambda) = \frac{p(\lambda)}{q(\lambda)},$$

with $p = p_0 \prod_{i=1}^{N}(\lambda - \xi_i)$ and $q = q_0 \prod_{i=1}^{N}(\lambda - \zeta_i)$ the solution to the integral in eq. (3.16) can be expressed as

$$\frac{1}{\pi} \int_{\mathbb{R}} \log \left| \rho(\lambda)^{-1} \right| d\lambda = \sum_{i=1}^{N} \zeta_i - \sum_{i=1}^{N} \xi_i.$$

Furthermore, the roots of the minimum phase polynomial p such that $pp^\star = T_N^2 + \alpha^2$ have the explicit expression

$$\xi_i = \frac{1}{2} \left(X_\alpha^{1/N} e^{j\theta_k} + X_\alpha^{-1/N} e^{-j\theta_k} \right),$$

with $X_\alpha = \alpha + \sqrt{\alpha^2 + 1}$ and $\theta_k = \frac{\pi}{2N}(2k-1)$.

This is the main reason behind the interest of imposing a Tchebyshev shape for the global response S_{22} since it is possible now to write the restriction on α and β such that eq. (3.16) is saturated, namely the value of the integral is maximised. We have

$$\begin{aligned} h &= \sum_{k=1}^{N} \zeta_k - \sum_{k=1}^{N} \xi_k \\ &= \frac{1}{2} \left(X_\beta^{1/N} \sum_{k=1}^{N} e^{j\theta_k} + X_\beta^{-1/N} \sum_{k=1}^{N} e^{-j\theta_k} \right) - \frac{1}{2} \left(X_\alpha^{1/N} \sum_{k=1}^{N} e^{j\theta_k} + X_\alpha^{-1/N} \sum_{k=1}^{N} e^{-j\theta_k} \right), \end{aligned}$$

with $X_\beta = \beta + \sqrt{\beta^2+1}$. By means of trigonometric relations, the previous sums can be expressed as

$$\sum_{k=1}^{N} e^{j\theta_k} = \sin\left(\frac{\pi}{2N}\right)^{-1}, \qquad \sum_{k=1}^{N} e^{-j\theta_k} = -\sin\left(\frac{\pi}{2N}\right)^{-1}.$$

Also note that $\frac{1}{2}\left(X_\beta^{1/N} - X_\beta^{-1/N}\right) = \sinh\left(\frac{1}{N}\text{arcsinh}(\beta)\right)$. Therefore we have

$$h \cdot sin\left(\frac{\pi}{2N}\right) = sinh\left(\frac{1}{N} arcsinh\left(\beta\right)\right) - sinh\left(\frac{1}{N} arcsinh\left(\alpha\right)\right).$$

Now solving for β

$$\beta = \sinh\left(N \cdot arcsinh\left(sinh\left(\frac{1}{N} arcsinh\left(\alpha\right)\right) + h \cdot sin\left(\frac{\pi}{2N}\right)\right)\right).$$

Now suppose that $\mathbb{I}$ represent a single compact interval of the real line and T_N is the Tchebyshev polynomial in the interval $\mathbb{I}$ such that $-1 \leq T_N(\omega) \leq 1$ for all $\omega \in \mathbb{I}$. Finally the reflection coefficient S_{22} is expressed

$$|S_{22}(\omega)|^2 = \frac{T_N(\omega)^2 + \alpha^2}{T_N(\omega)^2 + \beta(\alpha)^2}.$$

Additionally we have

$$\max_{\omega \in \mathbb{I}} |S_{22}(\omega)|^2 = \frac{1+\alpha^2}{1+\beta(\alpha)^2}. \tag{D.1}$$

The optimal parameter α is obtained by minimising the function in eq. (D.1) with respect to α. Additionally, as show in next section, this function has a unique minimum in α.

D.2 Generalised oscillating responses

Instead of proving now the unicity of the optimum α in eq. (D.1), we consider now a more generic transfer function where some finite transmission zeros are allowed in the global system. This can be easily done by replacing the polynomial function $T_N(\omega)$ with the rational filtering function $\frac{\hat{p}(\omega)}{\hat{r}(\omega)}$ where $\hat{r}(\omega)$ has roots at the desired transmission zeros and $\hat{p}(\omega)$ is the Tchebyshev polynomial of degree N with the weight $\hat{r}$. This filtering function oscillates between -1 and 1 within the passband. We provide next the definition of a function $F(\omega)$ which represents the reflection coefficient of the global system.

Definition D.2.1 (). *Let $f(\omega) = \left|\frac{p(\omega)}{r(\omega)}\right|^2$ be a rational function with p and r non-zero polynomials and r of degree less than the degree of p ($deg(r) < deg(p)$). We define a function $F_{\alpha,\beta}(\omega)$ as the minimal phase realization of $|F_{\alpha,\beta}(\omega)|^2$:*

$$|F_{\alpha,\beta}(\omega)|^2 = F_{\alpha,\beta}(\omega)F^*_{\alpha,\beta}(\omega) = \frac{f(\omega)+\alpha}{f(\omega)+\beta} = \frac{n}{e} \qquad 0 \leq \alpha < \beta < \infty.$$

It implies that $F_{\alpha,\beta}(\omega)$ and $(F_{\alpha,\beta}(\omega))^{-1}$ are analytic in the lower half plane $\omega \in \mathbb{C}^-$. Then $F_{\alpha,\beta}(\omega)$ can be computed assigning the roots of n and e that lies in the upper half plane of ω to $F_{\alpha,\beta}(\omega)$ and the conjugate roots to $F^*_{\alpha,\beta}(\omega)$.

We are now disposed to state the main problem in this appendix, namely the function $F_{\alpha,\beta}(\omega)$ within the passband $\omega \in \mathbb{I}$ where β is obtained as a function $\beta = \beta(\alpha)$ ensuring that the load can be de-embedding.

Problem D.2.1 (Oscillating transfer function)**.**

$$\textit{Find:} \qquad \min_{\alpha} \frac{f_1 + \alpha}{f_1 + \beta(\alpha)} \qquad f_1 = \max_{\omega \in \mathbb{I}} \left| \frac{\hat{p}(\omega)}{\hat{r}(\omega)} \right|^2.$$

Next we argue towards the proof of unicity of the optimal value α. Note that this time we do not have an explicit relation between α and β, nevertheless we can define an implicit function $\Phi_{\alpha,\beta}(\omega)$ such that the de-embedding condition is satisfied, namely the derivative of the global reflection at a point ω_0 on the frequency axis matches the derivative of the reflection provided by the load, denoted here by η.

Lemma D.2.1. *Define $\Phi_{\alpha,\beta}(\omega) = arg\,(F_{\alpha,\beta}(\omega))$, η a strictly negative constant and ω_0 strictly positive such as $r(\omega_0) = 0$. Then for every value $\alpha > 0$ there exist an unique value β that satisfies:*

$$\frac{d}{d\omega}\Phi_{\alpha,\beta}(\omega)\bigg|_{\omega=\omega_0} = \eta. \tag{D.2}$$

Proof. $F_{\alpha,\beta}(\omega)$ is minimum-phase and both $F_{\alpha,\beta}(\omega)$ and $ln(F_{\alpha,\beta}(\omega))$ are analytic for $\omega \in \mathbb{C}^-$. In that case the phase ($\Phi_{\alpha,\beta}(\omega) = Im\{ln|F(\omega)|\}$) can be obtained from the Hilbert transform up to an unknown constant.

$$\Phi_{\alpha,\beta}(\omega) = \frac{1}{\pi} \oint_{\mathbb{R}} \frac{Re\{ln(F_{\alpha,\beta}(\tau))\}}{\tau - \omega} d\tau + C = \frac{1}{\pi} \oint_{\mathbb{R}} \frac{ln|F_{\alpha,\beta}(\tau)|}{\tau - \omega} d\tau + C. \tag{D.3}$$

The condition necessary and sufficient for the Hilbert transform to exist is that $ln|F_{\alpha,\beta}(\tau)| \in L^p(\mathbb{R})$ for $1 \leq p \leq \infty$. Therefore we study the integrability of $ln|F_{\alpha,\beta}(\tau)|$. We express the integral in two parts, one over $|\tau| > M$ and another over $|\tau| \leq M$ where $M >> \beta$.

$$\int_{|\tau|>M} ln|F_{\alpha,\beta}(\tau)|d\tau + \int_{|\tau|\leq M} ln|F_{\alpha,\beta}(\tau)|d\tau = \int_{\mathbb{R}} ln|F_{\alpha,\beta}(\tau)|d\tau.$$

The function $ln|F_{\alpha,\beta}(\tau)|$ can be expressed as:

$$ln|F_{\alpha,\beta}(\tau)| = ln\,(1 - (1 - |F_{\alpha,\beta}(\tau)|)) = ln\left(1 - \frac{\beta - \alpha}{\left|\frac{p(\tau)}{r(\tau)}\right|^2 + \beta}\right). \tag{D.4}$$

Since $|F_{\alpha,\beta}(\tau)| \to 1$ as $\tau \to \infty$, in the first integral the logarithm is equivalent to the function $g_{\alpha,\beta}(\tau)$.

$$g_{\alpha,\beta}(\tau) = -\left(1 - \frac{\alpha}{\beta}\right) \frac{1}{1 + \frac{1}{\beta}\left|\frac{p(\tau)}{r(\tau)}\right|^2}.$$

This function is bounded by $G_{\varepsilon,\alpha}(\tau) = \frac{1}{1+\frac{\varepsilon}{\beta}\tau^2}$ as $\tau \to \infty$ where ε is the relation between the leading coefficients of p and r and $G_{\varepsilon,\alpha}(\tau) \in L^2$. Therefore $g_{\alpha,\beta}(\tau) \in L^2$. Conversely,

in the second integral we found logarithmic singularities if $\alpha = 0$ and $f(\tau) = 0$. However this singularities are integrable. Thus we conclude that the Hilbert transform in (D.3) exists almost everywhere. Moreover, the derivative and the Hilbert transform operators can be inverted provided that the integral in (D.3) exist [88].

$$\frac{\partial \Phi_{\alpha,\beta}(\omega)}{\partial \omega} = \frac{1}{\pi} \oint_{\mathbb{R}} \frac{\frac{\partial}{\partial \tau} ln|F(\tau)|}{\tau - \omega} d\tau. \tag{D.5}$$

From (D.4), it can be seen that the function $ln|F_{\alpha,\beta}(\tau)|$ has a double zero at $\tau = \omega_0$ since $r(\omega_0) = 0$. In this case, the singularity at $\tau = \omega_0$ in equation (D.5) cancels and the principal value integral becomes a classic integral of a real negative function. Thus it is possible to integrate by parts in (D.5).

$$\frac{\partial \Phi_{\alpha,\beta}(\omega)}{\partial \omega} = \frac{1}{\pi} \int_{\mathbb{R}} \frac{\frac{\partial}{\partial \tau} ln|F(\tau)|}{\tau - \omega} d\tau = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{ln|F_{\alpha,\beta}(\tau)|}{(\tau - \omega)^2} d\tau. \tag{D.6}$$

Let denote by $H(\alpha, \beta)$ the left hand side of (D.2).

$$H(\alpha, \beta) = \left. \frac{\partial \Phi_{\alpha,\beta}(\omega)}{\partial \omega} \right|_{\omega=\omega_0} = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{ln|F_{\alpha,\beta}(\tau)|}{(\tau - \omega_0)^2} d\tau, \tag{D.7}$$

where $|F_{\alpha,\beta}(\tau)|$ has continuous partial derivatives with respect to the parameters α and β. In addition, if we define the function $T_\epsilon(\tau) \in L^2$ as:

$$T_\epsilon(\tau) = \frac{1}{f(\tau) + \epsilon} \quad with \quad 0 < \epsilon < \alpha, \tag{D.8}$$

then the partial derivatives of $|F_{\alpha,\beta}(\tau)|$ remain bounded by $T_\epsilon(\tau)$.

$$\frac{\partial}{\partial \alpha} |F_{\alpha,\beta}(\tau)| = \frac{1}{f(\tau) + \alpha} < T_\epsilon(\tau) \qquad 0 < \epsilon < \alpha,$$
$$\frac{\partial}{\partial \beta} |F_{\alpha,\beta}(\tau)| = \frac{1}{f(\tau) + \beta} < T_\epsilon(\tau) \qquad 0 < \epsilon < \beta.$$

Under this conditions, the *Leibniz lemma* claims that the passage of the limit under the integral sign is licit and therefore, the derivative can be passed under the integral sign. Thus:

$$\frac{\partial}{\partial \beta} H(\alpha, \beta) = -\frac{1}{2\pi} \int_{\mathbb{R}} \frac{d\tau}{(\tau - \omega_0)^2 (f(\tau) + \beta)},$$

where the previous integral is strictly positive. Hence if we fix α, the function $H(\alpha, \beta)$ will be strictly decreasing with β. Then we compute the values $H(\alpha, \beta)$ when $\beta = \alpha$ and when $\beta \to \infty$:

$$H(\alpha, \alpha) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{ln\,(1)}{(\tau - \omega_0)^2} d\tau = 0,$$
$$\lim_{\beta \to \infty} H(\alpha, \beta) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{ln\,(0)}{(\tau - \omega_0)^2} d\tau = -\infty.$$

We conclude that $H(\alpha, \beta)$ with α fixed is strictly monotonous from 0 to $-\infty$ if β varies in the interval $\alpha < \beta < \infty$. Therefore for a given η negative, there exist an unique value β that satisfies $H(\alpha, \beta) = \eta$. □

Remark D.2.1. *It is also possible that $\omega_0 = \infty$ and it will be necessary, for instance if $r(\omega)$ has no finite roots. In this case condition (D.2) takes an alternative form as follows:*

$$\frac{d}{d\omega}\Phi_{\alpha,\beta}\left(\frac{1}{\omega}\right)\bigg|_{\omega=0} = -\eta.$$

The proof remains similar but considering the function $H(\alpha,\beta)$ as in (D.9) and the same arguments are valid .

$$H(\alpha,\beta) = \frac{d}{d\omega}\Phi_{\alpha,\beta}\left(\frac{1}{\omega}\right)\bigg|_{\omega=0} = -\frac{1}{2\pi}\int_{\mathbb{R}} ln|F_{\alpha,\beta}(\tau)|d\tau. \tag{D.9}$$

Moreover this function $\Phi_{\alpha,\beta}(\omega)$ presents some interesting properties which are necessary in next section.

Lemma D.2.2. *Let $\Phi_{\alpha,\beta}(\omega)$ be defined as before and $\beta = \beta(\alpha)$ such as condition (D.2) is satisfied. Then β is a function $\beta = \beta(\alpha) \in \mathcal{C}^\infty$ with the following properties:*

1. $\beta'(\alpha) > 1$.[1]
2. $\beta''(\alpha) < 0$.
3. $\lim_{\alpha\to\infty} \beta'(\alpha) = 1$.

Proof. Define again $H(\alpha,\beta)$ as in (D.7). Thus since $H(\alpha,\beta)$ is differentiable and $\frac{\partial}{\partial\beta}H(\alpha,\beta) \neq 0$, by introducing (D.7) in (D.2), the expression $H(\alpha,\beta) - \eta = 0$ defines an implicit function $\beta(\alpha)$ within the range $0 < \alpha < \beta$. Let define now the function $h(x)$ as the integral:

$$h(x) = \int_{\mathbb{R}} \frac{d\tau}{(\tau-\omega_0)^2(f(\tau)+x)} \qquad x > 0.$$

The integrand of $h(x)$ is again bounded by the function $T_\epsilon(\tau)$ (D.8).

$$\frac{d}{dx}\left(\frac{1}{(\tau-\omega_0)^2(f(\tau)+x)}\right) < \frac{1}{f(\tau)+\epsilon} \qquad 0 < \epsilon < x.$$

Thus we can apply the *Leibniz rule* to obtain $h'(x)$.

$$h'(x) = -\int_{\mathbb{R}} \frac{d\tau}{(\tau-\omega_0)^2(f(\tau)+x)^2} \qquad x > 0.$$

This function is strictly negative since the integral is strictly positive. Then $h(x)$ is strictly decreasing. This approach can be iterated since the successive derivatives are all bounded by $T_\epsilon(\tau)$ and consequently $h(x) \in \mathbb{C}^\infty$. We obtain:

$$\frac{d^n}{dx^n} = (-1)^n \int_{\mathbb{R}} \frac{n!}{(\tau-\omega_0)^2(f(\tau)+x)^{n+1}} d\tau \qquad x > 0. \tag{D.10}$$

[1] The notation $\beta'(\alpha)$ and $\beta''(\alpha)$ stands for $\frac{d}{d\alpha}\beta(\alpha)$ and $\frac{d^2}{d\alpha^2}\beta(\alpha)$ respectively.

Now the partial derivatives of $H(\alpha,\beta)$ can be expressed in term of $h(\alpha)$ and $h(\beta)$:

$$\frac{\partial}{\partial\alpha}H(\alpha,\beta)=\frac{h(\alpha)}{2\pi},$$
$$\frac{\partial}{\partial\beta}H(\alpha,\beta)=-\frac{h(\beta)}{2\pi}.$$

By the theorem of the implicit function we compute $\beta'(\alpha)$ as:

$$\beta'(\alpha)=-\left(\frac{\partial H(\alpha,\beta)}{\partial\beta}\right)^{-1}\frac{\partial H(\alpha,\beta)}{\partial\alpha}=\frac{h(\alpha)}{h(\beta)},$$

where $h(x)$ is positive and decreasing. This implies that $\beta'(\alpha)>1$ and if $h(x)\in\mathcal{C}^\infty$ then $\beta(\alpha)\in\mathcal{C}^\infty$. This proves (a).

Next let write $\beta'(\alpha)$ as the relation between $h(\alpha)$ and $h(\beta)$ and compute $\beta''(\alpha)$ by differentiating again with respect to α:

$$\begin{aligned}h(\alpha)&=\beta'(\alpha)h(\beta),\\ h'(\alpha)&=\beta''(\alpha)h(\beta)+h'(\beta)\left(\beta'(\alpha)\right)^2.\end{aligned}\tag{D.11}$$

Introducing (D.11) and dividing by $h(\alpha)^2$:

$$\begin{aligned}\beta''(\alpha)h(\beta)&=h'(\alpha)-h'(\beta)\left(\frac{h(\alpha)}{h(\beta)}\right)^2,\\ \frac{\beta''(\alpha)h(\beta)}{h(\alpha)^2}&=\frac{h'(\alpha)}{h(\alpha)^2}-\frac{h'(\beta)}{h(\beta)^2},\end{aligned}\tag{D.12}$$

where $h(x)>0$. Consider now the function $\chi(x)=\frac{h'(x)}{h(x)^2}$. This is a strictly negative function due to the negative sign of $h'(x)$ and its derivative is:

$$\chi'(x)=\frac{h''(x)h(x)-2h'(x)^2}{h(x)^3},$$

Using (D.10):

$$\chi'(x)=2\frac{\int_{\mathbb{R}}\frac{d\tau}{(\tau-\omega_0)^2(f(\tau)+x)^3}\int_{\mathbb{R}}\frac{d\tau}{(\tau-\omega_0)^2(f(\tau)+x)}-\left(\int_{\mathbb{R}}\frac{d\tau}{(\tau-\omega_0)^2(f(\tau)+x)^2}\right)^2}{\left(\int_{\mathbb{R}}\frac{d\tau}{(\tau-\omega_0)^2(f(\tau)-x)}\right)^3}.$$

If we finally define the functions $|\varphi_1(\tau)|$ and $|\varphi_2(\tau)|$

$$|\varphi_1(\tau)|=\frac{1}{(\tau-\omega_0)(f(\tau)+x)^{3/2}}\qquad|\varphi_2(\tau)|=\frac{1}{(\tau-\omega_0)(f(\tau)+x)^{1/2}}.$$

the we obtain that $\chi'(x)\geq 0$ by the *Schwarz inequality*:

$$\left(\int_{\mathbb{R}}|\varphi_1(\tau)|^2d\tau\right)^{\frac{1}{2}}\left(\int_{\mathbb{R}}|\varphi_2(\tau)|^2d\tau\right)^{\frac{1}{2}}\geq\int_{\mathbb{R}}|\varphi_1(\tau)||\varphi_2(\tau)|d\tau.$$

Moreover, equality in the *Schwarz lemma* holds only if φ_1^2 and φ_2^2 are linearly dependent for some non-zero constant λ. Assume $\lambda\varphi_1^2 = \varphi_2^2$. Then

$$\begin{aligned}\frac{1}{(\tau-\omega_0)^2(f(\tau)+x)} &= \lambda\frac{1}{(\tau-\omega_0)^2(f(\tau)+x)^3}\\ \frac{(\tau-\omega_0)^2(f(\tau)+x)^3}{(\tau-\omega_0)^2(f(\tau)+x)} &= \lambda\\ (f(\tau)+x)^2 &= \lambda.\end{aligned}$$

It implies that the function $(f(\tau)+x)^2$ is equal to a constant for all $\tau \in \mathbb{R}$. This results in a contradiction since $f(\tau) = \left|\frac{p(\tau)}{r(\tau)}\right|$ and the degree of p is strictly greater that the degree of r. Then the assumption $\lambda\varphi_1^2 = \varphi_2^2$ must be false. As a result $\chi'(x)$ is strictly positive and therefore $\chi(x) = \frac{h'(x)}{h(x)^2}$ is strictly increasing. Returning now to the equation (D.12) and considering that $\chi(\alpha) - \chi(\beta) < 0$ it follows that $\beta''(\alpha) < 0$. It yields (b).

Finally consider the function $\frac{\alpha}{\beta(\alpha)}$ and compute the derivative $\frac{d}{d\alpha}\left(\frac{\alpha}{\beta(\alpha)}\right)$.

$$\frac{d}{d\alpha}\left(\frac{\alpha}{\beta(\alpha)}\right) = \frac{\beta(\alpha)-\alpha\beta'(\alpha)}{\beta(\alpha)^2} = \frac{N(\alpha)}{\beta(\alpha)^2}. \tag{D.13}$$

Next take the derivative of the numerator $N(\alpha)$ in (D.13):

$$N'(\alpha) = -\alpha\beta''(\alpha).$$

Since $\alpha > 0$ and $\beta''(\alpha) < 0$, $N(\alpha)$ is strictly increasing and therefore the derivative of $\frac{\alpha}{\beta(\alpha)}$ can not vanish more that once. This implies that the function $\frac{\alpha}{\beta(\alpha)}$ can not have more than one minimum and considering that it is bounded $0 < \frac{\alpha}{\beta(\alpha)} < 1$, then the limit of $\frac{\alpha}{\beta(\alpha)}$ as $\alpha \to \infty$ exists. Now consider again the expression (D.2) along with (D.6):

$$\int_{\mathbb{R}} \frac{ln\left(\frac{f(\tau)+\alpha}{f(\tau)+\beta(\alpha)}\right)}{(\tau-\omega_0)^2} d\tau = 2\pi\eta. \tag{D.14}$$

We split (D.14) in two parts, the first one over an interval X near ω_0

$$X = \left\{\tau : \omega_0 - 1 < \tau < \omega_0 - \frac{1}{M+1}\right\}.$$

The second part over the complement of X.

$$\int_X \frac{ln\left(\frac{f(\tau)+\alpha}{f(\tau)+\beta(\alpha)}\right)}{(\tau-\omega_0)^2} d\tau + \int_{\mathbb{R}\setminus X} \frac{ln\left(\frac{f(\tau)+\alpha}{f(\tau)+\beta(\alpha)}\right)}{(\tau-\omega_0)^2} d\tau = 2\pi\eta.$$

We know that both integrals are negative, then

$$\int_X \frac{ln\left(\frac{f(\tau)+\alpha}{f(\tau)+\beta(\alpha)}\right)}{(\tau-\omega_0)^2} d\tau \leq 2\pi\eta. \tag{D.15}$$

The integrand in (D.15) is bounded by $G_{\alpha_0}(\tau) \in L^2$ for all $\tau \in X$.

$$G_{\alpha_0}(\tau) = \left| \frac{lg\left(\frac{\alpha_0}{\beta(\alpha_0)}\right)}{(\tau-\omega_0)^2} \right| \qquad 0 < \alpha_0 < \alpha.$$

Then we can apply the *Lebesgue dominated convergence theorem*:

$$\lim_{\alpha\to\infty} \int_X \frac{ln\left(\frac{f(\tau)+\alpha}{f(\tau)+\beta(\alpha)}\right)}{(\tau-\omega_0)^2} d\tau = \int_X \frac{\lim\limits_{\alpha\to\infty} ln\left(\frac{\alpha}{\beta(\alpha)}\right)}{(\tau-\omega_0)^2} d\tau = M \lim_{\alpha\to\infty} ln\left(\frac{\alpha}{\beta(\alpha)}\right).$$

We denote $K = \lim\limits_{\alpha\to\infty} ln\left(\frac{\alpha}{\beta(\alpha)}\right)$. Then from (D.15) we have:

$$MK \leq 2\pi\eta.$$

If we assume $K \neq 0$ and we take $M > \frac{2\pi\eta}{K}$. Then $MK \leq 2\pi\eta$ and $MK > 2\pi\eta$, a contradiction. We conclude that $K = 0$ what implies that $\lim\limits_{\alpha\to\infty}\left(\frac{\alpha}{\beta(\alpha)}\right) = 1$. To complete the prove, we know that $\lim\limits_{\alpha\to\infty} \alpha = \infty$ and $\beta(\alpha) > \alpha$. Moreover $\beta'(\alpha)$ is decreasing and greater than 1, this implies that $\lim\limits_{\alpha\to\infty} \beta'(\alpha)$ exists. Furthermore α and $\beta(\alpha)$ are both differentiable for every α greater than a fixed positive value L. Under this conditions we can apply the *l'Hôpital rule for an indetermination* $\frac{\infty}{\infty}$.

$$\lim_{\alpha\to\infty} \frac{\alpha}{\beta(\alpha)} = \lim_{\alpha\to\infty} \frac{1}{\beta'(\alpha)} = \frac{1}{\lim\limits_{\alpha\to\infty} \beta'(\alpha)} = 1.$$

Hence, (c) holds. □

D.3 Unicity of the solution to problem D.2.1

We are now disposed to proof the unicity of the value α optimal for problem D.2.1.

Lemma D.3.1. *Consider the function $\beta = \beta(\alpha)$ such as condition (D.2) is satisfied for a given $f(\omega)$. Fix a constant $f_1 \geq 0$, then the function $\psi(\alpha)$:*

$$\psi(\alpha) = \frac{f_1 + \alpha}{f_1 + \beta(\alpha)} \qquad \alpha \geq 0$$

is minimum at an unique finite value $\alpha = \alpha_0$.

Proof. Denote by $\Psi(\alpha)$ the numerator of the derivative of $\psi(\alpha)$.

$$\psi'(\alpha) = \frac{f_1 + \beta(\alpha) - \beta'(\alpha)(f_1+\alpha)}{(f_1+\beta(\alpha))^2} = \frac{\Psi(\alpha)}{(f_1+\beta(\alpha))^2}.$$

Since $(f_1 + \beta(\alpha)) \neq 0$ then $\psi'(\alpha)$ if and only if $\Psi = 0$. Now compute $\Psi(\alpha)$.

$$\Psi'(\alpha) = -\beta''(\alpha)(f_1+\alpha),$$

where $\beta''(\alpha) < 0$ and $(f_1 + \alpha) > 0$. Therefore $\Psi(\alpha)$ is strictly increasing and vanish at most once. It allows us to distinguish between three possible cases:

1. $\Psi(\alpha_0) = 0$ for $0 < \alpha_0 < \infty$.
2. $\Psi(\alpha) > 0$ for all $\alpha > 0$.
3. $\Psi(\alpha) < 0$ for all $\alpha > 0$.

In the first case, the function $\psi(\alpha)$ has an unique minimum at $\alpha = \alpha_0 \in]0, \infty[$. If (ii) holds, $\psi(\alpha)$ is an increasing function for $\alpha > 0$. Then the minimum occurs for $\alpha = 0$. Finally, it can be proved that (iii) never happens. Compute:

$$\lim_{\alpha \to \infty} \Psi(\alpha) = \lim_{\alpha \to \infty} (f1 + \beta(\alpha) - \beta'(\alpha)(f_1 + \alpha)) = \lim_{\alpha \to \infty} (\beta(\alpha) - \alpha).$$

Since by definition $\beta > \alpha$, it is possible to conclude that $\lim_{\alpha \to \infty} \Psi(\alpha) > 0$ and (iii) cannot hold.

References

[88] J. N. Pandey, "The Hilbert Transform of Schwartz Distributions," *Proceedings of the American Mathematical Society*, vol. 89, no. 1, pp. 86–90, 1983. [Online]. Available: http://www.jstor.org/stable/2045069

□

Appendix E:

Manufactured prototypes

In chapter 15 we have shown the measurements of two manufactured prototypes, both of them made of plastic using additive manufacturing techniques and subsequently metallized. In this appendix we provide a compilation of the illustrations as well as tables with the dimensions corresponding to of each of the structures presented in chapter 15.

E.1 Six-poles dual-mode filter.

The first prototype included in this chapter is the waveguide filter corresponding to the 3D model shown in fig. 15.5. This filter has a 3d-printed plastic body while the inner surface has been metallised by silver painting. The parameters which define the filter structure are depicted in fig. E.1 while the dimensions (in mm) are listed in table E.1. The obtained device is shown in fig. E.2 where the transparent plastic body can be appreciated.

The device shown in fig. E.2 presents a mono block structure where no assembling of spare pieces is required. This mono block manufacturing is possible by means of a printed plastic support filling the inner volume of the cavities which is removed afterwards. When this type of manufacturing is used, the transparent body helps to perform an adequate cleaning and metallising of the interior of the structure since it allows to visualize any possible imperfection. The result of the metallisation process is illustrated in fig. E.3 where the filter structure shown in fig. 15.5 can be perfectly recognised. This technique has already been used in the literature, for instance in [89]. Moreover, fig. E.4 shows the tuning screws used to compensate the small manufacturing tolerances. These screws allow to fine tune each coupling and each resonant mode in the structure. This filter provides, after tuning, the response shown in fig. 15.6.

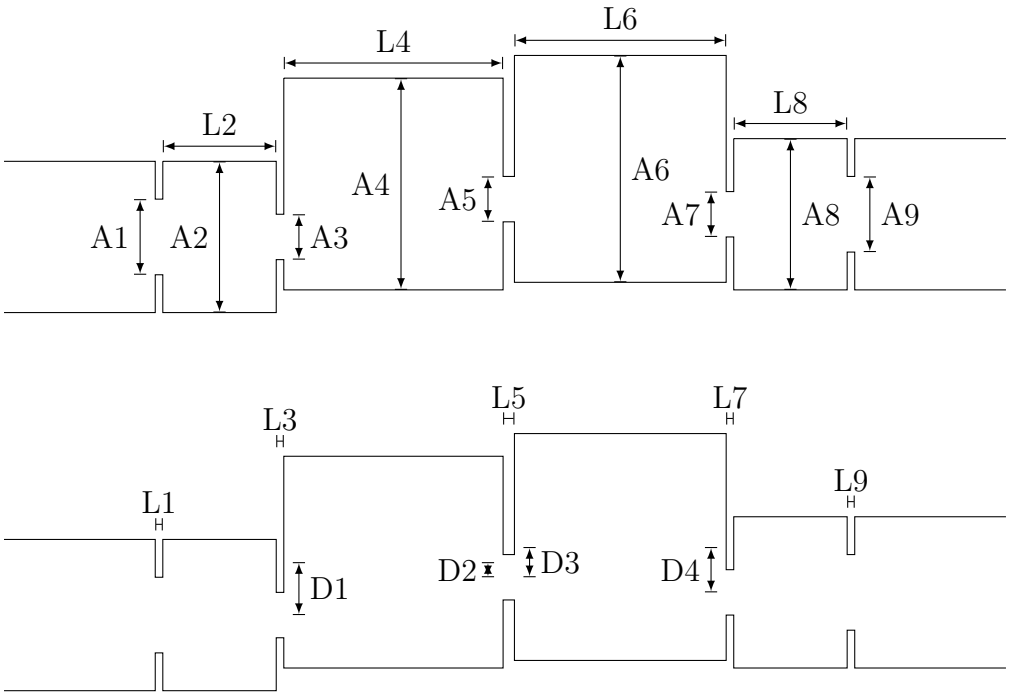


Figure E.1: Dimmensions of the manufactured dual-mode filter.

Parameter	Value	Parameter	Value	Parameter	Value
A1	9.903	L1	1.000	D1	5.534
A2	19.050	L2	14.617	D2	5.019
A3	7.174	L3	1.000	D3	2.815
A4	30.338	L4	27.875	D4	6.180
A5	7.384	L5	1.555		
A6	28.654	L6	28.609		
A7	7.234	L7	1.000		
A8	19.050	L8	14.621		
A9	9.824	L9	1.000		

Table E.1: Values in mm of the dimensions indicated in fig. E.1.

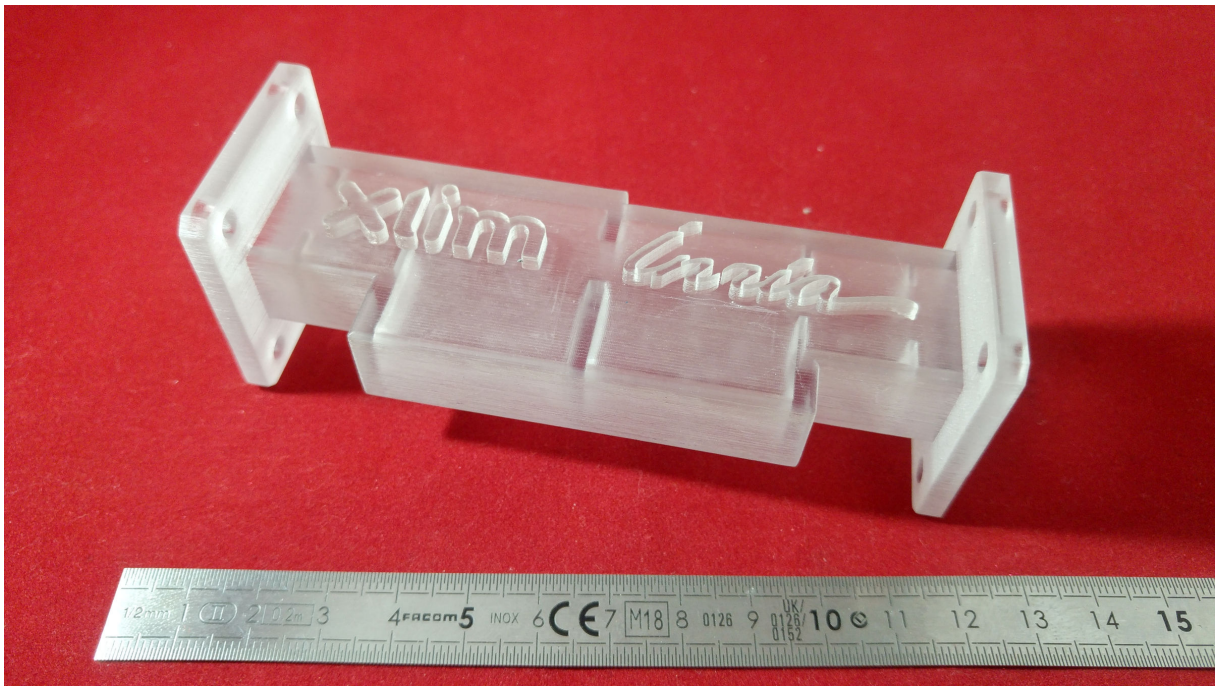


Figure E.2: Prototype of reference filter: plastic body

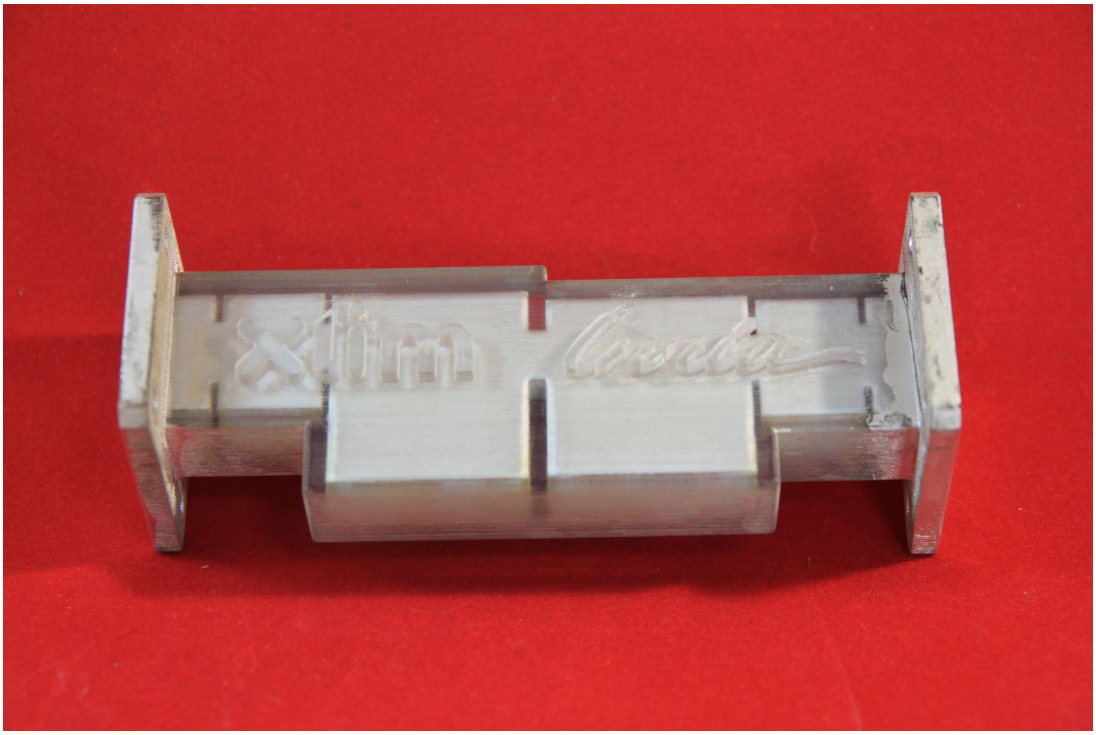


Figure E.3: Prototype of reference filter: front view

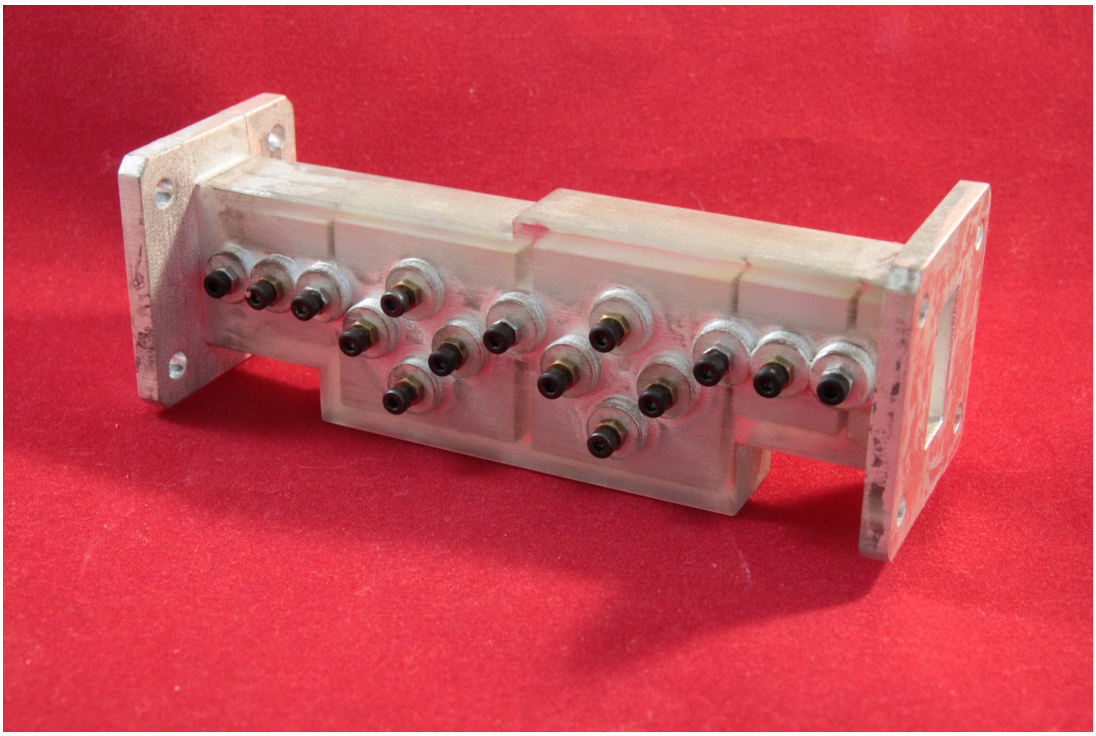


Figure E.4: Prototype of reference filter: back view

E.2 Manifold multiplexer

The second manufactured prototype corresponds to the structure shown in fig. 15.17, which has also been implemented by means of additive manufacturing techniques.

We have on the one side in fig. E.13 the dimensions of the box that covers the external volume of the final device, including the flasks, the coupling and assembly screws and other external elements. On the other side we provide the final dimensions of the waveguide structure which constitutes the internal volume of the multiplexer. These dimensions are denoted with the parameters shown in fig. E.14a. Note that we use the same parameters name for all channels. The values of the cited parameters for each of the channel filters are listed in table E.2. Additionally we indicate in fig. E.14b the dimension obtained as the result of the manifold synthesis following the procedure presented in chapter 13.

We can see in figs. E.5 to E.7 the 3d-printed plastic body which is ready to be metallised. It is important to note that, unlike the previous case and due to the high complexity of the triplexer structure, we have decided to manufacture the device in two halves, using assembly screws to join both halves posteriorly. This allows us greater control in the process of metallization of the device, since by applying the silver paint on each of the halves separately, we get better access to each of the internal surfaces of the device. Figures E.8 and E.9 illustrate the inner metallisation of both halves of the device while figs. E.10 to E.12 show the final assembled device, after the inner surface is metallised and with the tuning screws inserted.

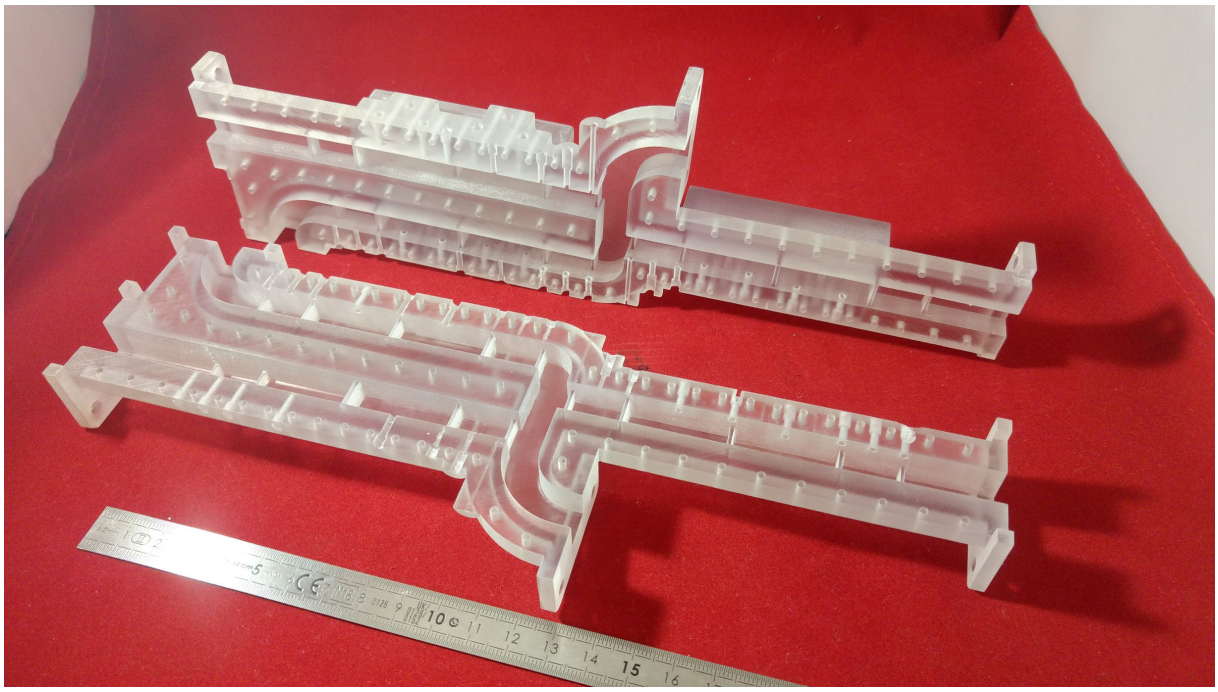


Figure E.5: Triplexer prototype: inside view

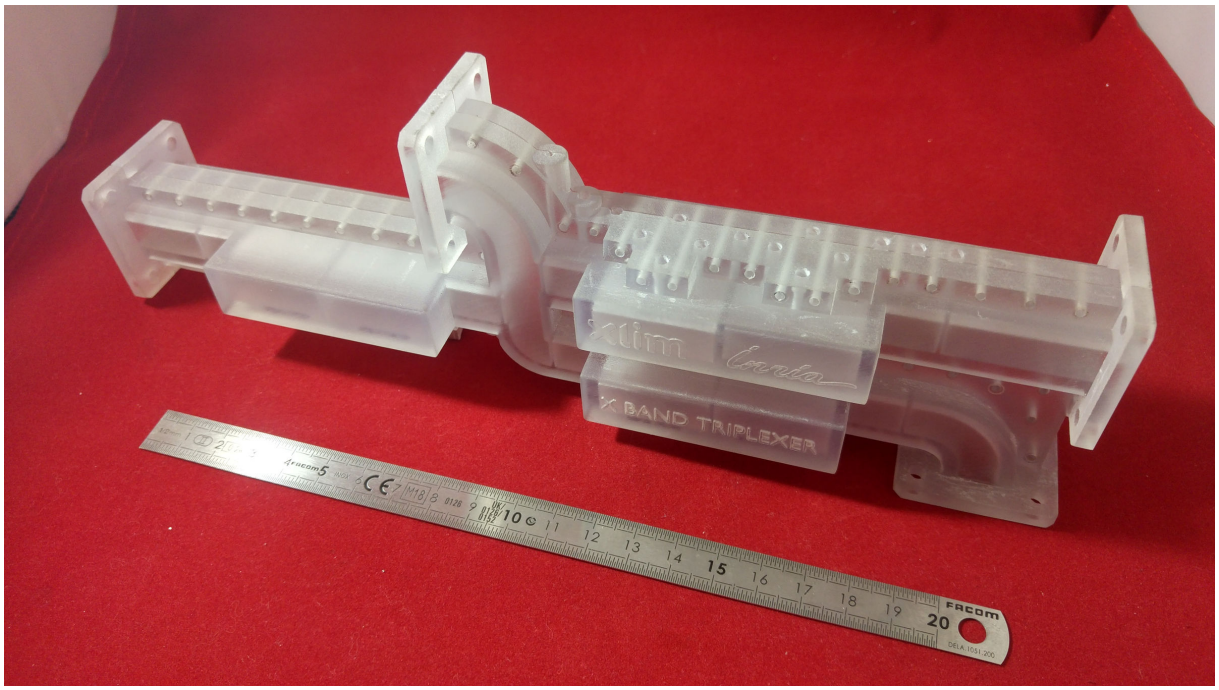


Figure E.6: Triplexer prototype: front view

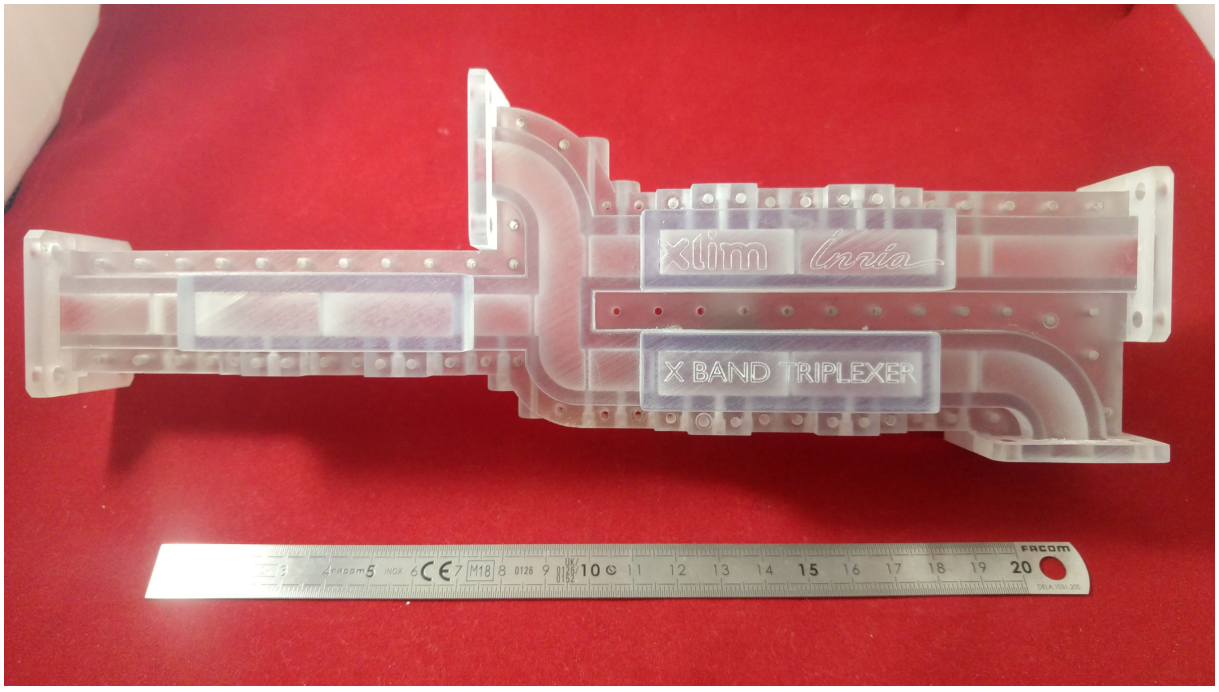


Figure E.7: Triplexer prototype: top view

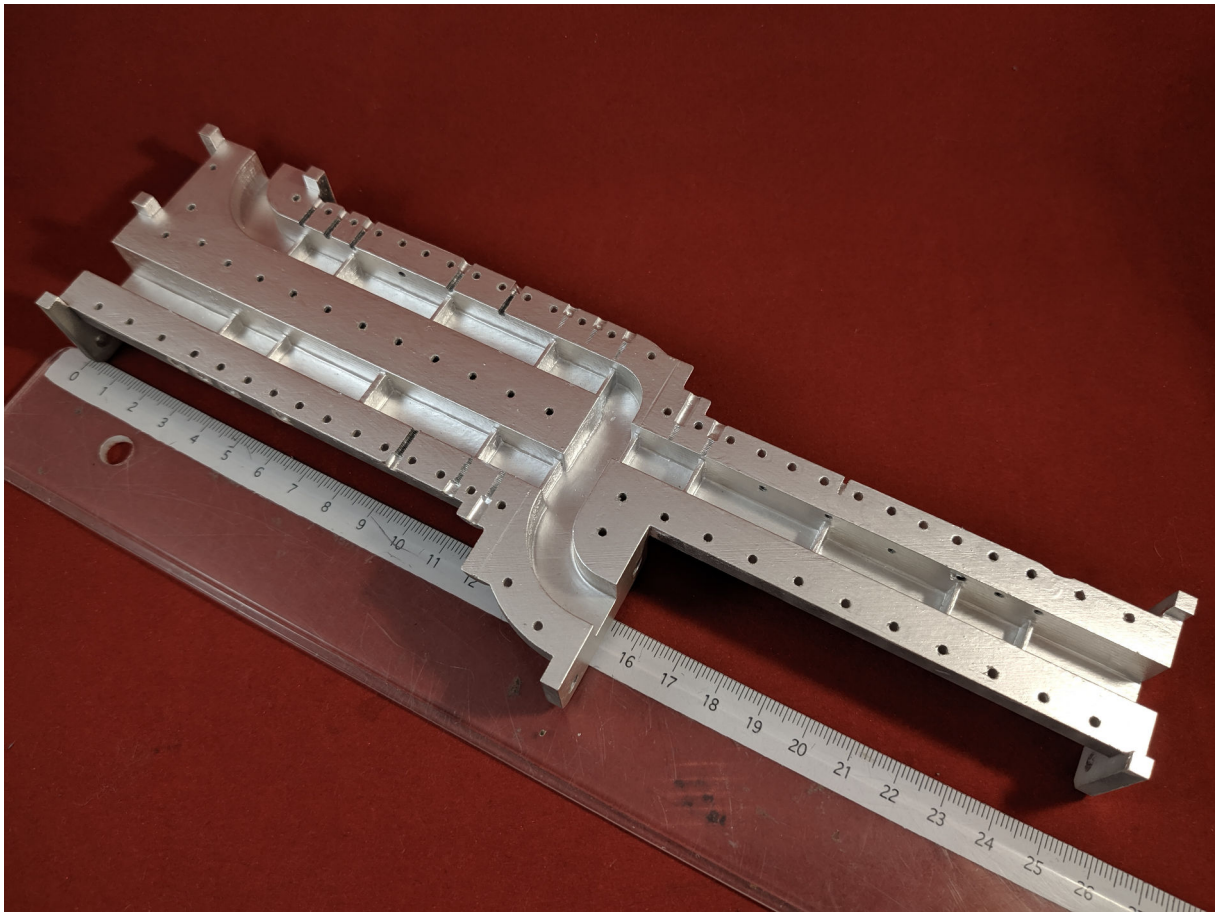


Figure E.8: Triplexer prototype: metallisation of upper section.

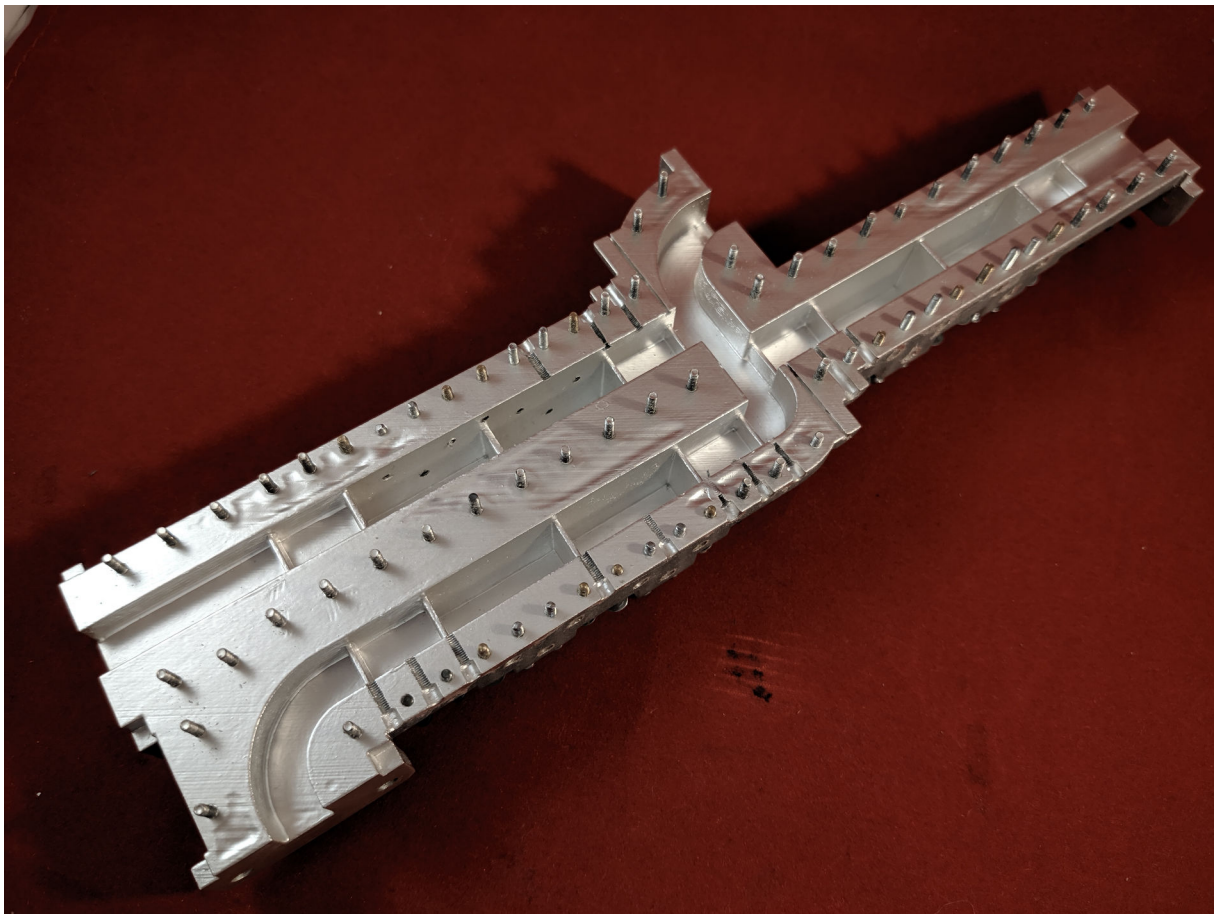


Figure E.9: Triplexer prototype: metallisation of lower section.

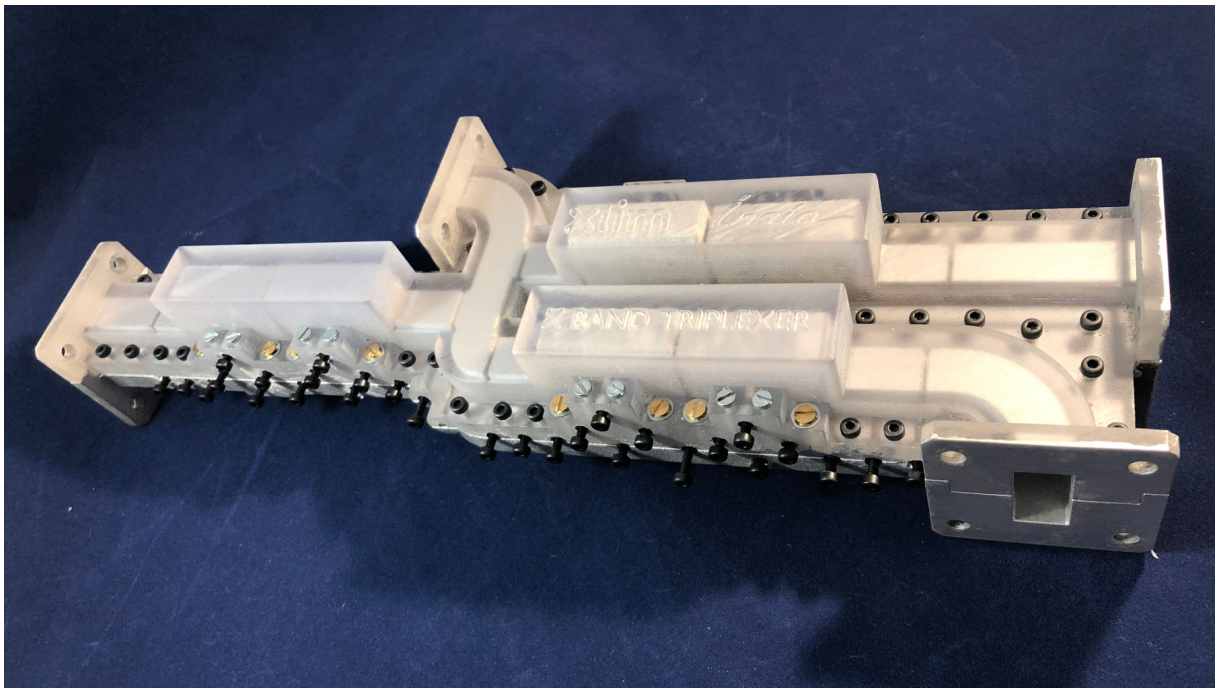


Figure E.10: Triplexer final structure: front view.

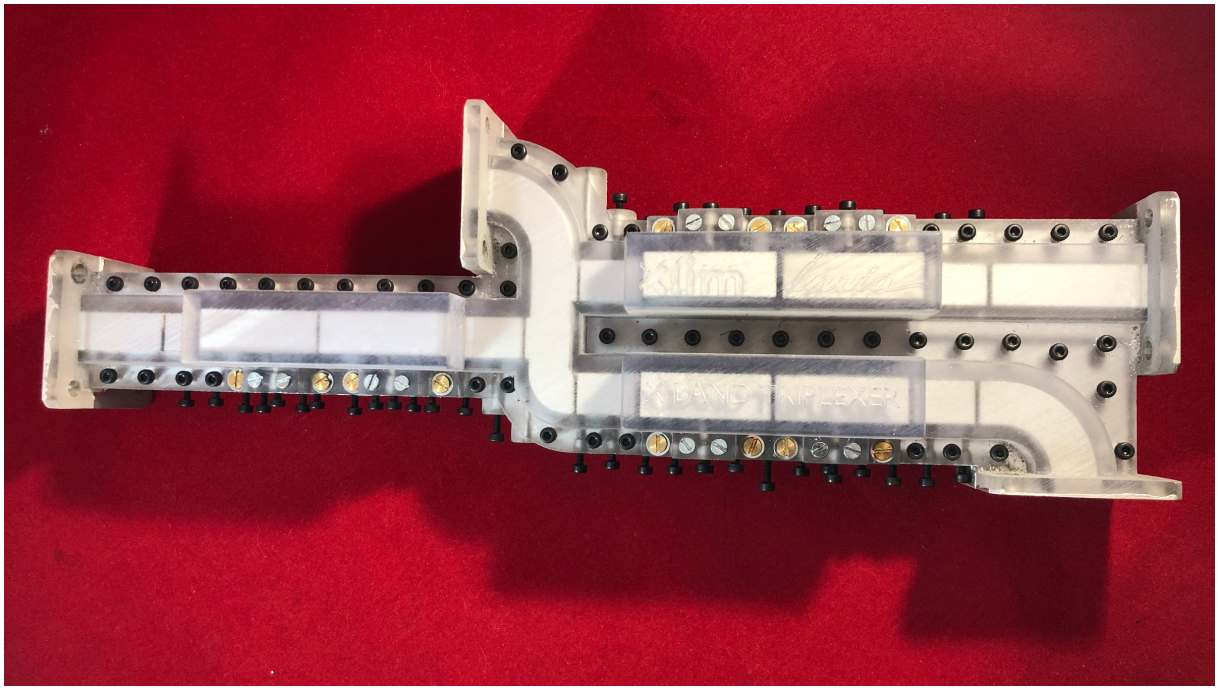


Figure E.11: Triplexer final structure: top view.

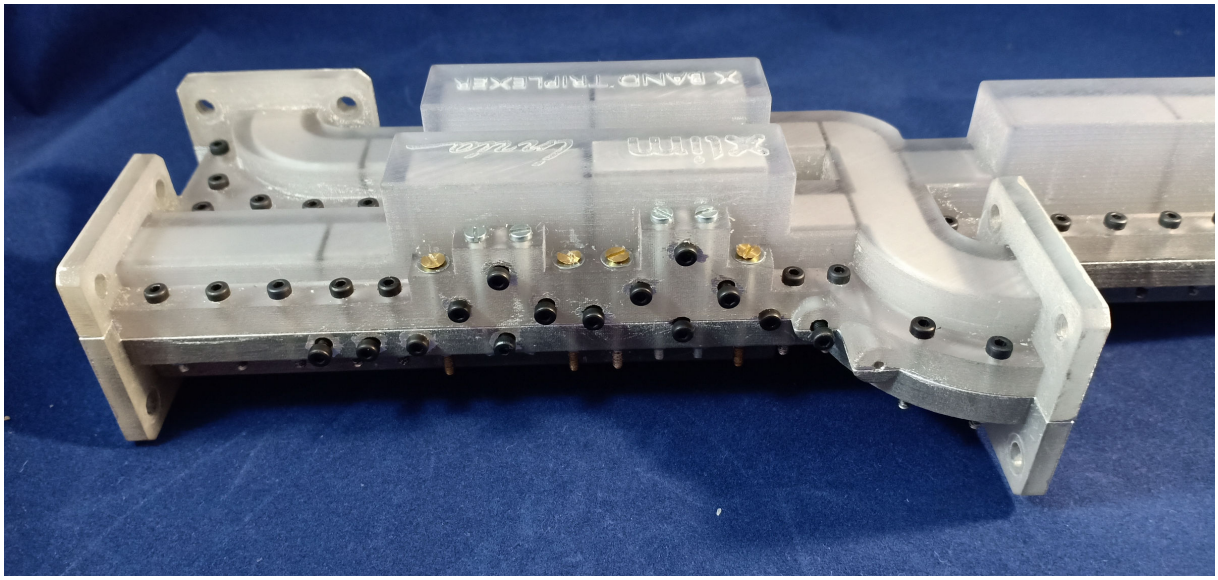
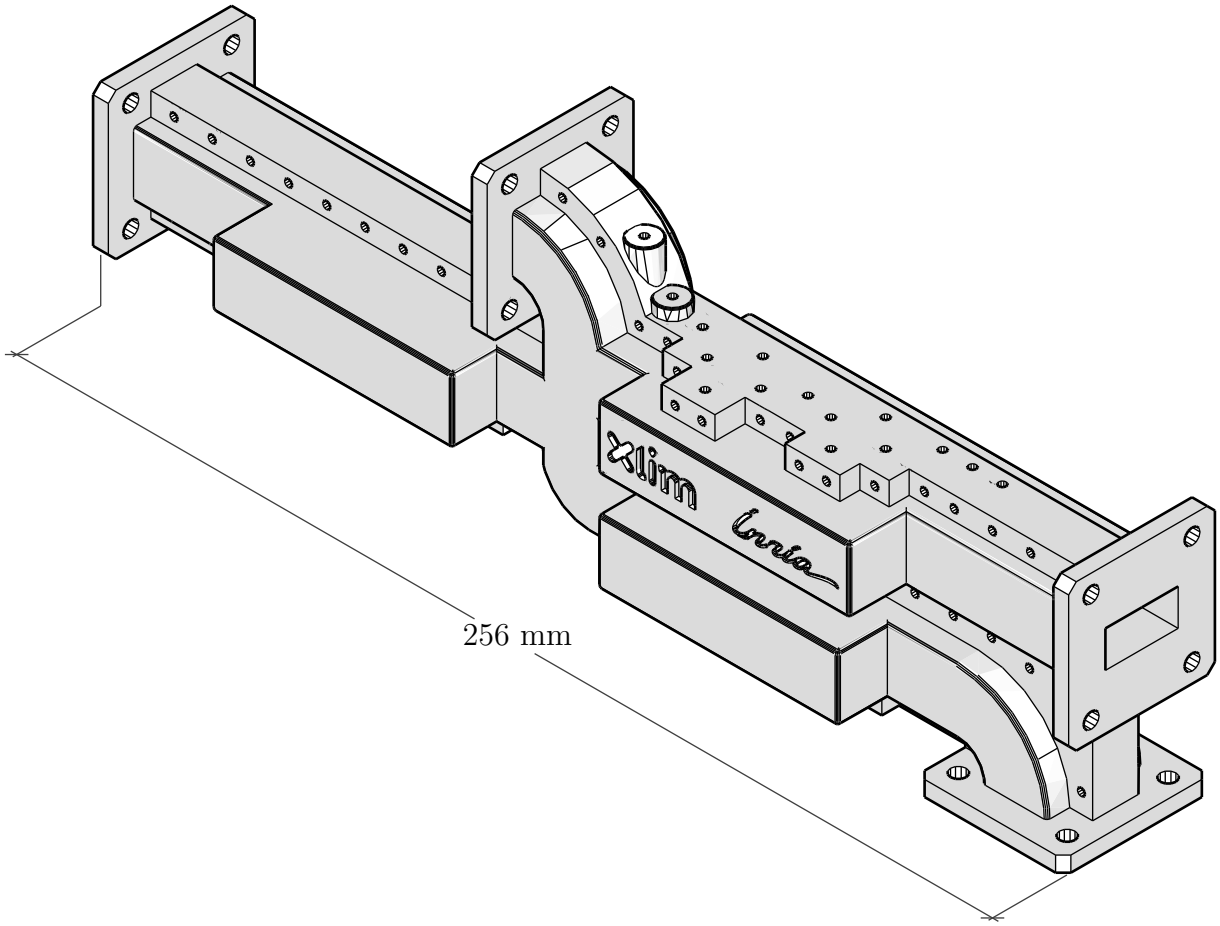
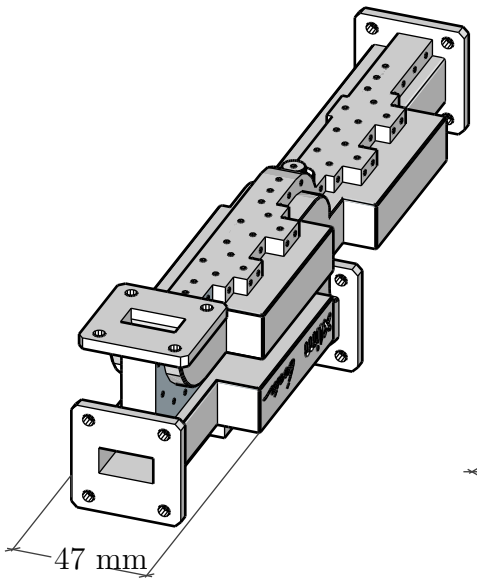


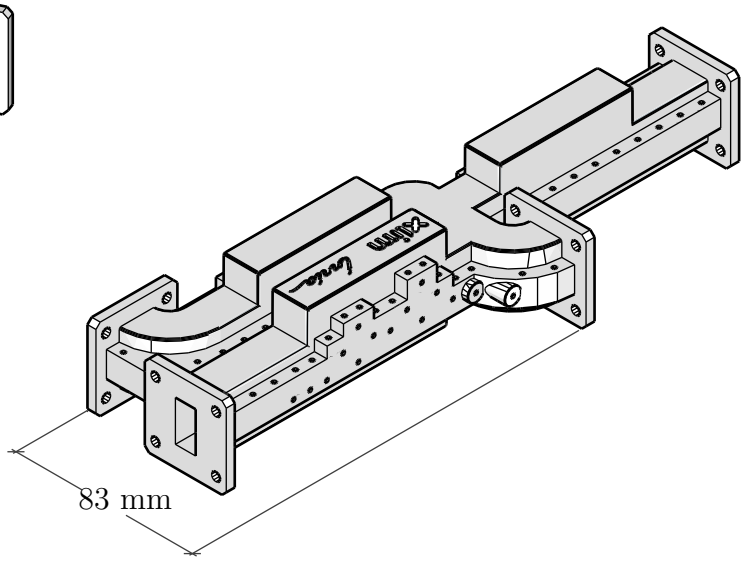
Figure E.12: Triplexer final structure: back view.



(a) Front view

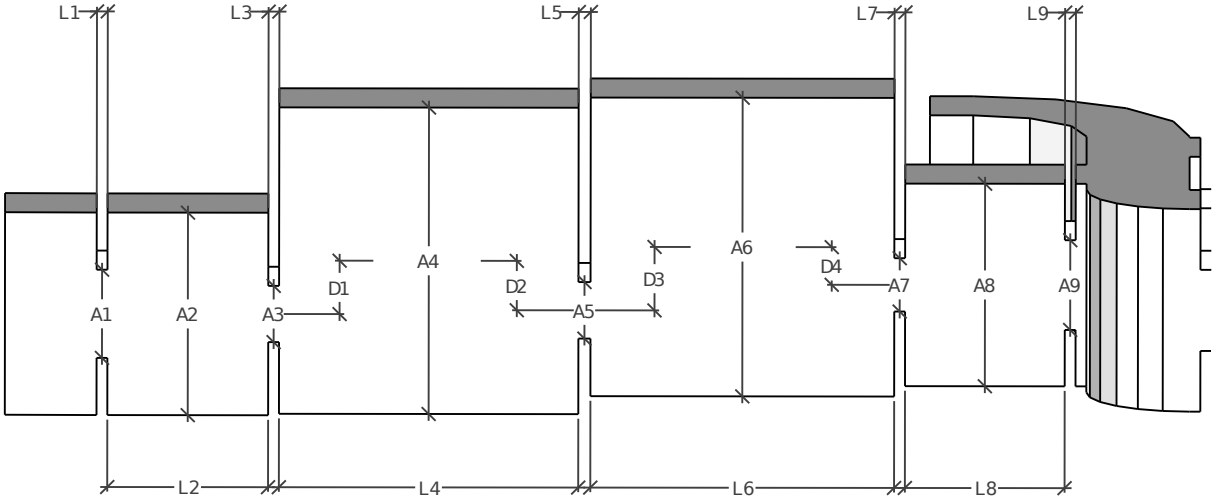


(b) Side view

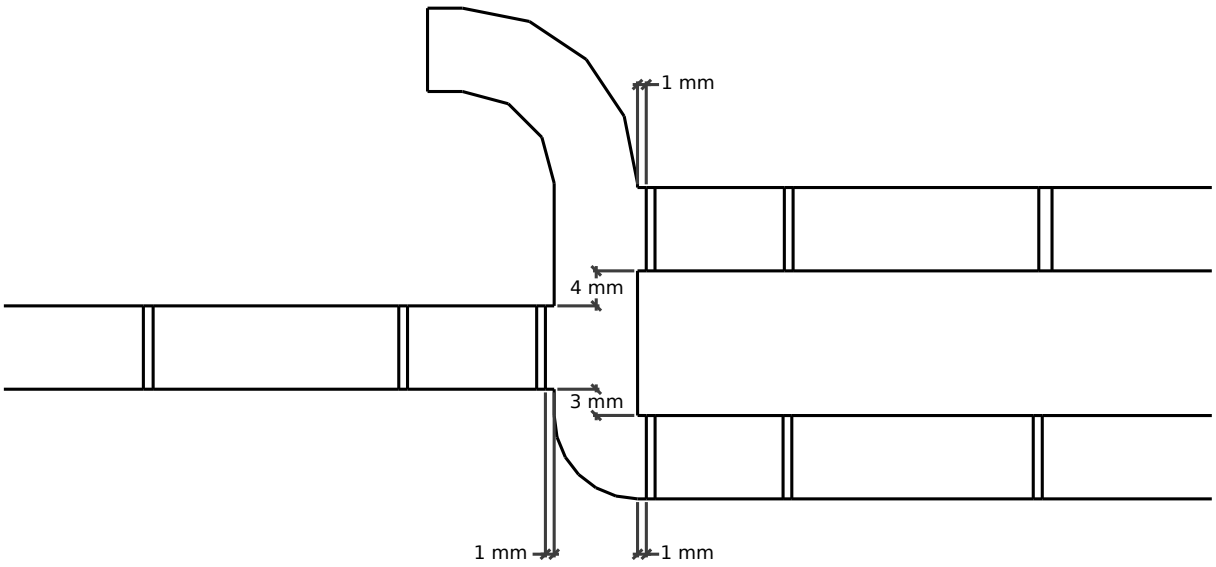


(c) Top view

Figure E.13: Designed fishbone structure



(a) Channel filter parameters



(b) Manifold dimensions

Figure E.14: Definition of the dimension parameters in the triplexer structure

Parameter	Value	Parameter	Value	Parameter	Value
A1	9.708	L1	1.000	D1	7.236
A2	19.050	L2	14.648	D2	1.929
A3	7.301	L3	1.000	D3	4.599
A4	28.683	L4	28.665	D4	6.340
A5	7.207	L5	1.523		
A6	30.310	L6	28.083		
A7	6.959	L7	1.000		
A8	19.050	L8	14.784		
A9	9.372	L9	1.000		

(a) Channel 1

Parameter	Value	Parameter	Value	Parameter	Value
A1	8.296	L1	1.000	D1	5.024
A2	19.050	L2	14.864	D2	4.643
A3	5.290	L3	1.000	D3	5.969
A4	28.821	L4	27.669	D4	3.540
A5	5.683	L5	1.117		
A6	28.088	L6	28.076		
A7	5.014	L7	1.000		
A8	19.050	L8	14.761		
A9	8.440	L9	1.000		

(b) Channel 2

Parameter	Value	Parameter	Value	Parameter	Value
A1	7.914	L1	1.000	D1	4.714
A2	19.050	L2	14.457	D2	4.284
A3	5.009	L3	1.000	D3	6.187
A4	28.212	L4	27.104	D4	5.296
A5	5.105	L5	1.032		
A6	27.453	L6	27.596		
A7	4.746	L7	1.000		
A8	19.050	L8	14.627		
A9	7.629	L9	1.000		

(c) Channel 3

Table E.2: Values of the parameters shown in fig. E.14a

References

[89] E. Laplanche, A. Delage, A. Haidar, W. Feuray, J. Sence, A. Perigaud, O. Tantot, N. Delhote, S. Verdeyme, S. Bila, and L. Carpentier, "Recent development in additive manufacturing of passive hardware and conformal printing," in *European Microwave Conference (EuMW 2018)*. Madrid, Spain: Society for Industrial and Applied Mathematics, 2018.

Glossaries

List of notation

List of symbols

List of sets

Index

List of chapters

Contents

List Of Definitions

List Of Theorems

List Of Problems

RÉSUMÉ

Le probléme de l'adaptation d'impédances en électronique et particulièrement en ingénierie des hyper fréquences consiste à minimiser la réflexion de la puissance qui doit être transmise, par un générateur, à une charge donnée dans une bande de fréquence. Les exigences d'adaptation et de filtrage dans les systèmes de communication classiques sont généralement satisfaites en utilisant un circuit d'adaptation suivi d'un filtre. Nous proposons ici de concevoir des filtres d'adaptation qui intègrent à la fois les exigences de filtrage et d'adaptation dans un seul appareil et augmentent ainsi l'efficacité globale et la compacité du système.

Dans ce travail, le problè d'adaptation est formulé en introduisant un problème convexe d'optimisation dans le cadre établi par la théorie de d'adaptation de Fano et Youla. De ce contexte, au moyen de techniques modernes de programmation semi-définies non linéaires, un problème convexe, et donc avec une optimalité garantie, est obtenu.

Enfin, pour démontrer les avantages fournis par la théorie développée au-delà de la synthèse de filtres avec des charges complèxes variables en fréquence, nous examinons deux applications pratiques récurrentes dans la conception de ce type de dispositifs. Ces applications correspondent, d'une part, à l'adaptation d'un réseau d'antennes dans le but de maximiser l'efficacité du rayonnement, et, d'autre part, à la synthèse de multiplexeurs où chacun des filtres de canal est adapté au reste du dispositif, notamment les filtres correspondant aux autres canaux.

Mots-clés: Synthèse de filtres, adaptation large bande, limites, Nevanlinna-Pick, interpolation Schur, optimisation convexe, antennes, manifold, multiplexeurs.

ABSTRACT

The problem of impedance matching in electronics and particularly in RF engineering consists on minimising the reflection of the power that is to be transmitted, by a generator, to a given load within a frequency band. The matching and filtering requirements in communication systems are usually satisfied by using a matching circuit followed by a filter. We propose here to integrate both, matching and filtering requirements, in a single device and thereby increase the overall efficiency and compactness of the system.

In this work, the matching problem is formulated by introducing convex optimisation on the framework established by the matching theory of Fano and Youla. As a result, by means of modern non-linear semi-definite programming techniques, a convex problem, and therefore with guaranteed optimality, is achieved.

Finally, to demonstrate the advantages provided by the developed theory beyond the synthesis of filters with frequency varying loads, we consider two practical applications which are recurrent in the design of communication devices. These applications are, on the one hand, the matching of an array of antennas with the objective of maximizing the radiation efficiency, and on the other hand the synthesis of multiplexers where each of the channel filters is matched to the rest of the device, including the filters corresponding to the other channels.

Keywords: Filter synthesis, broadband matching, bounds, Nevanlinna-Pick, Schur interpolation, convex optimisation, antennas, manifold, multiplexers.