Thèse de Doctorat

Université de Limoges

Université de Limoges
ED 653 : Sciences et Ingénierie (SI)
Faculté des Sciences et Techniques – Institut de Recherche XLIM

THÈSE
pour obtenir le grade de
Docteur de l'Université de Limoges
Dicipline: Mathématiques appliquées

Présentée et soutenue par
Kouegnon Dénis MITCHOZOUNNOU

Le 07 mai 2026

Accélération inertielle en optimisation continue : ajustement du coefficient de viscosité α et convergence vers le flot proximal.

Thèse dirigée par le Prof. Samir ADLY et M. Olivier PROT, Maître de Conférences

JURY :

Président du jury:	M. Charles DOSSAL (Professeur INSA/IMT Toulouse)
Rapporteurs:	M. Mounir HADDOU (Professeur INSA Rennes) Mme Aude RONDEPIERRE (Professeur INSA/IMT Toulouse)
Examinateurs:	M. Charles DOSSAL (Professeur INSA/IMT Toulouse) M. Jean-François AUJOL (Professeur à l'IMB de Bordeaux) M. Francisco SILVA (MCF-HDR, Xlim/Limoges) M. Mounir HADDOU (Professeur INSA Rennes) Mme Aude RONDEPIERRE (Professeur INSA/IMT Toulouse)
Directeurs:	M. Samir ADLY (Professeur à Xlim/Limoges) M. Olivier PROT (MCF, Xlim/Limoges)

Dedication: I dedicate this work to my mother and father.

Licence CC BY-NC-ND 3.0

« Dieu, toujours, fait la géométrie. »
Le philosophe Platon

Licence CC BY-NC-ND 3.0

Remerciements

Je tiens à exprimer ma profonde gratitude à toutes les personnes qui ont accompagné et soutenu ce travail de thèse.
Mes premiers remerciements vont à mes directeurs de thèse, le Professeur Samir Adly et M. Olivier Prot, maître de conférences, pour leur encadrement exceptionnel et la confiance qu'ils m'ont accordée tout au long de ce parcours. Leurs conseils avisés, leur disponibilité et leur rigueur scientifique ont été déterminants dans la réalisation de ce travail. Je les remercie pour la liberté intellectuelle qu'ils m'ont laissée, tout en guidant mes réflexions avec bienveillance.
Je souhaite exprimer ma reconnaissance aux rapporteurs qui ont accepté d'évaluer ce manuscrit et dont les remarques ont grandement contribué à son amélioration. Je remercie chaleureusement les membres du jury pour l'intérêt qu'ils ont porté à mes recherches et la qualité de nos échanges scientifiques.
Ma gratitude va également à l'ensemble des membres de l'équipe MOD pour leur accueil chaleureux et l'environnement de travail stimulant qu'ils ont su créer. Je garde un souvenir précieux des nombreuses discussions scientifiques et des séminaires enrichissants qui ont jalonné ces années de recherche.
Je pense avec bonheur à mes collègues et amis du laboratoire, compagnons de route indispensables, avec qui j'ai partagé tant de moments formateurs. Leurs suggestions, leur soutien moral et notre entraide quotidienne ont été des piliers essentiels durant cette aventure.
Une pensée particulière va à ma famille, dont le soutien indéfectible a été mon ancrage tout au long de ce doctorat. À mes parents, M. Assogba Mitchozounnou et Mme Mahoussi Tchango, je dis merci pour leur confiance absolue, leurs sacrifices et les valeurs qu'ils m'ont transmises.
Je remercie également M. Noukpo César et M. Noukpo Salvador, ainsi que ma fiancée, Mme Pauline Assogba, qui ont été présents depuis le début de cette aventure. Je remercie mes frères (M. Michozounnou Pierre et M. Mitchozounnou Hossou) et mes sœurs (Mme Danmitondé Julie et Mme Mitchozounnou Agathe) pour leurs encouragements constants et leur capacité à me ramener toujours à l'essentiel.
Je n'oublie pas mes amis extérieurs au monde universitaire, qui m'ont offert des échappées bienfaisantes et un précieux équilibre. Leurs questions parfois naïves mais toujours pertinentes m'ont souvent aidé à clarifier ma pensée et à prendre du recul sur mon travail.
Je tiens à saluer le rôle fondamental des institutions et organismes de financement sans lesquels cette thèse n'aurait pu voir le jour. Leur soutien matériel et logistique a été une condition sine qua non de la réalisation de ce projet.
Je garde une gratitude particulière pour les enseignants qui, depuis mes premières années

Licence CC BY-NC-ND 3.0

d'étude, ont su éveiller et nourrir ma curiosité scientifique. Leurs exemples ont été des sources d'inspiration déterminantes dans mon parcours académique.
Je remercie également le personnel administratif et technique, dont le travail souvent invisible mais essentiel a facilité mon quotidien de chercheur. Leur professionnalisme et leur bonne humeur ont contribué à créer un environnement de travail agréable et efficace.
Je n'oublierai jamais tous ceux, connus ou anonymes, qui, par une simple conversation, une remarque pertinente ou un geste de soutien, ont contribué à l'aboutissement de ce travail. Chaque interaction, si brève soit-elle, a laissé une empreinte dans ce cheminement intellectuel.
Enfin, je voudrais dédier ces remerciements à toutes les personnes qui croient en la recherche fondamentale et en sa capacité à faire progresser la connaissance. Leur confiance dans le travail des chercheurs est un moteur essentiel pour notre communauté scientifique.
Que toutes celles et tous ceux qui se reconnaîtront dans ces lignes trouvent ici l'expression de ma sincère reconnaissance. Ce travail est aussi le leur, par la part d'eux-mêmes qu'ils ont bien voulu y investir.

Licence CC BY-NC-ND 3.0

Droits d'auteurs

Cette création est mise à disposition selon le Contrat :

« Attribution-Pas d'Utilisation Commerciale-Pas de modification 3.0 France »
disponible en ligne : http://creativecommons.org/licenses/by-nc-nd/3.0/fr/



Licence CC BY-NC-ND 3.0

Table of contents

1 Introduction **12**

1.1 Introduction (version française) . . . 13

1.2 Introduction (English version) . . . 17

2 Background and notations **23**

2.1 Hilbert Space . . . 24

2.2 Compactness and Distributions . . . 26

2.3 Convexity in Hilbert Spaces . . . 30

2.4 Operators and Bessel Functions . . . 38

2.4.1 Operators . . . 38

2.4.2 Bessel functions . . . 39

3 From the inertial system to proximal flow **41**

3.1 Introduction . . . 42

3.1.1 Main results . . . 44

3.1.2 Chapter Contents . . . 46

3.2 Direct Approach . . . 46

3.3 Time-Scaling Approach . . . 50

3.4 Limit dynamics of (ISIHD): Convex case . . . 57

3.5 Limit dynamics of (ISIHD): Non-convex quadratic case . . . 69

3.6 Connection Between Continuous Dynamics and Discrete Algorithms . . . 73

3.7 Numerical Experiments . . . 75

3.8 Conclusion and Perspectives . . . 80

4 From the inertial system to first-order flow: Non-potential operators **82**

4.1 Introduction . . . 83

4.1.1 Problematic and challenges . . . 83

4.1.2 Approach . . . 84

4.1.3 Main results . . . 85

Licence CC BY-NC-ND 3.0

4.1.4 Chapter Contents 85
4.2 Limit Dynamics of (AVD): Non-Potential Operator Case 86
4.3 Limiting Dynamics of (DIN-AVD): Non-Potential Operator Case 96
4.4 Limit Dynamics of (ISIHD): Non-Potential Operator Case 104
4.5 Numerical Experiment 116
4.6 Conclusion and Perspectives 121

5 Study of the acceleration parameter **123**
5.1 Introduction 124
5.1.1 Main results 125
5.1.2 Chapter Contents 126
5.2 Analysis of the (AVD) dynamics: the quadratic convex case 127
5.3 Numerical experiments 142
5.4 Conclusion and Perspectives 149

6 Bibliography **151**
References 152

Licence CC BY-NC-ND 3.0

Table of Figures

3.1 Left: Phase portrait of $x_1(t)$ as a function of $x_2(t)$ for (AVD), (DIN-AVD), and (ISIHD). Right: t as a function of $\|\nabla f(x(t))\|$ in semi-log scale for (AVD), (DIN-AVD), and (ISIHD). 76

3.2 t as a function of $f(x(t))$ (left) and t as a function of $t^2 f(x(t))$ (right) in semi-log scale for (AVD), (DIN-AVD), and (ISIHD). 76

3.3 Left: Phase portrait for (AVD), (DIN-AVD), and (ISIHD). Right: t as a function of $\|\nabla f(x(t))\|$ in semi-logarithmic scale. 77

3.4 t as a function of $f(x(t))$ (left) and t as a function of $t^2 f(x(t))$ (right) in semi-logarithmic scale. 77

3.5 Left: 3D surface of $f(x, y)$ with the trajectories. Right: Contour plot of $f(x, y)$ with the trajectories of (AVD), (DIN-AVD), and (ISIHD). 78

3.6 Left: $t \mapsto \|\nabla f(x(t))\|$. Right: $t \mapsto t^2 f(x(t))$. Hessian damping significantly reduces oscillations and improves convergence. 78

3.7 Trajectories $y_\alpha(s)$ for different values of α and the limiting dynamics (black). 79

3.8 Phase portraits $y_1(s)$ vs $y_2(s)$ for varying β_0. Left: $\alpha = 5$. Right: $\alpha = 100$. 80

4.1 Comparison of the three inertial systems (AVD), $(\mathrm{DIN}-\mathrm{AVD})$ and (ISIHD) for the non-potential cocoercive operator introduced in Example 4.5.1. Left: trajectories $t \mapsto \|V(z(t))\|_2$. Right: pairwise differences $t \mapsto \|V_1(t) - V_2(t)\|_2$. 118

4.2 Rescaled trajectories $y_\alpha(s)$ for the (AVD) system and their limit dynamics $\dot{y}(s) + V(y(s)) = 0$ as $\alpha \to \infty$. Left: trajectory error $\|y_\alpha(s) - y(s)\|_2$ in rescaled time s. Right: $\|V(y_\alpha(s))\|_2$ for various values of α, compared with the limit flow (black dashed line). 119

4.3 Rescaled trajectories $y_\alpha(s)$ for the $(\mathrm{DIN}-\mathrm{AVD})$ system and their limit dynamics $\dot{y}(s) + V(y(s)) + \beta_0 \frac{d}{ds} V(y(s)) = 0$ as $\alpha \to \infty$, with $\beta_0 = 0.1$. Left: trajectory error $\|y_\alpha(s) - y(s)\|_2$ in rescaled time s. Right: $\|V(y_\alpha(s))\|_2$ for various values of α, compared with the limit flow (black dashed line). 119

4.4 Rescaled trajectories $y_\alpha(s)$ for the (ISIHD) system and their limit dynamics $\dot{y}(s) + V(y(s) + \beta_0 \dot{y}(s)) = 0$ as $\alpha \to \infty$, with $\beta_0 = 0.1$. Left: trajectory error $\|y_\alpha(s) - y(s)\|_2$ in rescaled time s. Right: $\|V(y_\alpha(s))\|_2$ for various values of α, compared with the limit flow (black dashed line). 119

Licence CC BY-NC-ND 3.0

4.5 (Left) Uniform trajectory error $\max_{s\in[0,S]} \|y_\alpha(s) - y(s)\|_2$ as a function of α for (AVD), $(\mathrm{DIN}-\mathrm{AVD})$ and (ISIHD). (Right) L^2-error on the rescaled velocities for (ISIHD), namely $\|\dot{y}_\alpha - \dot{y}\|_{L^2(0,S)}$ as a function of α (log–log scale), illustrating the L^2-convergence of the (ISIHD) velocities towards the limit first-order flow. 120
4.6 Pairwise differences between the three limit flows in rescaled time: $\|y_{\mathrm{DIN}}(s) - y_{\mathrm{AVD}}(s)\|_2$ (red), $\|y_{\mathrm{ISI}}(s) - y_{\mathrm{AVD}}(s)\|_2$ (green), and $\|y_{\mathrm{DIN}}(s) - y_{\mathrm{ISI}}(s)\|_2$ (black). The (AVD) limit departs significantly from the two corrected flows, while (DIN-AVD) and (ISIHD) remain close but not identical. 121

5.1 Left: the function $J_{10.5}$ and its asymptotic behavior. Right: the difference between the function $J_{10.5}$ and its asymptotic expression. 131
5.2 Left: the function $Y_{10.5}$ and its asymptotic behavior. Right: the difference between the function $Y_{10.5}$ and its asymptotic expression. 131
5.3 Convergence rate for f_1 (semi-logarithmic scale). 143
5.4 Optimality of the rate for f_1. 143
5.5 Convergence rate for f_2 (semi-logarithmic scale). 144
5.6 Optimality of the rate for f_2. 145
5.7 Convergence rate for f_3 (semi-logarithmic scale). 145
5.8 Optimality of the rate for f_3. 146
5.9 Convergence rate of the trajectories on a semi-logarithmic scale (for f_1). 147
5.10 Convergence rate of the trajectories on a semi-logarithmic scale (for f_2). 147
5.11 Convergence rate of the trajectories on a semi-logarithmic scale (for f_3). 148
5.12 Comparison between Heavy Ball and (AVD) for $\alpha = 10$ and $\alpha = 25$. 148
5.13 Comparison between Heavy Ball and (AVD) for $\alpha = 50$ and $\alpha = 100$. 148

Licence CC BY-NC-ND 3.0

Licence CC BY-NC-ND 3.0

1

Introduction

Contents

1.1 Introduction (version française) 13
1.2 Introduction (English version) 17

Licence CC BY-NC-ND 3.0

1.1 Introduction (version française)

Au cœur de nombreuses méthodes modernes en apprentissage automatique et en ingénierie se trouve la minimisation des fonctions convexes. En general, on considère le problème de minimisation, dans un espace de Hilbert $\mathcal{H}$, d'une fonction $f : \mathcal{H} \to \mathbb{R}$ convexe, différentiable, à gradient lipschitzien, et admettant au moins un minimiseur.
L'un des grands défis contemporains est de comprendre comment accélérer efficacement les méthodes d'optimisation dites du premier ordre, c'est-à-dire celles qui utilisent uniquement le gradient. Ce défi est à la fois théorique, pour analyser les vitesses de convergence, et pratique, pour concevoir des algorithmes rapides et robustes. Cette thèse s'inscrit dans cette dynamique et propose une analyse des méthodes accélérées, en mettant constamment en parallèle les algorithmes discrets et les équations différentielles continues qui en révèlent la structure profonde.

La famille la plus simple de ces méthodes est celle de la descente de gradient, basée sur la suite $(x_k)_k$ définie par recurrence

$$x_{k+1} = x_k - s\nabla f(x_k),$$

où $s > 0$ est le taux d'apprentissage et $-\nabla f(x_k)$ la direction de descente. Ce schéma garantit, sous certaines conditions sur le taux d'apprentissage, une décroissance progressive des valeurs de la fonction f (lemme de descente) et permet également d'atteindre le taux de convergence suivant [1]:

$$f(x_k) - \inf(f) \leq \mathcal{O}\left(\frac{1}{k}\right),$$

où $\inf(f)$ est le minimum de f, et, de plus, la suite $(x_k)_k$ converge faiblement vers un minimiseur de la fonction. Cependant, malgré sa simplicité et sa robustesse, cette méthode reste lente: il faut souvent un grand nombre d'itérations pour atteindre une précision satisfaisante.

En 1964, Polyak propose une amélioration avec la méthode dite de la Boule Pesante (Heavy Ball) [2]. Cette méthode introduit un terme d'inertie censé lisser la trajectoire et renforcer la vitesse de déplacement dans les directions favorables:

$$x_{k+1} = x_k - s\nabla f(x_k) + \beta(x_k - x_{k-1}),$$

Licence CC BY-NC-ND 3.0

où $\beta > 0$ est le paramètre d'inertie. Ce mécanisme permet souvent de réduire les oscillations et d'accélérer la convergence, notamment sur des problèmes mal conditionnés.

Une avancée a lieu en 1983 avec les travaux de Nesterov [3], qui introduit une méthode accélérée pour les fonctions convexes à gradient lipschitzien. L'idée essentielle de Nesterov repose sur une extrapolation linéaire des itérées passées (Voir [3]):

$$\begin{cases} y_k = x_k + \left(1 - \dfrac{\alpha}{k}\right)(x_k - x_{k-1}), \\ x_{k+1} = y_k - s\nabla f(y_k). \end{cases}$$

Ce mécanisme d'extrapolation a inspiré une grande partie des méthodes modernes d'optimisation, tant en théorie que dans les applications. Pour $\alpha = 3$, l'algorithme de Nesterov atteint une vitesse de convergence asymptotique de $\mathcal{O}\left(\frac{1}{k^2}\right)$ [3], [4], mais la convergence de la suite elle-même reste une question ouverte jusqu'à ce jour. Lorsque $\alpha > 3$, la convergence s'améliore encore avec $f(x_k) - \inf(f) = o\left(\frac{1}{k^2}\right)$ lorsque $k \to +\infty$, comme démontré par Attouch et Peypouquet [5], et les itérées $(x_k)_k$ convergent faiblement vers un minimiseur de f, selon les résultats de Chambolle et Dossal [6]. Le cas sous-critique, correspondant à $0 < \alpha \leq 3$, a également été étudié dans [7] et [8].

Un tournant se produit en 2015 lorsque Su, Boyd et Candès [9] montrent que l'algorithme de Nesterov peut être interprété comme la discrétisation d'une équation différentielle du second ordre. Cette dynamique continue est donnée par

$$(\text{AVD}) \ : \ \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0, \quad t > 0.$$

Comme dans le cas discret, le comportement asymptotique de la dynamique (AVD) dépend fortement du paramètre α. Dans le régime sous-critique $0 < \alpha \leq 3$, on obtient une convergence en $\mathcal{O}(1/t^{2\alpha/3})$ [7], [8]. Le cas critique, correspondant à $\alpha = 3$, coïncide avec la méthode de Nesterov et conduit à une convergence en $\mathcal{O}(1/t^2)$. Enfin, dans le régime sur-critique, pour $\alpha > 3$, on obtient une convergence en $o(1/t^2)$ [5] et les trajectoires convergent faiblement vers un minimiseur de f [6]. Ces résultats soulignent le rôle du paramètre α.
Ce lien entre dynamique discrète et continue ouvre un champ de recherche riche. Il a également motivé l'étude de variantes avec amortissement dépendant du Hessien, donnant naissance à des dynamiques dites de haute résolution (voir [10], [11], [12], [13], [14], [15], [16]), telles que

$$(\text{DIN-AVD}) : \ \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta(t)\nabla^2 f(x(t))\dot{x}(t) + b(t)\nabla f(x(t)) = 0,$$

premièrement introduit par Attouch, Peypouquet et Redont dans [13] ou encore la dynamique

Licence CC BY-NC-ND 3.0

implicite introduit par Alesca, Laszlo, et Pinta dans [16]

$$(\text{ISIHD}):\ \ddot{x}(t)+\frac{\alpha}{t}\dot{x}(t)+\nabla f(x(t)+\beta(t)\dot{x}(t))=0.$$

Ces dynamiques plus riches permettent d'analyser la décroissance de la fonction, du gradient, des oscillations et de la structure géométrique de la trajectoire.

La présente thèse est structurée autour d'une analyse de ces dynamiques accélérées. Tout d'abord, le comportement asymptotique de la dynamique (ISIHD) est analysé lorsque le paramètre d'amortissement α tend vers l'infini. L'analyse est ensuite étendue aux trois dynamiques dans un cadre non potentiel. Enfin, la dynamique (AVD) est étudiée dans le cas de fonctions quadratiques convexes. Chaque chapitre apporte des réponses à la fois théoriques et numériques, soulignant le rôle du paramètre α.

Le Chapitre 2 présente les définitions, notations et outils mathématiques utilisés dans la suite du manuscrit.

Le Chapitre 3 élargit la perspective en étudiant le comportement limite de la dynamique accélérée (ISIHD) lorsque le paramètre α tend vers l'infini. Cette analyse révèle une transition: la dynamique inertielle du second ordre se réduit à un flot proximal. Plus précisément, en introduisant le changement d'échelle

$$t=\sqrt{2(\alpha+1)s}$$

et en posant $y_\alpha(s)=x(t)$, on montre que la suite $(y_\alpha)_\alpha$ converge uniformément sur tout compact de $[0,+\infty)$ vers la solution de la dynamique de type proximal:

$$\dot{y}(s)+\nabla f(y(s)+\beta_0\dot{y}(s))=0.$$

Cette équation peut se réécrire sous la forme

$$\dot{y}(s)=-\frac{1}{\beta_0}y(s)+\frac{1}{\beta_0}\mathsf{prox}_{\beta_0 f}(y(s))=-\nabla\mathcal{M}_{\beta_0 f}(y(s)),$$

où $\mathsf{prox}_{\beta_0 f}$ et $\mathcal{M}_{\beta_0 f}$ désignent respectivement l'opérateur proximal et l'enveloppe de Moreau de f. L'analyse est menée premièrement pour les fonctions convexes, et ensuite étendue au cas non convexe sous une hypothèse quadratique en dimension finie. Des estimations d'énergie uniformes et le théorème d'Ascoli–Arzelà permettent d'établir la convergence.

Le Chapitre 4 poursuit cette investigation dans le cadre plus général des opérateurs non potentiels, où le gradient ∇f est remplacé par un opérateur λ-cocoercif V. Cette extension est essentielle pour traiter des problèmes variationnels incluant des contraintes ou des régularisations non différentiables. Les trois dynamiques principales (AVD), (DIN-AVD) et (ISIHD) sont

Licence CC BY-NC-ND 3.0

reformulées en remplaçant ∇f par V. L'analyse de leur comportement limite lorsque $\alpha \to +\infty$ révèle des phénomènes de réduction à des flot du premier ordre. Pour la dynamique (AVD):

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + V(x(t)) = 0,$$

le changement d'échelle

$$t = \sqrt{2(\alpha+1)s}$$

conduit à une convergence vers le flot monotone standard:

$$\dot{y}(s) + V(y(s)) = 0.$$

Pour la dynamique (DIN-AVD), enrichie d'un amortissement hessien explicite:

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta_0\frac{\alpha+1}{t}\frac{d}{dt}V(x(t)) + V(x(t)) = 0,$$

la limite produit un flot régularisé de type Newton–Levenberg–Marquardt:

$$\dot{y}(s) + \beta_0\frac{d}{ds}V(y(s)) + V(y(s)) = 0.$$

Enfin, pour la dynamique (ISIHD) avec amortissement implicite:

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + V\left(x(t) + \beta_0\frac{\alpha+1}{t}\dot{x}(t)\right) = 0,$$

la limite conduit à un flot monotone implicite à la structure particulièrement riche:

$$\dot{y}(s) + V(y(s) + \beta_0\dot{y}(s)) = 0.$$

La démonstration de ces convergences repose sur des techniques variationnelles et des inégalités énergétiques adaptées aux opérateurs cocoercifs. L'analyse requiert le développement d'estimations uniformes en α pour les trajectoires et leurs dérivées, ainsi que l'utilisation du théorème d'Ascoli–Arzelà pour établir la convergence forte. Dans le cas non potentiel, la monotonie de l'opérateur V remplace la convexité de la fonction f.

Le chapitre 5 aborde la question relative au choix du paramètre α. Il est consacré à une analyse de la dynamique visqueuse accélérée (AVD) dans le cas des fonctions quadratiques convexes. Ce choix se justifie à la fois par la possibilité d'obtenir des solutions explicites et par le caractère représentatif de ce cadre pour de nombreux problèmes applicatifs. On

Licence CC BY-NC-ND 3.0

considère la fonction quadratique

$$f(x) = \frac{1}{2}\langle Qx, x\rangle + \langle b, x\rangle + d_0,$$

où Q est une matrice symétrique semi-définie positive, ce qui permet une diagonalisation du système dynamique associé. La solution explicite de l'équation différentielle correspondante fait intervenir des fonctions de Bessel de première espèce J_ν et de seconde espèce Y_ν, d'ordre

$$\nu = \frac{\alpha - 1}{2}.$$

L'analyse asymptotique de ces solutions montre que l'écart à l'optimalité décroît comme $O(t^{-\alpha})$, et ce taux est démontré comme étant optimal. Le paramètre α joue ainsi le rôle d'un accélérateur de convergence. Des expériences numériques viennent étayer ces résultats, illustrant à la fois le taux de convergence obtenu et son caractère optimal.

En synthèse, cette thèse étudie les mécanismes d'accélération en optimisation. Elle examine les liens entre algorithmes discrets et dynamiques continues, analyse le rôle des paramètres d'amortissement et d'extrapolation, et considère une extension de ces cadres au cas non potentiel. Les travaux présentés comprennent des résultats théoriques et numériques relatifs à la conception d'algorithmes et à l'analyse de leur comportement.

Contributions principales

En d'autres termes, cette thèse apporte plusieurs résultats sur l'analyse des dynamiques accélérées et leur comportement asymptotique dans les cadres convexe, quadratique convexe, non convexe quadratique et des opérateurs non potentiels. Ces travaux prolongent et complètent des résultats classiques et se résument en trois points:

- Étude de la limite de perturbation singulière $\alpha \to +\infty$ pour le système (ISIHD), conduisant à un flot proximal (voir Chapitre 3, Théorèmes 3.2.1, 3.4.1 et 3.5.1);
- Extensions au cadre des opérateurs cocoercifs non potentiels pour les systèmes (AVD), (DIN-AVD) et (ISIHD), avec implications asymptotiques (voir Chapitre 4, Théorèmes 4.2.1, 4.3.1 et 4.4.1);
- Établissement du taux de convergence $O(t^{-\alpha})$ et discussion de son optimalité pour le système (AVD) dans le cas quadratique convexe (voir Chapitre 5, Théorème 5.2.1).

1.2 Introduction (English version)

At the core of many modern methods in machine learning and engineering lies the minimization of convex functions. In general, one studies the minimization problem, in a Hilbert space $\mathcal{H}$,

Licence CC BY-NC-ND 3.0

of a convex function $f : \mathcal{H} \to \mathbb{R}$ that is differentiable with a Lipschitz continuous gradient and whose set of minimizers is nonempty. One of the major contemporary challenges is to understand how to efficiently accelerate first-order optimization methods, that is, methods relying solely on gradient information. This challenge is both theoretical, in terms of analyzing convergence rates, and practical, in terms of designing fast and robust algorithms. This thesis fits within this line of research and provides a detailed analysis of accelerated methods, constantly drawing parallels between discrete algorithms and the continuous differential equations that reveal their underlying structure.

The simplest family of such methods is gradient descent, based on the iterative sequence $(x_k)_k$ defined by

$$x_{k+1} = x_k - s\nabla f(x_k),$$

where $s > 0$ is the step size and $-\nabla f(x_k)$ the descent direction. Under suitable conditions on the step size, this scheme ensures a progressive decrease in the function values of f (descent lemma) and moreover achieves the following convergence rate [1]:

$$f(x_k) - \inf(f) \leq \mathcal{O}\left(\frac{1}{k}\right),$$

where $\inf(f)$ denotes the minimum of f, and, in addition, the sequence $(x_k)_k$ converges weakly to a minimizer of the function. However, despite its simplicity and robustness, this method remains slow: it often requires a large number of iterations to reach a satisfactory level of accuracy.

In 1964, Polyak proposed an improvement known as the heavy ball method [2]. This method introduces an inertial term designed to smooth the trajectory and reinforce the speed of movement in favorable directions:

$$x_{k+1} = x_k - s\nabla f(x_k) + \beta(x_k - x_{k-1}),$$

where $\beta > 0$ regulates the amount of memory preserved from one iteration to the next. This mechanism often reduces oscillations and accelerates convergence, particularly on ill-conditioned problems.

An even more significant breakthrough occurred in 1983 with the work of Nesterov [3], who introduced an accelerated method for convex functions with Lipschitz continuous gradients. The essential idea of Nesterov relies on a clever extrapolation of past iterates (see [3]):

$$\begin{cases} y_k = x_k + \left(1 - \dfrac{\alpha}{k}\right)(x_k - x_{k-1}), \\ x_{k+1} = y_k - s\nabla f(y_k). \end{cases}$$

Licence CC BY-NC-ND 3.0

This extrapolation mechanism has inspired a large part of modern optimization methods, both in theory and in applications. For $\alpha = 3$, Nesterov's algorithm achieves an asymptotic convergence rate of $\mathcal{O}\left(\frac{1}{k^2}\right)$ [3], [4], although the convergence of the sequence itself remains an open question to this day. When $\alpha > 3$, the convergence improves further, with $f(x_k) - \inf(f) = o\left(\frac{1}{k^2}\right)$ as $k \to +\infty$, as shown by Attouch and Peypouquet [5], and the iterates $(x_k)_k$ converge weakly to a minimizer of f, according to the results of Chambolle and Dossal [6]. The subcritical case, corresponding to $0 < \alpha \leq 3$, has also been studied in [7] and [8].

A turning point occurred in 2015, when Su, Boyd, and Candès [9] showed that Nesterov's algorithm can be interpreted as the discretization of a second-order differential equation. This continuous-time dynamics is given by

$$(\text{AVD}) : \ \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0, \quad t > 0.$$

As in the discrete case, the asymptotic behaviour of the dynamics (AVD) depends strongly on the parameter α. In the subcritical regime $0 < \alpha \leq 3$, one obtains a convergence rate of $\mathcal{O}(1/t^{2\alpha/3})$ [7], [8]. The critical case, corresponding to $\alpha = 3$, coincides with Nesterov's method and yields a convergence rate of $\mathcal{O}(1/t^2)$. Finally, in the supercritical regime $\alpha > 3$, one obtains a convergence rate of $o(1/t^2)$ [5], and the trajectories converge weakly to a minimizer of f [6]. These results underscore the role of the parameter α.

This connection between discrete and continuous dynamics opens a rich research direction. It has also motivated the study of variants with Hessian-dependent damping, giving rise to so-called high-resolution dynamics (see [10], [11], [12], [13], [14], [15], [16]), such as

$$(\text{DIN-AVD}) : \ \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta(t)\nabla^2 f(x(t))\dot{x}(t) + b(t)\nabla f(x(t)) = 0,$$

first introduced by Attouch, Peypouquet, and Redont in [13] or the implicit dynamics introduced by Alesca, Laszlo, and Pinta in [16]:

$$(\text{ISIHD}) : \ \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t) + \beta(t)\dot{x}(t)) = 0.$$

These richer dynamics allow for a refined analysis of the decay of the function value, the gradient, the oscillations, and the geometric structure of the trajectory.

The present thesis is organized around a comprehensive analysis of these accelerated dynamics. First, the asymptotic behavior of the (ISIHD) dynamics is analyzed as the damping parameter α tends to infinity. The analysis is then extended to the three dynamics in a non-potential framework. Finally, the (AVD) dynamics are investigated in the case of convex quadratic functions. Each chapter provides both theoretical and numerical contributions, highlighting the role of the parameter α.

Licence CC BY-NC-ND 3.0

Chapter 2 introduces the definitions, notations, and mathematical tools used throughout the manuscript.

Chapter 3 broadens the perspective by studying the limiting behaviour of the accelerated dynamics (ISIHD) as the parameter α tends to infinity. This analysis reveals a transition: the second-order inertial dynamics reduces to a proximal flow. More precisely, introducing the rescaling

$$t = \sqrt{2(\alpha+1)s}$$

and setting $y_\alpha(s) = x(t)$, one shows that the sequence $(y_\alpha)_\alpha$ converges uniformly on every compact subset of $[0, +\infty)$ to the solution of the proximal-type dynamics:

$$\dot{y}(s) + \nabla f(y(s) + \beta_0 \dot{y}(s)) = 0.$$

This equation can be rewritten as

$$\dot{y}(s) = -\frac{1}{\beta_0} y(s) + \frac{1}{\beta_0} \mathrm{prox}_{\beta_0 f}(y(s)) = -\nabla \mathcal{M}_{\beta_0 f}(y(s)),$$

where $\mathrm{prox}_{\beta_0 f}$ and $\mathcal{M}_{\beta_0 f}$ denote respectively the proximal operator and the Moreau envelope of f. The analysis is carried out rigorously for convex functions, and extended to the non-convex case under a quadratic assumption in finite dimension. Uniform energy estimates and the Ascoli–Arzelà theorem are used to establish the convergence.

Chapter 4 continues this investigation in the more general framework of non-potential operators, where the gradient ∇f is replaced by a λ-cocoercive operator V. This extension is essential for addressing variational problems involving constraints or nondifferentiable regularizations. The three main dynamics (AVD), (DIN-AVD), and (ISIHD) are reformulated by replacing ∇f with V. The analysis of their limiting behaviour as $\alpha \to +\infty$ reveals reduction phenomena to first-order flows. For the (AVD) dynamics:

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + V(x(t)) = 0,$$

the rescaling

$$t = \sqrt{2(\alpha+1)s}$$

leads to convergence towards the standard monotone flow:

$$\dot{y}(s) + V(y(s)) = 0.$$

Licence CC BY-NC-ND 3.0

For the (DIN-AVD) dynamics, enriched with an explicit Hessian-type damping:

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta_0 \frac{\alpha+1}{t}\frac{d}{dt}V(x(t)) + V(x(t)) = 0,$$

the singular limit yields a regularized flow of Newton–Levenberg–Marquardt type:

$$\dot{y}(s) + \beta_0 \frac{d}{ds}V(y(s)) + V(y(s)) = 0.$$

Finally, for the (ISIHD) dynamics with implicit damping:

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + V\left(x(t) + \beta_0 \frac{\alpha+1}{t}\dot{x}(t)\right) = 0,$$

the singular limit leads to an implicit monotone flow with particularly rich structure:

$$\dot{y}(s) + V(y(s) + \beta_0 \dot{y}(s)) = 0.$$

The proof of these convergence results relies on variational techniques and energy inequalities tailored to cocoercive operators. The analysis requires developing uniform-in-α estimates for the trajectories and their derivatives, as well as using the Ascoli–Arzelà theorem to establish strong convergence. In the non-potential setting, the monotonicity of the operator V plays the role previously held by the convexity of the function f.

Chapter 5 addresses the issue related to the choice of the parameter α. This chapter is devoted to a refined analysis of the accelerated viscous dynamics (AVD) in the case of convex quadratic functions. This choice is motivated both by the possibility of obtaining explicit solutions and by the representativeness of this setting for a wide range of applied problems. This work considers the quadratic function

$$f(x) = \frac{1}{2}\langle Qx, x\rangle + \langle b, x\rangle + d_0,$$

where Q is a symmetric positive semidefinite matrix, which allows for a diagonalization of the associated dynamical system. The explicit solution of the corresponding differential equation involves Bessel functions of the first kind J_ν and of the second kind Y_ν, of order

$$\nu = \frac{\alpha - 1}{2}.$$

An asymptotic analysis of these solutions shows that the optimality gap decays as $O(t^{-\alpha})$, and this rate is proven to be optimal. The parameter α therefore acts as a convergence accelerator. Numerical experiments support these results, illustrating both the convergence rate and its optimality.

Licence CC BY-NC-ND 3.0

In summary, this thesis investigates acceleration mechanisms in optimization. It examines the connections between discrete algorithms and continuous dynamical systems, analyzes the role of damping and extrapolation parameters, and considers extensions to the non-potential setting. The work includes theoretical and numerical results related to algorithm design and the analysis of their behavior.

Main Contributions

In other words, this thesis presents several results on the analysis of accelerated dynamics and their asymptotic behavior in the convex, quadratic convex, non-convex quadratic, and non-potential operator settings. These works extend and complement classical results and can be summarized in three points:

- Study of the singular perturbation limit $\alpha \to +\infty$ for the (ISIHD) system, leading to a proximal flow (see Chapter 3, Theorems 3.2.1, 3.4.1, and 3.5.1).
- Extensions to the cocoercive/non-potential operator setting for the (AVD), (DIN-AVD), and (ISIHD) systems, with asymptotic implications (see Chapter 4, Theorems 4.2.1, 4.3.1, and 4.4.1).
- Establishment of the convergence rate $O(t^{-\alpha})$ and discussion of its optimality for the (AVD) system in the quadratic convex case (see Chapter 5, Theorem 5.2.1).

Licence CC BY-NC-ND 3.0

2

Background and notations

Contents

2.1 Hilbert Space **24**
2.2 Compactness and Distributions **26**
2.3 Convexity in Hilbert Spaces **30**
2.4 Operators and Bessel Functions **38**
2.4.1 Operators 38
2.4.2 Bessel functions 39

Licence CC BY-NC-ND 3.0

This chapter presents the main mathematical tools that will serve as the foundation of this thesis. It establishes the notation and recalls several notions from functional analysis and convex analysis required for the formulation and study of the problems considered in the following chapters. The optimization questions addressed subsequently are set within a framework involving algebraic, topological, and convex structures, which motivates the preliminary introduction of these notions.
The chapter notably presents:

- Hilbert spaces, which provide the functional framework for the variational formulations under study;
- several tools from convex analysis, such as the subdifferential, the proximal operator, and Fenchel conjugation, used in the modeling and analysis of optimization problems;
- a review of the theory of monotone and cocoercive operators, which plays a role in the study of convergence properties and the iterative methods considered later.

The purpose of this chapter is to collect the definitions and results used in the sequel, without aiming at exhaustiveness. Proofs are included only when they are directly used in subsequent developments.

2.1 Hilbert Space

Hilbert spaces form a particular class of vector spaces that arise naturally in functional analysis, especially in the study of function spaces and notions of orthogonality. The definitions and results presented in this section are classical and are mainly based on [17]. The base field is denoted by $\mathbb{K}$ and is assumed to be either $\mathbb{R}$ or $\mathbb{C}$. The definition of a Hilbert space is recalled and an example is provided, after which the notion of weak convergence and some of its properties are introduced.

Licence CC BY-NC-ND 3.0

Definition 2.1.1 (Hilbert Space)**.** A *Hilbert space* is a $\mathbb{K}$-vector space equipped with an inner product $\langle \cdot, \cdot \rangle$ that is complete with respect to the induced norm $\|x\| = \sqrt{\langle x, x \rangle}$ ([17, Chapter 5]). Recall that a normed vector space is complete if every Cauchy sequence converges.

Completeness ensures that any Cauchy sequence with respect to the induced norm converges in the space under consideration. An example of a Hilbert space is then provided.

Example 2.1.1. Let Ω be an open subset of $\mathbb{R}^n$. The space $L^2(\Omega, \mathbb{R})$, equipped with the inner product

$$\langle f, g \rangle_{L^2} = \int_\Omega f(x)\, g(x)\, dx,$$

is a Hilbert space. The induced norm is given by

$$\|f\|_{L^2} = \left(\int_\Omega |f(x)|^2\, dx \right)^{1/2}.$$

This space provides an example of an infinite-dimensional Hilbert space and arises recurrently in the study of partial differential equations.

In addition to the convergence associated with the norm, referred to as strong convergence, a weaker notion of convergence is defined in Hilbert spaces, called weak convergence, whose definition is given below.

Definition 2.1.2 (Weak convergence)**.** A sequence (x_n) in a Hilbert space $\mathcal{H}$ *weakly converges* to x, denoted $x_n \rightharpoonup x$, if [17]:

$$\forall y \in \mathcal{H}, \quad \lim_{n \to \infty} \langle x_n, y \rangle = \langle x, y \rangle$$

This definition leads to several general properties of weak convergence, which are summarized below.

Proposition 2.1.1. Key properties of weak convergence [17]:

- Strong convergence implies weak convergence.
- In finite dimension, weak and strong convergence coincide.
- The norm is lower semicontinuous with respect to weak convergence; that is, if $(x_n)_n \subset \mathcal{H}$ converges weakly to x, then
$$\|x\| \leq \liminf_{n \to \infty} \|x_n\|.$$
- From any bounded sequence, one may extract a weakly convergent subsequence.

This proposition highlights the relationship between strong convergence and weak convergence, as well as the difference in behavior between finite-dimensional and infinite-dimensional spaces.

Licence CC BY-NC-ND 3.0

2.2 Compactness and Distributions

This section gathers several results and notions from functional analysis that are used in the following chapters. It first presents a fixed point theorem and an integral inequality, then introduces notions of compactness for families of continuous functions, in particular equicontinuity and the Ascoli–Arzelà theorem. Finally, it recalls elements of distribution theory, including the definition of distributions, differentiation in the sense of distributions, and the associated notions of convergence.

The following fixed point theorem provides a condition for the existence and uniqueness of solutions for some contractive mappings in a Hilbert space.

Theorem 2.2.1 (Banach Fixed-Point Theorem)**.** Let $\mathcal{H}$ be a Hilbert space (real or complex) and a contraction mapping T defined from $\mathcal{H}$ into $\mathcal{H}$, i.e., there exists a constant $k \in [0, 1)$ such that:

$$\|T(x) - T(y)\| \leq k\|x - y\| \quad \text{for all } x, y \in \mathcal{H}.$$

Then there exists a unique fixed point $x^* \in \mathcal{H}$ such that $T(x^*) = x^*$.

After recalling the fixed point theorem, an integral inequality is presented to provide a tool for bounding functions in the study of some integral and differential equations.

Lemma 2.2.2. (Integral Inequality [18, Lemma A.5]) Let $m \in L^1(0, T; \mathbb{R})$ with $m \geq 0$ almost everywhere on $(0, T)$, and let $a > 0$. Let $\varphi \in C([0, T], \mathbb{R})$ satisfy

$$\frac{1}{2}\varphi^2(t) \leq \frac{1}{2}a^2 + \int_0^t m(s)\varphi(s)\, ds \quad \forall\, t \in [0, T].$$

Then:

$$|\varphi(t)| \leq a + \int_0^t m(s)\, ds \quad \forall\, t \in [0, T].$$

To characterize the compactness of families of continuous functions, it is necessary to define the notion of equicontinuity, which controls the simultaneous variation of all functions within the family.

Definition 2.2.1 (Equicontinuity)**.** Let $\mathcal{H}$ be a Hilbert space. Consider $E = C([a, b], \mathcal{H})$ with the uniform norm $\|f\|_\infty = \sup_{t\in[a,b]} \|f(t)\|$. A family $\mathcal{F} \subset E$ is *equicontinuous* at $t \in [a, b]$ if:

$$\forall \varepsilon > 0,\ \exists \delta > 0 \text{ such that } \forall f \in \mathcal{F},\ \forall s \in [a, b],\ |t - s| < \delta \implies \|f(t) - f(s)\| < \varepsilon.$$

This notion of equicontinuity allows for the characterization of the compactness of function families and sets the stage for the statement of the Ascoli–Arzelà theorem.

Theorem 2.2.3 (Ascoli–Arzelà)**.** Let $\mathcal{H}$ be a finite-dimensional Hilbert space. A family $\mathcal{F} \subset C([a, b], \mathcal{H})$ is relatively compact if and only if:

1. $\mathcal{F}$ is equicontinuous,

Licence CC BY-NC-ND 3.0

2. For all $t \in [a, b]$, the set $\{f(t) \mid f \in \mathcal{F}\}$ is bounded in $\mathcal{H}$.

The following part of the section presents a review of distribution theory. These concepts generalize the classical notions of function and differentiation, providing a suitable framework for the study of weak convergence and L^2 spaces.

Definition 2.2.2. Let Ω be an open subset of $\mathbb{R}^n$.

- The *support* of a function $f : \Omega \to \mathbb{R}$ is defined by

$$\operatorname{supp}(f) = \overline{\{x \in \Omega : f(x) \neq 0\}}.$$

- The space of test functions $\mathcal{D}(\Omega)$ consists of all C^∞ functions with compact support contained in Ω.
- A sequence $(\varphi_n)_{n\in\mathbb{N}}$ in $\mathcal{D}(\Omega)$ is said to converge to $\varphi \in \mathcal{D}(\Omega)$ in $\mathcal{D}(\Omega)$ if the following conditions are satisfied:
  1. *Common support*: There exists a compact set $K \subset \Omega$ such that $\operatorname{supp}(\varphi_n) \subset K$ for all $n \in \mathbb{N}$.
  2. *Uniform convergence of all derivatives*: For every multi-index $\alpha \in \mathbb{N}^n$,

$$\partial^\alpha \varphi_n \xrightarrow[n\to\infty]{} \partial^\alpha \varphi \quad \text{uniformly on } \Omega,$$

  that is,

$$\lim_{n\to\infty} \sup_{x\in\Omega} \left|\partial^\alpha \varphi_n(x) - \partial^\alpha \varphi(x)\right| = 0.$$

Definition 2.2.2 introduces the notion of convergence in the space of test functions, which will be useful for defining distributions.

Definition 2.2.3 (Distribution on Ω)**.** A *distribution* on the open set $\Omega \subset \mathbb{R}^n$ is a continuous linear form

$$T : \mathcal{D}(\Omega) \to \mathbb{K}, \quad \phi \mapsto T(\phi) := \langle T, \phi \rangle,$$

meaning that for any sequence $(\varphi_n)_{n\in\mathbb{N}}$ in $\mathcal{D}(\Omega)$ converging to 0,

$$\langle T, \varphi_n \rangle \longrightarrow 0 \quad \text{as } n \to \infty.$$

The set of all distributions on Ω is denoted by $\mathcal{D}'(\Omega)$.

The following examples illustrate regular distributions and the Dirac distribution, and establish the connection between classical functions and distributions.

Example 2.2.1.

- *Regular distributions:* Any locally integrable function $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ defines a distribution by

$$\langle T_f, \phi \rangle = \int_{\mathbb{R}^n} f(x)\phi(x)\, dx.$$

Licence CC BY-NC-ND 3.0

- *Dirac distribution:* The Dirac distribution at $a \in \mathbb{R}^n$ is defined by

$$\langle \delta_a, \phi \rangle = \phi(a).$$

The definition of distributions naturally leads to differentiation in the sense of distributions, which extends the classical notion of derivative to this more general framework and paves the way for the following definition.

Definition 2.2.4. Let $\alpha \in \mathbb{N}^n$ and T be a distribution on Ω. The derivative of order α of T, denoted $\partial^\alpha T$, is defined by

$$\langle \partial^\alpha T, \phi \rangle = (-1)^{|\alpha|} \langle T, \partial^\alpha \phi \rangle$$

for all test functions $\phi \in \mathcal{D}(\Omega)$.

The following examples illustrate the application of this definition to classical distributions, such as the Heaviside function and the Dirac distribution.

Example 2.2.2.

- The derivative H' of the Heaviside function H ($H(x) = 0$ for $x < 0$, $H(x) = 1$ for $x \geq 0$) is the Dirac distribution at 0: $H' = \delta_0$.
- The derivative of the Dirac distribution at $a \in \mathbb{R}$ is given by $\langle \delta_a', \phi \rangle = -\phi'(a)$.

To conclude this section, the notion of convergence of sequences of distributions is presented to characterize the limiting behavior of sequences of elements in $\mathcal{D}'(\Omega)$, in accordance with differentiation in the sense of distributions.

Definition 2.2.5 (Convergence of sequences of distributions)**.** Let $(T_n)_{n\geq 0}$ and T be distributions on Ω. It is said that $(T_n)_{n\geq 0}$ converges to T in $\mathcal{D}'(\Omega)$ if:

$$\langle T_n, \phi \rangle \to \langle T, \phi \rangle, \quad \forall \phi \in \mathcal{D}(\Omega).$$

The following proposition gathers several properties of distributions and their modes of convergence, linking differentiation and distributional convergence.

Proposition 2.2.1. Let Ω be an open subset of $\mathbb{R}^n$.

1. Every distribution is differentiable (in the sense of distributions).
2. If $\alpha \in \mathbb{N}^n$ and $(T_n)_{n\geq 0}$ is a sequence of distributions converging to a distribution T, then the sequence $(\partial^\alpha T_n)_{n\geq 0}$ converges to $\partial^\alpha T$.
3. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $L^2(\Omega)$ and let $f \in L^2(\Omega)$. Then

$$\left(f_n \rightharpoonup f \text{ in } L^2(\Omega)\right) \iff \left(f_n \to f \text{ in } \mathcal{D}'(\Omega) \text{ and } (f_n)_{n\in\mathbb{N}} \text{ is bounded in } L^2(\Omega)\right).$$

Proof. Proof of 1: Let $T \in \mathcal{D}'(\Omega)$ and $\phi \in \mathcal{D}(\Omega)$. By definition,

$$\langle \partial^\alpha T, \phi \rangle = (-1)^{|\alpha|} \langle T, \partial^\alpha \phi \rangle.$$

Licence CC BY-NC-ND 3.0

This definition makes sense since $\partial^\alpha \phi \in \mathcal{D}(\Omega)$. Linearity is immediate, and continuity follows from the continuity of T: if $\phi_n \to 0$ in $\mathcal{D}(\Omega)$, then $\partial^\alpha \phi_n \to 0$ in $\mathcal{D}(\Omega)$, hence

$$\langle \partial^\alpha T, \phi_n \rangle = (-1)^{|\alpha|} \langle T, \partial^\alpha \phi_n \rangle \longrightarrow 0.$$

Proof of 2: Let us assume that $T_n \to T$ in $\mathcal{D}'(\Omega)$. For every $\phi \in \mathcal{D}(\Omega)$,

$$\langle \partial^\alpha T_n, \phi \rangle = (-1)^{|\alpha|} \langle T_n, \partial^\alpha \phi \rangle.$$

Passing to the limit gives

$$(-1)^{|\alpha|} \langle T_n, \partial^\alpha \phi \rangle \longrightarrow (-1)^{|\alpha|} \langle T, \partial^\alpha \phi \rangle = \langle \partial^\alpha T, \phi \rangle,$$

which proves that $\partial^\alpha T_n \to \partial^\alpha T$ in $\mathcal{D}'(\Omega)$.

Proof of 3:
Direct implication: Assume that $f_n \rightharpoonup f$ in $L^2(\Omega)$.

- *Boundedness.* A weakly convergent sequence in a Banach space is bounded (consequence of Banach–Steinhaus principle).
- *Distributional convergence.* For all $\phi \in \mathcal{D}(\Omega) \subset L^2(\Omega)$,

$$\langle T_{f_n}, \phi \rangle = \langle f_n, \phi \rangle_{L^2} \longrightarrow \langle f, \phi \rangle_{L^2} = \langle T_f, \phi \rangle,$$

  so $f_n \to f$ in $\mathcal{D}'(\Omega)$.

Reverse implication: Assume $f_n \to f$ in $\mathcal{D}'(\Omega)$ and that $(f_n)_{n\in\mathbb{N}}$ is bounded in $L^2(\Omega)$. Then there exists $M > 0$ such that $\|f_n - f\|_{L^2} \leq M$. Let $g \in L^2(\Omega)$. By density of $\mathcal{D}(\Omega)$ in $L^2(\Omega)$, for every $\varepsilon > 0$, there exists $\tilde{g} \in \mathcal{D}(\Omega)$ such that

$$\|g - \tilde{g}\|_{L^2} < \frac{\varepsilon}{2M}.$$

By distributional convergence, there exists $N \in \mathbb{N}$ such that for all $n \geq N$,

$$|\langle T_{f_n - f}, \tilde{g} \rangle| = |\langle f_n - f, \tilde{g} \rangle_{L^2}| < \frac{\varepsilon}{2}.$$

Licence CC BY-NC-ND 3.0

Then, for any $n \geq N$,

$$\begin{aligned} |\langle f_n - f, g\rangle_{L^2}| &= |\langle T_{f_n - f}, g\rangle| \\ &\leq |\langle T_{f_n - f}, \tilde{g}\rangle| + |\langle T_{f_n - f}, g - \tilde{g}\rangle| \\ &\leq \frac{\varepsilon}{2} + \|f_n - f\|_{L^2} \, \|g - \tilde{g}\|_{L^2} \\ &\leq \frac{\varepsilon}{2} + M \cdot \frac{\varepsilon}{2M} = \varepsilon. \end{aligned}$$

Thus $f_n \rightharpoonup f$ in $L^2(\Omega)$.

□

2.3 Convexity in Hilbert Spaces

Some basic notions of convex analysis within the framework of Hilbert spaces are collected here. They are used in the formulation and analysis of the optimization problems considered subsequently, particularly when these problems involve potentially non-differentiable functions. The concepts of convex sets and convex functions are introduced first, followed by various characterizations of convexity. Tools such as the subdifferential, Fenchel conjugation, the proximal operator, and the Moreau envelope are then presented. These notions allow for the formulation of optimality conditions suitable for a general functional framework.

Definition 2.3.1 (Convex Set)**.** Let $\mathcal{H}$ be a Hilbert space. A subset $C \subseteq \mathcal{H}$ is called *convex* if:

$$\forall x, y \in C, \ \forall \lambda \in [0,1], \quad \lambda x + (1-\lambda) y \in C.$$

The previous definition is illustrated below with some classical examples of convex sets in a Hilbert space.

Example 2.3.1. Classical examples include:

- Open balls: $B(x_0, r) = \{x \in \mathcal{H} \mid \|x - x_0\| < r\}$
- Vector subspaces
- Half-spaces: $\{x \in \mathcal{H} \mid \langle a, x\rangle \leq b\}$ with $a \in \mathcal{H}$, $b \in \mathbb{R}$

Convexity is also preserved under several standard operations, as noted in the following remark.

Remark 2.3.1 (Stability)**.** The following operations preserve convexity:

- *Finite sum:* If C_1, C_2 are convex, then $C_1 + C_2$ is convex
- *Intersection:* Any intersection of convex sets is convex
- *Affine image:* If $A : \mathcal{H} \to \mathcal{H}$ is affine and C is convex, then $A(C)$ is convex
- *Preimage under affine map:* $A^{-1}(C)$ is convex if C is convex

The notion of a convex set allows for the introduction of convex functions, defined naturally from the convexity of their epigraph.

Licence CC BY-NC-ND 3.0

Definition 2.3.2. Let $\mathcal{H}$ be a Hilbert space. A function $f : \mathcal{H} \to \mathbb{R} \bigcup \{+\infty\}$ is called *convex* if its epigraph

$$\text{epi}(f) = \{(x, \alpha) \in \mathcal{H} \times \mathbb{R} \mid f(x) \leq \alpha\}$$

is a convex set in $\mathcal{H} \times \mathbb{R}$.

The following remark clarifies the relationship between this definition and the geometric interpretation of function convexity.

Remark 2.3.2. This geometric definition directly connects the convexity of functions to that of sets. The epigraph represents the set of points lying above the graph of f, and its convexity intuitively captures the idea that line segments between points on the graph lie above the graph itself.

The geometric definition of convexity can be reformulated in an equivalent algebraic form, which is often more suitable for applications.

Proposition 2.3.1 (Equivalent Characterization)**.** For a function $f : \mathcal{H} \to \mathbb{R} \bigcup \{+\infty\}$, the following statements are equivalent:

1. f is convex in the sense of the epigraph definition
2. $\text{Dom}(f) = \{x \in \mathcal{H} \mid f(x) < +\infty\}$ is convex, and

$$\forall x, y \in \text{Dom}(f),\ \forall \lambda \in [0, 1], \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y)$$

Proof. The proof relies on two observations:

- If the epigraph is convex, then for $(x, f(x))$ and $(y, f(y))$ in $\text{epi}(f)$, their convex combination must also belong to $\text{epi}(f)$.
- Conversely, the algebraic condition implies that any convex combination of points in the epigraph remains in the epigraph.

□

In the case where the function under consideration is differentiable, convexity can be characterized using analytical properties involving the gradient. To formulate these differential characterizations, the notion of Fréchet differentiability in a Hilbert space is recalled first.

Definition 2.3.3 (Fréchet Differential)**.** A function $f : \mathcal{H} \to \mathbb{R}$ is *Fréchet-differentiable* (or differentiable) at $x \in \mathcal{H}$ if there exists a continuous linear map $T : \mathcal{H} \to \mathbb{R}$ such that:

$$\lim_{\|h\| \to 0} \frac{|f(x + h) - f(x) - T(h)|}{\|h\|} = 0.$$

In this case, by the Riesz representation theorem, there exists a unique vector denoted $\nabla f(x) \in \mathcal{H}$ satisfying:

$$T(h) = \langle \nabla f(x), h \rangle, \quad \forall h \in \mathcal{H}.$$

Licence CC BY-NC-ND 3.0

$\nabla f(x)$ is called the *gradient* of f at x, and T is often denoted $\mathrm{D}f(x)$.

The following remark specifies the directional form of the differential for a Fréchet-differentiable function.

Remark 2.3.3. For a Fréchet-differentiable function at x, the differential can be expressed as:

$$\langle \nabla f(x), y \rangle = \lim_{t \to 0^+} \frac{f(x+ty) - f(x)}{t}, \quad \forall y \in \mathcal{H},$$

where this limit (directional derivative) exists and is linear in y.

These concepts allow for the statement of a differential characterization of convexity, linking the properties of the function to those of its gradient, as expressed in the following theorem.

Theorem 2.3.1 (Differential Characterization [19])**.** For a differentiable function f on $\mathcal{H}$, the following statements are equivalent:

1. f is convex
2. $\forall x, y \in \mathcal{H}$, $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$
3. ∇f is monotone: $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0 \quad \forall x, y \in \mathcal{H}$

Proof. We prove the equivalences by successive implications:

(1) ⇒ (2): Convexity ⇒ Subgradient Inequality

Let f be convex. For $x, y \in \mathcal{H}$ and $\lambda \in (0, 1]$, convexity gives:

$$f(x + \lambda(y - x)) \leq (1 - \lambda) f(x) + \lambda f(y).$$

Rearranging:

$$\frac{f(x + \lambda(y - x)) - f(x)}{\lambda} \leq f(y) - f(x).$$

Taking the limit $\lambda \to 0^+$, one obtains:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

since $\lim_{\lambda \to 0^+} \frac{f(x+\lambda(y-x)) - f(x)}{\lambda} = \mathrm{D}f(x)(y - x) = \langle \nabla f(x), y - x \rangle$.

(2) ⇒ (3): Inequality ⇒ Gradient Monotonicity

Assume (2) holds. For $x, y \in \mathcal{H}$, two inequalities are obtained:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$$

Licence CC BY-NC-ND 3.0

The sum of these inequalities gives:

$$0 \geq \langle \nabla f(x) - \nabla f(y), y - x \rangle$$

which is equivalent to $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0$.

(3) ⇒ (2): Monotonicity ⇒ Inequality

Define the auxiliary function:

$$m : [0,1] \to \mathbb{R}, \quad t \mapsto f(x + t(y - x))$$

for fixed $x, y \in \mathcal{H}$. By composition, m is differentiable on $[0,1]$ with derivative:

$$m'(t) = \langle \nabla f(x + t(y - x)), y - x \rangle.$$

By applying hypothesis (3) to m, the following result is obtained:

$$m(1) - m(0) = \int_0^1 m'(t)dt \geq m'(0),$$

which can be rewritten exactly as:

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle.$$

Thus, condition (2) is satisfied.

(2) ⇒ (1): Inequality ⇒ Convexity

Let $x, y \in \mathcal{H}$ with $x \neq y$ and $\lambda \in (0,1)$. Set $x_\lambda = (1 - \lambda)x + \lambda y$.

Apply hypothesis (2) to the pairs (x_λ, x) and (x_λ, y):

$$f(x) \geq f(x_\lambda) + \langle \nabla f(x_\lambda), x - x_\lambda \rangle$$
$$f(y) \geq f(x_\lambda) + \langle \nabla f(x_\lambda), y - x_\lambda \rangle$$

Observe that:

$$x - x_\lambda = \lambda(x - y), \quad y - x_\lambda = (1 - \lambda)(y - x).$$

Licence CC BY-NC-ND 3.0

Multiply the first inequality by $(1-\lambda)$ and the second by λ, then sum:

$$\begin{aligned}(1-\lambda)f(x)+\lambda f(y) &\geq (1-\lambda)f(x_\lambda)+\lambda f(x_\lambda)\\ &\quad+\langle\nabla f(x_\lambda),(1-\lambda)\lambda(x-y)+\lambda(1-\lambda)(y-x)\rangle\\ &= f(x_\lambda)+0\end{aligned}$$

Finally,

$$f((1-\lambda)x+\lambda y)\leq(1-\lambda)f(x)+\lambda f(y),$$

which proves that f is convex. □

A particular class of convex functions is introduced to strengthen the convexity properties by adding a uniform curvature constraint.

Definition 2.3.4 (Strongly Convex Function)**.** Let $f:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}$ and $\mu>0$. The function f is said to be *strongly convex with modulus* μ if:

1. its domain $\mathrm{Dom}(f)=\{x\in\mathcal{H}\mid f(x)<+\infty\}$ is convex;
2. for all $x,y\in\mathcal{H}$ and all $\lambda\in[0,1]$, one has:

$$f(\lambda x+(1-\lambda)y)\leq\lambda f(x)+(1-\lambda)f(y)-\frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2.$$

The following remark clarifies the analytical interpretation of strong convexity when the function is differentiable.

Remark 2.3.4. This condition strengthens classical convexity by adding a quadratic term $-\frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2$, which imposes *uniform curvature* on the function. When f is differentiable on $\mathcal{H}$, this property is equivalent to:

$$f(y)\geq f(x)+\langle\nabla f(x),y-x\rangle+\frac{\mu}{2}\|y-x\|^2,\quad\forall x,y\in\mathcal{H},$$

or, equivalently,

$$\langle\nabla f(x)-\nabla f(y),x-y\rangle\geq\mu\|x-y\|^2,\quad\forall x,y\in\mathcal{H}.$$

To handle potentially non-differentiable convex functions, convex analysis introduces dual constructions that allow the reformulation of optimization problems. Fenchel conjugation, as presented in [19], is one of these tools.

Definition 2.3.5 (Convex Conjugate)**.** Let $f:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}$ be a proper function, i.e. $f\not\equiv+\infty$. The *convex conjugate* of f is defined by:

$$f^*(z)=\sup_{x\in\mathcal{H}}\big(\langle z,x\rangle-f(x)\big).$$

Licence CC BY-NC-ND 3.0

The mapping $f \mapsto f^*$ is called the *Legendre-Fenchel transform*.

The formal properties of the convex conjugate are specified by the following theorem, which establishes its convexity, lower semicontinuity, and the relationships with the biconjugate as well as the Fenchel

Theorem 2.3.2. [19] Let $f : \mathcal{H} \to \mathbb{R} \bigcup \{+\infty\}$ be a proper function (that is, $\mathrm{Dom}(f) \neq \emptyset$).

1. f^* is always convex and lower semi-continuous
2. *Fenchel-Young inequality*: $\forall x, z \in \mathcal{H}$, $\langle z, x \rangle \leq f(x) + f^*(z)$
3. *Biconjugate*: f^{**} is the greatest convex lower semi-continuous function minorizing f
4. $f^{**} = f$ if and only if f is convex and lower semi-continuous.

Recall that a function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous (l.s.c.) if for all $x \in \mathcal{H}$, $f(x) \leq \liminf_{z \to x} f(z)$.

The general properties stated in the theorem allow for the illustration of the behavior of convex conjugates in concrete cases, such as L^p norms or quadratic forms, presented in the following example.

Example 2.3.2. Here are some examples of convex conjugates:

1. L^p *norms*: Let $p \in (1, +\infty)$ and q its Hölder conjugate (i.e. $\frac{1}{p} + \frac{1}{q} = 1$). Then:

$$\left(\frac{1}{p}\|\cdot\|_p^p\right)^* = \frac{1}{q}\|\cdot\|_q^q$$

2. *Quadratic form*: Let $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite and $f(x) = \frac{1}{2}\langle Qx, x \rangle$. Then:

$$f^*(z) = \frac{1}{2}\langle Q^{-1}z, z \rangle$$

The subdifferential notion generalizes the gradient to non-differentiable convex functions and allows for the formulation of optimality conditions.

Definition 2.3.6 (Subdifferential)**.** For a convex function $f : \mathcal{H} \to \mathbb{R} \bigcup \{+\infty\}$, at a point x where f is finite, the *subdifferential* of f at x is the set, denoted $\partial f(x)$, of slopes z of affine minorants of f that coincide with it at x:

$$\partial f(x) = \{z \in \mathcal{H} : f(y) \geq f(x) + \langle z, y - x \rangle, \forall y \in \mathcal{H}\}.$$

The definition of the subdifferential establishes a precise link between a convex function and its conjugate, as expressed by the following duality result.

Proposition 2.3.2 (Subdifferential Duality [19])**.** Let f be a convex l.s.c. function. Then, for all $(x, s) \in \mathcal{H} \times \mathcal{H}$:

$$s \in \partial f(x) \Leftrightarrow x \in \partial f^*(s).$$

Licence CC BY-NC-ND 3.0

Proof. The following equivalences hold:

$$\begin{aligned} s \in \partial f(x) &\Leftrightarrow f^*(s) + f(x) = \langle s, x \rangle \\ &\Leftrightarrow f^*(s) + f^{**}(x) = \langle s, x \rangle \quad \text{(since } f \text{ is convex)} \\ &\Leftrightarrow x \in \partial f^*(s) \end{aligned}$$

□

The proximal operator also constitutes a tool for nonsmooth optimization, allowing the extension of differential concepts to the non-regular case.

Definition 2.3.7 (Proximal Operator)**.** The proximal operator of f with parameter $\lambda > 0$ is denoted $\text{prox}_{\lambda f}$ and defined for all $x \in \mathcal{H}$ by:

$$\text{prox}_{\lambda f}(x) = \texttt{argmin}_{u \in \mathcal{H}} \left(f(u) + \frac{1}{2\lambda} \|u - x\|^2 \right)$$

This definition is complemented by some classical examples, which illustrate the behavior of the proximal operator for convex functions frequently encountered in optimization.

Example 2.3.3.

1. *Zero function*: If $f(x) = 0$, then $\text{prox}_f(x) = x$
2. *Indicator function*: If $f(x) = \chi_{\mathcal{N}}(x)$ where $\mathcal{N} \subseteq \mathbb{R}^n$ is a closed convex set, then:

$$\text{prox}_f(x) = P_{\mathcal{N}}(x)$$

   where $P_{\mathcal{N}}$ is the Euclidean projection onto $\mathcal{N}$.
3. ℓ_1 *norm*: If $f(x) = \alpha \|x\|_1$, then prox_f is the soft-thresholding operator:

$$(\text{prox}_f(x))_i = \begin{cases} x_i - \alpha & \text{if } x_i \geq \alpha \\ 0 & \text{if } |x_i| \leq \alpha \\ x_i + \alpha & \text{if } x_i \leq -\alpha \end{cases}$$

The variational definition of the proximal operator enables the statement of general results on existence, uniqueness, and characterization, formulated in the following proposition.

Proposition 2.3.3 (Existence and Characterization [20])**.** If $f : \mathcal{H} \to \mathbb{R} \bigcup \{+\infty\}$ is a convex l.s.c. function, then:

1. $\text{prox}_{\lambda f}(x)$ exists and is unique for all $x \in \mathcal{H}$
2. $u = \text{prox}_{\lambda f}(x) \Leftrightarrow 0 \in \lambda \partial f(u) + u - x$
3. $u = \text{prox}_{\lambda f}(x) \Leftrightarrow u = (\text{Id} + \lambda \partial f)^{-1}(x)$

Licence CC BY-NC-ND 3.0

The preceding characterizations lead to metric properties of the proximal operator, in particular non-expansiveness and continuity.

Proposition 2.3.4 (Contractivity Properties [20])**.** Let $f : \mathcal{H} \to \mathbb{R} \bigcup \{+\infty\}$ be a convex l.s.c. function. Then:

1. *Strongly non-expansive*:

$$\langle \operatorname{prox}_f(x) - \operatorname{prox}_f(y), x - y \rangle \geq \|\operatorname{prox}_f(x) - \operatorname{prox}_f(y)\|^2$$

2. *Lipschitz continuous*: prox_f is Lipschitz continuous with constant 1:

$$\|\operatorname{prox}_f(x) - \operatorname{prox}_f(y)\| \leq \|x - y\|$$

The proximal operator is closely related to convex duality. This relationship is formalized by Moreau's identity, which links the proximal operators of a function and its conjugate.

Theorem 2.3.3 (Moreau Identity [21])**.** Let $f : \mathcal{H} \to \mathbb{R} \bigcup \{+\infty\}$ be a convex l.s.c. function and $\lambda > 0$. Then:

$$x = \operatorname{prox}_{\lambda f}(x) + \lambda \operatorname{prox}_{f^*/\lambda}(x/\lambda)$$

The proximal operator is closely related to a regularization operation for convex functions, defined by the Moreau envelope.

Definition 2.3.8 (Moreau Envelope)**.** For $\lambda > 0$, the *Moreau envelope* of a convex l.s.c. function $f : \mathcal{H} \to \mathbb{R}$ is:

$$\mathcal{M}_{\lambda f}(x) = \min_{y \in \mathcal{H}} \left(f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right).$$

The regularization properties of the Moreau envelope, regarding its differentiability and the Lipschitz behavior of its gradient, follow directly from its definition [20], [21].

Theorem 2.3.4 (Regularization Properties)**.** Let $f : \mathcal{H} \to \mathbb{R}$ be a convex l.s.c. function. The Moreau envelope of f has the following remarkable properties:

1. *Differentiability*: $\mathcal{M}_{\lambda f}$ is $\mathcal{C}^1$ with

$$\nabla \mathcal{M}_{\lambda f}(x) = \frac{1}{\lambda}(x - \operatorname{prox}_{\lambda f}(x)). \tag{2.1}$$

2. *Lipschitz continuity*:

$$\|\nabla \mathcal{M}_{\lambda f}(x) - \nabla \mathcal{M}_{\lambda f}(y)\| \leq \frac{1}{\lambda} \|x - y\|. \tag{2.2}$$

3. f and $\mathcal{M}_{\lambda f}$ have the same minimizers and the same minimum value.

Licence CC BY-NC-ND 3.0

The set of tools introduced in this section allows for the formulation of equivalent optimality conditions for a broad class of convex optimization problems.

Theorem 2.3.5 (Optimality Condition)**.** Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be proper, convex, and lower semi-continuous. The following statements are equivalent:

1. x^* is a global minimizer
2. $0 \in \partial f(x^*)$ (subdifferential condition)
3. $\mathrm{prox}_{\lambda f}(x^*) = x^*$ (proximal characterization)
4. If f is differentiable at x^*, then $\nabla f(x^*) = 0$

The following definition introduces two notions describing the geometry of a convex function around its minimizers. These notions will be useful in the introduction of Chapter 5.

Definition 2.3.9. (Geometric properties near minimizers [22]) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex differentiable function, $S := \mathrm{argmin}(f) \neq \emptyset$.

1. Let $\gamma \geq 1$. The function f satisfies the hypothesis $H_1(\gamma)$ if, for any minimizer $x^* \in S$, there exists $\eta > 0$ such that

$$\forall x \in B(x^*, \eta), \quad f(x) - \inf(f) \leq \frac{1}{\gamma} \langle \nabla f(x), x - x^* \rangle.$$

2. Let $r \geq 1$. The function f satisfies the growth condition $H_2(r)$ if, for any minimizer $x^* \in S$, there exist $K > 0$ and $\varepsilon > 0$ such that

$$\forall x \in B(x^*, \varepsilon), \quad K\, d(x, X^*)^r \leq f(x) - \inf(f).$$

2.4 Operators and Bessel Functions

This section gathers several definitions and results concerning, on the one hand, some operators defined on a Hilbert space, and on the other hand, a review of Bessel functions. It is organized into two subsections corresponding to these two sets of notions, which are introduced successively.

2.4.1 Operators

Let $\mathcal{H}$ be a Hilbert space equipped with its inner product $\langle \cdot, \cdot \rangle$ and the associated norm $\| \cdot \|$. The first definition in this subsection concerns the notion of a monotone operator, which serves as the starting point for the definitions introduced subsequently.

Definition 2.4.1 (Monotone Operator)**.** An operator $A : \mathcal{H} \to \mathcal{H}$ is *monotone* if:

$$\langle A(x) - A(y), x - y \rangle \geq 0 \quad \forall x, y \in \mathcal{H}$$

Licence CC BY-NC-ND 3.0

The following examples are provided to illustrate the preceding definition.

Example 2.4.1.
1. The gradient of a convex differentiable function is monotone
2. Symmetric positive semi-definite linear operators are monotone
3. The proximal operator $\mathrm{prox}_{\lambda f}$ of a proper, convex, and l.s.c function f, with parameter $\lambda > 0$, is monotone.

Based on the notion of a monotone operator, a more restrictive definition is introduced below, based on the addition of a quantitative condition.

Definition 2.4.2 (Strongly Monotone Operator)**.** An operator $A : \mathcal{H} \to \mathcal{H}$ is *strongly monotone* with modulus $\mu > 0$ if:

$$\langle A(x) - A(y), x - y\rangle \geq \mu\|x - y\|^2 \quad \forall x, y \in \mathcal{H}$$

Another property of operators, distinct from strong monotonicity, is introduced in the following definition.

Definition 2.4.3 (Cocoercive Operator)**.** An operator $A : \mathcal{H} \to \mathcal{H}$ is *cocoercive* with parameter $\lambda > 0$ if:

$$\langle A(x) - A(y), x - y\rangle \geq \lambda\|A(x) - A(y)\|^2 \quad \forall x, y \in \mathcal{H}$$

Below are some classical examples of cocoercive operators.

Example 2.4.2.
1. The gradient of a convex differentiable function with a Lipschitz continuous gradient is a cocoercive operator.
2. The proximal operator $\mathrm{prox}_{\lambda f}$ of a proper, convex, and l.s.c function f, with parameter $\lambda > 0$, is a cocoercive operator.

The following remark highlights an immediate consequence of the preceding definition.

Remark 2.4.1. Cocoercivity is a stronger property than Lipschitz continuity because every λ-cocoercive operator is Lipschitz continuous with constant $1/\lambda$. Indeed, by Cauchy-Schwarz and cocoercivity:

$$\lambda\|A(x) - A(y)\|^2 \leq \langle A(x) - A(y), x - y\rangle \leq \|A(x) - A(y)\| \cdot \|x - y\|$$

Dividing by $\|A(x) - A(y)\|$ (if not zero) yields the result.

2.4.2 Bessel functions

Bessel functions naturally arise in the study of some differential equations after an appropriate change of variables. In particular, in the quadratic case of convex functions, the exact solution of the (AVD) system can be obtained by reducing the dynamics to a Bessel differential equation (see Chapter 5, Lemmas 5.2.2 and 5.2.3). Bessel functions are solutions of the

Licence CC BY-NC-ND 3.0

Bessel differential equation, which is frequently encountered in physics and engineering. They possess analytical and asymptotic properties that allow the description of oscillatory behavior and the derivation of precise asymptotic expansions.
But what exactly are Bessel functions?

Definition 2.4.4.

- The *Bessel function of the first kind of order* $\mu > 0$, denoted by J_μ, is defined by the entire series

$$J_\mu(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,\Gamma(k+\mu+1)} \left(\frac{x}{2}\right)^{2k+\mu}, \tag{2.3}$$

  where Γ denotes Euler's gamma function, defined for $z > 0$ by

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t}\, dt,$$

  which extends the factorial in the sense that $\Gamma(n+1) = n!$ for all $n \in \mathbb{N}$.
- For $\mu > 0$ non-integer, the *Bessel function of the second kind of order* μ, also called the *Neumann function*, denoted by Y_μ, is defined for $x > 0$ by

$$Y_\mu(x) = \frac{J_\mu(x)\cos(\mu\pi) - J_{-\mu}(x)}{\sin(\mu\pi)}. \tag{2.4}$$

  This definition extends by continuity to integer values of μ.

The following theorem concerns the solutions of the differential equation associated with the functions introduced above.

Theorem 2.4.1. Let $\mu > 0$. The functions J_μ and Y_μ are two linearly independent solutions of the Bessel ordinary differential equation

$$x^2 y'' + xy' + (x^2 - \mu^2)y = 0 \quad \text{on } (0, +\infty). \tag{2.5}$$

They form a basis of the real vector space of solutions of (2.5) on $(0, +\infty)$. In particular, any solution y of (2.5) can be written uniquely as

$$y(x) = A\, J_\mu(x) + B\, Y_\mu(x), \qquad A, B \in \mathbb{R}.$$

Proof. For a rigorous proof of this theorem, the reader is referred to [23]. □

Licence CC BY-NC-ND 3.0

3

From the inertial system to proximal flow

Contents

3.1 Introduction **42**
3.1.1 Main results 44
3.1.2 Chapter Contents 46
3.2 Direct Approach **46**
3.3 Time-Scaling Approach **50**
3.4 Limit dynamics of (ISIHD): Convex case **57**
3.5 Limit dynamics of (ISIHD): Non-convex quadratic case **69**
3.6 Connection Between Continuous Dynamics and Discrete Algorithms . . . **73**
3.7 Numerical Experiments **75**
3.8 Conclusion and Perspectives **80**

Licence CC BY-NC-ND 3.0

3.1 Introduction

This chapter is appointed to the dynamic analysis of accelerated gradient methods. The setting is a real Hilbert space $\mathcal{H}$, endowed with the inner product $\langle \cdot, \cdot \rangle$ and the associated norm $\| \cdot \|$, and the unconstrained minimization problem

$$\min_{x \in \mathcal{H}} f(x),$$

where $f : \mathcal{H} \to \mathbb{R}$ is the objective function. The following assumptions are made on f:

(H1) *Differentiability*: f is continuously differentiable on $\mathcal{H}$.

(H2) *Local Lipschitzness*: the gradient ∇f is locally Lipschitz on bounded subsets of $\mathcal{H}$. More precisely, for all $R > 0$, there exists $L_R > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L_R \|x - y\|, \quad \forall x, y \in B(0, R).$$

(H3) *Lower boundedness*: $\inf(f) > -\infty$.

(H4) *Existence of minimizers*: the set $S = \{x \in \mathcal{H} : f(x) = \inf(f)\}$ is nonempty.

(H5) *Absence of convexity*: unless otherwise stated, f is not assumed convex.

An important contribution to the dynamic understanding of Nesterov's method is due to Su, Boyd, and Candès [9], who showed that Nesterov's accelerated gradient method can be obtained as a temporal discretization of the dynamics

$$\text{(AVD)} : \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0.$$

This dynamics has been studied in detail in [9], as well as by Attouch, Chbani, Peypouquet, and Redont [24], and by May [25]. The temporal discretization of (AVD) leads to the following Nesterov accelerated gradient algorithm:

$$\begin{cases} y_k = x_k + \left(1 - \dfrac{\alpha}{k}\right)(x_k - x_{k-1}), \\ x_{k+1} = y_k - s\nabla f(y_k). \end{cases} \tag{3.1}$$

In this algorithm, the gradient step is applied at y_k, obtained by linear extrapolation from x_k and x_{k-1}. Note the particular choice of the extrapolation coefficient $0 < \alpha_k := 1 - \frac{\alpha}{k} < 1$, which approaches 1 from below as $k \to +\infty$.

For convex functions f with Lipschitz gradients, this algorithm has been extensively studied. Nesterov's classical algorithm corresponds to $\alpha = 3$ and achieves an asymptotic convergence rate of $\mathcal{O}(\frac{1}{k^2})$ (see [3], [4]). When $\alpha > 3$, the convergence rate improves to $f(x_k) - \inf(f) = o(\frac{1}{k^2})$ as $k \to +\infty$, as shown by Attouch and Peypouquet [5], and the iterates converge, as proved by Chambolle and Dossal [6]. The subcritical case $\alpha \leq 3$ was examined by Apidopoulos, Aujol

Licence CC BY-NC-ND 3.0

and Dossal in [7] and by Attouch, Chbani, Riahi in [8]. For strongly convex functions, it has been shown in [9], [24], [26] that the convergence rate of the function values is $\mathcal{O}\left(\frac{1}{t^{2\alpha/3}}\right)$, and it can become arbitrarily fast (in powers of $\frac{1}{t}$) for sufficiently large α.

Recent advances have led to a finer understanding of Nesterov's method through more precise continuous dynamic interpretations. The traditional low-resolution ODE (ordinary differential equation) approach, followed by Su, Boyd, and Candès, introduces the (AVD) dynamics. However, this approach does not adequately capture an essential property of accelerated gradient methods: the asymptotically fast convergence of the gradients to zero. To overcome this limitation, the high-resolution ODE provides a more precise dynamic interpretation, as explored in detail by H. Attouch and J. Fadili in [10].

The high-resolution method, commonly used in fluid mechanics to account for multi-scale physical phenomena, is applied here to optimization algorithms. In the case of Nesterov's method, the ansatz $x_k := x(t_k)$ is adopted, where $t \mapsto x(t)$ is a smooth curve defined for $t \geq t_0 > 0$ and $t_k = h(k - \frac{\alpha}{2})$, and f is assumed to be twice differentiable (f is $\mathcal{C}^2$). Nesterov's algorithm (3.1) can be rewritten as

$$\frac{x_{k+1} - 2x_k + x_{k-1}}{s} = -\frac{\alpha}{k\sqrt{s}} \frac{x_k - x_{k-1}}{\sqrt{s}} - \nabla f(y_k). \tag{3.2}$$

Taylor expansions give:

$$\frac{x_{k+1} - 2x_k + x_{k-1}}{h^2} = \ddot{x}(t_k) + O(h^2), \tag{3.3}$$

$$\frac{x_k - x_{k-1}}{h} = \dot{x}(t_k) - \frac{h}{2}\ddot{x}(t_k) + O(h^2), \tag{3.4}$$

$$\nabla f(y_k) = \nabla f(x(t_k)) + h\Big(1 - \frac{\alpha}{k}\Big)\nabla^2 f(x(t_k))\dot{x}(t_k) + O(h^2). \tag{3.5}$$

Substituting (3.3)–(3.5) into (3.2) gives the high-resolution ODE with time step $h = \sqrt{s}$ (see [10] for more details):

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \sqrt{s}\,\nabla^2 f(x(t))\dot{x}(t) + \left(1 + \frac{\alpha\sqrt{s}}{2t}\right)\nabla f(x(t)) = 0. \tag{3.6}$$

This equation is obtained by neglecting terms of order $s = h^2$ and retaining only those of order $h = \sqrt{s}$. This dynamics is a particular case of the more general system

$$\text{(DIN-AVD)}: \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta(t)\nabla^2 f(x(t))\dot{x}(t) + b(t)\nabla f(x(t)) = 0,$$

which plays an important role in various fields such as mechanics, shock theory, PDEs, and control. The term (DIN) refers to its interpretation as an inertial version of the continuous regularized Newton method [11]. Hessian-driven damping is important to reduce

Licence CC BY-NC-ND 3.0

oscillations and ensure fast asymptotic convergence of the gradients, an essential feature of Nesterov's accelerated gradient.

The (DIN-AVD) dynamics was introduced in [12] and systematically studied by Attouch, Bahlag, Chbani, and Riahi in [13], Attouch, Chbani, Fadili, and Riahi in [15], and Shi, Du, Jordan, and Su in [14].

In (DIN-AVD), the Hessian-driven damping term appears explicitly. A related system considers an implicit form of Hessian-driven damping, introduced by Alesca, Laszlo, and Pinta [16]. This ODE, called (ISIHD) for *Inertial System with Implicit Hessian Damping*, has the general form:

$$(\text{ISIHD}): \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f\left(x(t) + \beta(t)\dot{x}(t)\right) = 0.$$

The term "implicit" comes from the observation that, as $t \to +\infty$, $\dot{x}(t) \to 0$, allowing the Taylor expansion

$$\nabla f\left(x(t) + \beta(t)\dot{x}(t)\right) \approx \nabla f(x(t)) + \beta(t)\nabla^2 f(x(t))\dot{x}(t),$$

thus implicitly introducing Hessian damping. Similarly to explicit Hessian-driven damping, this implicit form also regularizes oscillations.

This chapter focuses on the dynamics (ISIHD), studying in particular its asymptotic behaviour when the parameter α tends towards infinity.

3.1.1 Main results

After introducing the theoretical framework, a natural question arises: *what is the behavior of the trajectories of the dynamics* (AVD), (DIN-AVD), *and* (ISIHD) *when the damping parameter* α *becomes very large*? In other words, the impact of this parameter on the convergence speed and on the properties of the solutions is analyzed.

To this end, consider $x_0, x_1 \in \mathcal{H}$ and the following Cauchy problem:

$$(\text{AVD}): \quad \begin{cases} \ddot{x}_\alpha(t) + \dfrac{\alpha}{t}\dot{x}_\alpha(t) + \nabla f(x_\alpha(t)) = 0, \\ x_\alpha(t_0) = x_0, \quad \dot{x}_\alpha(t_0) = x_1. \end{cases}$$

By introducing the temporal change of variable $t = \sqrt{2(\alpha+1)s}$ and setting

$$y_\alpha(s) = x_\alpha\Big(\sqrt{2(\alpha+1)s}\Big),$$

Licence CC BY-NC-ND 3.0

it has been shown in [27] that as $\alpha \to +\infty$, the family $(y_\alpha)_\alpha$ converges uniformly (on any bounded interval $[\delta, T]$ with $0 < \delta < T$) to the solution of the *continuous gradient flow*:

$$\text{(GF)} \quad \begin{cases} \dot{y}(s) + \nabla f(y(s)) = 0, \quad s \geq 0, \\ y(0) = x_0. \end{cases}$$

This result highlights a *singular transition* phenomenon: when α becomes large, the second-order inertial dynamics transforms into a first-order gradient flow. This formally explain the transition from a convergence rate of $1/t$ for steepest descent to $1/t^2$ for the Nesterov method, essentially depending on having α large enough. Moreover, if f is convex, [27, Theorem 3] shows that for all $T > 0$, there exists a constant C_T such that

$$\sup_{s\in[0,T]} \|\tilde{y}_\alpha(s) - y(s)\| \leq \frac{C_T}{\sqrt{\alpha+1}}.$$

In [28], this study was extended to systems with explicit Hessian damping:

$$\text{(DIN-AVD)} : \ddot{x}_\alpha(t) + \frac{\alpha}{t}\dot{x}_\alpha(t) + \beta_0\frac{\alpha+1}{t}\nabla^2 f(x_\alpha(t))\dot{x}_\alpha(t) + \nabla f(x_\alpha(t)) = 0,$$

where $\beta_0 > 0$. With the same change of variable $y_\alpha(s) = x_\alpha\Big(\sqrt{2(\alpha+1)s}\Big)$, it was shown, in the finite-dimensional setting and for convex f, that as $\alpha \to +\infty$, $(y_\alpha)_\alpha$ converges uniformly on compact subsets of $[0, +\infty)$ to the solution of

$$\dot{y}(s) + \beta_0\nabla^2 f(y(s))\dot{y}(s) + \nabla f(y(s)) = 0, \quad y(0) = x_0, \tag{3.7}$$

i.e., the continuous regularized Newton–Levenberg–Marquardt method.

This chapter extends the analysis to the *inertial system with implicit Hessian damping*. More precisely, the (ISIHD) system considered is defined by

$$\ddot{x}_\alpha(t) + \frac{\alpha}{t}\dot{x}_\alpha(t) + \nabla f\left(x_\alpha(t) + \beta_0\frac{\alpha+1}{t}\dot{x}_\alpha(t)\right) = 0,$$

with $x_\alpha(t_0) = x_0$ and $\dot{x}_\alpha(t_0) = x_1$. One of the main results of this chapter establishes that, under the same change of variables $y_\alpha(s) = x_\alpha\Big(\sqrt{2(\alpha+1)s}\Big)$, one obtains:

(i) y_α satisfies the equation

$$\frac{2s}{\alpha+1}\ddot{y}_\alpha(s) + \dot{y}_\alpha(s) + \nabla f(y_\alpha(s) + \beta_0\dot{y}_\alpha(s)) = 0;$$

(ii) If f is convex (or non-convex quadratic) and $x_1 = 0$, then as $\alpha \to +\infty$, $(y_\alpha)_\alpha$ converges

Licence CC BY-NC-ND 3.0

uniformly on compact subsets of $[0, +\infty)$ to y, the solution of

$$\dot{y}(s) + \nabla f(y(s) + \beta_0 \dot{y}(s)) = 0, \quad y(0) = x_0.$$

The analysis assumes $x_1 = 0$, since a nonzero initial velocity strongly complicates the control of energy. The implicit limiting dynamics obtained here is moreover very different from the explicit case (3.7), illustrating the richness of the (ISIHD) model.

3.1.2 Chapter Contents

This chapter addresses the asymptotic analysis of inertial systems with vanishing viscosity.
Section 3.2 provides a first direct approach to study the limit of the trajectories of the (ISIHD) system when the parameter α tends to infinity, showing uniform convergence towards the initial position.
Section 3.3 introduces a time-scaling method applied to the (ISIHD) system, allowing the dynamics to be reformulated for asymptotic analysis.
Section 3.4 examines the singular limit of the reparametrized system in the convex case. The reparametrized trajectories converge to a first-order proximal flow, which reduces to the classical gradient flow when $\beta_0 = 0$.
Section 3.5 extends this study to the non-convex case under the quadratic assumption. Using uniform estimates and the Ascoli–Arzelà theorem, convergence towards the proximal flow is established.
Section 3.6 connects the continuous dynamics to its discrete version by the I-NAG algorithm, interpreted as a discretization of the gradient flow with implicit damping.
Finally, Section 3.8 presents numerical illustrations confirming the theoretical results and comparing the second-order inertial dynamics with its first-order limit.

3.2 Direct Approach

This section focuses on the limit of the trajectories of the (ISIHD) system as the parameter α tends to infinity. This approach, called the *direct approach*, consists in analyzing the asymptotic behavior of the trajectory without resorting to time rescaling or complex reparametrizations. Before introducing the main result, recall that f is assumed to be convex and differentiable, with a gradient that is Lipschitz on bounded sets. In addition, a positive function $\beta(t)$, independent of α, is considered to model the implicit Hessian damping.
The following theorem quantitatively describes the limit of the trajectories as $\alpha \to +\infty$.

Licence CC BY-NC-ND 3.0

Theorem 3.2.1. Let f be a convex differentiable function with a gradient that is Lipschitz on bounded sets. Consider the differential equation (ISIHD):

$$\ddot{x}_\alpha(t) + \frac{\alpha}{t}\dot{x}_\alpha(t) + \nabla f\Big(x_\alpha(t) + \beta(t)\dot{x}_\alpha(t)\Big) = 0, \quad t \geq t_0 > 0, \tag{3.8}$$

with initial conditions $x_\alpha(t_0) = x_0$, $\dot{x}_\alpha(t_0) = x_1$, and $\beta : [t_0, +\infty) \to \mathbb{R}^+$ a positive function independent of α. Then:

(i) For each $t \geq t_0$, $x_\alpha(t) \to x_0$ strongly in $\mathcal{H}$ as $\alpha \to +\infty$.

(ii) The convergence is uniform on bounded subsets of $[t_0, +\infty)$. More precisely, for any $T > t_0$,

$$\sup_{t\in[t_0,T]} \|x_\alpha(t) - x_0\| \leq \frac{M_T}{\alpha - 1},$$

with

$$M_T = t_0\|x_1\| + T^2\Big(\|\nabla f(x_0)\| + L_r(T + \beta_T)\sqrt{2(f(x_0) - \inf(f)) + \|x_1\|^2}\Big),$$

where $\beta_T := \sup\{\beta(t),\ t \in [t_0, T]\}$ and L_r is the Lipschitz constant of ∇f on the ball centered at the origin in $\mathcal{H}$ of radius

$$r = \|x_0\| + (T + \beta_T)\sqrt{2(f(x_0) - \inf(f)) + \|x_1\|^2}.$$

Proof. As an ingredient, the fact that the total energy is decreasing along the solution trajectories of (ISIHD) is used.

Step 1 (Energy analysis): Let $x_\alpha : [t_0, +\infty) \to \mathcal{H}$ be the solution trajectory of (ISIHD) satisfying the initial conditions $x_\alpha(t_0) = x_0$ and $\dot{x}_\alpha(t_0) = x_1$. Consider the energy function

$$\mathcal{E}_\alpha(t) = f(x_\alpha(t)) - \inf(f) + \frac{1}{2}\|\dot{x}_\alpha(t)\|^2.$$

By the chain rule, its time derivative is

$$\dot{\mathcal{E}}_\alpha(t) = \langle \nabla f(x_\alpha(t)), \dot{x}_\alpha(t)\rangle + \langle \ddot{x}_\alpha(t), \dot{x}_\alpha(t)\rangle.$$

Substituting the differential equation (3.8) yields:

$$\dot{\mathcal{E}}_\alpha(t) = -\frac{\alpha}{t}\|\dot{x}_\alpha(t)\|^2 + \langle \nabla f(x_\alpha(t)) - \nabla f\left(x_\alpha(t) + \beta(t)\dot{x}_\alpha(t)\right), \dot{x}_\alpha(t)\rangle.$$

The first term is non-positive since $\frac{\alpha}{t} > 0$. Concerning the second term, the convexity of f implies the monotonicity of its gradient, hence

$$\langle \nabla f(x_\alpha(t)) - \nabla f(x_\alpha(t) + \beta(t)\dot{x}_\alpha(t)), \dot{x}_\alpha(t)\rangle \leq 0.$$

Licence CC BY-NC-ND 3.0

Thus, $\dot{\mathcal{E}}_\alpha(t) \leq 0$, which implies the energy decay: $\mathcal{E}_\alpha(t) \leq \mathcal{E}_\alpha(t_0)$ for all $t \geq t_0$.

Step 2 (Preliminary uniform bounds): From the energy inequality, it follows that::

$$\|\dot{x}_\alpha(t)\| \leq \sqrt{2(f(x_0) - \inf(f)) + \|x_1\|^2} =: c_0.$$

By integration, one obtains:

$$\|x_\alpha(t) - x_0\| \leq \int_{t_0}^{t} \|\dot{x}_\alpha(s)\|\, ds \leq c_0(t - t_0).$$

Consequently, for all $t \in [t_0, T]$, $x_\alpha(t)$ remains in the closed ball centered at the origin with radius $r_0 = \|x_0\| + Tc_0$.

Step 3 (Control of the inertial term): Set $\beta_T = \sup\{\beta(t),\ t \in [t_0, T]\}$. Then, for $t \in [t_0, T]$, one has:

$$\|\beta(t)\dot{x}_\alpha(t)\| \leq \beta_T c_0 =: r_1.$$

Therefore, $x_\alpha(t) + \beta(t)\dot{x}_\alpha(t)$ belongs to the ball centered at the origin of radius $r = r_0 + r_1$.

Step 4 (Lipschitz estimate): Since ∇f is Lipschitz on bounded sets, there exists a constant $L_r > 0$ such that for all $t \in [t_0, T]$ and $\alpha \geq 3$:

$$\|\nabla f(x_\alpha(t) + \beta(t)\dot{x}_\alpha(t)) - \nabla f(x_0)\| \leq L_r \|x_\alpha(t) + \beta(t)\dot{x}_\alpha(t) - x_0\|. \tag{3.9}$$

This yields the uniform bound:

$$\sup_{\alpha \geq 3,\, t \in [t_0, T]} \|\nabla f(x_\alpha(t) + \beta(t)\dot{x}_\alpha(t))\| \leq \|\nabla f(x_0)\| + L_r(T + \beta_T)c_0 =: c_T. \tag{3.10}$$

Step 5 (Rewriting the equation and integration): Equation (3.8) can be rewritten in integral form. First, observe that:

$$\begin{aligned} \frac{d}{dt}(t^\alpha \dot{x}_\alpha(t)) &= \alpha t^{\alpha-1}\dot{x}_\alpha(t) + t^\alpha \ddot{x}_\alpha(t) \\ &= -t^\alpha \nabla f(x_\alpha(t) + \beta(t)\dot{x}_\alpha(t)) \quad \text{(by (3.8))}. \end{aligned}$$

Integrating from t_0 to t yields:

$$t^\alpha \dot{x}_\alpha(t) - t_0^\alpha x_1 = -\int_{t_0}^{t} s^\alpha \nabla f(x_\alpha(s) + \beta(s)\dot{x}_\alpha(s))\, ds,$$

hence

$$\dot{x}_\alpha(t) = -\frac{1}{t^\alpha}\int_{t_0}^{t} s^\alpha \nabla f(x_\alpha(s) + \beta(s)\dot{x}_\alpha(s))\, ds + \frac{t_0^\alpha}{t^\alpha} x_1.$$

Licence CC BY-NC-ND 3.0

Step 6 (Velocity estimate): Taking the norm and using (3.10) gives:

$$\begin{aligned}\|\dot{x}_\alpha(t)\| &\leq \frac{1}{t^\alpha}\int_{t_0}^{t} s^\alpha \|\nabla f(x_\alpha(s)+\beta(s)\dot{x}_\alpha(s))\|\, ds + \frac{t_0^\alpha}{t^\alpha}\|x_1\| \\ &\leq c_T \frac{1}{t^\alpha}\int_{t_0}^{t} s^\alpha\, ds + \frac{t_0^\alpha}{t^\alpha}\|x_1\| \\ &= c_T \frac{1}{(\alpha+1)t^\alpha}(t^{\alpha+1}-t_0^{\alpha+1}) + \frac{t_0^\alpha}{t^\alpha}\|x_1\| \\ &\leq c_T \frac{T}{\alpha+1} + \frac{t_0^\alpha}{t^\alpha}\|x_1\|.\end{aligned}$$

Step 7 (Position estimate): Integrating again and using the previous estimate:

$$\begin{aligned}\|x_\alpha(t)-x_0\| &\leq \int_{t_0}^{t}\|\dot{x}_\alpha(s)\|\, ds \\ &\leq c_T\frac{T^2}{\alpha+1} + t_0^\alpha\|x_1\|\int_{t_0}^{t}\frac{1}{s^\alpha}\, ds \\ &= c_T\frac{T^2}{\alpha+1} + t_0^\alpha\|x_1\|\cdot\frac{1}{1-\alpha}(t^{1-\alpha}-t_0^{1-\alpha}) \\ &\leq c_T\frac{T^2}{\alpha+1} + t_0\|x_1\|\frac{1}{\alpha-1} \\ &\leq \frac{1}{\alpha-1}\left(c_T T^2 + t_0\|x_1\|\right).\end{aligned}$$

Thus,

$$\sup_{t\in[t_0,T]}\|x_\alpha(t)-x_0\| \leq \frac{M_T}{\alpha-1},$$

with $M_T = t_0\|x_1\| + T^2(\|\nabla f(x_0)\| + L_r(T+\beta_T)c_0)$. This proves point (ii).

Step 8 (Conclusion): Point (i) follows immediately from (ii), since for any fixed $t \geq t_0$, taking $T = t$ gives:

$$\|x_\alpha(t)-x_0\| \leq \frac{M_t}{\alpha-1} \to 0 \quad \text{as } \alpha \to +\infty.$$

□

This direct approach does not provide much information about the behavior of the trajectories of (ISIHD) for large values of α, suggesting a singular perturbation problem. To further investigate this issue, a time rescaling technique, as described in Section 3.3 below, is employed.

Licence CC BY-NC-ND 3.0

3.3 Time-Scaling Approach

Consider the inertial dynamic system with implicit Hessian damping (ISIHD):

$$\text{(ISIHD)}: \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f\Big(x(t) + \beta(t)\dot{x}(t)\Big) = 0, \tag{3.11}$$

where $\beta(t)$ is a positive function. To analyze the limit as α becomes large, a time rescaling $t = \tau(s)$ is introduced, inspired by the approach used for systems with explicit Hessian damping in [28]. It is assumed that $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ is increasing, twice differentiable, and satisfies $\lim_{s\to+\infty} \tau(s) = +\infty$.
Set

$$y(s) := x(\tau(s)).$$

Applying the chain rule yields:

$$\dot{y}(s) = \dot{\tau}(s)\dot{x}(\tau(s)), \tag{3.12}$$

$$\ddot{y}(s) = \ddot{\tau}(s)\dot{x}(\tau(s)) + \dot{\tau}(s)^2\ddot{x}(\tau(s)). \tag{3.13}$$

Substituting $t = \tau(s)$ into (3.11):

$$\ddot{x}(\tau(s)) + \frac{\alpha}{\tau(s)}\dot{x}(\tau(s)) + \nabla f\Big(x(\tau(s)) + \beta(\tau(s))\dot{x}(\tau(s))\Big) = 0. \tag{3.14}$$

Combining (3.12), (3.13) and (3.14) yields the reparametrized dynamics:

$$\ddot{y}(s) + \left(\alpha\frac{\dot{\tau}(s)}{\tau(s)} - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)}\right)\dot{y}(s) + \dot{\tau}(s)^2\nabla f\left(y(s) + \frac{\beta(\tau(s))}{\dot{\tau}(s)}\dot{y}(s)\right) = 0. \tag{3.15}$$

The reparametrized viscous damping coefficient is defined as:

$$\tilde{\gamma}(s) := \alpha\frac{\dot{\tau}(s)}{\tau(s)} - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)} = \frac{\alpha\dot{\tau}(s)^2 - \tau(s)\ddot{\tau}(s)}{\tau(s)\dot{\tau}(s)}. \tag{3.16}$$

To ensure that the damping is asymptotically vanishing and compatible with Su-Boyd-Candès type accelerated dynamics, the following condition is imposed:

$$\tilde{\gamma}(s) = \frac{C(\alpha)}{s}, \tag{3.17}$$

where $C(\alpha) > 0$ is a constant. This condition leads to the following differential equation for $\tau(s)$:

$$\frac{\alpha\dot{\tau}(s)^2 - \tau(s)\ddot{\tau}(s)}{\tau(s)\dot{\tau}(s)} = \frac{C(\alpha)}{s}. \tag{3.18}$$

Licence CC BY-NC-ND 3.0

The following lemma characterizes the general solutions of this ODE.

Lemma 3.3.1. The solutions of the equation

$$-\tau(s)\ddot{\tau}(s) + \alpha\dot{\tau}(s)^2 = \frac{C(\alpha)}{s}\,\tau(s)\dot{\tau}(s) \tag{3.19}$$

are given by

$$\tau(s) = \left(\frac{-C(\alpha)+1}{(\alpha-1)\left(C(\alpha)K_2 s^{C(\alpha)} - K_1 s - K_2 s^{C(\alpha)}\right)}\right)^{\frac{1}{\alpha-1}} s^{\frac{C(\alpha)}{\alpha-1}}, \tag{3.20}$$

where K_1 and K_2 are arbitrary constants, $\alpha \neq 1$, and $C(\alpha)$ is a positive constant.

Proof. Step 1: Definition of quantities. Set

$$A = \left(\frac{1-C(\alpha)}{\alpha-1}\right)^{\frac{1}{\alpha-1}}, \quad D(s) = (C(\alpha)-1)K_2 s^{C(\alpha)} - K_1 s. \tag{3.21}$$

Then, the function $\tau(s)$ is defined by:

$$\tau(s) = A\, D(s)^{-\frac{1}{\alpha-1}} s^{\frac{C(\alpha)}{\alpha-1}}. \tag{3.22}$$

Step 2: First derivative computation. Differentiating $\tau(s)$ with respect to s and applying the product and chain rules gives:

$$\begin{aligned}\dot{\tau}(s) &= -A\frac{1}{\alpha-1}\dot{D}(s)D(s)^{-\frac{1}{\alpha-1}-1}s^{\frac{C(\alpha)}{\alpha-1}} + A\frac{C(\alpha)}{\alpha-1}D(s)^{-\frac{1}{\alpha-1}}s^{\frac{C(\alpha)}{\alpha-1}-1} \\ &= \tau(s)\left[-\frac{1}{\alpha-1}\frac{\dot{D}(s)}{D(s)} + \frac{C(\alpha)}{\alpha-1}\frac{1}{s}\right],\end{aligned} \tag{3.23}$$

where

$$\dot{D}(s) = C(\alpha)(C(\alpha)-1)K_2 s^{C(\alpha)-1} - K_1.$$

Step 3: Second derivative computation. Differentiating (3.23) again yields:

$$\ddot{\tau}(s) = \dot{\tau}(s)\left[-\frac{\dot{D}(s)}{(\alpha-1)D(s)} + \frac{C(\alpha)}{(\alpha-1)s}\right] + \tau(s)\frac{d}{ds}\left[-\frac{\dot{D}(s)}{(\alpha-1)D(s)} + \frac{C(\alpha)}{(\alpha-1)s}\right].$$

The second term expands as:

$$\frac{d}{ds}\left[-\frac{\dot{D}(s)}{(\alpha-1)D(s)} + \frac{C(\alpha)}{(\alpha-1)s}\right] = -\frac{\ddot{D}(s)D(s) - \dot{D}^2(s)}{(\alpha-1)D^2(s)} - \frac{C(\alpha)}{(\alpha-1)s^2},$$

Licence CC BY-NC-ND 3.0

where

$$\ddot{D}(s) = C(\alpha)(C(\alpha)-1)^2 K_2 s^{C(\alpha)-2}.$$

Thus,

$$\ddot{\tau}(s) = \dot{\tau}(s)\left[-\frac{\dot{D}(s)}{(\alpha-1)D(s)} + \frac{C(\alpha)}{(\alpha-1)s}\right] + \tau(s)\left[-\frac{\ddot{D}(s)D(s)-\dot{D}^2(s)}{(\alpha-1)D^2(s)} - \frac{C(\alpha)}{(\alpha-1)s^2}\right].$$

Step 4: Definition of the expression $T(s)$. The goal is to verify that $\tau(s)$ satisfies the differential equation (3.19):

$$T(s) = -\tau(s)\ddot{\tau}(s) + \alpha\dot{\tau}^2(s) - \frac{C(\alpha)}{s}\tau(s)\dot{\tau}(s).$$

Substitute the expressions of $\dot{\tau}(s)$ and $\ddot{\tau}(s)$ into $T(s)$.

Term 1 ($-\tau(s)\ddot{\tau}(s)$): From the expression of $\ddot{\tau}(s)$, it follows that:

$$\begin{aligned}
-\tau(s)\ddot{\tau}(s) &= -\tau(s)\Bigg[\dot{\tau}(s)\left(-\frac{\dot{D}(s)}{(\alpha-1)D(s)} + \frac{C(\alpha)}{(\alpha-1)s}\right) \\
&\quad + \tau(s)\left(-\frac{\ddot{D}(s)D(s)-\dot{D}^2(s)}{(\alpha-1)D^2(s)} - \frac{C(\alpha)}{(\alpha-1)s^2}\right)\Bigg] \\
&= -\tau^2(s)\left(-\frac{\dot{D}(s)}{(\alpha-1)D(s)} + \frac{C(\alpha)}{(\alpha-1)s}\right)^2 \\
&\quad - \tau^2(s)\left(-\frac{\ddot{D}(s)D(s)-\dot{D}^2(s)}{(\alpha-1)D^2(s)} - \frac{C(\alpha)}{(\alpha-1)s^2}\right).
\end{aligned} \tag{3.24}$$

Term 2 ($\alpha\dot{\tau}^2(s)$): From (3.23):

$$\alpha\dot{\tau}^2(s) = \alpha\tau^2(s)\left(-\frac{\dot{D}(s)}{(\alpha-1)D(s)} + \frac{C(\alpha)}{(\alpha-1)s}\right)^2. \tag{3.25}$$

Term 3 ($-\frac{C(\alpha)}{s}\tau(s)\dot{\tau}(s)$): Again from (3.23):

$$-\frac{C(\alpha)}{s}\tau(s)\dot{\tau}(s) = -\frac{C(\alpha)}{s}\tau^2(s)\left(-\frac{\dot{D}(s)}{(\alpha-1)D(s)} + \frac{C(\alpha)}{(\alpha-1)s}\right). \tag{3.26}$$

Licence CC BY-NC-ND 3.0

Step 5: Combination of terms. Adding (3.24), (3.25) and (3.26) yields:

$$T(s) = \tau^2(s)\left[(\alpha-1)\left(-\frac{\dot{D}(s)}{(\alpha-1)D(s)}+\frac{C(\alpha)}{(\alpha-1)s}\right)^2 + \frac{\ddot{D}(s)D(s)-\dot{D}^2(s)}{(\alpha-1)D^2(s)}+\frac{C(\alpha)}{(\alpha-1)s^2}-\frac{C(\alpha)}{s}\left(-\frac{\dot{D}(s)}{(\alpha-1)D(s)}+\frac{C(\alpha)}{(\alpha-1)s}\right)\right].$$

Expanding and grouping terms leads to massive simplifications:

$$\begin{aligned} T(s) &= \tau^2(s)\left[\left(\frac{\dot{D}^2(s)}{(\alpha-1)D^2(s)}-\frac{\dot{D}^2(s)}{(\alpha-1)D^2(s)}\right)+\left(-\frac{2C(\alpha)\dot{D}(s)}{(\alpha-1)sD(s)}+\frac{C(\alpha)\dot{D}(s)}{(\alpha-1)sD(s)}\right)\right. \\ &\quad \left.+\left(\frac{C(\alpha)^2}{(\alpha-1)s^2}-\frac{C(\alpha)^2}{(\alpha-1)s^2}\right)+\left(\frac{\ddot{D}(s)}{(\alpha-1)D(s)}+\frac{C(\alpha)}{(\alpha-1)s^2}\right)\right] \\ &= \tau^2(s)\left[-\frac{C(\alpha)\dot{D}(s)}{(\alpha-1)sD(s)}+\frac{\ddot{D}(s)}{(\alpha-1)D(s)}+\frac{C(\alpha)}{(\alpha-1)s^2}\right] \\ &= \frac{\tau^2(s)}{(\alpha-1)s^2D(s)}(-C(\alpha)s\dot{D}(s)+s^2\ddot{D}(s)+C(\alpha)D(s)) \end{aligned}$$

Step 6: Verification of vanishing. Replacing $D(s)$, $\dot{D}(s)$ and $\ddot{D}(s)$ by their expressions:

$$\begin{aligned} -C(\alpha)\,\dot{D}(s)\,s+\ddot{D}(s)\,s^2+C(\alpha)\,D(s) &= -C(\alpha)\Big(C(\alpha)(C(\alpha)-1)K_2s^{C(\alpha)-1}-K_1\Big)s \\ &\quad + C(\alpha)(C(\alpha)-1)^2K_2s^{C(\alpha)} \\ &\quad + C(\alpha)\Big((C(\alpha)-1)K_2s^{C(\alpha)}-K_1s\Big) \\ &= \Big[-C(\alpha)^2(C(\alpha)-1)+C(\alpha)(C(\alpha)-1)^2 \\ &\quad + C(\alpha)(C(\alpha)-1)\Big]K_2s^{C(\alpha)}. \\ &= 0 \end{aligned}$$

Thus, $T(s)=0$ for all s.

Conclusion. The function $\tau(s)$ defined by (3.22) indeed satisfies the differential equation

$$-\tau(s)\ddot{\tau}(s)+\alpha\dot{\tau}^2(s)-\frac{C(\alpha)}{s}\tau(s)\dot{\tau}(s)=0,$$

completing the proof. □

Remark 3.3.1. This lemma provides a description of the time scale functions able to generate the desired damping behaviour, which will facilitate the analysis of the convergence properties of

Licence CC BY-NC-ND 3.0

the system and its link with discrete optimisation algorithms. The general solution, depending on two arbitrary constants K_1 and K_2, encompasses a wide range of possible behaviors. By choosing these constants, one can isolate simpler particular solutions, for example, power-law parameterizations $\tau(s) = \lambda s^p$. This type of solution is particularly interesting due to its mathematical simplicity and its direct connection to known optimization schemes.

corollary 3.3.2 (Monomial solution)**.** A solution of the ODE (3.19) can be expressed in monomial form $\tau(s) = \lambda s^p$ (with $\lambda > 0$ and $p \in]0, 1]$) if and only if the constants K_1 and K_2 in the general solution (3.20) are chosen as follows:

1. $K_2 = 0$,
2. $K_1 = \lambda^{-(\alpha-1)} \left(\dfrac{C(\alpha) - 1}{\alpha - 1} \right)$.

In this case, the exponent p is related to the constant $C(\alpha)$ by:

$$p = \frac{C(\alpha) - 1}{\alpha - 1}, \tag{3.27}$$

and conversely, $C(\alpha) = 1 + p(\alpha - 1)$.

Proof. From the previous lemma,

$$\tau(s) = A \left[D(s) \right]^{-\frac{1}{\alpha-1}} s^{\frac{C(\alpha)}{\alpha-1}},$$

where

$$A = \left(\frac{1 - C(\alpha)}{\alpha - 1} \right)^{\frac{1}{\alpha-1}}, \quad D(s) = (C(\alpha) - 1) K_2 s^{C(\alpha)} - K_1 s,$$

with $\alpha \geq 3$. The necessary and sufficient condition for $\tau(s)$ to be a monomial is that $D(s)$ is also a monomial. Let us examine the two terms of $D(s)$:

Necessary condition: For $D(s) = (C(\alpha) - 1) K_2 s^{C(\alpha)} - K_1 s$ to be a monomial, the two terms must have the same exponent or one of the coefficients must vanish. The case $C(\alpha) = 1$ is excluded (Since $\tau(s)$ would always be equal to zero and therefore does not tend to $+\infty$ as $s \to +\infty$, which makes it not a suitable choice.), so the only possibility is to cancel one of the terms:

- If $K_1 = 0$, then $D(s) = (C(\alpha) - 1) K_2 s^{C(\alpha)}$ and $\tau(s) = A[(C(\alpha) - 1) K_2]^{-\frac{1}{\alpha-1}}$, which is constant ($p = 0$).
- If $K_2 = 0$, then $D(s) = -K_1 s$ and one obtains the desired form.

Thus, the condition $K_2 = 0$ is necessary to have $p > 0$.

Licence CC BY-NC-ND 3.0

Computation with $K_2 = 0$: It holds that $D(s) = -K_1 s$. For $D(s) > 0$ (necessary for the definition of $\tau(s)$), it is required that $K_1 < 0$. Then:

$$\begin{aligned}\tau(s) &= A(-K_1 s)^{-\frac{1}{\alpha-1}} s^{\frac{C(\alpha)}{\alpha-1}} \\ &= A(-K_1)^{-\frac{1}{\alpha-1}} s^{\frac{C(\alpha)-1}{\alpha-1}}.\end{aligned}$$

The following identifications are made:

$$p = \frac{C(\alpha)-1}{\alpha-1}, \quad \lambda = A(-K_1)^{-\frac{1}{\alpha-1}}.$$

Constraints on p: To ensure $0 < p \leq 1$, it is required that:

$$0 < \frac{C(\alpha)-1}{\alpha-1} \leq 1 \quad \Leftrightarrow \quad 1 < C(\alpha) \leq \alpha.$$

Expression for K_1: From the expression of λ, it follows that:

$$\begin{aligned}\lambda = \left(\frac{1-C(\alpha)}{\alpha-1}\right)^{\frac{1}{\alpha-1}} (-K_1)^{-\frac{1}{\alpha-1}} &\Leftrightarrow (-K_1)^{-\frac{1}{\alpha-1}} = \lambda \left(\frac{\alpha-1}{1-C(\alpha)}\right)^{\frac{1}{\alpha-1}} \\ &\Leftrightarrow -K_1 = \lambda^{-(\alpha-1)} \left(\frac{1-C(\alpha)}{\alpha-1}\right) \\ &\Leftrightarrow K_1 = \lambda^{-(\alpha-1)} \left(\frac{C(\alpha)-1}{\alpha-1}\right).\end{aligned}$$

This completes the proof. □

Remark 3.3.2. Substituting $\tau(s) = \lambda s^p$ into the reparametrized system (3.15) yields:

$$\ddot{y}(s) + \frac{\alpha_p}{s}\dot{y}(s) + \lambda^2 p^2 s^{2(p-1)} \nabla f\left(y(s) + \frac{\beta(\lambda s^p)}{\lambda p s^{p-1}}\dot{y}(s)\right) = 0, \tag{3.28}$$

where

$$\alpha_p = 1 + (\alpha - 1)p.$$

This equation shows how a power-law reparametrization affects both the damping and the gradient of the system. The damping describes how the velocity decays over time, while the time rescaling modifies the argument inside the gradient. The time-dependent factor $\lambda^2 p^2 s^{2(p-1)}$ acts as a dynamic preconditioner, able to accelerate or slow down convergence depending on the value of p. Table 3.1 illustrates the effect of the parameter p on both damping and gradient scaling. This reparametrization has several advantages. It preserves the asymptotic vanishing damping behavior of the original system. Furthermore, the parameter p provides an

Licence CC BY-NC-ND 3.0

additional degree of freedom to adjust the dissipation of inertial energy as needed for analysis. In addition, it also allows unifies several well-known optimization dynamics. Indeed, the case $p = 1$ corresponds to classical non-accelerated gradient descent. For $p = \frac{1}{2}$, one recovers the inertial behavior of Nesterov's accelerated scheme. The choice $p = \frac{2}{3}$ corresponds to the regime providing an optimal convergence rate for functions with Lipschitz Hessian, while $p = \frac{1}{3}$ is associated with some higher-order acceleration methods. Thus, the power-law $\tau(s) = \lambda s^p$ establishes a natural connection between second-order inertial models and first-order gradient schemes.

Finally, the simplicity of this monomial form facilitates the analytical study of asymptotic properties and the comparison between the continuous dynamics and their discrete counterparts. As shown in Corollary 3.3.2, this particular solution is obtained from the general solution of the Lemma by setting $K_2 = 0$. This choice eliminates possible oscillations and ensures a pure power-law behavior, while the constant K_1 is determined by the scaling parameter λ. Therefore, although it is a particular case, the family of monomial solutions occupies a key position in the analysis, clarifying the link between damping, reparametrization, and inertial acceleration.

Table 3.1: *Effect of the power law exponent p on the damping coefficient and gradient scaling (for $\alpha = 3$ and $\lambda = 1$).*

Exponent p	$\alpha_p = 1 + (\alpha - 1)p$	damping $\tilde{\gamma}(s) = \alpha_p/s$	Scaling $\lambda^2 p^2 s^{2(p-1)}$
1	3.00	$3/s$	constant $(= 1)$
2/3	7/3	$(7/3)s^{-1}$	$(4/9)s^{-2/3}$
1/2	2.00	$2/s$	$(1/4)s^{-1}$
1/3	5/3	$(5/3)s^{-1}$	$(1/9)s^{-4/3}$

Remark 3.3.3 (*Choice of p*)**.** The choice $p = \frac{1}{2}$ for the power law $\tau(s) = \lambda s^{1/2}$ is of particular interest, both theoretically and practically. Indeed, for $\alpha = 3$, one obtains $\alpha_p = 2$, which corresponds to a damping $\tilde{\gamma}(s) = \frac{2}{s}$ in the s variable. Returning to the original time $t = \lambda s^{1/2}$, one then recovers the damping $\frac{3}{t}\dot{x}(t)$, typical of Nesterov-type accelerated systems. This behavior ensures a convergence rate of $\mathcal{O}(1/t^2)$, characteristic of these optimal dynamics. Thus, the case $p = 1/2$ establishes a direct link between the continuous formulation studied here and Nesterov's method. It serves as a reference point for precisely comparing the model with the well-known dynamics (AVD) and $(\mathrm{DIN}-\mathrm{AVD})$, and allows a clearer examination of the effect of the time rescaling.

By setting $p = 1/2$ in equation (3.28), the system becomes:

$$\ddot{y}(s) + \frac{\alpha+1}{2s}\dot{y}(s) + \frac{\lambda^2}{4s}\nabla f\Big(y(s) + \frac{2\sqrt{s}\,\beta(\lambda\sqrt{s})}{\lambda}\dot{y}(s)\Big) = 0. \tag{3.29}$$

Licence CC BY-NC-ND 3.0

The particular choice $\lambda = \sqrt{2(\alpha+1)}$ balances the coefficients of the first and second-order differential terms, ensuring $\frac{\lambda^2}{4} = \frac{\alpha+1}{2}$. The equation then takes the form:

$$\ddot{y}(s) + \frac{\alpha+1}{2s}\dot{y}(s) + \frac{\alpha+1}{2s}\nabla f\left(y(s) + \sqrt{\frac{2s}{\alpha+1}}\,\beta(\sqrt{2(\alpha+1)}s)\dot{y}(s)\right) = 0. \tag{3.30}$$

To obtain a simpler autonomous system for analysis, it is imposed that the coefficient of the velocity inside the gradient argument be constant. A function $\beta(t)$ is then sought such that:

$$\sqrt{\frac{2s}{\alpha+1}}\beta(\sqrt{2(\alpha+1)}s) = \beta_0, \tag{3.31}$$

where β_0 is a positive constant. The solution to this relation is:

$$\beta(t) = \beta_0\frac{\alpha+1}{t}. \tag{3.32}$$

Substituting this choice of β into (3.30) and multiplying the entire equation by the integrating factor $\frac{2s}{\alpha+1}$ to eliminate the singularity, this yields:

$$\frac{2s}{\alpha+1}\ddot{y}(s) + \dot{y}(s) + \nabla f(y(s) + \beta_0\dot{y}(s)) = 0. \tag{3.33}$$

This equation highlights a second-order system with constant damping, where the gradient is evaluated at a fixed offset point β_0. This structure simplifies both theoretical analysis and numerical discretization.

3.4 Limit dynamics of (ISIHD): Convex case

The previous analysis highlighted the importance of the two parameters of the system: the damping coefficient α and the scaling exponent p. The behavior of the system is now studied as α becomes very large, i.e., as $\alpha \to +\infty$. In this limit, the second-order inertial term $\frac{2s}{\alpha+1}\ddot{y}(s)$ vanishes in the rescaled equation (3.33). A simplified first-order dynamics is then obtained:

$$\dot{y}(s) + \nabla f\left(y(s) + \beta_0\dot{y}(s)\right) = 0. \tag{3.34}$$

When the function f is convex, the existence and uniqueness of a global solution for this differential equation are guaranteed by the Cauchy–Lipschitz theorem. Indeed, using subdifferential

Licence CC BY-NC-ND 3.0

duality (Proposition 2.3.2), equation (3.34) can be reformulated in a so-called *proximal* form:

$$\begin{aligned}\nabla f\left(y(s)+\beta_0\dot{y}(s)\right)=-\dot{y}(s) &\iff y(s)+\beta_0\dot{y}(s)\in\partial f^*(-\dot{y}(s))\\ &\iff \frac{1}{\beta_0}y(s)\in\left(\mathrm{Id}+\frac{1}{\beta_0}\partial f^*\right)(-\dot{y}(s))\\ &\iff -\dot{y}(s)=\operatorname{prox}_{\frac{1}{\beta_0}f^*}\left(\frac{1}{\beta_0}y(s)\right)\\ &\iff \dot{y}(s)=-\frac{1}{\beta_0}y(s)+\frac{1}{\beta_0}\operatorname{prox}_{\beta_0 f}(y(s)).\end{aligned}\tag{3.35}$$

The resulting equation (3.35) is an ordinary differential equation with a Lipschitz vector field, since the proximal operator itself is Lipschitz continuous (Proposition 2.3.4). Hence, the Cauchy–Lipschitz theorem applies directly, ensuring well-posedness of the problem.

Theorem 3.4.1. Let $f:\mathcal{H}\to\mathbb{R}$ be a convex differentiable function with a locally Lipschitz gradient, and let $\beta_0>0$ and $t_0>0$. Take arbitrary initial data $x_0,x_1\in\mathcal{H}$. For each $\alpha\geq 3$, consider the trajectory $x_\alpha:[t_0,+\infty)\to\mathcal{H}$ defined as the solution of the second-order differential equation

$$\ddot{x}_\alpha(t)+\frac{\alpha}{t}\dot{x}_\alpha(t)+\nabla f\left(x_\alpha(t)+\beta_0\frac{\alpha+1}{t}\dot{x}_\alpha(t)\right)=0,\tag{3.36}$$

with initial conditions $x_\alpha(t_0)=x_0$ and $\dot{x}_\alpha(t_0)=x_1$.

Define the rescaled trajectories in time $y_\alpha:[s_{0,\alpha},+\infty)\to\mathcal{H}$ by

$$y_\alpha(s):=x_\alpha\left(\sqrt{2(\alpha+1)s}\right),\quad \text{where } s_{0,\alpha}:=\frac{t_0^2}{2(\alpha+1)}.\tag{3.37}$$

(i). For each $\alpha\geq 3$, the function y_α satisfies the second-order differential equation

$$\frac{2s}{\alpha+1}\ddot{y}_\alpha(s)+\dot{y}_\alpha(s)+\nabla f\left(y_\alpha(s)+\beta_0\dot{y}_\alpha(s)\right)=0,\quad \forall s\geq s_{0,\alpha},\tag{3.38}$$

with initial conditions

$$y_\alpha(s_{0,\alpha})=x_0,\quad \dot{y}_\alpha(s_{0,\alpha})=\frac{\alpha+1}{t_0}x_1.\tag{3.39}$$

(ii). Now assume $x_1=0$. Extend y_α to the entire interval $[0,+\infty)$ by defining

$$\tilde{y}_\alpha(s):=\begin{cases}x_0, & \text{if } s\in[0,s_{0,\alpha}],\\ y_\alpha(s), & \text{if } s\geq s_{0,\alpha}.\end{cases}\tag{3.40}$$

Licence CC BY-NC-ND 3.0

Then, as $\alpha \to +\infty$, the sequence $(\tilde{y}_\alpha)_\alpha$ converges uniformly on any compact subset of $[0, +\infty)$ to the unique solution $y : [0, +\infty) \to \mathcal{H}$ of the first-order differential equation

$$\dot{y}(s) + \nabla f\left(y(s) + \beta_0 \dot{y}(s)\right) = 0, \quad y(0) = x_0. \tag{3.41}$$

Furthermore, for any $T > 0$, strong convergence in L^2 holds:

$$\dot{\tilde{y}}_\alpha \to \dot{y} \quad \text{strongly in } L^2([0, T], \mathcal{H}). \tag{3.42}$$

To establish a complete proof of the theorem, the following lemmas are introduced, which constitute the steps of the proof.

Lemma 3.4.2. On the interval $[0, s_{0,\alpha}]$, the trajectory $\tilde{y}_\alpha$ defined by (3.40) and the solution y of (3.41) satisfy

$$\|\tilde{y}_\alpha(s) - y(s)\| \leq \frac{t_0}{\sqrt{2(\alpha+1)}} \left(f(x_0) - \inf(f)\right)^{1/2} =: C_1(\alpha), \quad \forall s \in [0, s_{0,\alpha}]. \tag{3.43}$$

Proof. From relations (2.1) and (3.35), the limit equation (3.41) can be equivalently written as

$$\dot{y}(s) = -\frac{1}{\beta_0}\left(y(s) - \operatorname{prox}_{\beta_0 f}(y(s))\right) = -\nabla \mathcal{M}_{\beta_0 f}(y(s)), \tag{3.44}$$

where $\mathcal{M}_{\beta_0 f}$ denotes the Moreau envelope of f, defined for $\beta_0 > 0$ by

$$\mathcal{M}_{\beta_0 f}(y) := \min_{x \in \mathcal{H}} \left\{ f(x) + \frac{1}{2\beta_0}\|x - y\|^2 \right\}.$$

For any $s \in [0, s_{0,\alpha}]$, one obtains

$$\|\tilde{y}_\alpha(s) - y(s)\| = \|x_0 - y(s)\| = \left\| \int_0^s \dot{y}(\tau)\, d\tau \right\| \leq \int_0^{s_{0,\alpha}} \|\dot{y}(\tau)\|\, d\tau.$$

Applying Hölder's inequality gives:

$$\|\tilde{y}_\alpha(s) - y(s)\| \leq s_{0,\alpha}^{1/2} \left(\int_0^{s_{0,\alpha}} \|\dot{y}(\tau)\|^2\, d\tau \right)^{1/2}.$$

Using (3.44) yields the energy relation:

$$\begin{aligned} \int_0^{s_{0,\alpha}} \|\dot{y}(\tau)\|^2\, d\tau &= -\int_0^{s_{0,\alpha}} \langle \dot{y}(\tau), \nabla \mathcal{M}_{\beta_0 f}(y(\tau)) \rangle\, d\tau \\ &= \mathcal{M}_{\beta_0 f}(y(0)) - \mathcal{M}_{\beta_0 f}(y(s_{0,\alpha})) \\ &\leq f(x_0) - \inf(f). \end{aligned} \tag{3.45}$$

Licence CC BY-NC-ND 3.0

Then

$$\|\tilde{y}_\alpha(s) - y(s)\| \leq s_{0,\alpha}^{1/2}(f(x_0) - \inf(f))^{1/2} = \frac{t_0}{\sqrt{2(\alpha+1)}}(f(x_0) - \inf(f))^{1/2},$$

which proves (3.43).

Moreover, the bound $C_1(\alpha)$ tends to zero as $\alpha \to +\infty$. And in particular, for $s = s_{0,\alpha}$

$$\|\tilde{y}_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\| = \|y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\| = \|x_0 - y(s_{0,\alpha})\| \leq C_1(\alpha). \tag{3.46}$$

□

Estimates on the initial interval allow control of the deviation between y_α and y near $s = 0$. To extend this control to all $s \geq s_{0,\alpha}$, uniform estimates on the energy of the trajectories y_α are required, which are established in the following lemma.

Lemma 3.4.3. The trajectory $\tilde{y}_\alpha$ defined by (3.40) satisfies the following inequalities:

$$\int_0^{+\infty} \|\dot{\tilde{y}}_\alpha(\tau)\|^2 \, d\tau \leq 2C_2, \tag{3.47}$$

$$\frac{\sqrt{s}}{\sqrt{\alpha+1}}\|\dot{y}_\alpha(s)\| \leq C_2^{1/2}, \quad \forall s \geq s_{0,\alpha}, \tag{3.48}$$

where

$$C_2 := \frac{1}{2}\|x_1\|^2 + f(x_0) - \inf(f).$$

Proof. Let $s \geq s_{0,\alpha}$. The inner product of equation (3.38) with $\dot{y}_\alpha(s)$ gives:

$$\left\langle \dot{y}_\alpha(s), \frac{2s}{\alpha+1}\ddot{y}_\alpha(s) + \dot{y}_\alpha(s) + \nabla f\big(y_\alpha(s) + \beta_0\dot{y}_\alpha(s)\big)\right\rangle = 0. \tag{3.49}$$

Using the monotonicity of the gradient ∇f (which follows from the convexity of f), the last term is decomposed:

$$\begin{aligned}\langle \dot{y}_\alpha(s), \nabla f(y_\alpha(s) + \beta_0\dot{y}_\alpha(s))\rangle &= \underbrace{\langle \dot{y}_\alpha(s), \nabla f(y_\alpha(s) + \beta_0\dot{y}_\alpha(s)) - \nabla f(y_\alpha(s))\rangle}_{\geq 0 \text{ (by monotonicity)}} \\ &\quad + \langle \dot{y}_\alpha(s), \nabla f(y_\alpha(s))\rangle \\ &\geq \frac{d}{ds} f(y_\alpha(s)).\end{aligned}$$

Substituting this estimate into (3.49) gives the differential inequality:

$$\frac{s}{\alpha+1}\frac{d}{ds}\|\dot{y}_\alpha(s)\|^2 + \|\dot{y}_\alpha(s)\|^2 + \frac{d}{ds} f(y_\alpha(s)) \leq 0. \tag{3.50}$$

Licence CC BY-NC-ND 3.0

Integrating (3.50) over $[s_{0,\alpha}, s]$ yields:

$$\int_{s_{0,\alpha}}^{s} \frac{\tau}{\alpha+1}\frac{d}{d\tau}\|\dot{y}_\alpha(\tau)\|^2\,d\tau + \int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2\,d\tau + f(y_\alpha(s)) - f(y_\alpha(s_{0,\alpha})) \leq 0.$$

Integrating the first term by parts gives:

$$\left[\frac{\tau}{\alpha+1}\|\dot{y}_\alpha(\tau)\|^2\right]_{s_{0,\alpha}}^{s} - \frac{1}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2\,d\tau + \int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2\,d\tau \leq f(x_0) - f(y_\alpha(s)).$$

Simplifying, one obtains:

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{\alpha}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2\,d\tau \leq \frac{1}{2}\|x_1\|^2 + f(x_0) - f(y_\alpha(s)). \tag{3.51}$$

Since $f(y_\alpha(s)) \geq \inf(f)$ and $\frac{\alpha}{\alpha+1} > \frac{1}{2}$ for all $\alpha > 1$, it follows that the desired bounds:

$$\int_0^{+\infty} \|\ddot{y}_\alpha(\tau)\|^2\,d\tau \leq 2C_2, \qquad \frac{\sqrt{s}}{\sqrt{\alpha+1}}\|\dot{y}_\alpha(s)\| \leq C_2^{1/2}.$$

□

Building on these uniform estimates for $\|\dot{y}_\alpha\|$ and the energy, one can prove the convergence of the trajectories y_α towards the solution y of the proximal flow.

Proof of Theorem 3.4.1. The goal is to estimate the difference $\|\tilde{y}_\alpha(s) - y(s)\|$ by a quantity depending on α that tends to zero as $\alpha \to +\infty$, uniformly for s in compact subsets of $[0, +\infty)$. From the monotonicity of the gradient mapping ∇f, it holds that

$$\langle \nabla f\left(y_\alpha(s) + \beta_0\dot{y}_\alpha(s)\right) - \nabla f\left(y(s) + \beta_0\dot{y}(s)\right), \left(y_\alpha(s) + \beta_0\dot{y}_\alpha(s)\right) - \left(y(s) + \beta_0\dot{y}(s)\right)\rangle \geq 0.$$

Substituting the expressions of the gradients from the respective differential equations, this gives

$$\left\langle -\frac{2s}{\alpha+1}\ddot{y}_\alpha(s) - \dot{y}_\alpha(s) + \dot{y}(s),\ y_\alpha(s) - y(s) + \beta_0(\dot{y}_\alpha(s) - \dot{y}(s)) \right\rangle \geq 0.$$

Expanding the inner product yields

$$-\frac{2s}{\alpha+1}\langle \ddot{y}_\alpha(s), y_\alpha(s) - y(s)\rangle - \frac{2s\beta_0}{\alpha+1}\langle \ddot{y}_\alpha(s), \dot{y}_\alpha(s) - \dot{y}(s)\rangle \geq \frac{1}{2}\frac{d}{ds}\|y_\alpha(s) - y(s)\|^2 + \beta_0\|\dot{y}_\alpha(s) - \dot{y}(s)\|^2. \tag{3.52}$$

Licence CC BY-NC-ND 3.0

Integrating inequality (3.52) over the interval $[s_{0,\alpha}, s]$, one obtains

$$\frac{1}{2}\|y_\alpha(s)-y(s)\|^2+\beta_0\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)-\dot{y}(\tau)\|^2\,d\tau \leq \frac{1}{2}\|x_0-y(s_{0,\alpha})\|^2+T_1(s,\alpha)+T_2(s,\alpha), \quad (3.53)$$

where the remaining terms are defined by

$$T_1(s,\alpha) := -\frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s}\langle \ddot{y}_\alpha(\tau),\ \tau(y_\alpha(\tau)-y(\tau))\rangle\ d\tau,$$
$$T_2(s,\alpha) := -\frac{2\beta_0}{\alpha+1}\int_{s_{0,\alpha}}^{s}\langle \ddot{y}_\alpha(\tau),\ \tau(\dot{y}_\alpha(\tau)-\dot{y}(\tau))\rangle\ d\tau.$$

To establish convergence, it remains to show that $T_1(s,\alpha)$ and $T_2(s,\alpha)$ vanish in the limit $\alpha \to +\infty$. The analysis of these terms will be carried out in two steps.

Estimation of the term $T_1(s,\alpha)$: The integration by parts is performed as follows

$$\int_{s_{0,\alpha}}^{s}\langle \ddot{y}_\alpha(\tau),\ \tau(y_\alpha(\tau)-y(\tau))\rangle\ d\tau = B(s,\alpha)-A(s,\alpha).$$

where

$$A(s,\alpha) = \int_{s_{0,\alpha}}^{s}\langle \dot{y}_\alpha(\tau),\ (y_\alpha(\tau)-y(\tau))+\tau(\dot{y}_\alpha(\tau)-\dot{y}(\tau))\rangle\ d\tau$$
$$B(s,\alpha) = [\langle \dot{y}_\alpha(\tau),\ \tau(y_\alpha(\tau)-y(\tau))\rangle]_{s_{0,\alpha}}^{s}$$

Estimation of the term $A(s,\alpha)$: Using the Cauchy-Schwarz inequality,

$$\begin{aligned}|A(s,\alpha)| &\leq \int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|\cdot\|y_\alpha(\tau)-y(\tau)\|\,d\tau \\ &\quad + s\int_{s_{0,\alpha}}^{s}|\langle \dot{y}_\alpha(\tau),\,\dot{y}_\alpha(\tau)-\dot{y}(\tau)\rangle|\ d\tau \\ &=: I_1+I_2.\end{aligned}$$

The second term I_2 is estimated next. For any τ, the identity

$$\begin{aligned}2\,|\langle \dot{y}_\alpha(\tau),\,\dot{y}_\alpha(\tau)-\dot{y}(\tau)\rangle| &= |2\|\dot{y}_\alpha(\tau)\|^2-2\,\langle \dot{y}_\alpha(\tau),\,\dot{y}(\tau)\rangle| \\ &\leq 2\|\dot{y}_\alpha(\tau)\|^2+2\|\dot{y}_\alpha(\tau)\|\cdot\|\dot{y}(\tau)\| \\ &\leq 3\|\dot{y}_\alpha(\tau)\|^2+\|\dot{y}(\tau)\|^2.\end{aligned}$$

Licence CC BY-NC-ND 3.0

Consequently,

$$|A(s,\alpha)| \leq \int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \cdot \|y_\alpha(\tau) - y(\tau)\|\, d\tau + s\left(\frac{3}{2}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2\, d\tau + \frac{1}{2}\int_{s_{0,\alpha}}^{s} \|\dot{y}(\tau)\|^2\, d\tau\right). \tag{3.54}$$

From estimate (3.47), it follows that:

$$\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2\, d\tau \leq 2C_2.$$

Moreover, the energy of y is bounded using the Lyapunov decay:

$$\begin{aligned}\int_{s_{0,\alpha}}^{s} \|\dot{y}(\tau)\|^2\, d\tau &= -\int_{s_{0,\alpha}}^{s} \langle \dot{y}(\tau),\, \nabla \mathcal{M}_{\beta_0 f}(y(\tau))\rangle\, d\tau \\ &= \mathcal{M}_{\beta_0 f}(y(s_{0,\alpha})) - \mathcal{M}_{\beta_0 f}(y(s)) \\ &\leq f(x_0) + \frac{1}{2\beta_0}\|x_0 - y(s_{0,\alpha})\|^2 - \inf(f) \\ &\leq C_2 + \frac{C_1^2(\alpha)}{2\beta_0}.\end{aligned}$$

Plugging these results into (3.54) yields:

$$\frac{2}{\alpha+1}|A(s,\alpha)| \leq \frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \cdot \|y_\alpha(\tau) - y(\tau)\|\, d\tau + \frac{8sC_2}{\alpha+1} + \frac{sC_1^2(\alpha)}{\beta_0(\alpha+1)}. \tag{3.55}$$

Estimate of the term $B(s,\alpha)$: Attention is now turned to the boundary contributions appearing in the integration by parts. Applying the Cauchy–Schwarz inequality and the identity for $\dot{y}_\alpha(s_{0,\alpha})$ yields

$$\begin{aligned}\frac{2}{\alpha+1}\Big|\, \langle \dot{y}_\alpha(s_{0,\alpha}),\ s_{0,\alpha}(y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha}))\rangle\, \Big| &\leq \frac{2s_{0,\alpha}}{\alpha+1}\, \|\dot{y}_\alpha(s_{0,\alpha})\| \cdot \|y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\| \\ &= \frac{2s_{0,\alpha}}{\alpha+1} \cdot \left\|\frac{\alpha+1}{t_0}x_1\right\| \cdot \|y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\| \\ &= \frac{t_0}{\alpha+1}\|x_1\| \cdot \|y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\| \\ &\leq \frac{t_0\|x_1\|}{\alpha+1}C_1(\alpha),\end{aligned} \tag{3.56}$$

where the last inequality follows from estimate (3.43).

The second boundary term is estimated next. Applying the Cauchy–Schwarz inequality again,

$$|\langle \dot{y}_\alpha(s),\, s(y_\alpha(s) - y(s))\rangle| \leq s\, \|\dot{y}_\alpha(s)\|\, \|y_\alpha(s) - y(s)\|. \tag{3.57}$$

Licence CC BY-NC-ND 3.0

Dividing by $(\alpha+1)$ and using (3.48) gives

$$\frac{2}{\alpha+1}|\langle \dot{y}_\alpha(s),\, s(y_\alpha(s)-y(s))\rangle| \leq 2\sqrt{\frac{s}{\alpha+1}}\cdot C_2^{1/2}\cdot \|y_\alpha(s)-y(s)\|. \tag{3.58}$$

To bound the right-hand side, Young's inequality $2ab \leq a^2+b^2$, valid for all $a,b\in\mathbb{R}$, is applied. Setting

$$a=\frac{1}{2}\|y_\alpha(s)-y(s)\|, \qquad b=2\sqrt{\frac{s}{\alpha+1}}\,C_2^{1/2},$$

(3.58) then yields:

$$\frac{2}{\alpha+1}|\langle \dot{y}_\alpha(s),\, s(y_\alpha(s)-y(s))\rangle| \leq \frac{1}{4}\|y_\alpha(s)-y(s)\|^2+\frac{4s}{\alpha+1}C_2. \tag{3.59}$$

Combining (3.56) and (3.59) leads to the final bound:

$$\frac{2}{\alpha+1}|B(s,\alpha)| \leq \frac{t_0\|x_1\|}{\alpha+1}C_1(\alpha)+\frac{1}{4}\|y_\alpha(s)-y(s)\|^2+\frac{4s}{\alpha+1}C_2. \tag{3.60}$$

Altogether, combining the bounds for $A(s,\alpha)$ and $B(s,\alpha)$, the following estimate for the full remainder term is obtained:

$$T_1(s,\alpha) \leq \frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|\cdot\|y_\alpha(\tau)-y(\tau)\|\,d\tau+\frac{1}{4}\|y_\alpha(s)-y(s)\|^2+C(\alpha,s) \tag{3.61}$$

where

$$C(\alpha,s):=\frac{t_0\|x_1\|}{\alpha+1}C_1(\alpha)+\frac{12s}{\alpha+1}C_2+\frac{sC_1^2(\alpha)}{\beta_0(\alpha+1)}.$$

Estimate of the term $T_2(s,\alpha)$: The second remainder term $T_2(s,\alpha)$ is now analyzed using integration by parts. First, observe that

$$\begin{aligned}
2\int_{s_{0,\alpha}}^{s}\tau\langle\ddot{y}_\alpha(\tau),\dot{y}_\alpha(\tau)\rangle d\tau &= \int_{s_{0,\alpha}}^{s}\tau\frac{d}{d\tau}\|\dot{y}_\alpha(\tau)\|^2\,d\tau \\
&= s\|\dot{y}_\alpha(s)\|^2-s_{0,\alpha}\|\dot{y}_\alpha(s_{0,\alpha})\|^2-\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|^2\,d\tau \\
&\geq -\frac{\alpha+1}{2}\|x_1\|^2-\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|^2\,d\tau,
\end{aligned} \tag{3.62}$$

where the last inequality uses the expression of $\dot{y}_\alpha(s_{0,\alpha})$. Multiplying by $-\frac{\beta_0}{\alpha+1}$ and invoking (3.47), one obtains:

$$-\frac{2\beta_0}{\alpha+1}\int_{s_{0,\alpha}}^{s}\tau\langle\ddot{y}_\alpha(\tau),\dot{y}_\alpha(\tau)\rangle d\tau \leq \frac{\beta_0}{2}\|x_1\|^2+\frac{2\beta_0C_2}{\alpha+1}. \tag{3.63}$$

Licence CC BY-NC-ND 3.0

An estimate is now provided for the cross term involving $\dot{y}(\tau)$. Using integration by parts:

$$\begin{aligned}\int_{s_{0,\alpha}}^{s} \tau\langle \ddot{y}_\alpha(\tau), \dot{y}(\tau)\rangle \, d\tau &= \langle s\dot{y}_\alpha(s), \dot{y}(s)\rangle - \langle s_{0,\alpha}\dot{y}_\alpha(s_{0,\alpha}), \dot{y}(s_{0,\alpha})\rangle \\ &\quad - \int_{s_{0,\alpha}}^{s} \Big\langle \dot{y}_\alpha(\tau), \tau\ddot{y}(\tau) + \dot{y}(\tau)\Big\rangle \, d\tau \\ &\leq s\|\dot{y}_\alpha(s)\| \, \|\dot{y}(s)\| + \frac{t_0}{2}\|x_1\| \, \|\dot{y}(s_{0,\alpha})\| \\ &\quad + \int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \, \|\tau\ddot{y}(\tau) + \dot{y}(\tau)\| \, d\tau.\end{aligned}$$

Let $x^* \in \texttt{argmin}(f)$, hence also a minimizer of $\mathcal{M}_{\beta_0 f}$. Then:

$$\begin{aligned}\|\dot{y}(s_{0,\alpha})\| &= \|\nabla\mathcal{M}_{\beta_0 f}(y(s_{0,\alpha})) - \nabla\mathcal{M}_{\beta_0 f}(x^*)\| \\ &\leq \frac{1}{\beta_0}\|y(s_{0,\alpha}) - x^*\| \leq \frac{1}{\beta_0}\big(C_1(\alpha) + \|x_0 - x^*\|\big) =: \frac{C_3(\alpha)}{\beta_0}.\end{aligned}$$

Define the control quantities:

$$M_0(s) := \sup_{0\leq\tau\leq s} \|\tau\ddot{y}(\tau) + \dot{y}(\tau)\|, \qquad M_1(s) := \sup_{0\leq\tau\leq s} \|y(\tau) - x^*\|. \tag{3.64}$$

This yields the bound:

$$\begin{aligned}\frac{2\beta_0}{\alpha+1}\int_{s_{0,\alpha}}^{s} \tau\langle \ddot{y}_\alpha(\tau), \dot{y}(\tau)\rangle d\tau &\leq 2M_1(s)\frac{s\|\dot{y}_\alpha(s)\|}{\alpha+1} + \frac{t_0\|x_1\|}{\alpha+1}C_3(\alpha) \\ &\quad + \frac{2\beta_0}{\alpha+1}M_0(s)\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \, d\tau.\end{aligned}$$

Using Hölder's inequality and (3.47) yields the estimate:

$$\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \, d\tau \leq \sqrt{s - s_{0,\alpha}}\left(\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2 d\tau\right)^{1/2} \leq \sqrt{2sC_2}. \tag{3.65}$$

Combining all results and using (3.48) gives:

$$T_2(s,\alpha) \leq 2M_1(s)\sqrt{\frac{sC_2}{\alpha+1}} + \frac{t_0\|x_1\|}{\alpha+1}C_3(\alpha) + \frac{\beta_0\sqrt{8sC_2}}{\alpha+1}M_0(s) + \frac{\beta_0}{2}\|x_1\|^2 + \frac{2\beta_0 C_2}{\alpha+1}. \tag{3.66}$$

Finally, by combining (3.46), (3.61), and (3.66) with the estimate (3.53), the following result

Licence CC BY-NC-ND 3.0

is obtained:

$$\frac{1}{2}\|y_\alpha(s)-y(s)\|^2+\beta_0\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)-\dot{y}(\tau)\|^2d\tau\leq\frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|\cdot\|y_\alpha(\tau)-y(\tau)\|\,d\tau + \frac{1}{4}\|y_\alpha(s)-y(s)\|^2+K(s,\alpha),$$

where

$$\begin{aligned}K(s,\alpha)=&\frac{t_0\|x_1\|}{\alpha+1}C_1(\alpha)+\frac{12s}{\alpha+1}C_2+\frac{sC_1^2(\alpha)}{\beta_0(\alpha+1)}\\&+2M_1(s)\sqrt{\frac{sC_2}{\alpha+1}}+\frac{t_0\|x_1\|}{\alpha+1}C_3(\alpha)+\frac{\beta_0\sqrt{8sC_2}}{\alpha+1}M_0(s)\\&+\frac{2\beta_0C_2}{\alpha+1}+\frac{C_1^2(\alpha)}{2}+\frac{\beta_0\|x_1\|^2}{2}.\end{aligned}$$

Rearranging the terms yields:

$$\frac{1}{2}\|y_\alpha(s)-y(s)\|^2+2\beta_0\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)-\dot{y}(\tau)\|^2d\tau\leq\frac{4}{\alpha+1}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|\,\|y_\alpha(\tau)-y(\tau)\|\,d\tau + 2K(s,\alpha). \tag{3.67}$$

In particular, this implies:

$$\frac{1}{2}\|y_\alpha(s)-y(s)\|^2\leq\frac{4}{\alpha+1}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|\,\|y_\alpha(\tau)-y(\tau)\|\,d\tau+2K(s,\alpha). \tag{3.68}$$

Final convergence estimates: Applying Grönwall's inequality to (3.68) (Lemma 2.2.2) yields:

$$\|y_\alpha(s)-y(s)\|\leq\frac{4}{\alpha+1}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|\,d\tau+2\sqrt{K(s,\alpha)},\qquad\forall s\geq s_{0,\alpha}. \tag{3.69}$$

Combining this with Hölder's inequality and (3.47):

$$\|y_\alpha(s)-y(s)\|\leq\frac{4\sqrt{2sC_2}}{\alpha+1}+2\sqrt{K(s,\alpha)},\qquad\forall s\geq s_{0,\alpha}. \tag{3.70}$$

This yields the refined estimate for the reparametrized solution $\tilde{y}_\alpha(s)=y_\alpha(s)$ on $[s_{0,\alpha},+\infty[$:

$$\|\tilde{y}_\alpha(s)-y(s)\|\leq\frac{4\sqrt{2sC_2}}{\alpha+1}+K_1(s,\alpha)+\sqrt{2}C_1(\alpha),\qquad\forall s\geq s_{0,\alpha}, \tag{3.71}$$

Licence CC BY-NC-ND 3.0

where

$$K_1(s,\alpha) = \left[\frac{t_0\|x_1\|}{\alpha+1} C_1(\alpha) + \frac{12s}{\alpha+1} C_2 + \frac{sC_1^2(\alpha)}{\beta_0(\alpha+1)} + 2M_1(s)\sqrt{\frac{sC_2}{\alpha+1}} \right.$$
$$\left. + \frac{t_0\|x_1\|}{\alpha+1} C_3(\alpha) + \frac{\beta_0\sqrt{8sC_2}}{\alpha+1} M_0(s) + \frac{2\beta_0 C_2}{\alpha+1} + \frac{\beta_0\|x_1\|^2}{2} \right]^{1/2}.$$

Using (3.46), estimate (3.71) remains valid for all $s \geq 0$. But

$$\lim_{\alpha\to+\infty} K_1(s,\alpha) + \sqrt{2}C_1(\alpha) = 0 \quad \text{iff} \quad \frac{\beta_0\|x_1\|^2}{2} = 0.$$

Therefore, under the assumption $x_1 = 0$, it follows that the sequence $(\tilde{y}_\alpha)_\alpha$ converges uniformly to y on every compact subset of $[0, +\infty[$.

The strong convergence of the time derivatives in $L^2([0,T])$ is now established. Observe that

$$\int_0^T \|\dot{\tilde{y}}_\alpha(\tau) - \dot{y}(\tau)\|^2 d\tau = \int_0^{s_{0,\alpha}} \|\dot{\tilde{y}}_\alpha(\tau) - \dot{y}(\tau)\|^2 d\tau + \int_{s_{0,\alpha}}^T \|\dot{\tilde{y}}_\alpha(\tau) - \dot{y}(\tau)\|^2 d\tau.$$

The first term is bounded by the use of (3.44), combined with the $\frac{1}{\beta_0}$-Lipschitz property of $\nabla\mathcal{M}_{\beta_0 f}$ and the increasing nature of the function $M_1(s)$ (see (3.64)).

$$\int_0^{s_{0,\alpha}} \|\dot{y}(\tau)\|^2 d\tau \leq \frac{1}{\beta_0^2} \int_0^{s_{0,\alpha}} \|y(\tau) - x^*\|^2 d\tau \leq \frac{s_{0,\alpha}}{\beta_0^2} M_1(t_0^2),$$

which tends to zero as $\alpha \to +\infty$, since $s_{0,\alpha} = \frac{t_0^2}{2(\alpha+1)} \to 0$. For the second term, using (3.67) together with (3.47) gives:

$$\int_{s_{0,\alpha}}^T \|\dot{\tilde{y}}_\alpha(\tau) - \dot{y}(\tau)\|^2 d\tau \leq \frac{1}{\beta_0(\alpha+1)} \left(\int_{s_{0,\alpha}}^T \|y_\alpha(\tau) - y(\tau)\|^2 d\tau + 2C_2 \right) + \frac{1}{\beta_0} K(T,\alpha)$$
$$\leq \frac{1}{\beta_0(\alpha+1)} \left(T \sup_{s\in[0,T]} \|\tilde{y}_\alpha(s) - y(s)\|^2 + 2C_2 \right) + \frac{1}{\beta_0} K(T,\alpha).$$

Combining both estimates yields:

$$\lim_{\alpha\to+\infty} \int_0^T \|\dot{\tilde{y}}_\alpha(\tau) - \dot{y}(\tau)\|^2 d\tau = 0, \qquad \text{if } x_1 = 0. \tag{3.72}$$

This concludes the proof.

□

The preceding proof highlights two key aspects, which are further discussed in the follow-

Licence CC BY-NC-ND 3.0

ing remark.

Remark 3.4.1. 1. *On the necessity of the initial condition $x_1 = 0$:*

Contrary to systems where the damping explicitly depends on the Hessian, the study of the implicit dynamics (3.36) involves a more subtle correction term, related to the gradient evaluated at $x+\beta\dot{x}$. This structure makes the analysis more delicate, particularly when controlling some terms in the energy estimates during the limit $\alpha \to +\infty$.

The convergence proof relies on a differential inequality concerning the quantity $\|\tilde{y}_\alpha(s) - y(s)\|^2$ and on an L^2-norm estimate of the difference between the velocities $\|\dot{\tilde{y}}_\alpha(s) - \dot{y}(s)\|$. A critical term, related to the reparametrized initial condition $\dot{y}_\alpha(s_{0,\alpha})$, appears in these estimates and can only be uniformly controlled in α if it vanishes. The condition $\dot{y}_\alpha(s_{0,\alpha}) = 0$, equivalent to $x_1 = 0$ according to (3.39), is therefore a technical assumption.

This assumption prevents the appearance of an initial impulse that could perturb or cause divergence of the system in the large α limit. It is essential to ensure the consistency of Gronwall-type estimates and to obtain strong L^2 convergence of the velocities (3.42) as well as uniform convergence of the trajectories.

Although it may seem restrictive, the assumption $x_1 = 0$ (zero initial velocity) is standard in the study of inertial systems, whether continuous or discrete. It corresponds to a natural choice of initialization in accelerated methods and primarily aims to guarantee system stability rather than limit generality.

2. *Connection with Levenberg–Marquardt regularization:*

The limiting dynamics (3.34) bears similarities to the Levenberg–Marquardt method, often used to stabilize Newton's method. By expanding the gradient to first order, one obtains:

$$\nabla f\left(y(s) + \beta_0 \dot{y}(s)\right) \approx \nabla f(y(s)) + \beta_0 \nabla^2 f(y(s))\dot{y}(s).$$

Substituting this expression into (3.34) yields:

$$\left(\mathrm{Id} + \beta_0 \nabla^2 f(y(s))\right) \dot{y}(s) \approx -\nabla f(y(s)),$$

which formally corresponds to the equation used in the Levenberg–Marquardt method. However, this resemblance remains only formal. The limiting dynamics (3.34) is well-defined without assuming that the Hessian of f exists. It only requires the convexity of f and the Lipschitz regularity of its gradient. This robustness comes directly from the inertial structure of the original system, which naturally leads to this implicit form when α becomes very large.

Licence CC BY-NC-ND 3.0

3.5 Limit dynamics of (ISIHD): Non-convex quadratic case

Building on the convergence results obtained in the convex case (Theorem 3.4.1), the analysis is now extended to the more delicate setting of non-convex functions, restricted to finite-dimensional spaces. This extension requires particular attention to the behavior of the gradient when convergence is only weak.

To avoid technical difficulties related to weak convergence of gradients in a general non-convex setting and to conduct a study, the focus is placed here on the case of quadratic functions. The gradient of a quadratic function is an affine operator, hence continuous and bounded on any bounded set, which ensures good compatibility with weak convergence in L^2, essential for the asymptotic analysis. The quadratic setting provides a clear and precise framework to study the strong damping regime ($\alpha \to +\infty$), without the technical complications introduced by general nonlinearities.

These considerations lead us to establish the main convergence result for quadratic functions, whether convex or non-convex.

Remark 3.5.1 (Role of the Quadratic Assumption)**.** In this section, the quadratic assumption is *structural* rather than merely technical. In particular, when $\beta_0 > 0$, the affinity of the gradient ∇f for a quadratic function allows the passage to the limit and guarantees weak convergence of the solutions. Without this structure, these results would not hold, and additional assumptions would be required, such as monotonicity or compactness.

Theorem 3.5.1. [*Non-convex quadratic case in finite dimension*] Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a quadratic function, and let $x_0 \in \mathbb{R}^n$ with initial velocity $x_1 = 0$. For each $\alpha > 0$, let y_α be the reparametrized trajectory associated with the inertial system (3.36), defined on $[s_{0,\alpha}, +\infty)$, and define the extended function $\tilde{y}_\alpha : [0, +\infty) \to \mathbb{R}^n$ by

$$\tilde{y}_\alpha(s) = \begin{cases} x_0 & \text{if } s \in [0, s_{0,\alpha}], \\ y_\alpha(s) & \text{if } s \geq s_{0,\alpha}, \end{cases} \tag{3.73}$$

where $s_{0,\alpha} = \dfrac{t_0^2}{2(\alpha+1)}$.

Then, as $\alpha \to +\infty$, the sequence $(\tilde{y}_\alpha)_{\alpha>0}$ converges uniformly on any compact interval of $[0, +\infty)$ to the unique solution y of the proximal flow

$$\dot{y}(s) + \nabla f\left(y(s) + \beta_0 \dot{y}(s)\right) = 0, \quad y(0) = x_0. \tag{3.74}$$

The proof of this theorem is provided, beginning with the statement and proof of the following lemma.

Licence CC BY-NC-ND 3.0

Lemma 3.5.2. Under the assumption $x_1 = 0$, the trajectory $\tilde{y}_\alpha$ defined by (3.73) satisfies

$$\int_0^{+\infty} \|\dot{\tilde{y}}_\alpha(\tau)\|^2 \, d\tau \leq a_1, \quad \sup_{s\geq 0} \|\dot{\tilde{y}}_\alpha(s)\|^2 \leq \frac{a_1}{\beta_0}, \tag{3.75}$$

with $a_1 := 2(f(x_0) - \inf(f))$.

Proof. For $s \geq s_{0,\alpha}$, taking the inner product of equation (3.38) with $\dot{y}_\alpha(s) + \beta_0 \ddot{y}_\alpha(s)$, one obtains:

$$\Big\langle \frac{2s}{\alpha+1}\ddot{y}_\alpha(s) + \dot{y}_\alpha(s) + \nabla f\big(y_\alpha(s) + \beta_0 \dot{y}_\alpha(s)\big), \dot{y}_\alpha(s) + \beta_0 \ddot{y}_\alpha(s)\Big\rangle = 0.$$

This identity can be rewritten in differential form:

$$\begin{aligned}\Big(\frac{s}{\alpha+1} + \frac{\beta_0}{2}\Big)\frac{d}{ds}\|\dot{y}_\alpha(s)\|^2 + \frac{2s\beta_0}{\alpha+1}\|\ddot{y}_\alpha(s)\|^2 \\ + \|\dot{y}_\alpha(s)\|^2 + \frac{d}{ds} f\big(y_\alpha(s) + \beta_0 \dot{y}_\alpha(s)\big) = 0.\end{aligned} \tag{3.76}$$

Integrating (3.76) over $[s_{0,\alpha}, s]$ gives:

$$\begin{aligned}&\Big[\Big(\frac{\tau}{\alpha+1} + \frac{\beta_0}{2}\Big)\|\dot{y}_\alpha(\tau)\|^2\Big]_{s_{0,\alpha}}^{s} + \frac{\alpha}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2 d\tau \\ &\quad + \frac{2\beta_0}{\alpha+1}\int_{s_{0,\alpha}}^{s} \tau\|\ddot{y}_\alpha(\tau)\|^2 d\tau = f\big(y_\alpha(s_{0,\alpha}) + \beta_0\dot{y}_\alpha(s_{0,\alpha})\big) - f\big(y_\alpha(s) + \beta_0\dot{y}_\alpha(s)\big).\end{aligned} \tag{3.77}$$

Using the bound $\frac{\alpha}{\alpha+1} > \frac{1}{2}$, the following inequality is deduced:

$$\begin{aligned}&\Big(\frac{s}{\alpha+1} + \frac{\beta_0}{2}\Big)\|\dot{y}_\alpha(s)\|^2 + \frac{1}{2}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2 d\tau + \frac{2\beta_0}{\alpha+1}\int_{s_{0,\alpha}}^{s} \tau\|\ddot{y}_\alpha(\tau)\|^2 d\tau \\ &\qquad \leq f\Big(x_0 + \frac{\beta_0(\alpha+1)}{t_0}x_1\Big) - \inf(f) + \Big(\frac{1}{2} + \frac{\beta_0(\alpha+1)^2}{2t_0^2}\Big)\|x_1\|^2.\end{aligned} \tag{3.78}$$

Assuming $x_1 = 0$, one then obtains:

$$\Big(\frac{s}{\alpha+1} + \frac{\beta_0}{2}\Big)\|\dot{y}_\alpha(s)\|^2 + \frac{1}{2}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2 d\tau + \frac{2\beta_0}{\alpha+1}\int_{s_{0,\alpha}}^{s} \tau\|\ddot{y}_\alpha(\tau)\|^2 d\tau \leq f(x_0) - \inf(f).$$

From this, it follows that the uniform bounds hold:

$$\int_0^{s} \|\dot{\tilde{y}}_\alpha(\tau)\|^2 d\tau = \int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2 d\tau \leq a_1 := 2(f(x_0) - \inf(f)), \quad \|\dot{y}_\alpha(s)\|^2 \leq \frac{a_1}{\beta_0}. \tag{3.79}$$

Licence CC BY-NC-ND 3.0

Taking the limit $s \to +\infty$ finally yields (3.75):

$$\int_0^{+\infty} \|\dot{\tilde{y}}_\alpha(\tau)\|^2 d\tau \leq a_1, \quad \sup_{s\geq 0} \|\dot{\tilde{y}}_\alpha(s)\|^2 \leq \frac{a_1}{\beta_0}.$$

□

The theorem 3.5.1 can thus be established by relying on the estimates provided in Lemma 3.5.2.

Proof of Theorem 3.5.1. For all $s > 0$, it holds that:

$$\tilde{y}_\alpha(s) = \tilde{y}_\alpha(s_{0,\alpha}) + \int_{s_{0,\alpha}}^{s} \dot{\tilde{y}}_\alpha(\tau) d\tau = x_0 + \int_{s_{0,\alpha}}^{s} \dot{\tilde{y}}_\alpha(\tau) d\tau.$$

Combining the Cauchy–Schwarz inequality with estimate (3.75) from Lemma 3.5.2, one obtains:

$$\|\tilde{y}_\alpha(s) - x_0\| \leq \sqrt{|s - s_{0,\alpha}|} \left(\int_0^s \|\dot{\tilde{y}}_\alpha(\tau)\|^2 d\tau \right)^{1/2} \leq (\sqrt{T} + t_0)\sqrt{a_1}, \quad \forall s \in [0, T]. \quad (3.80)$$

Estimates (3.75) and (3.80) imply that $(\tilde{y}_\alpha)_\alpha$ and $(\dot{\tilde{y}}_\alpha)_\alpha$ are uniformly bounded in $\mathcal{C}_0(0, T; \mathcal{H})$ for all $T > 0$.

The equicontinuity of $(\tilde{y}_\alpha)_\alpha$ is established next. For any $s_1, s_2 > 0$, it holds that:

$$\|\tilde{y}_\alpha(s_2) - \tilde{y}_\alpha(s_1)\|^2 \leq |s_2 - s_1| \int_{\min(s_1,s_2)}^{\max(s_1,s_2)} \|\dot{\tilde{y}}_\alpha(\tau)\|^2 d\tau \leq a_1 |s_2 - s_1|. \quad (3.81)$$

This shows that the sequence is uniformly equicontinuous. Since $\mathcal{H}$ is finite-dimensional, the Ascoli–Arzelà theorem implies that the family $(\tilde{y}_\alpha)_\alpha$ is relatively compact in $\mathcal{C}_0(0, T; \mathcal{H})$ for the topology of uniform convergence.

Hence, $(\tilde{y}_\alpha)_\alpha$ admits at least one cluster point $\xi \in \mathcal{C}_0(0, T; \mathcal{H})$. To prove uniqueness, it suffices to show that ξ is a solution of (3.74).

Let $(\tilde{y}_{\alpha_k})_k$ be a subsequence such that

$$\sup_{0\leq s\leq T} \|\tilde{y}_{\alpha_k}(s) - \xi(s)\| \to 0 \quad \text{as } k \to +\infty. \quad (3.82)$$

Since $s_{0,\alpha_k} \to 0$ and $\tilde{y}_{\alpha_k}(s_{0,\alpha_k}) = x_0$, then

$$\xi(0) = x_0. \quad (3.83)$$

Passing to the limit in the dynamics. Let $\delta \in [0, T]$ be such that $\tilde{y}_{\alpha_k} = y_{\alpha_k}$ on $[\delta, T]$ for k sufficiently large. The two main terms of the dynamics are examined separately.

Licence CC BY-NC-ND 3.0

First derivative: The sequence $(\dot{\tilde{y}}_{\alpha_k})_k$ converges to $\dot{\xi}$ in $\mathcal{D}'(\delta, T; \mathcal{H})$. Moreover, from the energy estimates, it is bounded in $L^2(\delta, T; \mathcal{H})$. According to the proposition 2.2.1, it follows that

$$\dot{\tilde{y}}_{\alpha_k} \rightharpoonup \dot{\xi} \quad \text{weakly in } L^2(\delta, T; \mathcal{H}).$$

Inertial term: Since the factor $\frac{2s}{\alpha_k+1}$ tends uniformly to zero on $[\delta, T]$ as $k \to +\infty$, for any test function $\varphi \in \mathcal{D}(\delta, T; \mathcal{H})$ one obtains

$$\int_\sigma^T \Big\langle \frac{2\tau}{\alpha_k+1}\ddot{\tilde{y}}_{\alpha_k}(\tau), \varphi(\tau) \Big\rangle \, d\tau = \frac{2}{\alpha_k+1} \int_\sigma^T \Big\langle \tilde{y}_{\alpha_k}(\tau), \frac{d^2}{d\tau^2}\big(\tau\varphi(\tau)\big) \Big\rangle \, d\tau \longrightarrow 0,$$

which establishes convergence to zero in the sense of distributions. Moreover, the sequence $\left(\frac{2s}{\alpha_k+1}\ddot{\tilde{y}}_{\alpha_k}\right)_k$ remains bounded in $L^2(\delta, T; \mathcal{H})$, yielding weak convergence (According to the proposition 2.2.1)

$$\frac{2s}{\alpha_k+1}\ddot{\tilde{y}}_{\alpha_k} \rightharpoonup 0 \quad \text{in } L^2(\delta, T; \mathcal{H}).$$

Gradient term: From the governing equation, it follows that

$$\nabla f(\tilde{y}_{\alpha_k} + \beta_0 \dot{\tilde{y}}_{\alpha_k}) \rightharpoonup -\dot{\xi} \quad \text{in } L^2(\delta, T; \mathcal{H}).$$

This is precisely where the quadratic structure of f intervenes. When f is quadratic, its gradient ∇f is an affine operator. Thus, the weak convergences

$$\tilde{y}_{\alpha_k} \rightharpoonup \xi, \qquad \dot{\tilde{y}}_{\alpha_k} \rightharpoonup \dot{\xi} \quad \text{in } L^2(\delta, T; \mathcal{H})$$

imply

$$\tilde{y}_{\alpha_k} + \beta_0 \dot{\tilde{y}}_{\alpha_k} \rightharpoonup \xi + \beta_0 \dot{\xi} \quad \Longrightarrow \quad \nabla f(\tilde{y}_{\alpha_k} + \beta_0 \dot{\tilde{y}}_{\alpha_k}) \rightharpoonup \nabla f(\xi + \beta_0 \dot{\xi}) \quad \text{in } L^2(\delta, T; \mathcal{H}).$$

By uniqueness of the weak limit, it follows that the limit satisfies the equation:

$$\dot{\xi}(s) + \nabla f(\xi(s) + \beta_0 \dot{\xi}(s)) = 0.$$

Combined with the initial condition $\xi(0) = x_0$, this shows that $\xi = y$, which completes the proof.

□

The preceding proof relies on the affine structure of the gradient ∇f in the quadratic case. The following remark addresses this point.

Licence CC BY-NC-ND 3.0

Remark 3.5.2. *Case $\beta_0 = 0$ versus $\beta_0 > 0$:* When $\beta_0 = 0$, the limit dynamics reduce to the classical gradient flow:

$$\dot{y}(s) + \nabla f(y(s)) = 0.$$

In this context, it is not necessary for f to be quadratic: the uniform bounds obtained on $\dot{\tilde{y}}_\alpha$ via Lemma 3.5.2 are sufficient to pass to the limit using simply the continuity of the gradient ∇f (see [27]).
If $\beta_0 > 0$, the limit dynamics become implicit:

$$\dot{y}(s) + \nabla f\big(y(s) + \beta_0 \dot{y}(s)\big) = 0.$$

To justify the passage to the limit in the nonlinear term $\nabla f(\tilde{y}_\alpha + \beta_0 \dot{\tilde{y}}_\alpha)$, it is necessary that the Nemytskii operator

$$T_f : z \mapsto \nabla f \circ z$$

be weakly sequentially continuous in $L^2([\sigma, T]; \mathcal{H})$. This property is ensured when f is quadratic, since its gradient is affine and thus T_f is weakly sequentially continuous. This is why the quadratic assumption on f is made in the theorem for $\beta_0 > 0$. *Nevertheless, the result remains valid for any function f whose gradient makes the operator T_f weakly sequentially continuous in L^2.*

Condition on the initial velocity: As in the convex case, the condition $x_1 = 0$ is essential, even though it may seem restrictive. It eliminates the term in $\|x_1\|^2$ in the energy inequality (3.78), which allows uniform bounds in α and hence the application of compactness arguments by Lemma 3.5.2 and the Ascoli-Arzelà theorem.

3.6 Connection Between Continuous Dynamics and Discrete Algorithms

The discussion begins with a brief explanation of how the results obtained in the previous sections for continuous-time dynamics can be transferred to algorithms arising from time discretization. In general, given a continuous-time dynamical system that generates solution trajectories denoted by $x(t)$, a discrete algorithm is obtained by considering sequences $(x_k)_k$ such that $x_k \approx x(t_k)$, where $t_k = kh$ and $h > 0$ is the time step of the discretization. The key requirement is that $\sum t_k = +\infty$ in order to capture the asymptotic behaviour of the continuous dynamics. This condition is clearly satisfied in the case of a fixed time step as described above. The time discretization of the dynamical system (ISIHD):

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t) + \beta(t)\dot{x}(t)) = 0, \qquad \beta(t) = \beta_0 \frac{\alpha + 1}{t},$$

Licence CC BY-NC-ND 3.0

is given by:

$$\frac{x_{k+1}-2x_k+x_{k-1}}{h^2}+\frac{\alpha}{kh}\cdot\frac{x_k-x_{k-1}}{h}+\nabla f\Big(z_k+\beta_0\frac{\alpha+1}{kh^2}(x_k-x_{k-1})\Big)=0,$$

where z_k is chosen according to the Nesterov extrapolation (a point lying on the line through x_{k-1} and x_k). After simplification and setting $\beta_k=\beta_0\frac{\alpha+1}{kh^2}$, one obtains:

$$x_{k+1}=x_k+\Big(1-\frac{\alpha}{k}\Big)(x_k-x_{k-1})-h^2\nabla f\big(z_k+\beta_k(x_k-x_{k-1})\big).$$

Equivalently,

$$\text{(I-NAG)}_\alpha\quad\begin{cases}z_k=x_k+\Big(1-\dfrac{\alpha}{k}\Big)(x_k-x_{k-1}) & \text{(Nesterov extrapolation)}\\ v_k=z_k+\beta_k(x_k-x_{k-1}) & \text{(Implicit point for the gradient)}\\ x_{k+1}=z_k-s\nabla f(v_k) & \text{(Update)}\end{cases}$$

with $s=h^2$ and initial iterates $x_0,x_1\in\mathcal{H}$. The algorithm is referred to as I–NAG, standing for *Implicit Hessian–driven Nesterov Accelerated Gradient*. The term $\beta_k(x_k-x_{k-1})$ introduces an additional extrapolation, shifting the gradient evaluation point in the direction of the current velocity. When $\beta_0=0$, one recovers exactly the classical scheme $(\text{NAG})_\alpha$. Thus, I–NAG can be seen as an extension of the classical Nesterov Accelerated Gradient (NAG) method, aiming to improve stability and convergence properties, particularly in complex optimization landscapes. The addition of the implicit term allows the algorithm to adapt more effectively to the curvature of the objective function, offering potential advantages in both theory and practice over traditional NAG.

Theorem 3.4.1 establishes a connection between the continuous-time dynamics (ISIHD) and the implicit gradient flow. Assuming $x_1=0$ and defining $y_\alpha(s)=x_\alpha(\sqrt{2(\alpha+1)s})$, it follows that

$$y_\alpha(s)\to y(s)$$

as $\alpha\to+\infty$, where $y(s)$ solves

$$\dot{y}(s)+\nabla f(y(s)+\beta_0\dot{y}(s))=0,\qquad y(0)=x_0. \tag{3.84}$$

This continuous-time convergence suggests that a similar relationship may hold for the discrete algorithms. For large α, the iterates $(x_k)_k$ generated by $(\text{I-NAG})_\alpha$ are expected to approximate those $(y_n)_n$ produced by a discretization of the implicit gradient flow (3.84). The correspondence between the indices k (accelerated algorithm) and n (gradient algorithm)

Licence CC BY-NC-ND 3.0

is given by the time relation:

$$t_k = \sqrt{2(\alpha+1)s_n} \quad \Rightarrow \quad kh \approx \sqrt{2(\alpha+1)\,(nh_{\text{grad}})}.$$

If one uses the same time step h for both discretizations ($s_n = nh$), this simplifies to:

$$n \approx \frac{k^2 h}{2(\alpha+1)}.$$

Thus, for large α, the following heuristic expression can be written:

$$x_k \sim y_n \quad \text{with} \quad n = \left\lfloor \frac{k^2 h}{2(\alpha+1)} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ denotes the integer part. This relation shows that the accelerated algorithm (I-NAG)$_\alpha$ performs, in terms of approximation of the solution, the equivalent of $\mathcal{O}(k^2)$ iterations of the implicit gradient method in only k iterations.
A study of this correspondence and the analysis of convergence rates for the discrete iterations constitute an interesting research programme, which lies beyond the scope of this chapter.

3.7 Numerical Experiments

This section aims to illustrate the theoretical results of the previous sections through a series of numerical experiments. For this purpose, adaptive Runge–Kutta integrators from MATLAB [29] are used to solve the continuous dynamics.

Example 3.7.1. A comparison of the (AVD), (DIN-AVD), and (ISIHD) methods is performed on the convex and non-quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$x = (x_1, x_2) \mapsto f(x_1, x_2) = ax_1^4 + bx_2^4,$$

with $a = 1$ and $b = 10000$.

Licence CC BY-NC-ND 3.0

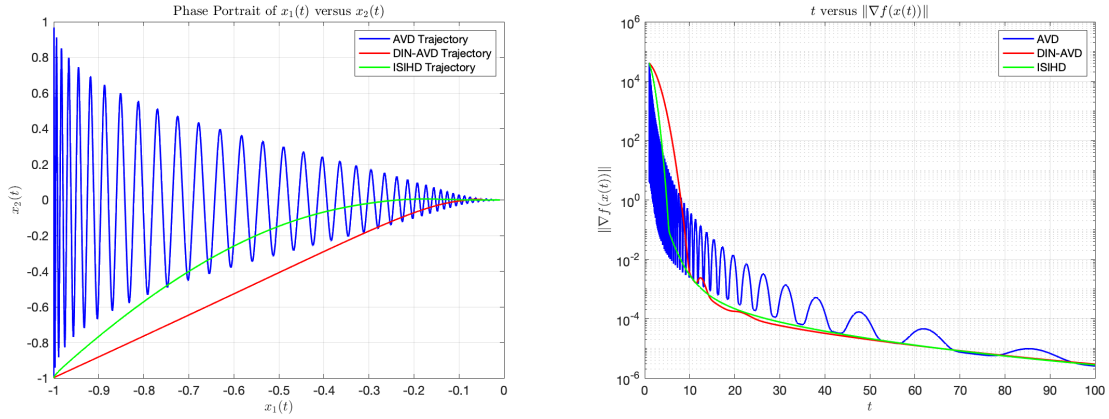


Figure 3.1: Left: Phase portrait of $x_1(t)$ as a function of $x_2(t)$ for (AVD), (DIN-AVD), and (ISIHD). Right: t as a function of $\|\nabla f(x(t))\|$ in semi-log scale for (AVD), (DIN-AVD), and (ISIHD).

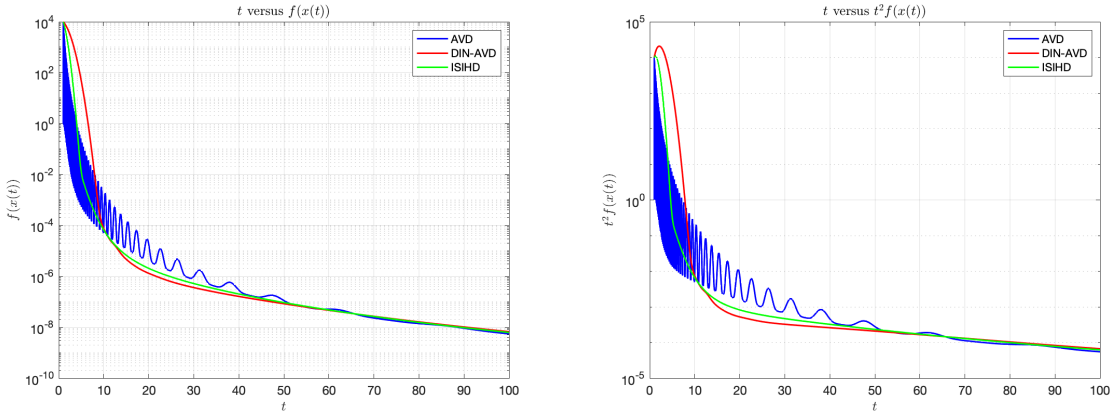


Figure 3.2: t as a function of $f(x(t))$ (left) and t as a function of $t^2 f(x(t))$ (right) in semi-log scale for (AVD), (DIN-AVD), and (ISIHD).

The phase portrait (Figure 3.1, left) shows that the (AVD) trajectory exhibits many oscillations. In contrast, the trajectories with Hessian damping ((DIN-AVD) and (ISIHD)) exhibit significantly fewer oscillations, demonstrating the effectiveness of this mechanism in stabilizing trajectories.

The semi-logarithmic plot of the gradient norm (Figure 3.1, right) confirms that the methods with Hessian damping ((DIN-AVD) and (ISIHD)) provide smoother and more stable convergence compared with (AVD).

The plots of the function values (Figure 3.2) show that the methods with Hessian damping lead to a more regular decay and reach lower values faster than (AVD), highlighting their improved convergence properties. Notably, (AVD) performs better in the time interval $t \in [80, 100]$ than (DIN-AVD) and (ISIHD) for this specific criterion.

Licence CC BY-NC-ND 3.0

Example 3.7.2. The methods are now compared on the convex, non-quadratic, and non-strongly convex function:

$$f(x, y) = x^4 + 5y^2 - 4x - 10y + 8.$$

The gradient of f is given by

$$\nabla f(x, y) = (4x^3 - 4, 10y - 10).$$

The unique minimizer is $x^* = (1, 1)$ and the minimal value is $\inf(f) = f(x^*) = 0$.

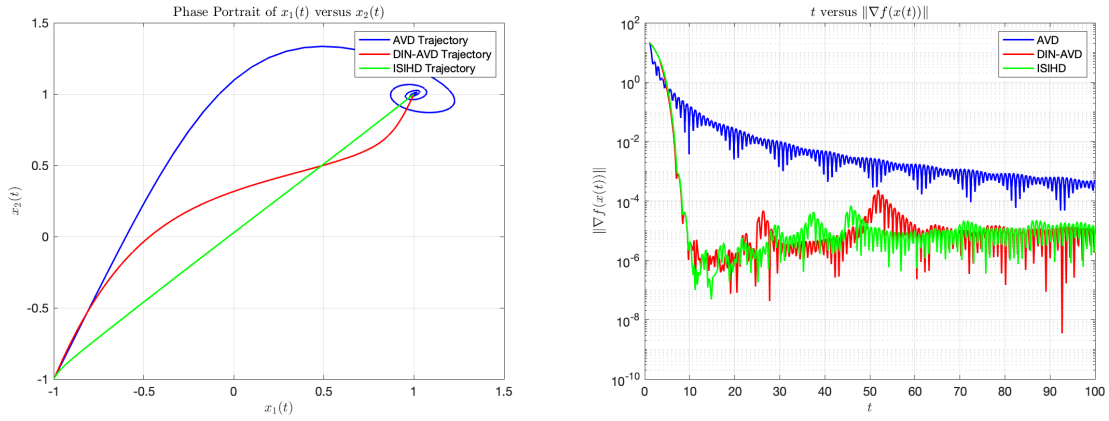


Figure 3.3: Left: Phase portrait for (AVD), (DIN-AVD), and (ISIHD). Right: t as a function of $\|\nabla f(x(t))\|$ in semi-logarithmic scale.

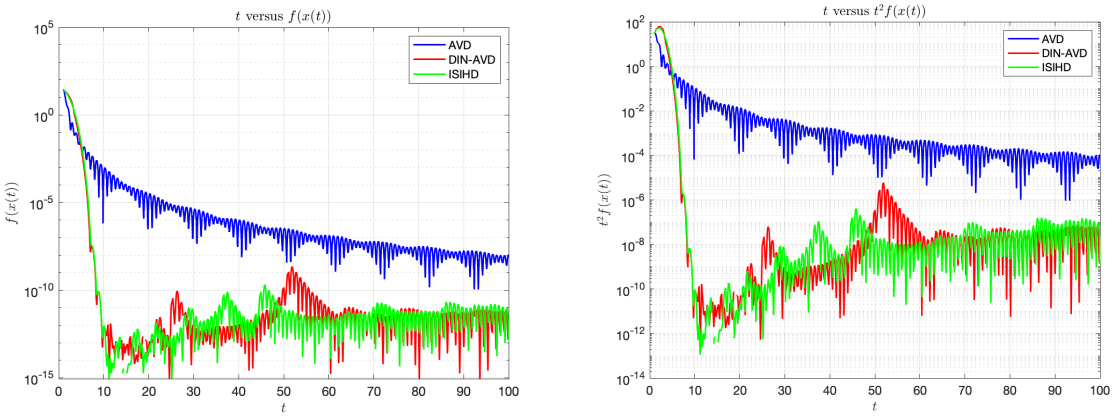


Figure 3.4: t as a function of $f(x(t))$ (left) and t as a function of $t^2 f(x(t))$ (right) in semi-logarithmic scale.

As in Example 3.7.1, Hessian damping significantly improves the convergence behaviour. The (DIN-AVD) and (ISIHD) methods display faster decay of the objective value and reduced oscillations compared with (AVD), confirming their robustness for convex non-quadratic functions.

Licence CC BY-NC-ND 3.0

Example 3.7.3. Consider the smooth convex function

$$f(x, y) = (x - 1)^4 + 2y^4 - 8y + 6,$$

which admits a unique global minimizer at $(x^*, y^*) = (1, 1)$.

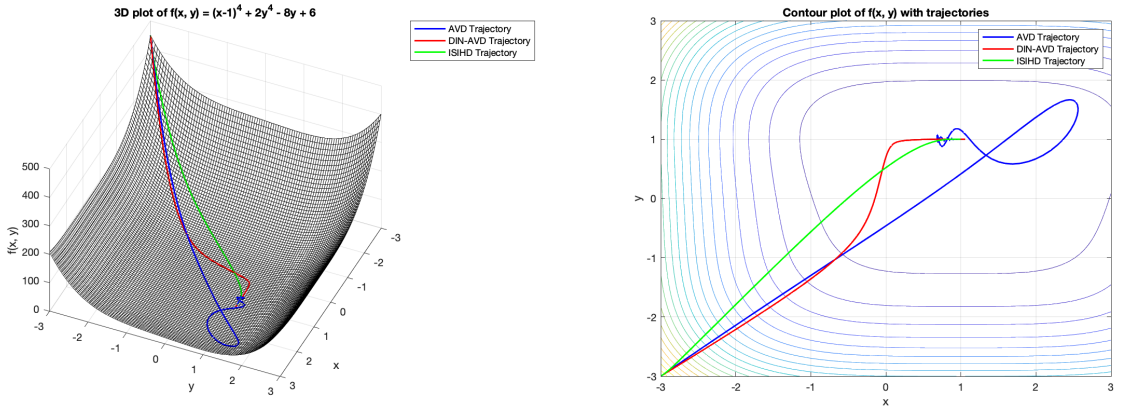


Figure 3.5: Left: 3D surface of $f(x, y)$ with the trajectories. Right: Contour plot of $f(x, y)$ with the trajectories of (AVD), (DIN-AVD), and (ISIHD).

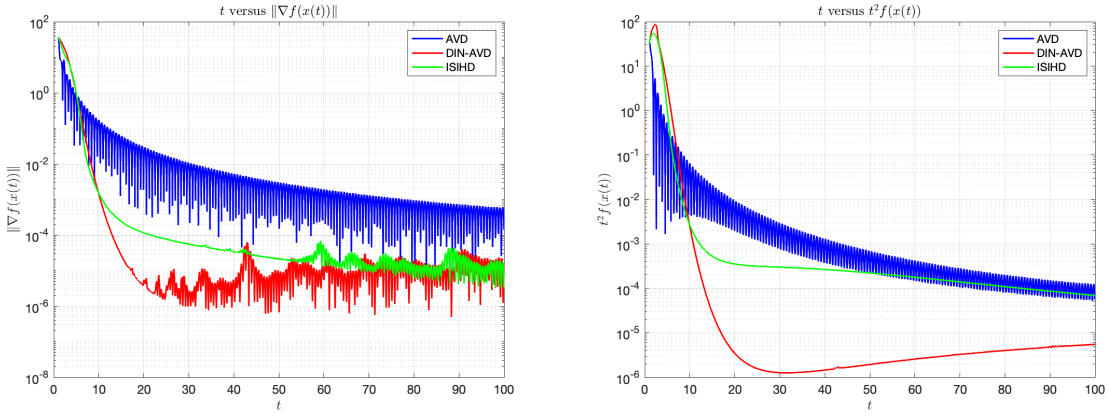


Figure 3.6: Left: $t \mapsto \|\nabla f(x(t))\|$. Right: $t \mapsto t^2 f(x(t))$. Hessian damping significantly reduces oscillations and improves convergence.

Figure 3.5 (left) displays the 3D surface of the function with the trajectories. The right-hand plot shows the contour plot. (ISIHD) exhibits faster convergence to the minimizer with significantly fewer oscillations. Figure 3.6 (left) shows the evolution of the gradient norm. (ISIHD) achieves a cleaner decay. The right-hand plot confirms accelerated convergence through the decay of the renormalized value $t^2 f(x(t))$. These results are aligned with the theoretical predictions.

Licence CC BY-NC-ND 3.0

Example 3.7.4. The convergence toward the limiting dynamics is illustrated here. The trajectories of the reparametrized system $y_\alpha(s)$ are plotted for various values of α ($\alpha = 3, 5, 8, 10, 15, 20, 25, 50, 100, 200$) and compared with the solution $y(s)$ of the limiting equation.

$$\dot{y}(s) + \nabla f\left(y(s) + \beta_0 \dot{y}(s)\right) = 0, \quad y(0) = x_0 = (-1, -1).$$

The test function is $f(x_1, x_2) = ax_1^4 + bx_2^4$ (with $a = 1$, $b = 10000$), from Example 3.7.1.

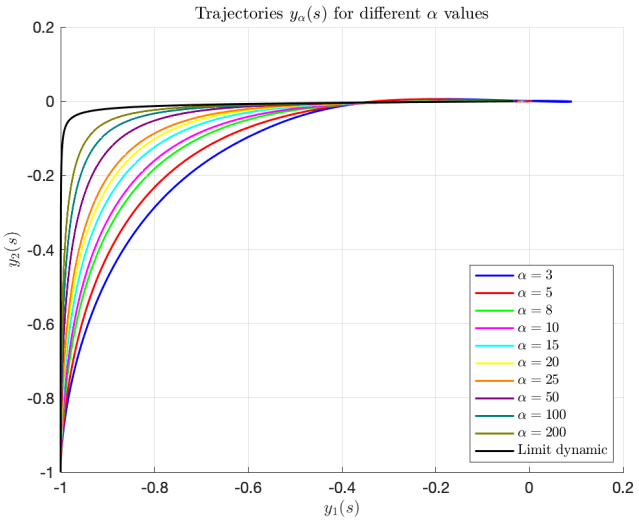


Figure 3.7: Trajectories $y_\alpha(s)$ for different values of α and the limiting dynamics (black).

Figure 3.7 confirms the theoretical asymptotic behaviour: as α increases, the trajectories converge to the limiting dynamics (in black).

Example 3.7.5. (Influence of the parameter β_0). The influence of the parameter β_0 on the trajectories of the reparametrized system is examined. The values $\alpha = 5$ and $\alpha = 100$ are fixed, while β_0 is varied from 0 to 5 (10 evenly spaced values). The test function is the ill-conditioned function $f(x_1, x_2) = x_1^4 + 10000x_2^4$.

The system being solved is:

$$\frac{2s}{\alpha + 1}\ddot{y}_\alpha(s) + \dot{y}_\alpha(s) + \nabla f\left(y_\alpha(s) + \beta_0 \dot{y}_\alpha(s)\right) = 0,$$

with initial conditions:

$$y_\alpha\left(s_{0,\alpha}\right) = (-1, -1), \quad \dot{y}_\alpha\left(s_{0,\alpha}\right) = (0, 0), \quad \text{where } s_{0,\alpha} = \frac{t_0^2}{2(\alpha + 1)}.$$

Licence CC BY-NC-ND 3.0

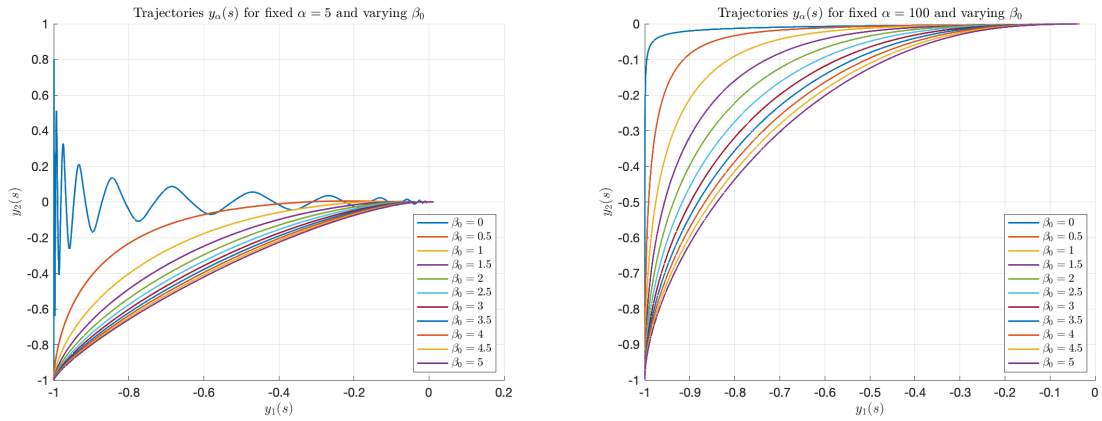


Figure 3.8: Phase portraits $y_1(s)$ vs $y_2(s)$ for varying β_0. Left: $\alpha = 5$. Right: $\alpha = 100$.

Figure 3.8 shows that increasing β_0 leads to smoother and more stable convergence paths. For small values of β_0, the trajectories oscillate more before stabilizing, whereas higher values damp these oscillations more rapidly and accelerate convergence.

3.8 Conclusion and Perspectives

This chapter investigated the asymptotic behaviour of inertial systems with vanishing viscous damping, with particular emphasis on the singular limit as the damping parameter $\alpha \to +\infty$. Two main approaches were employed.

The first approach consists in directly examining the limit of the trajectories without any time reparametrization. This analysis shows that, without a change of scale, the trajectory converges to the initial position when the parameter becomes very large.

The second approach relies on a suitable time reparametrization. Through this renormalization, it was shown that the accelerated inertial system with implicit Hessian damping (ISIHD) converges, after reparametrization, to a first-order flow described by an implicit gradient dynamics. The form of the limiting equation depends on the parameter β_0: for $\beta_0 = 0$, one recovers the classical gradient flow, whereas for $\beta_0 > 0$, an implicit gradient system emerges, incorporating the effect of the Hessian-driven damping. The analysis was first carried out in the convex setting under general assumptions, and then extended to the non-convex quadratic case in finite dimension, where the affine structure of the gradient ensures the well-posedness of the limit. The discrete algorithm I-NAG, inspired by the continuous model, is also introduced, with a formal correspondence established between its iterations and those of implicit gradient descent by a suitable time rescaling.

Several research perspectives naturally arise from this study. A first direction consists in relaxing the assumption $x_1 = 0$, which is essential in the implicit setting to obtain uniform bounds and compactness. Developing techniques to handle nonzero initial velocities would broaden

Licence CC BY-NC-ND 3.0

the applicability of the model. Another important question concerns the limiting behaviour of the implicit system for fully non-convex and non-quadratic functions, which would require new tools to deal with the lack of weak sequential continuity of the gradient. Finally, exploring the numerical properties and convergence guarantees of the I-NAG algorithm in practical applications represents a promising avenue for future work.

Licence CC BY-NC-ND 3.0

4

From the inertial system to first-order flow: Non-potential operators

Contents

4.1 Introduction **83**
4.1.1 Problematic and challenges 83
4.1.2 Approach 84
4.1.3 Main results 85
4.1.4 Chapter Contents 85
4.2 Limit Dynamics of (AVD): Non-Potential Operator Case **86**
4.3 Limiting Dynamics of (DIN-AVD): Non-Potential Operator Case **96**
4.4 Limit Dynamics of (ISIHD): Non-Potential Operator Case **104**
4.5 Numerical Experiment **116**
4.6 Conclusion and Perspectives **121**

Licence CC BY-NC-ND 3.0

4.1 Introduction

The previous chapter provided an in-depth exploration of inertial dynamical systems within the framework of convex optimization, analyzing the asymptotic properties of the dynamics (AVD), (DIN-AVD) and (ISIHD) under the assumption that the considered operator is derived from a potential, that is, ∇f with f a convex differentiable function on a Hilbert space $(\mathcal{H}; \langle\cdot,\cdot\rangle; \|\cdot\|)$. The results obtained highlighted the mechanisms behind accelerated convergence and rigorously established the passage from second-order systems to first-order flows as the damping parameter α tends to infinity.

However, this restriction to the potential framework proves limiting given the diversity of problems encountered in practice. Many important situations in variational analysis, optimal control theory, etc. involve operators that do not derive from a potential. This includes in particular variational inequalities of the form: find $x \in \mathcal{H}$ such that $\langle V(x), y - x\rangle \geq 0$ for all $y \in C$, where V is a monotone non-gradient operator and C a closed convex subset of $\mathcal{H}$, as well as fixed-point problems reformulated as zero-finding problems for monotone operators, projection-type operators in nonsmooth convex analysis, etc.

This chapter therefore aims to extend the scope of the analysis by considering an operator $V : \mathcal{H} \to \mathcal{H}$ *not necessarily potential*. This generalization allows us to address a much broader class of problems while preserving a sufficiently rich mathematical structure to permit a rigorous analysis, thereby opening the way to new applications in areas where non-potential operators are pervasive.

4.1.1 Problematic and challenges

The transition from the potential to the non-potential framework introduces substantial mathematical challenges. In the analysis presented in the previous chapter, the variational structure associated with the objective function f was exploited, using tools specific to the convex setting. In particular, an important relation employed was

$$\langle \nabla f(z(s)), \dot{z}(s)\rangle = \frac{d}{ds} f(z(s)) \tag{4.1}$$

valid for any differentiable function $z : \mathbb{R} \to \mathcal{H}$. This identity, a direct consequence of the chain rule, allowed us to relate in a simple way the inner product between the gradient and the velocity to the time derivative of the objective function.

In the present non-potential framework, this approach is no longer available because there is no function f such that $V = \nabla f$. The structural relation that greatly simplified the energy analysis is therefore lost. New strategies must now be developed to analyze the stability and convergence of the dynamical systems, working directly with the intrinsic properties of the operator V without relying on the existence of a potential.

Licence CC BY-NC-ND 3.0

This is where the notion of *cocoercivity* plays an important role in the analysis. Although it was defined in Chapter 2, let us recall that an operator $V : \mathcal{H} \to \mathcal{H}$ is said to be λ-cocoercive if it satisfies

$$\langle V(y) - V(x), y - x\rangle \geq \lambda \|V(y) - V(x)\|^2 \quad \forall x, y \in \mathcal{H}.$$

This property, although stronger than simple monotonicity, presents two decisive advantages for the study. First, it implies the monotonicity of V, which ensures the stability of the associated dynamical systems. Second, it guarantees uniform Lipschitz regularity, since Cauchy–Schwarz inequality combined with cocoercivity shows that V is $\frac{1}{\lambda}$-Lipschitz (Remark 2.4.1). These two aspects (stability and regularity) are essential for controlling the nonlinear terms in the differential equations and obtaining estimates uniform in α. Cocoercivity therefore constitutes the convenient assumption for extending the analysis of the previous chapter to the non-potential case, compensating for the absence of variational structure. It will allow us to construct appropriate Lyapunov functions and establish energy estimates that do not rely on relation (4.1) valid in the potential setting. However, although the results of this chapter are presented in the setting of cocoercive operators, they remain valid for any monotone and Lipschitz operator. Indeed, cocoercivity is not required; only monotonicity and Lipschitz continuity of the operator are necessary. Nevertheless, the results will be presented in the specific framework of cocoercive operators.

4.1.2 Approach

The methodological approach relies on the well-established technique of *time rescaling*, already identified as an essential tool in the analysis of the previous chapter. The change of variable

$$t = \tau(s) = \sqrt{2(\alpha + 1)s}$$

is not introduced arbitrarily: it answers to a deep mathematical structure, examined in detail in Chapter 3. It is the element that allows one to extract the limiting dynamics when $\alpha \to +\infty$. The absence of potential structure, however, makes the analysis considerably more complex. It requires the construction of new energy estimates tailored to the non-potential framework, as well as the use of finer control techniques.

Licence CC BY-NC-ND 3.0

The following three dynamics are studied successively:

$$\text{(AVD)}: \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + V(x(t)) = 0 \tag{4.2}$$

$$\text{(DIN-AVD)}: \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + V(x(t)) + \beta(t)\frac{d}{dt}V(x(t)) = 0 \tag{4.3}$$

$$\text{(ISIHD)}: \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + V(x(t) + \beta(t)\dot{x}(t)) = 0 \tag{4.4}$$

which generalize the systems of Chapter 3 to the non-potential context. The analysis relies on the following assumptions:

- $V : \mathcal{H} \to \mathcal{H}$ is λ-cocoercive for some $\lambda > 0$;
- the set of zeros $S = \{x^* \in \mathcal{H} : V(x^*) = 0\}$ is nonempty;
- $\beta(t) = \beta_0 \frac{\alpha+1}{t}$, with $\beta_0 > 0$ and $\alpha > 0$ fixed;
- x_0 and x_1 denote the initial position and velocity at the initial time $t_0 > 0$.

4.1.3 Main results

This chapter shows the convergence of the three non-potential inertial dynamics toward first-order flows, under precise conditions on the parameters:

- (AVD) converges to the classical monotone flow:

$$\dot{y}(s) + V(y(s)) = 0,\ y(0) = x_0;$$

- (DIN-AVD) converges, in finite dimension and under the assumption $\lambda > 2\beta_0$, to a Levenberg–Marquardt-type equation:

$$\dot{y}(s) + V(y(s)) + \beta_0 \frac{d}{ds}V(y(s)) = 0,\ y(0) = x_0;$$

- (ISIHD) converges, when $x_1 = 0$ and $2\beta_0 < \lambda$, to the implicit monotone flow:

$$\dot{y}(s) + V(y(s) + \beta_0\dot{y}(s)) = 0,\ y(0) = x_0.$$

These results substantially extend those of the previous chapter to the non-potential framework, while revealing new phenomena: critical parameter conditions, the need for additional structural assumptions, and strong convergence of derivatives in L^2 in the implicit case.

4.1.4 Chapter Contents

The chapter is structured as follows. Section 4.2 is focused on the study of the limiting dynamics for equation (AVD) in the case of a non-potential operator. Energy estimates

Licence CC BY-NC-ND 3.0

are developed, and the uniform convergence of the rescaled trajectories to the solution of a first-order differential equation is proved. Section 4.3 extends these results to the dynamics (DIN-AVD) in finite dimension. Under the strengthened cocoercivity condition ($\lambda > 2\beta_0$), convergence is established, and the asymptotic behavior of the system is studied. Section 4.4 addresses the implicit dynamics (ISIHD). The uniform convergence of the trajectories and the strong convergence of the derivatives in L^2 are proved under the additional assumptions that the initial velocity is zero ($x_1 = 0$) and that $2\beta_0 < \lambda$. Section 4.5 presents numerical experiments illustrating the theoretical results and highlighting qualitative differences between the inertial second-order system and its first-order limit. Finally, a brief discussion of open questions and potential research directions concludes the chapter.

This progressive organization makes it possible to address each dynamics with the appropriate level of complexity, from the simplest case (AVD) to the most technical one (ISIHD), while maintaining methodological coherence throughout the analysis.

4.2 Limit Dynamics of (AVD): Non-Potential Operator Case

This section investigates the asymptotic behavior of the inertial system (AVD) in the case of non-potential operators. Consideration is given to an operator V defined on a Hilbert space $\mathcal{H}$, assumed to be λ-cocoercive, and the time evolution governed by a second-order differential equation with vanishing damping is studied. The main objective is to understand the behavior of the trajectories $x_\alpha(t)$ as the damping parameter α becomes very large. To study this question, the time transformation is introduced

$$t = \sqrt{2s(\alpha + 1)}$$

allowing us to reparametrize the dynamical system. This change of variable, far from being merely formal, has a clear mathematical motivation: it compensates for the effect of the vanishing damping term $\frac{\alpha}{t}$ by adapting the time scale. The main result of this section, stated in Theorem 4.2.1 (which also holds for monotone and Lipschitz operators: cocoercivity is not required), shows that, under this new scaling, the reparametrized trajectories converge uniformly toward the solution of a standard monotone flow:

$$\dot{y}(s) + V(y(s)) = 0, \qquad y(0) = x_0.$$

This convergence highlights an essential link between accelerated heavy-ball-type dynamical systems with friction and classical monotone flows, thus extending to the non-potential setting well-established results in convex optimization.

Licence CC BY-NC-ND 3.0

Theorem 4.2.1 (Convergence of the (AVD) dynamics)**.** Let $\mathcal{H}$ be a real Hilbert space. Consider an operator $V : \mathcal{H} \to \mathcal{H}$ that is λ-cocoercive and admits at least one zero $x^* \in \mathcal{H}$. Consider the vanishing dynamics:

$$\ddot{x}_\alpha(t) + \frac{\alpha}{t}\dot{x}_\alpha(t) + V(x_\alpha(t)) = 0, \quad x_\alpha(t_0) = x_0, \quad \dot{x}_\alpha(t_0) = x_1, \tag{4.5}$$

with $\alpha > 0$ and $t_0 > 0$. Setting $t = \sqrt{2(\alpha+1)s}$, define $y_\alpha(s) = x_\alpha(\sqrt{2(\alpha+1)s})$.

1. The reparametrized trajectory y_α satisfies the differential equation:

$$\frac{2s}{\alpha+1}\ddot{y}_\alpha(s) + \dot{y}_\alpha(s) + V(y_\alpha(s)) = 0, \ \forall s \geq s_{0,\alpha} \tag{4.6}$$

   with initial conditions:

$$y_\alpha(s_{0,\alpha}) = x_0, \quad \dot{y}_\alpha(s_{0,\alpha}) = \frac{\alpha+1}{t_0}x_1, \quad \text{where } s_{0,\alpha} = \frac{t_0^2}{2(\alpha+1)}.$$

2. By extending y_α on $[0, +\infty)$ by $\tilde{y}_\alpha(s) = x_0$ for $s \in [0, s_{0,\alpha}]$ and $\tilde{y}_\alpha(s) = y_\alpha(s)$ for $s \geq s_{0,\alpha}$, the family $(\tilde{y}_\alpha)_{\alpha \geq 3}$ converges uniformly on every compact subset of $[0, +\infty)$ toward the unique solution y of:

$$\dot{y}(s) + V(y(s)) = 0, \quad y(0) = x_0. \tag{4.7}$$

The proof of this result is based on two estimates that ensure uniform control of the trajectories. These preparatory lemmas form the supports of the convergence analysis.

Lemma 4.2.2 (Stability of the limit solution)**.** Equation (4.7) admits a unique solution, and this solution y satisfies the following stability estimate:

$$\|y(s) - x^*\| \leq \|x_0 - x^*\| \qquad \text{for all } s \geq 0. \tag{4.8}$$

Proof. Since V is λ-cocoercive, it is in particular Lipschitz continuous with constant $\frac{1}{\lambda}$ (Remark 2.4.1). The right-hand side of $\dot{y}(s) = -V(y(s))$ is thus Lipschitz continuous, and the classical Cauchy–Lipschitz theorem for ordinary differential equations in Hilbert spaces ensures that problem (4.7) admits a unique global solution $y \in C^1([0, +\infty); \mathcal{H})$.

For the stability estimate, the proof is based on a classical Lyapunov function. Consider

$$u(s) = \frac{1}{2}\|y(s) - x^*\|^2.$$

Its derivative is

$$\dot{u}(s) = \langle \dot{y}(s), y(s) - x^* \rangle = -\langle V(y(s)), y(s) - x^* \rangle,$$

Licence CC BY-NC-ND 3.0

where the last equality follows from (4.7). Using the cocoercivity (or monotonicity) of V and the fact that $V(x^*) = 0$, we obtain

$$\dot{u}(s) = -\langle V(y(s)) - V(x^*), y(s) - x^* \rangle \leq -\lambda \|V(y(s)) - V(x^*)\|^2 \leq 0.$$

Thus u is decreasing, which directly yields the stability inequality (4.8). □

Once the stability properties for the limiting system have been established, uniform estimates for the approximate trajectories y_α are obtained. The following lemma provides control of their distance to a fixed point x^*, which is required for the proof of convergence stated in Theorem 4.2.1.

Lemma 4.2.3 (Uniform control of the approximated trajectories)**.** For all $s \geq s_{0,\alpha}$, the trajectory y_α satisfies the estimate:

$$\frac{1}{2}\|y_\alpha(s) - x^*\|^2 \leq C_2(s), \tag{4.9}$$

where

$$C_2(s) = \left(\frac{1}{2}\|x_1\|^2 + \frac{1}{2}\|x_0 - x^*\|^2\right) e^{(1+\frac{1}{\lambda})^2 s}.$$

Proof. Introduce the following modified energy function:

$$E_\alpha(s) = \frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{1}{2}\|y_\alpha(s) - x^*\|^2. \tag{4.10}$$

Evaluating its derivative:

$$\dot{E}_\alpha(s) = \frac{1}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{2s}{\alpha+1}\langle \ddot{y}_\alpha(s), \dot{y}_\alpha(s)\rangle + \langle y_\alpha(s) - x^*, \dot{y}_\alpha(s)\rangle.$$

The dynamics (4.6) gives:

$$\frac{2s}{\alpha+1}\ddot{y}_\alpha(s) = -\dot{y}_\alpha(s) - V(y_\alpha(s)),$$

which leads to

$$\dot{E}_\alpha(s) = \frac{1}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \langle -\dot{y}_\alpha(s) - V(y_\alpha(s)), \dot{y}_\alpha(s)\rangle + \langle y_\alpha(s) - x^*, \dot{y}_\alpha(s)\rangle.$$

After rearranging the terms:

$$\dot{E}_\alpha(s) = -\left(1 - \frac{1}{\alpha+1}\right)\|\dot{y}_\alpha(s)\|^2 + \langle y_\alpha(s) - x^* - V(y_\alpha(s)), \dot{y}_\alpha(s)\rangle.$$

Licence CC BY-NC-ND 3.0

Now estimate the scalar product term. The Lipschitz continuity of V and $V(x^*) = 0$ imply:

$$\|V(y_\alpha(s))\| \leq \frac{1}{\lambda}\|y_\alpha(s) - x^*\|.$$

Thus,

$$\|y_\alpha(s) - x^* - V(y_\alpha(s))\| \leq \|y_\alpha(s) - x^*\| + \|V(y_\alpha(s))\| \leq \left(1 + \frac{1}{\lambda}\right)\|y_\alpha(s) - x^*\|.$$

Applying Cauchy–Schwarz and Young's inequality:

$$\begin{aligned}|\langle y_\alpha(s) - x^* - V(y_\alpha(s)), \dot{y}_\alpha(s)\rangle| &\leq \left(1 + \frac{1}{\lambda}\right)\|y_\alpha(s) - x^*\|\,\|\dot{y}_\alpha(s)\| \\ &\leq \frac{(1+\frac{1}{\lambda})^2}{2}\|y_\alpha(s) - x^*\|^2 + \frac{1}{2}\|\dot{y}_\alpha(s)\|^2.\end{aligned}$$

Hence,

$$\dot{E}_\alpha(s) \leq -\left(1 - \frac{1}{\alpha+1}\right)\|\dot{y}_\alpha(s)\|^2 + \frac{(1+\frac{1}{\lambda})^2}{2}\|y_\alpha(s) - x^*\|^2 + \frac{1}{2}\|\dot{y}_\alpha(s)\|^2.$$

Simplifying:

$$\dot{E}_\alpha(s) \leq -\left(\frac{1}{2} - \frac{1}{\alpha+1}\right)\|\dot{y}_\alpha(s)\|^2 + \frac{(1+\frac{1}{\lambda})^2}{2}\|y_\alpha(s) - x^*\|^2.$$

Since for $\alpha \geq 3$, the quantity $\frac{1}{2} - \frac{1}{\alpha+1}$ is positive:

$$\dot{E}_\alpha(s) \leq \frac{(1+\frac{1}{\lambda})^2}{2}\|y_\alpha(s) - x^*\|^2.$$

Moreover, from the definition of $E_\alpha(s)$,

$$\frac{1}{2}\|y_\alpha(s) - x^*\|^2 \leq E_\alpha(s),$$

hence

$$\dot{E}_\alpha(s) \leq \left(1 + \frac{1}{\lambda}\right)^2 E_\alpha(s).$$

Gronwall's lemma yields, for all $s \geq s_{0,\alpha}$:

$$E_\alpha(s) \leq E_\alpha(s_{0,\alpha})e^{(1+\frac{1}{\lambda})^2(s-s_{0,\alpha})}.$$

Licence CC BY-NC-ND 3.0

Compute the initial value:

$$\begin{aligned} E_\alpha(s_{0,\alpha}) &= \frac{s_{0,\alpha}}{\alpha+1}\|\dot{y}_\alpha(s_{0,\alpha})\|^2 + \frac{1}{2}\|x_0 - x^*\|^2 \\ &= \frac{1}{2}\|x_1\|^2 + \frac{1}{2}\|x_0 - x^*\|^2, \end{aligned}$$

since $\frac{s_{0,\alpha}}{\alpha+1}\|\dot{y}_\alpha(s_{0,\alpha})\|^2 = \frac{1}{2}\|x_1\|^2$. Thus $E_\alpha(s_{0,\alpha})$ is bounded independently of α. Therefore,

$$E_\alpha(s) \leq C_2(s). \tag{4.11}$$

Combining (4.10) and (4.11),

$$\frac{1}{2}\|y_\alpha(s) - x^*\|^2 \leq E_\alpha(s) \leq C_2(s),$$

which completes the proof. □

Taking into account the established stability results, the proof of Theorem 4.2.1 can be undertaken. The method consists of directly comparing the trajectories $\tilde{y}_\alpha$ with the limit solution y, based on the co-coercivity of V and the previously established energy bounds.

Proof of Theorem 4.2.1. The main objective is to establish an upper bound for $\|\tilde{y}_\alpha(s) - y(s)\|$ by a quantity as a function of α that tends to 0 as $\alpha \to +\infty$, uniformly with respect to $s \geq 0$.

Analysis on the interval $[0, s_{0,\alpha}]$: Let $s \in [0, s_{0,\alpha}]$. By the definition of $\tilde{y}_\alpha$,

$$\|\tilde{y}_\alpha(s) - y(s)\| = \|x_0 - y(s)\| = \|y(s) - y(0)\| = \left\|\int_0^s \dot{y}(\tau)\, d\tau\right\| \leq \int_0^{s_{0,\alpha}} \|\dot{y}(\tau)\|\, d\tau.$$

From equation (4.7) satisfied by y, it follows that:

$$\int_0^{s_{0,\alpha}} \|\dot{y}(\tau)\|\, d\tau = \int_0^{s_{0,\alpha}} \|V(y(\tau))\|\, d\tau \leq \frac{1}{\lambda}\int_0^{s_{0,\alpha}} \|y(\tau) - x^*\|\, d\tau,$$

using that $V(x^*) = 0$ and the Lipschitz property of V with constant $\frac{1}{\lambda}$. Combining this with Lemma 4.2.2 yields:

$$\|\tilde{y}_\alpha(s) - y(s)\| \leq \frac{s_{0,\alpha}}{\lambda}\|x_0 - x^*\| =: C_1(\alpha) \quad \text{for all} \quad s \in [0, s_{0,\alpha}]. \tag{4.12}$$

A particular case is for $s = s_{0,\alpha}$:

$$\|\tilde{y}_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\| = \|x_0 - y(s_{0,\alpha})\| \leq C_1(\alpha). \tag{4.13}$$

Licence CC BY-NC-ND 3.0

Note that

$$C_1(\alpha) \to 0 \quad \text{as} \quad \alpha \to +\infty.$$

Analysis on $[s_{0,\alpha}, +\infty)$: Now consider the inner product of the equation (4.6) with the velocity $\dot{y}_\alpha$:

$$\left\langle \frac{2s}{\alpha+1}\ddot{y}_\alpha(s) + \dot{y}_\alpha(s) + V(y_\alpha(s)), \dot{y}_\alpha(s) \right\rangle = 0.$$

This can be rewritten as:

$$\frac{s}{\alpha+1}\frac{d}{ds}\|\dot{y}_\alpha(s)\|^2 + \|\dot{y}_\alpha(s)\|^2 + \langle V(y_\alpha(s)), \dot{y}_\alpha(s)\rangle = 0.$$

Integrating this relation over $[s_{0,\alpha}, s]$,

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{\alpha}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha\|^2 d\tau = \frac{\|x_1\|^2}{2} - \int_{s_{0,\alpha}}^{s} \langle V(y_\alpha), \dot{y}_\alpha\rangle \, d\tau. \tag{4.14}$$

An upper bound for the integral term is obtained using the Cauchy-Schwarz inequality followed by Young's inequality:

$$\begin{aligned}
\left|\int_{s_{0,\alpha}}^{s} \langle V(y_\alpha), \dot{y}_\alpha\rangle \, d\tau\right| &\le \int_{s_{0,\alpha}}^{s} |\langle V(y_\alpha), \dot{y}_\alpha\rangle| \, d\tau \\
&\le \int_{s_{0,\alpha}}^{s} \|V(y_\alpha)\| \cdot \|\dot{y}_\alpha\| \, d\tau \\
&\le \frac{1}{\lambda}\int_{s_{0,\alpha}}^{s} \|y_\alpha - x^*\| \cdot \|\dot{y}_\alpha\| \, d\tau \\
&\le \frac{1}{2\lambda^2}\int_{s_{0,\alpha}}^{s} \|y_\alpha - x^*\|^2 \, d\tau + \frac{1}{2}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha\|^2 \, d\tau \\
&\le \frac{sC_2(s)}{\lambda^2} + \frac{1}{2}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha\|^2 d\tau
\end{aligned} \tag{4.15}$$

where the last inequality follows from Lemma 4.2.3. From (4.14) and (4.15), it follows that:

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{\alpha-1}{2(\alpha+1)}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha\|^2 \, d\tau \le \frac{\|x_1\|^2}{2} + \frac{sC_2(s)}{\lambda^2}. \tag{4.16}$$

Note that $\frac{\alpha-1}{2(\alpha+1)} \ge \frac{1}{4}$ for $\alpha \ge 3$. Hence:

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{1}{4}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha\|^2 \, d\tau \le \frac{\|x_1\|^2}{2} + \frac{sC_2(s)}{\lambda^2} =: C_3(s). \tag{4.17}$$

Licence CC BY-NC-ND 3.0

This leads to two estimates:

$$\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha\|^2 \, d\tau \leq 4C_3(s) \tag{4.18}$$

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 \leq C_3(s) \tag{4.19}$$

With these bounds, the asymptotic behavior of the sequence $(\tilde{y}_\alpha)_\alpha$ with respect to the solution y of the limit system can now be analyzed. The cocoercivity (or monotonicity) of V implies

$$\langle V(y_\alpha(s)) - V(y(s)),\, y_\alpha(s) - y(s)\rangle \geq \lambda \|V(y_\alpha(s)) - V(y(s))\|^2 \geq 0.$$

Hence, for all $s \geq s_{0,\alpha}$,

$$\left\langle -\frac{2s}{\alpha+1}\ddot{y}_\alpha(s) - \dot{y}_\alpha(s) + \dot{y}(s),\, y_\alpha(s) - y(s) \right\rangle \geq 0.$$

Equivalently,

$$\left\langle -\frac{2s}{\alpha+1}\ddot{y}_\alpha(s), y_\alpha(s) - y(s) \right\rangle \geq \frac{1}{2}\frac{d}{ds}(\|y_\alpha(s) - y(s)\|^2).$$

Integrating over $[s_{0,\alpha}, s]$ gives

$$\frac{1}{2}\|y_\alpha(s) - y(s)\|^2 - \frac{1}{2}\|y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\|^2 + \frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s} \langle \ddot{y}_\alpha,\, \tau(y_\alpha - y)\rangle \, d\tau \leq 0,$$

which implies

$$\frac{1}{2}\|y_\alpha(s) - y(s)\|^2 \leq \frac{1}{2}\|x_0 - y(s_{0,\alpha})\|^2 + \frac{2}{\alpha+1}\left|\int_{s_{0,\alpha}}^{s} \langle \ddot{y}_\alpha,\, \tau(y_\alpha - y)\rangle \, d\tau\right|. \tag{4.20}$$

Integration by parts yields:

$$\int_{s_{0,\alpha}}^{s} \langle \ddot{y}_\alpha(\tau),\, \tau(y_\alpha(\tau) - y(\tau))\rangle \, d\tau = \underbrace{\{\langle \dot{y}_\alpha(\tau),\, \tau(y_\alpha(\tau) - y(\tau))\rangle\}_{s_{0,\alpha}}^{s}}_{B(s,\alpha)} - \underbrace{\int_{s_{0,\alpha}}^{s} \langle \dot{y}_\alpha(\tau),\, (y_\alpha(\tau) - y(\tau)) + \tau(\dot{y}_\alpha(\tau) - \dot{y}(\tau))\rangle \, d\tau}_{A(s,\alpha)}.$$

*Control of the terms:*Each term in this decomposition is estimated individually. The first term, $A(s,\alpha)$, is bounded using norm properties and standard inequalities:

$$|A(s,\alpha)| \leq \int_{s_{0,\alpha}}^{s} |\langle \dot{y}_\alpha(\tau),\, y_\alpha(\tau) - y(\tau)\rangle| \, d\tau + \int_{s_{0,\alpha}}^{s} \tau |\langle \dot{y}_\alpha(\tau),\, \dot{y}_\alpha(\tau) - \dot{y}(\tau)\rangle| \, d\tau. \tag{4.21}$$

Licence CC BY-NC-ND 3.0

For the second term, an expansion is performed and appropriate inequalities are applied:

$$\begin{aligned} 2|\langle \dot{y}_\alpha(\tau), \dot{y}_\alpha(\tau) - \dot{y}(\tau)\rangle| &= 2|\|\dot{y}_\alpha(\tau)\|^2 - \langle \dot{y}_\alpha(\tau), \dot{y}(\tau)\rangle| \\ &\leq 2\|\dot{y}_\alpha(\tau)\|^2 + 2\|\dot{y}_\alpha(\tau)\| \cdot \|\dot{y}(\tau)\| \\ &\leq 2\|\dot{y}_\alpha(\tau)\|^2 + \|\dot{y}_\alpha(\tau)\|^2 + \|\dot{y}(\tau)\|^2 = 3\|\dot{y}_\alpha(\tau)\|^2 + \|\dot{y}(\tau)\|^2 \end{aligned} \tag{4.22}$$

From (4.21) and (4.22), one obtains:

$$|A(s,\alpha)| \leq \int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|\|y_\alpha(\tau) - y(\tau)\|\, d\tau + s\int_{s_{0,\alpha}}^{s} \left(\frac{3}{2}\|\dot{y}_\alpha(\tau)\|^2 + \frac{1}{2}\|\dot{y}(\tau)\|^2\right) d\tau.$$

By (4.18),

$$\frac{3}{2}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2\, d\tau \leq 6C_3(s).$$

Furthermore, (4.8) (Lemma 4.2.2) together with the Lipschitz property of V gives

$$\int_{s_{0,\alpha}}^{s} \|\dot{y}(\tau)\|^2\, d\tau = \int_{s_{0,\alpha}}^{s} \|V(y(\tau))\|^2\, d\tau \leq \frac{s}{\lambda^2}\|x_0 - x^*\|^2.$$

Hence,

$$|A(s,\alpha)| \leq \int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|\|y_\alpha(\tau) - y(\tau)\| d\tau + 6sC_3(s) + \frac{s^2}{\lambda^2}\|x_0 - x^*\|^2. \tag{4.23}$$

The boundary terms are next bounded. First,

$$\begin{aligned} \frac{2}{\alpha+1}|\langle \dot{y}_\alpha(s_{0,\alpha}), s_{0,\alpha}(y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha}))\rangle| &\leq \frac{2}{\alpha+1} s_{0,\alpha}\|\dot{y}_\alpha(s_{0,\alpha})\| \cdot \|y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\| \\ &= \frac{2s_{0,\alpha}}{\alpha+1}\left\|\frac{\alpha+1}{t_0}x_1\right\| \cdot \|x_0 - y(s_{0,\alpha})\| \\ &\leq \frac{t_0\|x_1\|}{(\alpha+1)}C_1(\alpha). \end{aligned} \tag{4.24}$$

On the other hand, by Cauchy-Schwarz,

$$|\langle \dot{y}_\alpha(s), s(y_\alpha(s) - y(s))\rangle| \leq s\|\dot{y}_\alpha(s)\| \cdot \|y_\alpha(s) - y(s)\|. \tag{4.25}$$

Combining (4.19) and (4.25) yields for all $s \geq s_{0,\alpha}$:

$$\frac{2}{\alpha+1}|\langle \dot{y}_\alpha(s), s(y_\alpha(s) - y(s))\rangle| \leq 2\sqrt{\frac{s}{\alpha+1}}C_3^{\frac{1}{2}}(s) \cdot \|y_\alpha(s) - y(s)\|. \tag{4.26}$$

Licence CC BY-NC-ND 3.0

Applying the elementary inequality $2ab \leq a^2 + b^2$ with

$$a = \frac{1}{2}\|y_\alpha(s) - y(s)\| \quad \text{and} \quad b = 2\sqrt{\frac{s}{\alpha+1}C_3(s)}$$

gives:

$$\frac{2}{\alpha+1}|\langle \dot{y}_\alpha(s), s(y_\alpha(s) - y(s))\rangle| \leq \frac{1}{4}\|y_\alpha(s) - y(s)\|^2 + \frac{4s}{\alpha+1}C_3(s). \tag{4.27}$$

From (4.24) and (4.27), it follows that:

$$\frac{2}{\alpha+1}|B(s,\alpha)| \leq \frac{t_0\|x_1\|}{(\alpha+1)}C_1(\alpha) + \frac{1}{4}\|y_\alpha(s) - y(s)\|^2 + \frac{4s}{\alpha+1}C_3(s). \tag{4.28}$$

Combining estimates (4.13), (4.23) and (4.28), inequality (4.20) becomes:

$$\frac{1}{4}\|y_\alpha(s) - y(s)\|^2 \leq \frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \cdot \|y_\alpha(\tau) - y(\tau)\|\, d\tau + C(\alpha,s) + \frac{C_1^2(\alpha)}{2} \tag{4.29}$$

with

$$C(\alpha,s) = \frac{t_0\|x_1\|}{(\alpha+1)}C_1(\alpha) + \frac{4s}{\alpha+1}C_3(s) + \frac{12s}{\alpha+1}C_3(s) + \frac{2s^2}{(\alpha+1)\lambda^2}\|x_0 - x^*\|^2.$$

Finally,

$$\frac{1}{2}\|y_\alpha(s) - y(s)\|^2 \leq \frac{4}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \cdot \|y_\alpha(\tau) - y(\tau)\|\, d\tau + 2C(\alpha,s) + C_1^2(\alpha). \tag{4.30}$$

Introducing the equilibrium point x^* and using estimates (4.8), (4.9) and (4.18)

$$\begin{aligned}
\frac{1}{2}\|y_\alpha(s) - y(s)\|^2 &\leq \frac{4}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \cdot (\|y_\alpha(\tau) - x^*\| + \|x^* - y(\tau)\|)\, d\tau \\
&\quad + 2C(\alpha,s) + C_1^2(\alpha) \\
&\leq \frac{4}{\alpha+1}(\sqrt{2C_2(s)} + \|x_0 - x^*\|)\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| d\tau + 2C(\alpha,s) + C_1^2(\alpha) \\
&\leq \frac{8\sqrt{sC_3(s)}}{\alpha+1}(\sqrt{2C_2(s)} + \|x_0 - x^*\|) + 2C(\alpha,s) + C_1^2(\alpha)
\end{aligned}$$

where the last inequality follows from Hölder's inequality and (4.18).
Thus

$$\|y_\alpha(s) - y(s)\| \leq \frac{2K(s,\alpha)}{\sqrt{\alpha+1}} \tag{4.31}$$

Licence CC BY-NC-ND 3.0

where

$$K(s,\alpha) = \Big[2t_0\|x_1\|C_1(\alpha) + 32C_3(s) + \frac{4s^2\|x_0 - x^*\|^2}{\lambda^2} \\ +8\sqrt{sC_3(s)}\left(\sqrt{2C_2(s)} + \|x_0 - x^*\|\right)\Big]^{1/2} + \frac{t_0^2\|x_0 - x^*\|}{2\lambda\sqrt{\alpha+1}}.$$

Inequalities (4.12) and (4.31) imply that estimate (4.31) holds for all $s \geq 0$. Moreover, the right-hand side of (4.31) is uniformly bounded on every compact subset of $[0, +\infty)$ and tends to 0 as $\alpha \to +\infty$. This shows the uniform convergence of the sequence $(\tilde{y}_\alpha)_\alpha$ to y on every compact subset of $[0, +\infty)$. □

Based on this finding, several additional observations can be made regarding the scope and conditions of convergence.

Remark 4.2.1. Theorem 4.2.1 provides a theoretical justification for the use of acceleration methods in continuous-time optimization. It shows that the inertial term introduced by the term $\frac{\alpha}{t}\dot{x}_\alpha(t)$ can be compensated by a suitable time rescaling, which reveals the simple monotone flow in the limit. More precisely, this result establishes a link between Nesterov-type accelerated dynamical systems and the classical monotone flow. Moreover, although Theorem 4.2.1 is established in the setting of cocoercive operators, it remains valid for any monotone and Lipschitz operator (not necessarily cocoercive) admitting at least one zero.

This strong convergence result, uniform on compact sets, exhibits several noteworthy features. First, it does not require V to be the gradient of a potential function: monotonicity and Lipschitz continuity alone are sufficient to control the system, guarantee the existence of a solution to the limit equation (4.7) and guarantee convergence. Next, the obtained convergence rate, of order $O(1/\sqrt{\alpha+1})$, quantifies the distance between the accelerated trajectories and the limiting trajectory.

The energy estimates established in the preliminary lemmas constitute the main tool in the proof. The modified energy function $E_\alpha(s)$ captures the balance between rescaled cinetic and potential energies, providing uniform control of the trajectories. The estimation techniques used, based on monotonicity, Lipschitz continuity and differential inequalities, highlight the robustness of the approach.

From a practical perspective, this analysis opens the way for studying other acceleration schemes and for designing numerical algorithms inspired by this dynamical interpretation.

Finally, the structure of the proof is worth highlighting: it proceeds step by step, first establishing a priori estimates by suitable Lyapunov functions, and then leveraging them to control the deviation between solutions.

Licence CC BY-NC-ND 3.0

4.3 Limiting Dynamics of (DIN-AVD): Non-Potential Operator Case

A more elaborate accelerated dynamic is now studied, incorporating an additional regularization term depending on the time variation of the operator. This system, denoted (DIN-AVD), differs from the previously studied (AVD) model through the presence of the extra term

$$\beta_0 \frac{\alpha+1}{t} \frac{d}{dt} V(x_\alpha(t)),$$

which represents the effect of the derivative of the operator along the trajectories. The objective of this section is to shows a convergence result analogous to that of Section 4.2, adapted to this richer hybrid dynamic. Theorem 4.3.1 (which also holds for monotone and Lipschitz operators: cocoercivity is not required) shows that, under the same time transformation, the family of solutions $(\tilde{y}_\alpha)_\alpha$ converges to the solution of a modified differential equation with a term proportional to the derivative of V:

$$\dot{y}(s) + \beta_0 \frac{d}{ds}\big(V(y(s))\big) + V(y(s)) = 0,\ y(0) = x_0.$$

A technical condition arises in this analysis: the assumption $\lambda > 2\beta_0$ is required to ensure stability of the system and to obtain sufficient control of the energy quantities. It should also be noted that the result is, for now, established only in finite-dimensional Hilbert spaces. This restriction stems from compactness arguments used in the proof, which are essential for identifying the limiting behavior of the solutions. Finally, the preparatory lemma 4.3.3 provides an estimate on the evolution of the trajectories, serving as the starting point for the convergence analysis.

Theorem 4.3.1 (**Finite-dimensional case**). Let $\mathcal{H}$ be a finite-dimensional real Hilbert space, and let $V : \mathcal{H} \to \mathcal{H}$ be a λ-cocoercive operator admitting at least one equilibrium point $x^* \in \mathcal{H}$ such that $V(x^*) = 0$. Consider the inertial dynamics:

$$\ddot{x}_\alpha(t) + \frac{\alpha}{t}\dot{x}_\alpha(t) + V(x_\alpha(t)) + \beta_0 \frac{\alpha+1}{t}\frac{d}{dt}V(x_\alpha(t)) = 0, \quad x_\alpha(t_0) = x_0, \quad \dot{x}_\alpha(t_0) = x_1, \tag{4.32}$$

where $\alpha \geq 0$ and $t_0 > 0$ are given parameters. Define $y_\alpha(s) = x_\alpha(\sqrt{2(\alpha+1)s})$, yielding the reparametrized system:

$$\frac{2s}{\alpha+1}\ddot{y}_\alpha(s) + \dot{y}_\alpha(s) + V(y_\alpha(s)) + \beta_0 \frac{d}{ds}V(y_\alpha(s)) = 0, \quad \forall s \geq s_{0,\alpha}, \tag{4.33}$$

with initial conditions

$$y_\alpha(s_{0,\alpha}) = x_0, \quad \dot{y}_\alpha(s_{0,\alpha}) = \frac{\alpha+1}{t_0}x_1, \quad s_{0,\alpha} = \frac{t_0^2}{2(\alpha+1)}.$$

Licence CC BY-NC-ND 3.0

Let $\tilde{y}_\alpha(s) = y_\alpha(s)$ for $s \geq s_{0,\alpha}$ and $\tilde{y}_\alpha(s) = x_0$ for $s \in [0, s_{0,\alpha}]$, the extension of y_α.

Under the assumption $\lambda > 2\beta_0$, the family $(\tilde{y}_\alpha)_\alpha$ converges uniformly on every compact interval of $[0, +\infty)$ to the unique solution y of the limit system:

$$\dot{y}(s) + V(y(s)) + \beta_0 \frac{d}{ds} V(y(s)) = 0, \quad y(0) = x_0. \tag{4.34}$$

The proof of Theorem 4.3.1 relies on an energy estimate that is particularly sensitive to the additional regularization term. The following lemma, which forms the preliminary step, allows one to control the growth of solutions and to handle the specific technical difficulties of this enriched dynamic.

Lemma 4.3.2. Under the assumptions $\lambda > \beta_0$, equation (4.34) admits a unique solution $y \in C^1([0, \infty), \mathcal{H})$.

Proof. It is first shown that the operator $\mathrm{Id} + \beta_0 V$ is bijective. Fix $y \in \mathcal{H}$ and consider the operator

$$T_0 : \mathcal{H} \to \mathcal{H}, \qquad z \mapsto T_0(z) := y - \beta_0 V(z).$$

Since V is λ-cocoercive, it is therefore $1/\lambda$-Lipschitz. Hence, for all $z_1, z_2 \in \mathcal{H}$,

$$\|T_0(z_1) - T_0(z_2)\| = \beta_0 \|V(z_1) - V(z_2)\| \leq \frac{\beta_0}{\lambda} \|z_1 - z_2\|.$$

As $\frac{\beta_0}{\lambda} < 1$ (because $\beta_0 < \lambda$), the map T_0 is a contraction. By the Banach fixed point theorem, T_0 admits a unique fixed point $z_0 \in \mathcal{H}$, which is precisely the solution of

$$z_0 + \beta_0 V(z_0) = y.$$

Thus, $\mathrm{Id} + \beta_0 V$ is bijective, and its inverse is denoted by T, that is,

$$T := (\mathrm{Id} + \beta_0 V)^{-1}.$$

Moreover, for all $x, y \in \mathcal{H}$,

$$\begin{aligned} \langle (\mathrm{Id} + \beta_0 V)(x) - (\mathrm{Id} + \beta_0 V)(y),\, x - y \rangle &= \|x - y\|^2 + \beta_0 \langle V(x) - V(y),\, x - y \rangle \\ &\geq \|x - y\|^2, \end{aligned}$$

by the monotonicity of V. Therefore, by the Cauchy–Schwarz inequality,

$$\|x - y\| \leq \|(\mathrm{Id} + \beta_0 V)(x) - (\mathrm{Id} + \beta_0 V)(y)\|.$$

Licence CC BY-NC-ND 3.0

Hence, T is 1-Lipschitz. Define

$$w(s) := (\mathrm{Id} + \beta_0 V)(y(s)).$$

Then

$$\dot{w}(s) = \dot{y}(s) + \beta_0 \frac{d}{ds} V(y(s)).$$

Hence equation (4.34) rewrites as

$$\dot{w}(s) = -V(y(s)) = -V\big(T(w(s))\big).$$

Define the map

$$F : \mathcal{H} \to \mathcal{H}, \qquad F(w) := -V\big(T(w)\big).$$

Since T is Lipschitz and V is Lipschitz, their composition F is Lipschitz continuous on $\mathcal{H}$, with Lipschitz constant at most $1/\lambda$:

$$\|F(w_1) - F(w_2)\| \leq \frac{1}{\lambda}\|T(w_1) - T(w_2)\| \leq \frac{1}{\lambda}\|w_1 - w_2\|.$$

By the Cauchy–Lipschitz theorem in a Banach space, the ODE

$$\dot{w}(s) = F(w(s)), \qquad w(0) = w_0 := x_0 + \beta_0 V(x_0),$$

admits a unique global solution $w \in C^1([0, +\infty), \mathcal{H})$. Setting $y(s) := T(w(s))$ yields a solution of the original equation (4.34). Since T is Lipschitz and $w \in C^1$, then $y \in C^1([0, +\infty), \mathcal{H})$. Uniqueness of w implies uniqueness of y. This completes the proof. □

To complete the convergence analysis of the system $(V\text{–DIN}-\text{AVD})$, it is necessary to establish uniform estimates for the trajectory y_α. The following lemma provides these estimates.

Lemma 4.3.3. (Uniform energy estimate) Under the assumption $\lambda > 2\beta_0$ and for all $\alpha \geq \sigma := \frac{\lambda+2\beta_0}{\lambda-2\beta_0}$, the trajectory y_α solution of (4.33) satisfies the uniform estimate:

$$\frac{1}{2}\|y_\alpha(s) - x^*\|^2 \leq G_2(s), \tag{4.35}$$

for all $s \geq s_{0,\alpha}$, where the majorant function is given by

$$G_2(s) := \left(\frac{1}{2}\|x_1\|^2 + \frac{1}{2}\|x_0 - x^*\|^2\right) e^{(1+\frac{1}{\lambda})^2 s}.$$

Licence CC BY-NC-ND 3.0

Proof. Introduce the following energy functional:

$$E_\alpha(s) = \frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{1}{2}\|y_\alpha(s) - x^*\|^2. \tag{4.36}$$

Its derivative with respect to the rescaled time s is:

$$\dot{E}_\alpha(s) = \frac{1}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{2s}{\alpha+1}\langle \ddot{y}_\alpha(s), \dot{y}_\alpha(s)\rangle + \langle y_\alpha(s) - x^*, \dot{y}_\alpha(s)\rangle.$$

Using the dynamics (4.33),

$$\frac{2s}{\alpha+1}\ddot{y}_\alpha(s) = -\dot{y}_\alpha(s) - V(y_\alpha(s)) - \beta_0\frac{d}{ds}V(y_\alpha(s)),$$

One obtains:

$$\dot{E}_\alpha(s) = \frac{1}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 - \langle \dot{y}_\alpha(s) + V(y_\alpha(s)) + \beta_0\frac{d}{ds}V(y_\alpha(s)), \dot{y}_\alpha(s)\rangle + \langle y_\alpha(s) - x^*, \dot{y}_\alpha(s)\rangle.$$

Expanding yields:

$$\dot{E}_\alpha(s) = -\left(1 - \frac{1}{\alpha+1}\right)\|\dot{y}_\alpha(s)\|^2 + \langle y_\alpha(s) - x^* - V(y_\alpha(s)), \dot{y}_\alpha(s)\rangle - \beta_0\left\langle \frac{d}{ds}V(y_\alpha(s)), \dot{y}_\alpha(s)\right\rangle.$$

Using the Cauchy–Schwarz and Young inequalities, together with the facts that V is $1/\lambda$-Lipschitz and that $V(x^*) = 0$, it follows that

$$|\langle y_\alpha(s) - x^* - V(y_\alpha(s)), \dot{y}_\alpha(s)\rangle| \leq \frac{(1+\frac{1}{\lambda})^2}{2}\|y_\alpha(s) - x^*\|^2 + \frac{1}{2}\|\dot{y}_\alpha(s)\|^2.$$

Since $\mathcal{H}$ is finite-dimensional and V is Lipschitz (V is $\frac{1}{\lambda}$-Lipschitz), V is differentiable almost everywhere (by Rademacher's theorem). Hence,

$$\left\|\frac{d}{ds}V(y_\alpha(s))\right\| \leq \frac{1}{\lambda}\|\dot{y}_\alpha(s)\|, \quad \text{a.e. } s \geq s_{0,\alpha}. \tag{4.37}$$

Consequently, by the Cauchy–Schwarz inequality and using relation (4.37),

$$-\beta_0\left\langle \frac{d}{ds}V(y_\alpha(s)), \dot{y}_\alpha(s)\right\rangle \leq \frac{\beta_0}{\lambda}\|\dot{y}_\alpha(s)\|^2, \quad \text{a.e. } s \geq s_{0,\alpha}.$$

The combination of these estimates gives:

$$\dot{E}_\alpha(s) \leq -\left(\frac{1}{2} - \frac{1}{\alpha+1} - \frac{\beta_0}{\lambda}\right)\|\dot{y}_\alpha(s)\|^2 + \frac{(1+\frac{1}{\lambda})^2}{2}\|y_\alpha(s) - x^*\|^2, \quad \text{a.e. } s \geq s_{0,\alpha}.$$

Licence CC BY-NC-ND 3.0

The condition $\alpha \geq \sigma$ ensures that the coefficient $\frac{1}{2} - \frac{1}{\alpha+1} - \frac{\beta_0}{\lambda}$ is positive, thus:

$$\dot{E}_\alpha(s) \leq \frac{(1+\frac{1}{\lambda})^2}{2}\|y_\alpha(s) - x^*\|^2 \leq \left(1+\frac{1}{\lambda}\right)^2 E_\alpha(s), \quad \text{a.e. } s \geq s_{0,\alpha}.$$

By applying Gronwall's lemma,

$$E_\alpha(s) \leq E_\alpha(s_{0,\alpha}) e^{(1+\frac{1}{\lambda})^2(s-s_{0,\alpha})},$$

for all $s \geq s_{0,\alpha}$. Since

$$E_\alpha(s_{0,\alpha}) = \frac{1}{2}\|x_1\|^2 + \frac{1}{2}\|x_0 - x^*\|^2,$$

it follows that $E_\alpha(s)$ is uniformly bounded by $G_2(s)$ defined above. Thus,

$$\frac{1}{2}\|y_\alpha(s) - x^*\|^2 \leq E_\alpha(s) \leq G_2(s),$$

which concludes the proof. □

With this important energy estimate, the tools necessary to carry out the main proof of the theorem are now available. The goal is to show that the family $(\tilde{y}_\alpha)_\alpha$ satisfies an appropriate compactness property and that any convergent subsequence necessarily satisfies the associated limiting differential equation.

Proof of Theorem 4.3.1. Let us consider $\alpha \geq \sigma + 1 = \frac{2\lambda}{\lambda - 2\beta_0}$ and $s \in [s_{0,\alpha}, +\infty)$. Taking the inner product of equation (4.33) with the derivative $\dot{y}_\alpha(s)$:

$$\left\langle \dot{y}_\alpha(s),\ \frac{2s}{\alpha+1}\ddot{y}_\alpha(s) + \dot{y}_\alpha(s) + V(y_\alpha(s)) + \beta_0 \frac{d}{ds}V(y_\alpha(s)) \right\rangle = 0. \tag{4.38}$$

This leads to the differential identity:

$$\frac{s}{\alpha+1}\frac{d}{ds}\|\dot{y}_\alpha(s)\|^2 + \|\dot{y}_\alpha(s)\|^2 + \langle \dot{y}_\alpha(s), V(y_\alpha(s))\rangle + \beta_0 \left\langle \frac{d}{ds}V(y_\alpha(s)),\ \dot{y}_\alpha(s) \right\rangle = 0. \tag{4.39}$$

Integrating (4.39) over the interval $[s_{0,\alpha}, s]$ gives:

$$\int_{s_{0,\alpha}}^{s} \frac{\tau}{\alpha+1}\frac{d}{d\tau}\|\dot{y}_\alpha(\tau)\|^2\, d\tau + \int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2\, d\tau \leq \left| \int_{s_{0,\alpha}}^{s} \langle \dot{y}_\alpha, V(y_\alpha)\rangle\ d\tau \right| + \beta_0 \left| \int_{s_{0,\alpha}}^{s} \left\langle \frac{d}{d\tau}V(y_\alpha(\tau)),\ \dot{y}_\alpha(\tau) \right\rangle\ d\tau \right|.$$

Licence CC BY-NC-ND 3.0

An integration by parts applied to the first term yields:

$$\left[\frac{\tau}{\alpha+1}\|\dot{y}_\alpha(\tau)\|^2\right]_{s_{0,\alpha}}^{s} - \frac{1}{\alpha+1}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|^2\,d\tau + \int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|^2\,d\tau \leq \int_{s_{0,\alpha}}^{s}|\,\langle\dot{y}_\alpha, V(y_\alpha)\rangle\,|\,d\tau + \beta_0\int_{s_{0,\alpha}}^{s}\left|\left\langle\frac{d}{d\tau}V(y_\alpha(\tau)),\ \dot{y}_\alpha(\tau)\right\rangle\right|\,d\tau.$$

This expression simplifies to:

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{\alpha}{\alpha+1}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|^2\,d\tau \leq \frac{\|x_1\|^2}{2} + \int_{s_{0,\alpha}}^{s}\|V(y_\alpha)\|\cdot\|\dot{y}_\alpha\|\,d\tau + \beta_0\int_{s_{0,\alpha}}^{s}\left\|\frac{d}{d\tau}V(y_\alpha(\tau))\right\|\cdot\|\dot{y}_\alpha(\tau)\|\,d\tau. \quad (4.40)$$

Now, exploiting the regularity of the operator V. Combining (4.37) with the fact that V is $1/\lambda$-Lipschitz and that $V(x^*)=0$, this yields the estimate:

$$\begin{aligned}\int_{s_{0,\alpha}}^{s}\|V(y_\alpha)\|\cdot\|\dot{y}_\alpha\|\,d\tau + \beta_0\int_{s_{0,\alpha}}^{s}\left\|\frac{d}{d\tau}V(y_\alpha(\tau))\right\|\cdot\|\dot{y}_\alpha(\tau)\|\,d\tau &\leq \frac{1}{\lambda}\int_{s_{0,\alpha}}^{s}\|y_\alpha - x^*\|\cdot\|\dot{y}_\alpha\|\,d\tau + \beta_0\int_{s_{0,\alpha}}^{s}\frac{1}{\lambda}\|\dot{y}_\alpha(\tau)\|\cdot\|\dot{y}_\alpha(\tau)\|\,d\tau\\ &\leq \frac{1}{2\lambda^2}\int_{s_{0,\alpha}}^{s}\|y_\alpha - x^*\|^2\,d\tau + \frac{1}{2}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha\|^2\,d\tau + \frac{\beta_0}{\lambda}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha\|^2\,d\tau\\ &\leq \frac{sG_2(s)}{\lambda^2} + \left(\frac{1}{2}+\frac{\beta_0}{\lambda}\right)\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha\|^2\,d\tau \quad \text{(by Lemma 4.3.3).} \quad (4.41)\end{aligned}$$

Combining (4.40) and (4.41) leads to:

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \left(\frac{\alpha-1}{2(\alpha+1)} - \frac{\beta_0}{\lambda}\right)\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha\|^2\,d\tau \leq \frac{\|x_1\|^2}{2} + \frac{sG_2(s)}{\lambda^2}. \quad (4.42)$$

The imposed condition $\alpha \geq \sigma+1 = \frac{2\lambda}{\lambda-2\beta_0}$ ensures that:

$$\frac{\alpha-1}{2(\alpha+1)} - \frac{\beta_0}{\lambda} \geq \frac{(\lambda-2\beta_0)^2}{2(3\lambda-2\beta_0)} \geq \frac{(\lambda-2\beta_0)^2}{6\lambda} =: \frac{1}{\gamma^2_{\lambda,\beta_0}}.$$

This leads to the uniform estimate:

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{1}{\gamma^2_{\lambda,\beta_0}}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha\|^2\,d\tau \leq \frac{\|x_1\|^2}{2} + \frac{sG_2(s)}{\lambda^2} =: G_3(s). \quad (4.43)$$

Licence CC BY-NC-ND 3.0

It immediately follows from this inequality that:

$$\int_0^s \|\dot{\tilde{y}}_\alpha\|^2 \, d\tau = \int_{s_{0,\alpha}}^s \|\dot{y}_\alpha\|^2 \, d\tau \leq \gamma^2_{\lambda,\beta_0} G_3(s) \quad \text{and} \tag{4.44}$$

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 \leq G_3(s) \tag{4.45}$$

Note that the function G_3 is strictly positive, increasing, independent of the parameter α, and bounded on every compact subset of $[0,+\infty)$. Thus, for all $s \geq 0$,

$$\|\tilde{y}_\alpha(s) - x_0\| = \left\| \int_0^s \dot{\tilde{y}}_\alpha \, d\tau \right\| \leq \gamma_{\lambda,\beta_0}(\sqrt{s} + t_0) G_3^{1/2}(s)$$

by Hölder's inequality and relation (4.44). The sequence $(\tilde{y}_\alpha)_\alpha$ is therefore uniformly bounded in $\mathcal{C}_0(0,T,\mathcal{H})$ for every $T > 0$. Moreover, the sequence of time derivatives $(\dot{\tilde{y}}_\alpha)_\alpha$ is uniformly bounded in $L^2(0,T;\mathcal{H})$ according to (4.44). These properties imply the equicontinuity of $(\tilde{y}_\alpha)_\alpha$. Consequently, it admits at least one cluster point ξ in $\mathcal{C}_0(0,T,\mathcal{H})$ when $\mathcal{H}$ is finite-dimensional (by the Ascoli–Arzelà theorem).
To establish the uniqueness of this cluster point, it suffices to verify that ξ is a solution of the differential equation (4.34). Consider a subsequence $(\dot{\tilde{y}}_{\alpha_k})_k$ converging uniformly to ξ in $\mathcal{C}_0(0,T,\mathcal{H})$. Since $s_{0,\alpha_k} = \frac{t_0^2}{2(\alpha_k+1)} \to 0$ as $k \to +\infty$ and $\tilde{y}_{\alpha_k}(s_{0,\alpha_k}) = x_0$, then:

$$\xi(0) = x_0. \tag{4.46}$$

Analysis of the limit passage in the dynamics: Let $\delta \in]0,T[$ such that $\tilde{y}_{\alpha_k} = y_{\alpha_k}$ on $[\delta,T]$ for k sufficiently large. Let us examine the asymptotic behavior of each term in equation (4.33):

- *Convergence of* $V(\tilde{y}_{\alpha_k})$: The Lipschitz regularity of the operator V ensures that the sequence $(V(\tilde{y}_{\alpha_k}))_k$ converges uniformly to $V(\xi)$ in $\mathcal{C}_0(\delta,T,\mathcal{H})$.
- *Convergence of the first derivatives*: The boundedness of the sequence $(\dot{\tilde{y}}_{\alpha_k})_k$ (relation (4.45)) and the continuity of differentiation in the sense of distributions imply (Proposition 2.2.1)

$$\dot{\tilde{y}}_{\alpha_k} \rightharpoonup \dot{\xi} \quad \text{in } L^2(\delta,T;\mathcal{H}). \tag{4.47}$$

- *Convergence of* $\frac{d}{ds}V(\tilde{y}_{\alpha_k})$: Since the sequence $(\dot{\tilde{y}}_{\alpha_k})_k$ is uniformly bounded in $L^2(\delta,T;\mathcal{H})$, the sequence $\left(\frac{d}{ds}V(\tilde{y}_{\alpha_k})\right)_k$ is also uniformly bounded in $L^2(\delta,T;\mathcal{H})$ (by relation (4.37)). Then, using the continuity of the distributional derivative, one obtains (Proposition 2.2.1)

$$\frac{d}{ds}V(\tilde{y}_{\alpha_k}(s)) \rightharpoonup \frac{d}{ds}(V(\xi(s))) \quad \text{in } L^2(\delta,T;\mathcal{H}).$$

Licence CC BY-NC-ND 3.0

- *Inertial term*: The second–order term satisfies

$$\frac{2s}{\alpha_k+1}\ddot{\tilde{y}}_{\alpha_k} \rightharpoonup 0 \quad \text{in } L^2(\delta,T;\mathcal{H}). \tag{4.48}$$

Indeed, this inertial term remains bounded in $L^2(\delta,T;\mathcal{H})$ since $(\dot{\tilde{y}}_{\alpha_k})_k$, $(V(\tilde{y}_{\alpha_k}))_k$, and $\left(\frac{d}{ds}V(\tilde{y}_{\alpha_k})\right)_k$ are all bounded. Moreover, for any test function $\phi \in \mathcal{D}((\delta,T);\mathcal{H})$,

$$\int_\sigma^T \left\langle \frac{2\tau}{\alpha_k+1}\,\ddot{\tilde{y}}_{\alpha_k}, \phi \right\rangle d\tau = \frac{2}{\alpha_k+1}\int_\sigma^T \left\langle \tilde{y}_{\alpha_k}, \frac{d^2}{d\tau^2}(\tau\phi)\right\rangle d\tau$$

$$\text{converge to} \quad 0 \times \int_\sigma^T \left\langle \xi, \frac{d^2}{d\tau^2}(\tau\phi)\right\rangle d\tau = 0, \quad \text{when} \quad k \to +\infty.$$

Combining these convergences shows that the limit ξ satisfies:

$$\dot{\xi}(s) + V(\xi(s)) + \beta_0 \frac{d}{ds}V(\xi(s)) = 0, \quad \text{for all } s \in (\delta,T), \quad \text{with} \quad \xi(0) = x_0.$$

Since $\delta > 0$ is arbitrary, the identity holds on $(0,T)$, and uniqueness of solutions to (4.34) yields $\xi = y$. Hence $(\tilde{y}_\alpha)_\alpha$ converges uniformly on compact intervals to y, which completes the proof. □

The interpretation of the condition $\lambda > 2\beta_0$, as well as the implications of relaxing it, can be examined in greater detail, as indicated in the following remark.

Remark 4.3.1. Theorem 4.3.1 extends the results of Section 4.2 by broadening the analysis to an accelerated dynamics enriched with an explicit regulation mechanism. The addition of the term $\beta_0 \frac{d}{ds}V(y_\alpha(s))$ modifies the system structure by introducing feedback depending on the temporal variation of the operator V along the trajectories. As clarified in the previous section, it is important to recall that Theorem 4.3.1 also remains valid for any monotone and Lipschitz operator (not necessarily cocoercive) admitting at least one zero.

The condition $\lambda > 2\beta_0$ plays an important role but also has an intuitive interpretation: it ensures that the regulation does not dominate the stabilizing effect provided by the cocoercivity of V. If this inequality is not satisfied, the system behavior becomes more difficult to control, and convergence to the limiting dynamics may fail.

Analytically, the proof relies crucially on the assumption that the space $\mathcal{H}$ is finite-dimensional. This allows the application of the Ascoli–Arzelà theorem to extract a convergent subsequence from the family $(\tilde{y}_\alpha)_\alpha$. In infinite dimension, such an approach would require more subtle compactness tools, such as weak compactness or measurable convergence, making the analysis more delicate.

The estimates obtained in this section extend those of Section 4.2, but the presence of the regulation term requires specific adjustments. Lemma 4.3.3 highlights how the parameter β_0

Licence CC BY-NC-ND 3.0

influences the energy constants and restricts the admissible range for the parameter α. The energy functional $E_\alpha(s)$ is adapted to account for interactions between inertia, cocoercivity, and regulation.

Finally, extending this result to the infinite-dimensional setting remains an open question. It poses several challenges: absence of strong compactness, increased regularity requirements, and the need for more refined functional analysis tools. A natural perspective would be to investigate the connections between this continuous dynamic approach and the corresponding discrete algorithms, to better understand how continuous properties translate to algorithmic behavior.

4.4 Limit Dynamics of (ISIHD): Non-Potential Operator Case

This penultimate section is devoted to the study of a third family of accelerated dynamics, referred to as (ISIHD). It differs from the previous models by the way the operator V is involved: it is evaluated at a predicted position along the trajectory, given by $x_\alpha(t) + \beta_0 \frac{\alpha+1}{t}\dot{x}_\alpha(t)$. This evaluation at a predictive point endows the dynamics with an implicit character and introduces notable analytical difficulties, related to the functional dependence of the equation. Theorem 4.4.1 (which also holds for monotone and Lipschitz operators: cocoercivity is not required) establishes the conditions ensuring the convergence of this dynamics, after an appropriate time rescaling, toward an implicit monotone flow type limit equation:

$$\dot{y}(s) + V(y(s) + \beta_0 \dot{y}(s)) = 0,\ y(0) = x_0.$$

A technical hypothesis is required: the vanishing of the initial velocity $x_1 = 0$. Although restrictive, this condition simplifies the analysis by removing initial singularities and facilitating the control of inertial effects. Moreover, the constant β_0 must satisfy $2\beta_0 < \lambda$ in order to guarantee the validity of the energy estimates derived from Lemma 4.4.3. The resulting convergence is particularly strong: it ensures not only the uniform convergence of trajectories but also the strong convergence of velocities in L^2. This reflects a precise control over both positions and velocities along the approximated trajectories.

Theorem 4.4.1. Let $\mathcal{H}$ be a real Hilbert space. Consider an operator $V : \mathcal{H} \to \mathcal{H}$ satisfying the λ-cocoercivity property and admitting at least one equilibrium point $x^* \in \mathcal{H}$ such that $V(x^*) = 0$. Consider the following dynamical system:

$$\ddot{x}_\alpha(t) + \frac{\alpha}{t}\dot{x}_\alpha(t) + V\left(x_\alpha(t) + \beta_0 \frac{\alpha+1}{t}\dot{x}_\alpha(t)\right) = 0, \quad x_\alpha(t_0) = x_0, \quad \dot{x}_\alpha(t_0) = x_1, \quad (4.49)$$

Licence CC BY-NC-ND 3.0

with $\alpha > 0$ and $t_0 > 0$ as parameters. Define $y_\alpha(s) = x_\alpha(\sqrt{2(\alpha+1)s})$.

- Then y_α is a solution of:

$$\frac{2s}{\alpha+1}\ddot{y}_\alpha(s) + \dot{y}_\alpha(s) + V\left(y_\alpha(s) + \beta_0\dot{y}_\alpha(s)\right) = 0, \quad \forall s \geq s_{0,\alpha}, \tag{4.50}$$

  with initial conditions:

$$y_\alpha(s_{0,\alpha}) = x_0, \quad \dot{y}_\alpha(s_{0,\alpha}) = \frac{\alpha+1}{t_0}x_1, \quad \text{and} \quad s_{0,\alpha} = \frac{t_0^2}{2(\alpha+1)}.$$

- Let $\tilde{y}_\alpha(s) = y_\alpha(s)$ for $s \geq s_{0,\alpha}$ and $\tilde{y}_\alpha(s) = x_0$ for $s \in [0, s_{0,\alpha}]$, the extension of y_α. Suppose that $2\beta_0 < \lambda$ and $x_1 = 0$. Then the family $(\tilde{y}_\alpha)_\alpha$ converges uniformly on every compact subset of $[0, +\infty)$ to the unique solution y of the limit system:

$$\dot{y}(s) + V(y(s) + \beta_0\dot{y}(s)) = 0, \quad y(0) = x_0. \tag{4.51}$$

  Moreover, the sequence of derivatives $(\dot{y}_\alpha)$ converges strongly to $\dot{y}$ in the space $L^2\left([0,T], \mathcal{H}\right)$ as $\alpha \to +\infty$.

The proof of Theorem 4.4.1, delicate due to the evaluation of the operator at a predicted point, requires precise a priori control of the trajectories. The analysis begins with two preparatory lemmas, which provide a framework for understanding the sometimes counter-intuitive behavior of this dynamics.

Lemma 4.4.2. If $\beta_0 < \lambda$, then equation (4.51) admits a unique solution $y \in C^1([0, +\infty), \mathcal{H})$ satisfying the following regularity estimate:

$$\|\dot{y}(s)\| \leq \frac{1}{\lambda - \beta_0} M(s) \quad \text{for all } s \geq 0, \tag{4.52}$$

where

$$M(s) := \sup_{\tau \in [0,s]} \|y(\tau) - x^*\|. \tag{4.53}$$

Proof. Set

$$z(s) := y(s) + \beta_0\dot{y}(s) \quad \Longleftrightarrow \quad \dot{y}(s) = \frac{1}{\beta_0}(z(s) - y(s)).$$

Then the equation becomes

$$\frac{1}{\beta_0}(z(s) - y(s)) + V(z(s)) = 0 \quad \Longleftrightarrow \quad z(s) + \beta_0 V(z(s)) = y(s).$$

As demonstrated in the proof of Lemma 4.3.2, the operator $\mathrm{Id} + \beta_0 V$ is bijective and its inverse

Licence CC BY-NC-ND 3.0

$(\mathrm{Id} + \beta_0 V)^{-1}$ is 1-Lipschitz. Define

$$F(y) := \frac{1}{\beta_0}\Big((\mathrm{Id} + \beta_0 V)^{-1}(y) - y\Big),$$

so that equation (4.51) becomes

$$\dot{y}(s) = F(y(s)), \quad y(0) = y_0.$$

Since $(\mathrm{Id} + \beta_0 V)^{-1}$ is Lipschitz, F is globally Lipschitz on $\mathcal{H}$. Thus, equation (4.51) admits a unique global solution $y \in C^1([0, +\infty), \mathcal{H})$ by the Cauchy-Lipschitz theorem.

From equation (4.51), it follows that

$$\|\dot{y}(s)\| = \|V(y(s) + \beta_0 \dot{y}(s))\| \leq \frac{1}{\lambda}\|y(s) + \beta_0 \dot{y}(s) - x^*\|,$$

because $V(x^*) = 0$ and V is $1/\lambda$ Lipschitz. Using the triangle inequality, one obtains

$$\|\dot{y}(s)\| \leq \frac{1}{\lambda}\|y(s) - x^*\| + \frac{\beta_0}{\lambda}\|\dot{y}(s)\|.$$

Under the assumption $\beta_0 < \lambda$, it suffices to reorganize to obtain

$$\|\dot{y}(s)\| \leq \frac{1}{\lambda - \beta_0}\|y(s) - x^*\|.$$

Since this estimate holds for all $s \geq 0$, it directly yields the stated result. □

This control of the limit solution's velocity is a tool for the initial analysis. To carry out the main proof, it is also necessary to establish an analogous estimate for the trajectories of the approximated dynamics.

Lemma 4.4.3. Let $2\beta_0 < \lambda$. Then, for all $\alpha \geq \sigma := \frac{\lambda + 2\beta_0}{\lambda - 2\beta_0}$, the trajectory y_α solving (4.50) satisfies the uniform estimate:

$$\frac{1}{2}\|y_\alpha(s) - x^*\|^2 \leq D_2(s) \quad \text{for all } s \geq s_{0,\alpha}, \tag{4.54}$$

where

$$D_2(s) := \left(\frac{1}{2}\|x_1\|^2 + \frac{1}{2}\|x_0 - x^*\|^2\right) e^{(1+\frac{1}{\lambda})^2 s}.$$

Proof. Introduce the adapted energy functional:

$$E_\alpha(s) = \frac{s}{\alpha + 1}\|\dot{y}_\alpha(s)\|^2 + \frac{1}{2}\|y_\alpha(s) - x^*\|^2. \tag{4.55}$$

Licence CC BY-NC-ND 3.0

Its derivative reads:

$$\dot{E}_\alpha(s) = \frac{1}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{2s}{\alpha+1}\langle \ddot{y}_\alpha(s), \dot{y}_\alpha(s)\rangle + \langle y_\alpha(s) - x^*, \dot{y}_\alpha(s)\rangle.$$

From (4.50),

$$\frac{2s}{\alpha+1}\ddot{y}_\alpha(s) = -\dot{y}_\alpha(s) - V(y_\alpha(s) + \beta_0\dot{y}_\alpha(s)),$$

which gives

$$\dot{E}_\alpha(s) = -\Big(1 - \frac{1}{\alpha+1}\Big)\|\dot{y}_\alpha(s)\|^2 + \langle y_\alpha(s) - x^* - V(y_\alpha(s) + \beta_0\dot{y}_\alpha(s)), \dot{y}_\alpha(s)\rangle.$$

Using the Lipschitz regularity of V and $V(x^*) = 0$, one obtains

$$\|y_\alpha(s) - x^* - V(y_\alpha(s) + \beta_0\dot{y}_\alpha(s))\| \leq \left(1 + \frac{1}{\lambda}\right)\|y_\alpha(s) - x^*\| + \frac{\beta_0}{\lambda}\|\dot{y}_\alpha(s)\|.$$

The Cauchy-Schwarz inequality combined with Young's inequality gives

$$|\langle y_\alpha(s) - x^* - V(y_\alpha(s) + \beta_0\dot{y}_\alpha(s)), \dot{y}_\alpha(s)\rangle| \leq \frac{(1+\frac{1}{\lambda})^2}{2}\|y_\alpha(s) - x^*\|^2 + \left(\frac{1}{2} + \frac{\beta_0}{\lambda}\right)\|\dot{y}_\alpha(s)\|^2.$$

Thus,

$$\dot{E}_\alpha(s) \leq -\left(\frac{1}{2} - \frac{1}{\alpha+1} - \frac{\beta_0}{\lambda}\right)\|\dot{y}_\alpha(s)\|^2 + \frac{(1+\frac{1}{\lambda})^2}{2}\|y_\alpha(s) - x^*\|^2.$$

Under the technical condition $\alpha \geq \sigma = \frac{\lambda+2\beta_0}{\lambda-2\beta_0}$, the coefficient in front of $\|\dot{y}_\alpha(s)\|^2$ is negative. The following is thus obtained:

$$\dot{E}_\alpha(s) \leq \frac{(1+\frac{1}{\lambda})^2}{2}\|y_\alpha(s) - x^*\|^2 \leq \left(1 + \frac{1}{\lambda}\right)^2 E_\alpha(s).$$

The application of Grönwall's lemma then yields

$$E_\alpha(s) \leq E_\alpha(s_{0,\alpha})\, e^{(1+\frac{1}{\lambda})^2(s-s_{0,\alpha})}.$$

But

$$E_\alpha(s_{0,\alpha}) = \frac{1}{2}\|x_1\|^2 + \frac{1}{2}\|x_0 - x^*\|^2,$$

which allows us to conclude

$$\frac{1}{2}\|y_\alpha(s) - x^*\|^2 \leq E_\alpha(s) \leq D_2(s),$$

as stated. □

Licence CC BY-NC-ND 3.0

The proof of Theorem 4.4.1 can be carried out by relying on the two preceding lemmas.

Proof of Theorem 4.4.1. The main objective is to establish an upper bound for $\|\tilde{y}_\alpha(s) - y(s)\|$ which converges uniformly to 0 on every compact subset of $[0, +\infty)$ as $\alpha \to +\infty$.

Analysis on the initial interval $[0, s_{0,\alpha}]$: For all $s \in [0, s_{0,\alpha}]$, one has:

$$\|\tilde{y}_\alpha(s) - y(s)\| = \|x_0 - y(s)\| = \|y(0) - y(s)\| = \left\| \int_0^s \dot{y}(\tau)\, d\tau \right\| \leq \int_0^{s_{0,\alpha}} \|\dot{y}(\tau)\|\, d\tau$$

Using the regularity estimate (4.52) established in Lemma 4.4.2, it follows that

$$\|\tilde{y}_\alpha(s) - y(s)\| \leq \frac{s_{0,\alpha}}{\lambda - \beta_0} M(t_0^2) =: D_1(\alpha). \tag{4.56}$$

In particular, for $s = s_{0,\alpha}$, this yields:

$$\|\tilde{y}_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\| = \|y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\| = \|x_0 - y(s_{0,\alpha})\| \leq D_1(\alpha). \tag{4.57}$$

Moreover,

$$D_1(\alpha) \to 0 \quad \text{as} \quad \alpha \to +\infty,$$

which guarantees convergence on the initial interval.

Study on the interval $[s_{0,\alpha}, +\infty)$: The analysis now focuses on the behavior of the dynamics (4.50). Taking the inner product of the equation with $\dot{y}_\alpha(s)$ yields:

$$\left\langle \dot{y}_\alpha(s), \frac{2s}{\alpha+1}\ddot{y}_\alpha(s) + \dot{y}_\alpha(s) + V\left(y_\alpha(s) + \beta_0 \dot{y}_\alpha(s)\right) \right\rangle = 0. \tag{4.58}$$

Using the monotonicity of V (a consequence of its cocoercivity) to expand this inner product:

$$\begin{aligned} \langle \dot{y}_\alpha(s), V(y_\alpha(s) + \beta_0 \dot{y}_\alpha(s)) \rangle &= \underbrace{\langle \dot{y}_\alpha(s), V(y_\alpha(s) + \beta_0 \dot{y}_\alpha(s)) - V(y_\alpha(s)) \rangle}_{\geq 0 \text{ by monotonicity}} + \langle \dot{y}_\alpha(s), V(y_\alpha(s)) \rangle \\ &\geq \langle \dot{y}_\alpha(s), V(y_\alpha(s)) \rangle. \end{aligned}$$

This leads to the differential inequality:

$$\frac{s}{\alpha+1}\frac{d}{ds}\|\dot{y}_\alpha(s)\|^2 + \|\dot{y}_\alpha(s)\|^2 + \langle \dot{y}_\alpha(s), V(y_\alpha(s)) \rangle \leq 0. \tag{4.59}$$

Integrating (4.59) over $[s_{0,\alpha}, s]$ gives:

$$\int_{s_{0,\alpha}}^s \frac{\tau}{\alpha+1}\frac{d}{d\tau}\|\dot{y}_\alpha(\tau)\|^2\, d\tau + \int_{s_{0,\alpha}}^s \|\dot{y}_\alpha(\tau)\|^2\, d\tau \leq \left| \int_{s_{0,\alpha}}^s \langle V(y_\alpha), \dot{y}_\alpha \rangle\, d\tau \right|.$$

Licence CC BY-NC-ND 3.0

Performing integration by parts on the first term:

$$\left[\frac{\tau}{\alpha+1}\|\dot{y}_\alpha(\tau)\|^2\right]_{s_{0,\alpha}}^{s} - \frac{1}{\alpha+1}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|^2\,d\tau + \int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|^2\,d\tau \leq \left|\int_{s_{0,\alpha}}^{s}\langle V(y_\alpha),\dot{y}_\alpha\rangle\,d\tau\right|.$$

This simplifies to:

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{\alpha}{\alpha+1}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|^2\,d\tau \leq \frac{\|x_1\|^2}{2} + \left|\int_{s_{0,\alpha}}^{s}\langle V(y_\alpha),\dot{y}_\alpha\rangle\,d\tau\right|. \tag{4.60}$$

The integral term is estimated by successive applications of the Cauchy–Schwarz and Young inequalities:

$$\begin{aligned}
\left|\int_{s_{0,\alpha}}^{s}\langle V(y_\alpha),\dot{y}_\alpha\rangle\,d\tau\right| &\leq \int_{s_{0,\alpha}}^{s}|\langle V(y_\alpha),\dot{y}_\alpha\rangle|\,d\tau \\
&\leq \int_{s_{0,\alpha}}^{s}\|V(y_\alpha)\|\cdot\|\dot{y}_\alpha\|\,d\tau \\
&\leq \frac{1}{\lambda}\int_{s_{0,\alpha}}^{s}\|y_\alpha - x^*\|\cdot\|\dot{y}_\alpha\|\,d\tau \\
&\leq \frac{1}{2\lambda^2}\int_{s_{0,\alpha}}^{s}\|y_\alpha - x^*\|^2\,d\tau + \frac{1}{2}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha\|^2\,d\tau \\
&\leq \frac{sD_2(s)}{\lambda^2} + \frac{1}{2}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha\|^2 d\tau \quad \text{(by Lemma } 4.4.3\text{)}.
\end{aligned} \tag{4.61}$$

Combining (4.60) and (4.61) yields:

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{\alpha-1}{2(\alpha+1)}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha\|^2\,d\tau \leq \frac{\|x_1\|^2}{2} + \frac{sD_2(s)}{\lambda^2}. \tag{4.62}$$

Using $\frac{\alpha-1}{2(\alpha+1)} \geq \frac{\sigma-1}{2(\sigma+1)} = \frac{\beta_0}{\lambda}$ for $\alpha \geq \sigma = \frac{\lambda+2\beta_0}{\lambda-2\beta_0}$, the following estimate is obtained:

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 + \frac{\beta_0}{\lambda}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha\|^2\,d\tau \leq \frac{\|x_1\|^2}{2} + \frac{sD_2(s)}{\lambda^2} =: D_3(s). \tag{4.63}$$

It immediately follows that

$$\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha\|^2\,d\tau \leq \frac{\lambda}{\beta_0}D_3(s) \quad \text{and} \tag{4.64}$$

$$\frac{s}{\alpha+1}\|\dot{y}_\alpha(s)\|^2 \leq D_3(s) \tag{4.65}$$

Note that D_3 is strictly positive, independent of α, and bounded on every compact subset of $[0,+\infty)$.

Licence CC BY-NC-ND 3.0

Comparison of dynamics and control of the gap: The monotonicity of V allows us to establish:

$$\langle V\left(y_\alpha(s)+\beta_0\dot{y}_\alpha(s)\right)-V\left(y(s)+\beta_0\dot{y}(s)\right),y_\alpha(s)+\beta_0\dot{y}_\alpha(s)-(y(s)+\beta_0\dot{y}(s))\rangle\geq 0.$$

After substitution of the expressions of $V(y_\alpha(s)+\beta_0\dot{y}_\alpha(s))$ and $V(y(s)+\beta_0\dot{y}(s))$, the following relation holds:

$$\langle-\frac{2s}{\alpha+1}\ddot{y}_\alpha(s)-\dot{y}_\alpha(s)+\dot{y}(s),y_\alpha(s)-y(s)+\beta_0(\dot{y}_\alpha(s)-\dot{y}(s))\rangle\geq 0.$$

Expanding this inequality leads to:

$$\begin{aligned}-\frac{2s}{\alpha+1}\langle\ddot{y}_\alpha(s),\,y_\alpha(s)-y(s)\rangle-\frac{2s\beta_0}{\alpha+1}\langle\ddot{y}_\alpha(s),\dot{y}_\alpha(s)-\dot{y}(s)\rangle&\\ \geq\frac{1}{2}\frac{d}{ds}\|y_\alpha(s)-y(s)\|^2+\beta_0\|\dot{y}_\alpha(s)-\dot{y}(s)\|^2.&\end{aligned}$$

Integration of this inequality provides the result:

$$\frac{1}{2}\|y_\alpha(s)-y(s)\|^2+\beta_0\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha-\dot{y}\|^2\,d\tau\leq\frac{1}{2}\|x_0-y(s_{0,\alpha})\|^2+T_1(s,\alpha)+T_2(s,\alpha)\quad(4.66)$$

with the definitions:

$$\begin{aligned}T_1(s,\alpha)&=-\frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s}\langle\ddot{y}_\alpha(\tau),\,\tau(y_\alpha(\tau)-y(\tau))\rangle\,d\tau,\\ T_2(s,\alpha)&=-\frac{2\beta_0}{\alpha+1}\int_{s_{0,\alpha}}^{s}\langle\ddot{y}_\alpha(\tau),\,\tau(\dot{y}_\alpha(\tau)-\dot{y}(\tau))\rangle\,d\tau.\end{aligned}$$

The convergence analysis relies on precise control of the terms $T_1(\alpha,s)$ and $T_2(\alpha,s)$ by quantities tending to 0 as α increases.

Detailed estimation of the term $T_1(\alpha,s)$: Integration by parts yields the decomposition:

$$T_1(s,\alpha)=F_2(s,\alpha)+F_1(s,\alpha)$$

with

$$\begin{aligned}F_1(s,\alpha)&=\frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s}\langle\dot{y}_\alpha(\tau),\,(y_\alpha(\tau)-y(\tau))+\tau(\dot{y}_\alpha(\tau)-\dot{y}(\tau))\rangle\,d\tau,\\ F_2(s,\alpha)&=-\frac{2}{\alpha+1}\left\{\langle\dot{y}_\alpha(\tau),\,\tau(y_\alpha(\tau)-y(\tau))\rangle\right\}_{s_{0,\alpha}}^{s}.\end{aligned}$$

Licence CC BY-NC-ND 3.0

The Cauchy-Schwarz and Young inequalities yield:

$$\begin{aligned}|F_1(s,\alpha)| &\leq \frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \cdot \|y_\alpha(\tau)-y(\tau)\|\, d\tau \\ &\quad + \frac{2s}{\alpha+1}\int_{s_{0,\alpha}}^{s} \left(\|\dot{y}_\alpha(\tau)\|^2 + |\langle \dot{y}_\alpha(\tau), \dot{y}(\tau)\rangle|\right)\, d\tau \\ &\leq \frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \cdot \|y_\alpha(\tau)-y(\tau)\|\, d\tau \\ &\quad + \frac{s}{\alpha+1}\left[3\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2\, d\tau + \int_{s_{0,\alpha}}^{s} \|\dot{y}(\tau)\|^2\, d\tau\right].\end{aligned}$$

Using estimate (4.64) and the energy control:

$$\int_{s_{0,\alpha}}^{s} \|\dot{y}(\tau)\|^2\, d\tau \leq \frac{s}{(\lambda-\beta_0)^2} M^2(s) \quad \text{(from (4.52))},$$

the following estimate is obtained:

$$|F_1(s,\alpha)| \leq \frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha\| \cdot \|y_\alpha - y\|\, d\tau + \frac{3\lambda s D_3(s)}{\beta_0(\alpha+1)} + \frac{s^2}{(\alpha+1)(\lambda-\beta_0)^2} M^2(s). \quad (4.67)$$

Let us analyze the boundary terms $F_2(s,\alpha)$. By the Cauchy-Schwarz inequality and the properties of y_α:

$$\begin{aligned}\frac{2}{\alpha+1}|\langle \dot{y}_\alpha(s_{0,\alpha}), s_{0,\alpha}(y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha}))\rangle| &\leq \frac{2}{\alpha+1} s_{0,\alpha}\|\dot{y}_\alpha(s_{0,\alpha})\| \cdot \|y_\alpha(s_{0,\alpha}) - y(s_{0,\alpha})\| \\ &= \frac{2s_{0,\alpha}}{\alpha+1}\left\|\frac{\alpha+1}{t_0}x_1\right\| \cdot \|x_0 - y(s_{0,\alpha})\| \\ &\leq \frac{t_0\|x_1\|}{(\alpha+1)} D_1(\alpha) \quad \text{(from (4.57))}. \qquad (4.68)\end{aligned}$$

Moreover, estimate (4.65) combined with the Cauchy-Schwarz inequality gives:

$$\begin{aligned}\frac{2}{\alpha+1}\,|\langle \dot{y}_\alpha(s), s(y_\alpha(s) - y(s))\rangle| &\leq \frac{2s}{\alpha+1}\|\dot{y}_\alpha(s)\| \cdot \|y_\alpha(s) - y(s)\| \\ &\leq 2\sqrt{\frac{s}{\alpha+1}} D_3^{1/2}(s) \cdot \|y_\alpha(s) - y(s)\|. \qquad (4.69)\end{aligned}$$

Applying Young's inequality with

$$a = \frac{1}{2}\|y_\alpha(s) - y(s)\| \quad \text{and} \quad b = 2\sqrt{\frac{s}{\alpha+1}} D_3^{1/2}(s),$$

Licence CC BY-NC-ND 3.0

yields from (4.69) the estimate:

$$\frac{2}{\alpha+1}\,|\langle \dot{y}_\alpha(s), s(y_\alpha(s)-y(s))\rangle| \le \frac{1}{4}\|y_\alpha(s)-y(s)\|^2 + \frac{4s}{\alpha+1}D_3(s). \tag{4.70}$$

Relations (4.68) and (4.70) imply

$$|F_2(s,\alpha)| \le \frac{t_0\|x_1\|}{\alpha+1}D_1(\alpha) + \frac{1}{4}\|y_\alpha(s)-y(s)\|^2 + \frac{4s}{\alpha+1}D_3(s).$$

Combining these results gives the upper bound:

$$T_1(s,\alpha) \le \frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|\cdot\|y_\alpha(\tau)-y(\tau)\|\,d\tau + \frac{1}{4}\|y_\alpha(s)-y(s)\|^2 + K_0(\alpha,s) \tag{4.71}$$

where

$$K_0(\alpha,s) = \frac{t_0\|x_1\|}{\alpha+1}D_1(\alpha) + \frac{4s}{\alpha+1}D_3(s) + \frac{3\lambda s D_3(s)}{\beta_0(\alpha+1)} + \frac{s^2}{(\alpha+1)(\lambda-\beta_0)^2}M^2(s).$$

Detailed estimation of the term $T_2(\alpha,s)$: The term $T_2(\alpha,s)$ is analyzed using integration by parts:

$$\begin{aligned} 2\int_{s_{0,\alpha}}^{s} \tau\langle \ddot{y}_\alpha(\tau),\, \dot{y}_\alpha(\tau)\rangle\,d\tau &= \int_{s_{0,\alpha}}^{s} \tau\frac{d}{d\tau}\|\dot{y}_\alpha(\tau)\|^2\,d\tau \\ &= s\|\dot{y}_\alpha(s)\|^2 - s_{0,\alpha}\|\dot{y}_\alpha(s_{0,\alpha})\|^2 - \int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|^2\,d\tau \\ &\ge -\frac{\alpha+1}{2}\|x_1\|^2 - \int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|^2\,d\tau. \end{aligned} \tag{4.72}$$

Multiplying by $-\frac{\beta_0}{\alpha+1}$ and combining with (4.64), one obtains:

$$-\frac{2\beta_0}{\alpha+1}\int_{s_{0,\alpha}}^{s}\tau\langle \ddot{y}_\alpha(\tau),\, \dot{y}_\alpha(\tau)\rangle\,d\tau \le \frac{\beta_0}{2}\|x_1\|^2 + \frac{\lambda D_3(s)}{\alpha+1}. \tag{4.73}$$

Consider the integral term involving $\dot{y}(\tau)$:

$$\begin{aligned} \int_{s_{0,\alpha}}^{s}\tau\langle \ddot{y}_\alpha(\tau),\, \dot{y}(\tau)\rangle\,d\tau &= \langle s\dot{y}_\alpha(s),\, \dot{y}(s)\rangle - \langle s_{0,\alpha}\dot{y}_\alpha(s_{0,\alpha}),\, \dot{y}(s_{0,\alpha})\rangle \\ &\quad - \int_{s_{0,\alpha}}^{s}\langle \dot{y}_\alpha(\tau),\, \tau\ddot{y}(\tau)+\dot{y}(\tau)\rangle\,d\tau \\ &\le s\|\dot{y}_\alpha(s)\|\cdot\|\dot{y}(s)\| + \frac{t_0}{2}\|x_1\|\cdot\|\dot{y}(s_{0,\alpha})\| \\ &\quad + \int_{s_{0,\alpha}}^{s}\|\dot{y_\alpha}(\tau)\|\cdot\|\tau\ddot{y}(\tau)+\dot{y}(\tau)\|\,d\tau. \end{aligned}$$

Licence CC BY-NC-ND 3.0

The following estimate holds:

$$\|\dot{y}(s_{0,\alpha})\| \leq \frac{1}{\lambda - \beta_0} M(t_0^2) =: D_4 \quad \text{(from (4.52))}.$$

Introduce the control quantity:

$$M_0(s) = \sup_{0 \leq \tau \leq s} \|\tau \ddot{y}(\tau) + \dot{y}(\tau)\|.$$

This allows us to establish:

$$\frac{2\beta_0}{\alpha+1} \int_{s_{0,\alpha}}^{s} \tau \langle \ddot{y}_\alpha(\tau),\, \dot{y}(\tau) \rangle \, d\tau \leq \frac{2\beta_0 M(s)}{(\lambda - \beta_0)} \cdot \frac{s\|\dot{y}_\alpha(s)\|}{\alpha+1} + \frac{\beta_0 t_0 \|x_1\|}{\alpha+1} D_4 + \frac{2\beta_0}{\alpha+1} M_0(s) \int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| \, d\tau.$$

The Hölder inequality combined with (4.64) gives:

$$\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\| d\tau \leq \sqrt{s - s_{0,\alpha}} \left(\int_{s_{0,\alpha}}^{s} \|\dot{y}_\alpha(\tau)\|^2 d\tau \right)^{1/2} \overset{(4.64)}{\leq} \sqrt{\frac{\lambda s}{\beta_0}} D_3^{1/2}(s). \tag{4.74}$$

The combination of these results with (4.65) gives:

$$\frac{2\beta_0}{\alpha+1} \int_{s_{0,\alpha}}^{s} \tau \langle \ddot{y}_\alpha(\tau), \dot{y}(\tau) \rangle d\tau \leq \frac{2\beta_0 M(s)}{(\lambda - \beta_0)} \cdot \sqrt{\frac{s D_3(s)}{\alpha+1}} + \frac{\beta_0 t_0 \|x_1\|}{\alpha+1} D_4 + \frac{2\sqrt{\beta_0 \lambda s D_3(s)}}{\alpha+1} M_0(s).$$

The combination of these elements gives the complete bound:

$$T_2(s, \alpha) \leq K_1(s, \alpha) \tag{4.75}$$

where

$$K_1(s,\alpha) = \frac{2\beta_0 M(s)}{(\lambda - \beta_0)} \cdot \sqrt{\frac{s D_3(s)}{\alpha+1}} + \frac{\beta_0 t_0 \|x_1\|}{\alpha+1} D_4 + \frac{2\sqrt{\beta_0 \lambda s D_3(s)}}{\alpha+1} M_0(s) + \frac{\beta_0}{2} \|x_1\|^2 + \frac{\lambda D_3(s)}{\alpha+1}.$$

Synthesis and convergence: Combining the bounds (4.57), (4.71), and (4.75) with the inequality

Licence CC BY-NC-ND 3.0

(4.66) establishes:

$$\begin{aligned}\frac{1}{2}\|y_\alpha(s)-y(s)\|^2+\beta_0\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)-\dot{y}(\tau)\|^2\,d\tau&\\ \leq \frac{D_1^2(\alpha)}{2}+\frac{2}{\alpha+1}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|\,\|y_\alpha(\tau)-y(\tau)\|\,d\tau&\\ +\frac{1}{4}\|y_\alpha(s)-y(s)\|^2+K_0(s,\alpha)+K_1(s,\alpha).&\end{aligned}$$

The integral is bounded by introducing x^* and applying successively (4.54), the Hölder inequality, and (4.64):

$$\begin{aligned}\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|\|y_\alpha(\tau)-y(\tau)\|d\tau &\leq \int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|\cdot(\|y_\alpha(\tau)-x^*\|+\|x^*-y(\tau)\|)d\tau\\ &\leq \Big(\sqrt{2D_2(s)}+M(s)\Big)\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)\|\,d\tau\\ &\leq \Big(\sqrt{2D_2(s)}+M(s)\Big)\sqrt{\frac{\lambda s\cdot D_3(s)}{\beta_0}}=:D_5(s).\end{aligned}$$

Thus, setting $K(s,\alpha)=K_0(s,\alpha)+K_1(s,\alpha)$, one obtains:

$$\|y_\alpha(s)-y(s)\|^2+4\beta_0\int_{s_{0,\alpha}}^{s}\|\dot{y}_\alpha(\tau)-\dot{y}(\tau)\|^2\,d\tau\leq 2D_1^2(\alpha)+\frac{8D_5(s)}{\alpha+1}+4K(s,\alpha). \quad (4.76)$$

In particular,

$$\|\tilde{y}_\alpha(s)-y(s)\|\leq\sqrt{2}D_1(\alpha)+\left[\frac{8D_5(s)}{\alpha+1}+4K(s,\alpha)\right]^{1/2}\quad\text{for all } s\geq s_{0,\alpha}. \quad (4.77)$$

Inequality (4.77) remains valid for all $s\geq 0$ when combining (4.56) and (4.77). The supremum in (4.77) is positive and bounded on any compact subset of $[0,+\infty)$. Moreover,

$$\lim_{\alpha\to+\infty}\left(\sqrt{2}D_1(\alpha)+\left[\frac{8D_5(s)}{\alpha+1}+4K(s,\alpha)\right]^{1/2}\right)=0\quad\text{if and only if}\quad x_1=0.$$

This establishes the uniform convergence of $(\tilde{y}_\alpha)_\alpha$ to y on any compact subset of $[0,+\infty)$ when $x_1=0$.

Strong convergence in L^2: For the strong convergence of derivatives, observe that

$$\int_0^T\|\dot{\tilde{y}}_\alpha(\tau)-\dot{y}(\tau)\|^2\,d\tau=\int_0^{s_{0,\alpha}}\|\dot{\tilde{y}}_\alpha(\tau)-\dot{y}(\tau)\|^2\,d\tau+\int_{s_{0,\alpha}}^T\|\dot{\tilde{y}}_\alpha(\tau)-\dot{y}(\tau)\|^2\,d\tau.$$

Licence CC BY-NC-ND 3.0

Using (4.76) combined with (4.52), it follows that:

$$\int_0^T \|\dot{\tilde{y}}_\alpha(\tau) - \dot{y}(\tau)\|^2 \, d\tau \leq \int_0^{s_{0,\alpha}} \|\dot{y}(\tau)\|^2 \, d\tau + \frac{1}{4\beta_0}\left(2D_1^2(\alpha) + \frac{8D_5(T)}{\alpha+1} + 4K(T,\alpha)\right)$$
$$\leq \frac{s_{0,\alpha}}{(\lambda-\beta_0)^2} M^2(t_0^2) + \frac{D_1^2(\alpha)}{2\beta_0} + \frac{2D_5(T)}{\beta_0(\alpha+1)} + \frac{1}{\beta_0} K(T,\alpha).$$

Consequently,

$$\lim_{\alpha\to+\infty} \int_0^T \|\dot{\tilde{y}}_\alpha(\tau) - \dot{y}(\tau)\|^2 d\tau = 0 \quad \text{if} \quad x_1 = 0. \tag{4.78}$$

□

The proof of Theorem 4.4.1 highlights several structural aspects of the asymptotic behavior of (ISIHD), which are discussed in the following remark.

Remark 4.4.1. Theorem 4.4.1 represents a progress in the study of accelerated dynamical systems, addressing the delicate case of a dynamics with a forward-looking (anticipated) argument. The particular structure of the (ISIHD) equation, where the operator V is evaluated at a predictive point $x_\alpha(t)+\beta_0\frac{\alpha+1}{t}\dot{x}_\alpha(t)$, introduces specific analytical challenges that distinguish this approach from previously studied dynamics. Moreover, it is important to recall that Theorem 4.4.1 extends beyond cocoercive operators to any monotone and Lipschitz operator admitting at least one zero.

The assumption $x_1 = 0$, although restrictive, is necessary to control initial singularities and ensure the uniform energy estimates used in the convergence proof. The question of whether this condition can be relaxed constitutes an interesting avenue for further research.

The strong convergence of the derivatives in L^2, demonstrated in (4.78), goes beyond the simple convergence of trajectories. It ensures that the (ISIHD) dynamics closely follows the limiting solution both in position and velocity, which can be essential for the stability of numerical discretization schemes.

The implicit structure of the limiting equation (4.51), where $\dot{y}$ appears inside the argument of V, opens promising theoretical perspectives. The condition $2\beta_0 < \lambda$, as in the previous sections, reflects the sensitivity of this dynamics to the parameters. The anticipation introduced by $\beta_0\dot{y}_\alpha(s)$ creates a feedback loop that can amplify instabilities if β_0 is too large. This constraint ensures stability and allows uniform energy control.

The comparison with the previous sections highlights a hierarchy in the complexity of the dynamics: Section 4.2 dealt with the standard case, Section 4.3 introduced explicit regulation, and this section addresses the most sophisticated case with forward-looking evaluation. This progression illustrates the flexibility of the analytical framework in handling dynamics of increasing complexity.

Licence CC BY-NC-ND 3.0

Finally, the techniques developed, notably the precise control of the terms $T_1(\alpha, s)$ and $T_2(\alpha, s)$, demonstrate the combined effectiveness of integration by parts, energy estimates, and regularity controls.

4.5 Numerical Experiment

The aim of this section is to provide a set of numerical experiments that complement and illustrate the theoretical results established above. All rescaled continuous-time dynamics associated with the three inertial systems are integrated using suitably adapted Runge-Kutta schemes, ensuring a faithful numerical approximation of the corresponding flows. The experiments are designed to highlight both the qualitative differences between the systems and the convergence properties predicted by the asymptotic analysis.

Example 4.5.1. To illustrate the behaviour of the three inertial systems studied in this chapter, consider a simple non-potential cocoercive operator $V : \mathbb{R}^n \to \mathbb{R}^n$ of the form

$$V(x) = Bx + \nabla f(x), \qquad x \in \mathbb{R}^n,$$

where $B \in \mathbb{R}^{n\times n}$ is linear and $f : \mathbb{R}^n \to \mathbb{R}$ is smooth, convex, with globally Lipschitz continuous gradient. For the linear part, the choice is made as

$$B = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix}.$$

Then, for any $x, y \in \mathbb{R}^n$,

$$\begin{aligned} \langle Bx - By, x - y\rangle &= \langle B(x-y), x-y\rangle \\ &= (x_1 - y_1)^2 + (x_2 - y_2)^2 + \sum_{i=3}^{n}(x_i - y_i)^2 \\ &\geq \frac{1}{2}\Big(2\big((x_1 - y_1)^2 + (x_2 - y_2)^2\big) + (x_3 - y_3)^2 + \cdots + (x_n - y_n)^2\Big) \\ &= \frac{1}{2}\,\|B(x-y)\|^2, \end{aligned}$$

so B is $\frac{1}{2}$-cocoercive and, in particular, 2-Lipschitz continuous. The zero set $\{x \in \mathbb{R}^n : Bx = 0\}$ reduces to $\{0\}$.

Licence CC BY-NC-ND 3.0

To introduce a controlled nonlinearity while preserving cocoercivity, the system is modified by adding a smooth convex perturbation $f : \mathbb{R}^n \to \mathbb{R}$ with globally Lipschitz gradient, namely

$$f(x) = \varphi(x_1) + \frac{1}{2}\sum_{i=2}^{n} x_i^2, \qquad \varphi(t) = \frac{1}{\gamma}\big(\sqrt{1+\gamma t^2} - 1\big), \quad \gamma > 0.$$

A direct computation shows that

$$\varphi'(t) = \frac{t}{\sqrt{1+\gamma t^2}}, \qquad \varphi''(t) = \frac{1}{(1+\gamma t^2)^{3/2}} \in (0, 1],$$

so φ is convex but not strongly convex and $\sup_{t\in\mathbb{R}} \varphi''(t) = 1$. Therefore

$$\nabla f(x) = \begin{pmatrix} \dfrac{x_1}{\sqrt{1+\gamma x_1^2}} \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \nabla^2 f(x) = \operatorname{diag}\Big(\varphi''(x_1), 1, \ldots, 1\Big),$$

and

$$\|\nabla^2 f(x)\| = \max\{\varphi''(x_1), 1\} \leq 1 \quad \forall x \in \mathbb{R}^n.$$

Hence ∇f is globally 1-Lipschitz and therefore 1-cocoercive.

Cocoercivity and equilibrium: The operator B is $\frac{1}{2}$-cocoercive and ∇f is 1-cocoercive. By [30, Lemma A.1] on the sum of cocoercive operators,

$$V(x) = Bx + \nabla f(x)$$

is again cocoercive, with coefficient $\lambda = \frac{1}{3}$. In particular, V is monotone and globally Lipschitz on $\mathbb{R}^n$, and $x^* = 0$ is an equilibrium satisfying $V(x^*) = 0$.

Licence CC BY-NC-ND 3.0

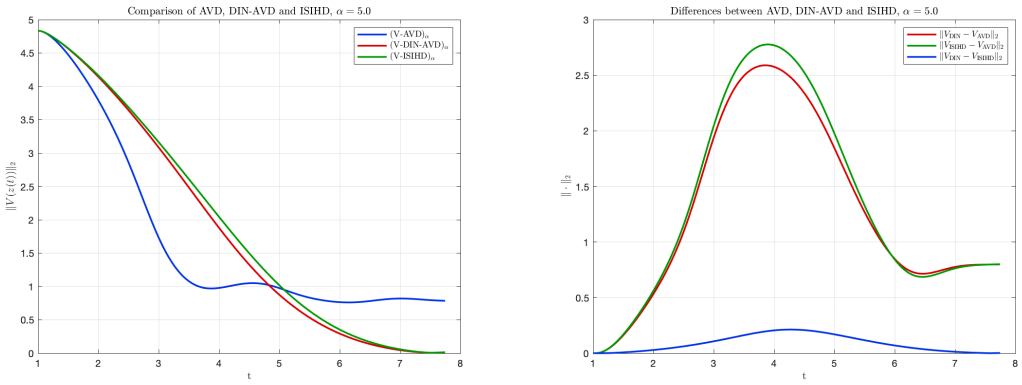


Figure 4.1: Comparison of the three inertial systems (AVD), $(\mathrm{DIN}-\mathrm{AVD})$ and (ISIHD) for the non-potential cocoercive operator introduced in Example 4.5.1. Left: trajectories $t \mapsto \|V(z(t))\|_2$. Right: pairwise differences $t \mapsto \|V_1(t) - V_2(t)\|_2$.

Comparison of the three inertial systems: Figure 4.1 (left) compares the trajectories of (AVD), $(\mathrm{DIN}-\mathrm{AVD})$ and (ISIHD) in the original time variable t, through $t \mapsto \|V(z(t))\|_2$. For a fixed α, the three systems exhibit distinct transient regimes: the (AVD) flow decays more rapidly at the beginning, while the two corrected flows ((DIN-AVD) and (ISIHD)) evolve more smoothly and remain close to each other. Figure 4.1 (right) shows the pairwise differences $t \mapsto \|V_1(t) - V_2(t)\|_2$, confirming that the corrected systems depart markedly from (AVD) and, although closer, still generate distinct trajectories.

Asymptotic behaviour through rescaling: The large-α behavior of the three systems is examined using the rescaled trajectories $y_\alpha(s)$ induced by the change of variables $t = \sqrt{2(\alpha+1)}s$. Figures 4.2–4.4 display, for each system, the evolution of $\|V(y_\alpha(s))\|_2$ (right panels) and the trajectory error $\|y_\alpha(s) - y(s)\|_2$ (left panels), where y denotes the corresponding limit flow derived in the previous sections. As α increases, the curves $\|V(y_\alpha(s))\|_2$ approach the dashed limit profiles, and the errors $\|y_\alpha(s) - y(s)\|_2$ decrease in rescaled time s, thereby providing a numerical confirmation of the asymptotic convergence results.

Licence CC BY-NC-ND 3.0

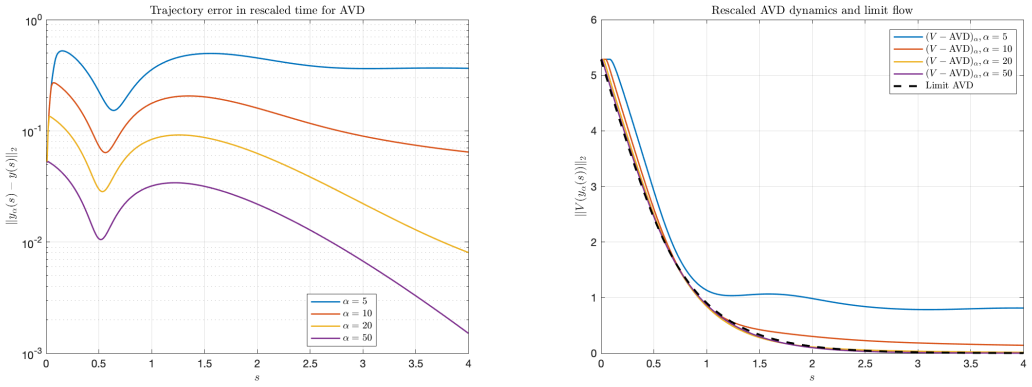


Figure 4.2: Rescaled trajectories $y_\alpha(s)$ for the (AVD) system and their limit dynamics $\dot{y}(s) + V(y(s)) = 0$ as $\alpha \to \infty$. Left: trajectory error $\|y_\alpha(s) - y(s)\|_2$ in rescaled time s. Right: $\|V(y_\alpha(s))\|_2$ for various values of α, compared with the limit flow (black dashed line).

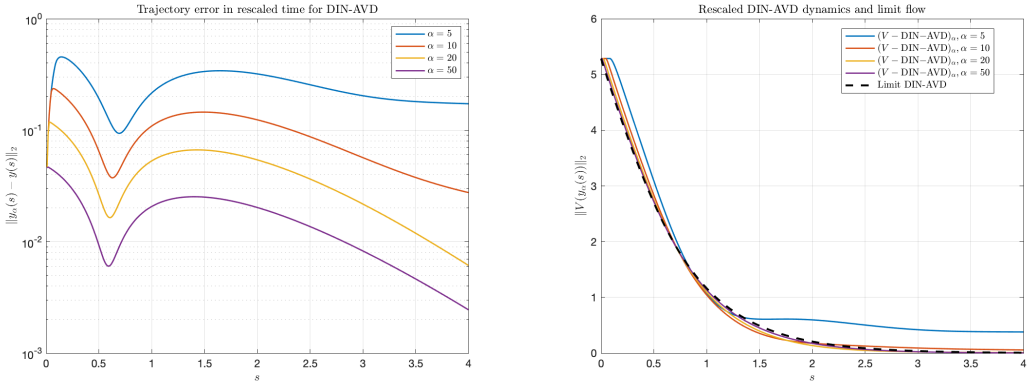


Figure 4.3: Rescaled trajectories $y_\alpha(s)$ for the $(\mathrm{DIN{-}AVD})$ system and their limit dynamics $\dot{y}(s) + V(y(s)) + \beta_0 \frac{d}{ds} V(y(s)) = 0$ as $\alpha \to \infty$, with $\beta_0 = 0.1$. Left: trajectory error $\|y_\alpha(s) - y(s)\|_2$ in rescaled time s. Right: $\|V(y_\alpha(s))\|_2$ for various values of α, compared with the limit flow (black dashed line).

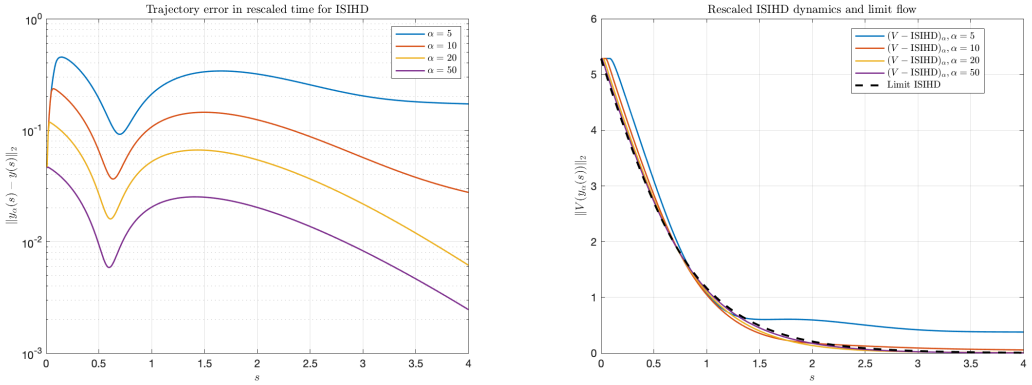


Figure 4.4: Rescaled trajectories $y_\alpha(s)$ for the (ISIHD) system and their limit dynamics $\dot{y}(s) + V(y(s) + \beta_0 \dot{y}(s)) = 0$ as $\alpha \to \infty$, with $\beta_0 = 0.1$. Left: trajectory error $\|y_\alpha(s) - y(s)\|_2$ in rescaled time s. Right: $\|V(y_\alpha(s))\|_2$ for various values of α, compared with the limit flow (black dashed line).

Licence CC BY-NC-ND 3.0

Uniform convergence with respect to α: Figure 4.5 (left) reports the uniform error

$$\max_{s\in[0,S]} \|y_\alpha(s) - y(s)\|_2$$

as a function of α for the three systems. As predicted by the rescaling analysis, this error decreases systematically with α, showing convergence of the trajectories towards their respective limit flows on each compact interval. The (AVD) dynamics exhibits a slower decay, while the corrected flows (DIN-AVD) and (ISIHD) achieve significantly smaller errors for the same α and remain very close to each other, reflecting the stabilising influence of the additional first-order correction terms.

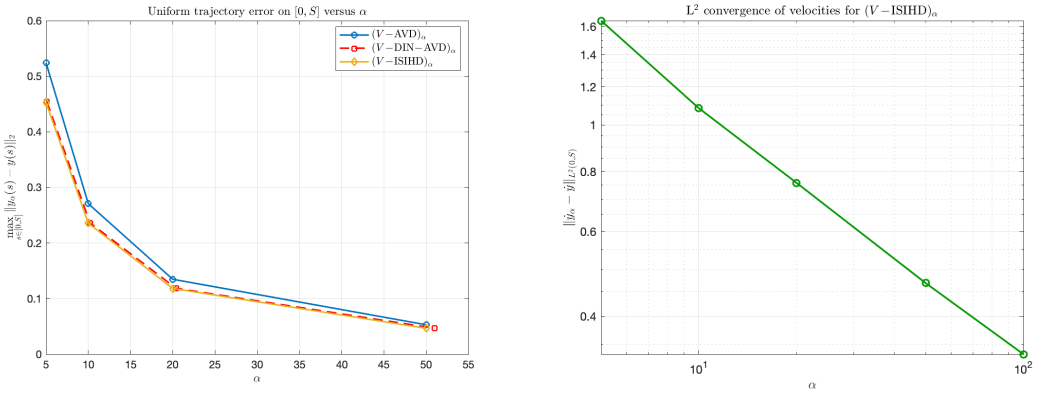


Figure 4.5: (Left) Uniform trajectory error $\max_{s\in[0,S]} \|y_\alpha(s) - y(s)\|_2$ as a function of α for (AVD), $(\mathrm{DIN}-\mathrm{AVD})$ and (ISIHD). (Right) L^2-error on the rescaled velocities for (ISIHD), namely $\|\dot{y}_\alpha - \dot{y}\|_{L^2(0,S)}$ as a function of α (log–log scale), illustrating the L^2-convergence of the (ISIHD) velocities towards the limit first-order flow.

L^2-convergence of velocities for (ISIHD): To illustrate the velocity convergence in Theorem 4.4.1, the rescaled velocities $\dot{y}_\alpha(s)$ associated with (ISIHD) are computed and compared with the limit velocity $\dot{y}(s)$ solving the implicit first-order system $\dot{y}(s) + V(y(s) + \beta_0\dot{y}(s)) = 0$. Figure 4.5 (right) displays $\|\dot{y}_\alpha - \dot{y}\|_{L^2(0,S)}$ as a function of α in log–log scale. The error decreases steadily as α increases, in agreement with the asymptotic analysis, and confirms that the rescaled (ISIHD) velocities converge to their limit profile in the $L^2(0, S)$ sense.

Comparison of the three limit flows: Finally, a comparison of the limit trajectories associated with the three systems is carried out. Under the assumptions of the corresponding convergence theorems (namely $2\beta_0 < \lambda$ for (DIN-AVD) and (ISIHD)), the rescaled trajectories y_α converge to three distinct first-order limit dynamics: the (AVD) limit y_{AVD}, the (DIN-AVD) limit y_{DIN}, and the (ISIHD) limit y_{ISI}. Figure 4.6 shows the quantities $s \mapsto \|y_{\mathrm{DIN}}(s) - y_{\mathrm{AVD}}(s)\|_2$, $\|y_{\mathrm{ISI}}(s) - y_{\mathrm{AVD}}(s)\|_2$ and $\|y_{\mathrm{DIN}}(s) - y_{\mathrm{ISI}}(s)\|_2$. The (AVD) limit flow differs significantly from the two corrected limits, which remain closer to each other. However, the (DIN-AVD) and (ISIHD) limits are not identical: their difference, although smaller in magnitude, is clearly

Licence CC BY-NC-ND 3.0

nonzero. This shows that the three inertial systems lead to distinct limit dynamics, despite sharing the same asymptotic time rescaling.

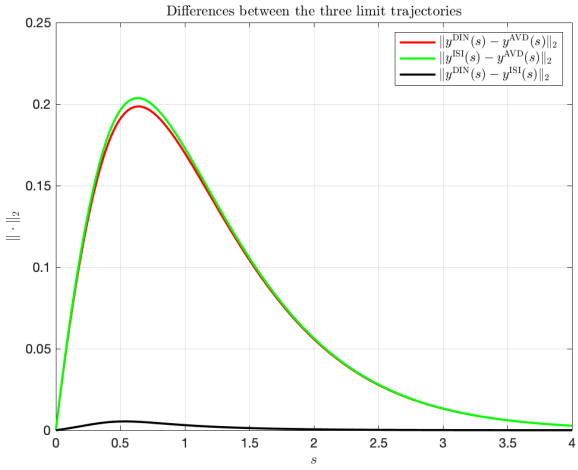


Figure 4.6: Pairwise differences between the three limit flows in rescaled time: $\|y_{\mathrm{DIN}}(s) - y_{\mathrm{AVD}}(s)\|_2$ (red), $\|y_{\mathrm{ISI}}(s) - y_{\mathrm{AVD}}(s)\|_2$ (green), and $\|y_{\mathrm{DIN}}(s) - y_{\mathrm{ISI}}(s)\|_2$ (black). The (AVD) limit departs significantly from the two corrected flows, while (DIN-AVD) and (ISIHD) remain close but not identical.

4.6 Conclusion and Perspectives

This study has systematically examined the asymptotic behavior of three classes of non-potential inertial dynamics as the damping parameter α tends to infinity. Furthermore, the three main convergence results remain valid for any monotone and Lipschitz operator, not only for cocoercive operators, provided that it admits at least one zero. The main results show that, under appropriate conditions, the trajectories of the systems (AVD), (DIN-AVD) and (ISIHD) converge uniformly to the solutions of first-order differential equations after a suitable time reparametrization. The time transformation $\tau(s) = \sqrt{2(\alpha+1)s}$ allows the limiting structure of these complex systems to be identified. It moderates the inertial effect and exposes the underlying first-order dynamics governing the asymptotic behavior. Convergence requires specific constraints on the parameters: $\lambda > 2\beta_0$ for the finite-dimensional (DIN-AVD) dynamics, and $x_1 = 0$ together with $2\beta_0 < \lambda$ for the implicit dynamics (ISIHD). These conditions ensure system stability and allow control of the inertial terms in the limiting process.

The methodology employed, based on suitable Lyapunov functions and refined energy estimates, has demonstrated its robustness for handling these different classes of dynamics. The analysis highlights how the structure of each system influences both the convergence conditions and the regularity of the limiting solutions. In particular, the strong convergence in L^2 for the dynamics (ISIHD) illustrates the additional regularity induced by the implicit nature of the equation.

Licence CC BY-NC-ND 3.0

Research Perspectives

Several research directions naturally emerge from this work, opening promising avenues for further theoretical and applied developments:

- *Extension to infinite-dimensional settings*: A natural generalization would be to extend the results for (DIN-AVD) to general Hilbert spaces. This would require developing compactness tools suitable for infinite dimensions and a fine analysis of the regularity properties of the solutions.
- *Development of numerical schemes*: Designing discretization methods that preserve the asymptotic properties identified in this study would be a significant contribution. Particular attention should be given to the compatibility between temporal discretizations and the limit $\alpha \to \infty$.
- *Generalization to more complex operators*: The analysis could be extended to multivalued operators, non-monotone operators, or operators with stronger nonlinearities, which would present a substantial theoretical challenge. Such generalizations would be relevant for modeling complex systems arising in mathematical physics and engineering.
- *Study of stochastic systems*: Extending the analysis to stochastic settings, where the operator V is subject to random perturbations, represents an interesting direction with applications in machine learning and stochastic optimization.
- *Analysis of transient phenomena*: Beyond asymptotic behavior, studying transient regimes and boundary layers emerging in the limit $\alpha \to \infty$ constitutes a relevant line of investigation.

Licence CC BY-NC-ND 3.0

5

Study of the acceleration parameter

Contents

5.1 Introduction **124**
5.1.1 Main results 125
5.1.2 Chapter Contents 126
5.2 Analysis of the (AVD) dynamics: the quadratic convex case **127**
5.3 Numerical experiments **142**
5.4 Conclusion and Perspectives **149**

Licence CC BY-NC-ND 3.0

5.1 Introduction

Consider the problem of minimizing a real-valued function $f : \mathcal{H} \to \mathbb{R}$, where $\mathcal{H}$ is a Hilbert space. Assume that f is differentiable with an L-Lipschitz continuous gradient and that it admits at least one minimizer $x^* \in \mathcal{H}$, so that the minimum

$$\inf(f) = \min_{x \in \mathcal{H}} f(x)$$

is well-defined. The introduction of second-order-in-time evolution equations with damping in optimization dates back to the work of B. Polyak [2], [31]. In 1964, he introduced the *heavy ball with friction* (HBF) method:

$$\text{(HBF)} \quad \ddot{x}(t) + \gamma_o \dot{x}(t) + \nabla f(x(t)) = 0, \tag{5.1}$$

where $\gamma_o > 0$ is a fixed positive number. This autonomous system is particularly useful for exploring local minima of f (see [32] for more details). HBF corresponds to taking a constant damping coefficient $\gamma(t) \equiv \gamma_o > 0$ in the more general inertial gradient system:

$$\text{(IGS)}_\gamma \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla f(x(t)) = 0, \ t \geq a_0 \tag{5.2}$$

with $a_0 > 0$ or $a_0 = 0$ depending on the singularity of $\gamma(t)$. In the context of mechanics and physics, $\gamma(t)$ represents a viscous friction coefficient. A positive damping coefficient $\gamma(t) > 0$ induces dissipation of the total system energy, playing a role in it optimization properties. For convex functions with low curvature, this friction can neutralize the inertial effect, preventing fast optimization convergence rates. Indeed, for a general convex function f, choosing a damping parameter $\gamma(t)$ that tends to zero as $t \to +\infty$ (vanishing damping) at a moderate rate allows the derivation of fast optimization methods. One can refer to [33], [34], [35], [36], where the role of the assumption

$$\int_{a_0}^{+\infty} \gamma(t)\, dt = +\infty$$

is demonstrated. In this context, a $\gamma(t)$ of order $1/t$ is the smallest admissible vanishing damping coefficient, which explains the importance of this case for achieving fast optimization methods. This highlights the interest of taking a damping coefficient of the form $\gamma(t) = \frac{\alpha}{t}$. Indeed, the corresponding dynamics is the Su–Boyd–Candès dynamics [9]:

$$\text{(AVD)} : \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0, \ t \geq t_0 \qquad x(t_0) = x_0, \quad \dot{x}(t_0) = x_1, \tag{5.3}$$

where $(x_0, x_1) \in \mathcal{H}^2$ and $t_0 > 0$ as the initial time for the studied dynamics, which is justified by the singularity of the damping coefficient $\frac{\alpha}{t}$ at the origin. The behavior of (AVD) is

Licence CC BY-NC-ND 3.0

influenced by the parameter α, and its analysis with respect to the values of this parameter leads to the distinction of three separate regimes. In the under-critical regime ($0 < \alpha \leq 3$), the convergence in function value follows a rate of $\mathcal{O}(1/t^{2\alpha/3})$ [7], [8]. The critical case ($\alpha = 3$), which corresponds to Nesterov's method, ensures optimal convergence of $\mathcal{O}(1/t^2)$ [9]. Finally, in the over-critical regime ($\alpha > 3$), accelerated convergence of $o(1/t^2)$ is observed [5], along with weak convergence of the trajectories $(x(t))$ to a minimizer of f [6]. This classification highlights the importance of the choice of α in accelerated optimization.

The analysis of the (AVD) dynamics constitutes the main focus of this chapter. Particular attention is given to the choice of the parameter α, which plays a role in ensuring fast convergence. The present work is conducted within this framework and aims to provide a detailed theoretical and numerical study of these dynamics.

5.1.1 Main results

In this chapter, the influence of the parameter α on the convergence properties of the accelerated dynamics

$$(\mathrm{AVD}): \quad \ddot{x}_\alpha(t) + \frac{\alpha}{t}\dot{x}_\alpha(t) + \nabla f(x_\alpha(t)) = 0,\ t \geq t_0, \qquad x_\alpha(t_0) = x_0, \quad \dot{x}_\alpha(t_0) = x_1$$

is investigated. The objective is to determine the values of α that yield the fastest convergence while preserving the stability of the trajectories.
The analysis focuses on the convergence rates when the objective function f is convex and quadratic. Recall that if f is a convex and differentiable function on $\mathbb{R}^n$, satisfying assumptions $H_1(\gamma)$ (with $\gamma \in [1, 2]$) and $H_2(2)$ (see Definition 2.3.9), and admitting a unique minimizer, then the solution $x_\alpha(\cdot)$ of the (AVD) system satisfies (see [22, Theorem 4.2])

$$f(x_\alpha(t)) - \inf(f) = \mathcal{O}\left(\frac{1}{t^\alpha}\right),$$

and in particular this holds for strongly convex quadratic functions. Moreover, this decay rate is optimal in the sense that, for any $\gamma \in (1, 2]$, it is attained for the function $f(x) = \|x\|^\gamma$ under the initial conditions $\|x_0\| < 1$ and $x_1 = 0$ (see [22, Proposition 4.3]). In the convex quadratic case (not strongly convex), where the minimizer is not necessarily unique, it is shown that the same convergence rate can still be achieved. More precisely, it is established that

$$f(x_\alpha(t)) - \inf(f) = \mathcal{O}\left(\frac{1}{t^\alpha}\right),$$

and that this rate remains optimal for convex quadratic functions (not necessarily strongly convex) admitting at least one minimizer, without any restriction on the initial conditions.

Licence CC BY-NC-ND 3.0

Moreover, strong convergence of the trajectories holds, with the estimate

$$\|x_\alpha(t) - x_{\alpha,\infty}\|^2 = \mathcal{O}\Big(\frac{1}{t^\alpha}\Big),$$

where $x_{\alpha,\infty}$ denotes a minimizer of the convex quadratic function f. These results indicate that, in the convex quadratic setting, larger values of the parameter α correspond to faster convergence in terms of both objective values and trajectories.
To derive these results, the analysis is restricted to the class of convex quadratic functions, which provides an analytically tractable framework for the study of accelerated dynamics. In this setting, the (AVD) dynamics can be analyzed in detail, in particular through the exact solution of the Su–Boyd–Candès differential equation. The study of these solutions involves Bessel functions, whose asymptotic behavior accounts for the acceleration mechanisms associated with the parameter α.
The convex quadratic framework is analytically tractable and frequently used in optimization theory. Quadratic functions appear naturally as local second-order approximations of general convex functions via Taylor expansion, providing a simplified model for studying the behavior of optimization algorithms.
Finally, this choice is motivated by practical considerations. A wide range of problems in machine learning and engineering can be formulated as the minimization of quadratic functions, including least-squares regression, principal component analysis (PCA), and several regularized linear models.
Altogether, the quadratic convex setting enables us to obtain sharp theoretical results while maintaining broad applicability, and provides a clear understanding of the role played by the parameter α in the (AVD) dynamics.

5.1.2 Chapter Contents

This chapter is organized into two sections. The first section (Section 5.2) presents a detailed study of the Su–Boyd–Candès accelerated dynamics (AVD) in the case of convex quadratic functions. The explicit solution of the associated differential equation is derived, and a complete asymptotic analysis is performed, which addresses the determination of the optimal parameter α for sufficiently large times. The second section (Section 5.3) presents numerical simulations illustrating the theoretical results.
This progressive structure enables us to investigate the choice of the parameter α from complementary perspectives, ranging from a refined theoretical analysis to practical validation through numerical experiments.

Licence CC BY-NC-ND 3.0

5.2 Analysis of the (AVD) dynamics: the quadratic convex case

This section is devoted to a detailed analysis of the accelerated dynamics (AVD) in the setting of quadratic convex functions. This particular framework allows us to derive explicit solutions and to carry out a precise study of the influence of the parameter α on the convergence properties of the dynamics. In particular, the analytical representation of the solutions in terms of Bessel functions makes it possible to establish sharp convergence rates for the accelerated dynamics. Throughout this section, the analysis is restricted to the class of convex quadratic functions defined by

$$f(x) = \frac{1}{2}\langle Qx, x\rangle + \langle b, x\rangle + d_0, \tag{5.4}$$

where $x \in \mathbb{R}^n$, $d_0 \in \mathbb{R}$ is a constant, $Q \in \mathbb{R}^{n\times n}$ is a symmetric positive semidefinite matrix ($Q = Q^\top \succeq 0$), and $b \in \mathbb{R}^n$ is a given vector. A necessary and sufficient condition for the function f to admit a finite minimum is that the vector b belongs to the range of the matrix Q. Indeed, if $b \notin \mathrm{Im}(Q)$, it is possible to construct a diverging minimizing sequence. More precisely, decomposing

$$b = b_1 + b_2, \quad \text{with } b_1 \in \mathrm{Im}(Q), \quad b_2 \in \mathrm{Im}(Q)^\perp \setminus \{0\},$$

(where $\mathrm{Im}(Q)^\perp$ is the orthogonal complement of $\mathrm{Im}(Q)$) and considering the sequence $(x_k)_{k\in\mathbb{N}}$ defined by $x_k = b_1 - kb_2$, one obtains

$$f(x_k) = \frac{1}{2}\langle b_1, Qb_1\rangle + \langle b, b_1\rangle + d_0 - k\|b_2\|^2 \xrightarrow[k\to+\infty]{} -\infty.$$

Thus, when $b \notin \mathrm{Im}(Q)$, it follows that $\inf(f) = -\infty$.

Under the assumption $b \in \mathrm{Im}(Q)$, the convergence rate of the objective function values is of order $\mathcal{O}(t^{-\alpha})$. This rate is optimal, in the sense that it cannot be improved. Moreover, the same order of convergence is obtained for the trajectories. The analysis relies on the spectral structure of the matrix Q, revealed through a preliminary decoupling of the dynamical system, as well as on the explicit characterization of the resulting trajectories. These results then allow us to state the main theorem in the quadratic convex setting.

Theorem 5.2.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2}\langle \mathrm{Q}x, x\rangle + \langle b, x\rangle + d_0,$$

Licence CC BY-NC-ND 3.0

where $d_0 \in \mathbb{R}$, $\mathrm{Q} \in \mathbb{R}^{n\times n}$ is a symmetric positive semi-definite matrix, and $b \in \mathrm{Im}(\mathrm{Q})$. Consider the accelerated dynamics

$$\ddot{x}_\alpha(t) + \frac{\alpha}{t}\dot{x}_\alpha(t) + \nabla f(x_\alpha(t)) = 0, \quad x_\alpha(t_0) = x_0, \quad \dot{x}_\alpha(t_0) = x_1. \tag{5.5}$$

Then, the solution $x_\alpha(\cdot)$ of (5.5) satisfies:

1. Convergence rate of function values (for $\alpha > 1$):

$$f(x_\alpha(t)) - \inf(f) = \mathcal{O}\left(\frac{1}{t^\alpha}\right).$$

2. Optimality of the rate: there exists no $\beta > \alpha$ such that

$$f(x_\alpha(t)) - \inf(f) = \mathcal{O}\left(\frac{1}{t^\beta}\right).$$

   In other words, the rate $\mathcal{O}(t^{-\alpha})$ is optimal for this dynamics.

3. Convergence rate of trajectories (for $\alpha \geq 2$):

$$\|x_\alpha(t) - x_{\alpha,\infty}\|^2 = \mathcal{O}\left(\frac{1}{t^\alpha}\right),$$

   where $x_{\alpha,\infty}$ denotes a minimizer of f.

The proof of Theorem 5.2.1 is based on a detailed analysis of the dynamics (5.5) in the convex quadratic case, made possible by a spectral decoupling of the system. To structure the proof and isolate the technical arguments, three preliminary lemmas are introduced.
The first lemma provides the explicit expression of the solutions of the linear differential equation associated with each spectral mode of the accelerated dynamics. It shows the involvement of Bessel functions in the description of the trajectories.

Lemma 5.2.2. Let $\alpha > 1$. Consider the ordinary differential equation

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \gamma x(t) = 0, \quad x(t_0) = a_0, \quad \dot{x}(t_0) = a_1, \tag{5.6}$$

where $a_0, a_1 \in \mathbb{R}$, $t_0 > 0$, and $\gamma \geq 0$. Then the solution $x(\cdot)$ is given by

$$\begin{cases} x(t) = \dfrac{1}{1-\alpha}\left(a_1 t_0^\alpha t^{1-\alpha} - a_1 t_0 + (1-\alpha)a_0\right), & \text{if } \gamma = 0, \\ x(t) = t^{\frac{1-\alpha}{2}}\left[A_1 J_\nu(t\sqrt{\gamma}) + A_2 Y_\nu(t\sqrt{\gamma})\right], & \text{if } \gamma > 0, \end{cases} \tag{5.7}$$

where J_ν and Y_ν are the Bessel functions (of order $\nu = \frac{\alpha-1}{2}$) of the first and second kind, respectively, and the constants A_1, A_2 are determined by the initial conditions.

Licence CC BY-NC-ND 3.0

Proof. The proof proceeds in two steps according to the nature of the constant γ, exploiting the particular structure of each case.

Case $\gamma = 0$: The equation reduces to:

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) = 0$$

By setting $U(t) = \dot{x}(t)$, a first-order linear ODE is obtained, with solution given by:

$$U(t) = a_1 t_0^{\alpha} t^{-\alpha}$$

Integration directly gives the announced form:

$$x(t) = \frac{1}{1-\alpha}[a_1 t_0^{\alpha} t^{1-\alpha} - a_1 t_0 + (1-\alpha)a_0] \tag{5.8}$$

Case $\gamma > 0$: The differential equation (5.6) is solved by applying a suitable transformation. Consider the transformation:

$$Z(u) = u^q x(u/\sqrt{\gamma}) \tag{5.9}$$

where $q \in \mathbb{R}$ will be determined later. Differentiating Z with respect to u gives:

$$\dot{Z}(u) = qu^{q-1}x(u/\sqrt{\gamma}) + \frac{u^q}{\sqrt{\gamma}}\dot{x}(u/\sqrt{\gamma})$$

Multiplying by u:

$$u\dot{Z}(u) = qu^q x(u/\sqrt{\gamma}) + \frac{u^{q+1}}{\sqrt{\gamma}}\dot{x}(u/\sqrt{\gamma}) \tag{5.10}$$

Differentiating (5.10) again:

$$u\ddot{Z}(u) + (1-q)\dot{Z}(u) = (q+1)\frac{u^q}{\sqrt{\gamma}}\dot{x}(u/\sqrt{\gamma}) + \frac{u^{q+1}}{\gamma}\ddot{x}(u/\sqrt{\gamma})$$

Multiplying by u and substituting $\ddot{x}(u/\sqrt{\gamma})$ from (5.6):

$$\begin{aligned} u^2\ddot{Z}(u) + (1-q)u\dot{Z}(u) &= (q+1)\frac{u^{q+1}}{\sqrt{\gamma}}\dot{x}(u/\sqrt{\gamma}) + \frac{u^{q+2}}{\gamma}\left(-\frac{\alpha\sqrt{\gamma}}{u}\dot{x}(u/\sqrt{\gamma}) - \gamma x(u/\sqrt{\gamma})\right) \\ &= (q+1-\alpha)\left(u\dot{Z}(u) - qZ(u)\right) - u^2 Z(u) \end{aligned}$$

This leads to:

$$u^2\ddot{Z}(u) + (\alpha - 2q)u\dot{Z}(u) + (u^2 + q^2 + q - q\alpha)Z(u) = 0$$

Licence CC BY-NC-ND 3.0

Selecting q such that

$$\alpha - 2q = 1 \quad \text{and} \quad q^2 + q - q\alpha = -q^2,$$

leads to

$$q = \frac{\alpha - 1}{2}.$$

The equation then reduces to the canonical Bessel form (replacing q by ν):

$$u^2\ddot{Z}(u) + u\dot{Z}(u) + (u^2 - \nu^2)Z(u) = 0 \tag{5.11}$$

The general solution of this equation can be expressed in terms of Bessel functions (see the theorem 2.4.1):

$$Z(u) = A_1 J_\nu(u) + A_2 Y_\nu(u) \tag{5.12}$$

By inverting the change of variable, the final expression is obtained:

$$x(t) = t^{-\frac{\alpha-1}{2}} Z(t\sqrt{\gamma}) = t^{\frac{1-\alpha}{2}} \left[A_1 J_\nu(t\sqrt{\gamma}) + A_2 Y_\nu(t\sqrt{\gamma})\right]$$

where A_1 and A_2 are constants determined by the initial conditions. □

The second lemma recalls the classical asymptotic behavior of Bessel functions as their argument tends to infinity. These expansions are used to establish the asymptotic decay of the solutions and to deduce the convergence rates.

Lemma 5.2.3 (Asymptotic behavior)**.** As $x \to +\infty$, the Bessel functions $J_\nu(x)$ and $Y_\nu(x)$ admit the following asymptotic expansions:

$$J_{\pm\nu}(x) = \sqrt{\frac{2}{\pi x}} \left[F_{\pm\nu}(x) + \mathcal{O}\Big(\frac{1}{x}\Big)\right], \tag{5.13}$$

$$Y_{\pm\nu}(x) = \sqrt{\frac{2}{\pi x}} \left[G_{\pm\nu}(x) + \mathcal{O}\Big(\frac{1}{x}\Big)\right], \tag{5.14}$$

where

$$F_{\pm\nu}(x) = \cos\Big(x \mp \frac{\nu\pi}{2} - \frac{\pi}{4}\Big), \qquad G_{\pm\nu}(x) = \sin\Big(x \mp \frac{\nu\pi}{2} - \frac{\pi}{4}\Big).$$

The signs $\pm$ and $\mp$ are correlated: for $J_{+\nu}$ (respectively $Y_{+\nu}$), take the upper sign in both symbols, i.e., "$+$" in $\pm$ and "$-$" in $\mp$; conversely, for $J_{-\nu}$ (respectively $Y_{-\nu}$), take the lower sign in both, i.e., "$-$" in $\pm$ and "$+$" in $\mp$.

The accuracy of these asymptotic approximations is illustrated in Figures 5.1 and 5.2.

Licence CC BY-NC-ND 3.0

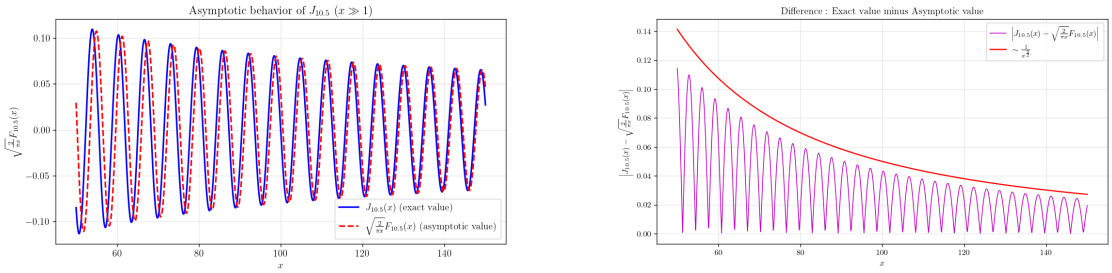


Figure 5.1: Left: the function $J_{10.5}$ and its asymptotic behavior. Right: the difference between the function $J_{10.5}$ and its asymptotic expression.

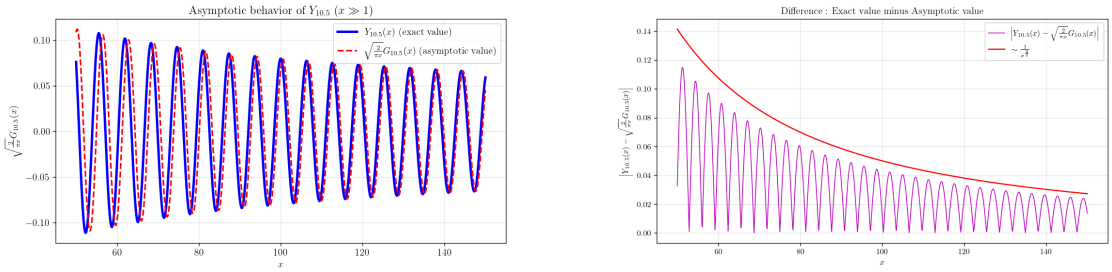


Figure 5.2: Left: the function $Y_{10.5}$ and its asymptotic behavior. Right: the difference between the function $Y_{10.5}$ and its asymptotic expression.

Proof. According to [23] (see page 205), the Bessel functions of the first and second kinds can be expressed as:

$$J_{\pm\nu}(x) = \sqrt{\frac{2}{\pi x}}\left[\cos\left(x \mp \frac{\nu\pi}{2} - \frac{\pi}{4}\right)P(x,\nu) - \sin\left(x \mp \frac{\nu\pi}{2} - \frac{\pi}{4}\right)Q(x,\nu)\right], \tag{5.15}$$

$$Y_{\pm\nu}(x) = \sqrt{\frac{2}{\pi x}}\left[\sin\left(x \mp \frac{\nu\pi}{2} - \frac{\pi}{4}\right)P(x,\nu) + \cos\left(x \mp \frac{\nu\pi}{2} - \frac{\pi}{4}\right)Q(x,\nu)\right], \tag{5.16}$$

where the functions P and Q are defined as follows:

$$P(x,\nu) = \frac{1}{2\Gamma(\nu+\frac{1}{2})}\int_0^{+\infty} e^{-u}u^{\nu-\frac{1}{2}}A(x,u)\,du,$$

$$Q(x,\nu) = \frac{1}{2i\Gamma(\nu+\frac{1}{2})}\int_0^{+\infty} e^{-u}u^{\nu-\frac{1}{2}}B(x,u)\,du,$$

Licence CC BY-NC-ND 3.0

with:

$$A(x,u) = \left(1+\frac{iu}{2x}\right)^{\nu-\frac{1}{2}} + \left(1-\frac{iu}{2x}\right)^{\nu-\frac{1}{2}},$$
$$B(x,u) = \left(1+\frac{iu}{2x}\right)^{\nu-\frac{1}{2}} - \left(1-\frac{iu}{2x}\right)^{\nu-\frac{1}{2}}.$$

To study the asymptotic behavior of these functions, the above expressions are expanded into power series. Let p be an integer such that $2p \geq \nu - \frac{1}{2}$. For large x, consider the function

$$K(u) = \left(1 \pm \frac{iu}{2x}\right)^{\nu-\frac{1}{2}}.$$

It admits a Taylor expansion around $u = 0$ with an integral remainder:

$$K(u) = \sum_{m=0}^{2p-1} \frac{\left(\frac{1}{2}-\nu\right)_m}{m!} \left(\frac{\pm u}{2ix}\right)^m + R_{2p-1}(u),$$

where $(a)_n$ denotes the rising Pochhammer symbol, defined by

$$(a)_n = \begin{cases} a(a+1)\dots(a+n-1), & n \geq 1, \\ 1, & n = 0. \end{cases}$$

The integral remainder $R_{2p-1}(u)$ can be written in its classical form as

$$R_{2p-1}(u) = \frac{\left(\frac{1}{2}-\nu\right)_{2p}}{(2p-1)!} \left(\pm\frac{1}{2ix}\right)^{2p} \int_0^u (u-t)^{2p-1} \left(1 \pm \frac{it}{2x}\right)^{\nu-\frac{1}{2}-2p} dt.$$

By performing the change of variable $t = \epsilon u$, $dt = u\, d\epsilon$, the remainder can be expressed in the normalized form:

$$R_{2p-1}(u) = \frac{\left(\frac{1}{2}-\nu\right)_{2p}}{(2p)!} \left(\pm\frac{u}{2ix}\right)^{2p} \theta_{\pm}(u),$$

where

$$\theta_{\pm}(u) = 2p \int_0^1 (1-\epsilon)^{2p-1} \left(1 \pm \frac{i\epsilon u}{2x}\right)^{\nu-\frac{1}{2}-2p} d\epsilon.$$

The coefficients $(\frac{1}{2}-\nu)_m$ arise directly from the successive derivatives of $K(u)$ and are expressed in terms of the Pochhammer symbol. The function $\theta_{\pm}(u)$ comes from the normalized integral remainder and allows the factor $(u/(2x))^{2p}$ to be isolated, making the order of the remainder explicit. Moreover, the complex symmetry $\overline{\theta_+}(u) = \theta_-(u)$ holds, and if $2p > \nu - \frac{1}{2}$, then

Licence CC BY-NC-ND 3.0

$|\theta_{\pm}(u)| \leq 1$, providing a strict control over the remainder.
Summing the expansions and setting $\theta_0 := \frac{\theta_+ + \theta_-}{2}$ gives:

$$\begin{aligned} A(x,u) &= \sum_{m=0}^{2p-1} \frac{(1+(-1)^m)\left(\frac{1}{2}-\nu\right)_m}{m!}\left(\frac{u}{2ix}\right)^m + \frac{\left(\frac{1}{2}-\nu\right)_{2p}}{(2p)!}\left(\frac{u}{2ix}\right)^{2p}(\theta_+(u)+\theta_-(u)) \\ &= 2\sum_{m=0}^{p-1} \frac{(-1)^m\left(\frac{1}{2}-\nu\right)_{2m}}{(2m)!}\left(\frac{u}{2x}\right)^{2m} + \frac{(-1)^p\, 2\theta_0(u)\left(\frac{1}{2}-\nu\right)_{2p}}{(2p)!}\left(\frac{u}{2x}\right)^{2p}. \end{aligned}$$

This expression for $A(x,u)$ allows us to derive the asymptotic expansion of $P(x,\nu)$ by integrating term by term:

$$\begin{aligned} P(x,\nu) &= \sum_{m=0}^{p-1} \frac{(-1)^m\left(\frac{1}{2}-\nu\right)_{2m}\int_0^{+\infty} e^{-u}u^{2m+\nu-\frac{1}{2}}\,du}{(2m)!\,\Gamma\left(\nu+\frac{1}{2}\right)(2x)^{2m}} \\ &\quad + \frac{(-1)^p\left(\frac{1}{2}-\nu\right)_{2p}}{(2p)!\,\Gamma\left(\nu+\frac{1}{2}\right)(2x)^{2p}}\int_0^{+\infty}\theta_0(u)e^{-u}u^{2p+\nu-\frac{1}{2}}\,du \\ &= \sum_{m=0}^{p-1} \frac{(-1)^m\left(\frac{1}{2}-\nu\right)_{2m}\Gamma\left(2m+\nu+\frac{1}{2}\right)}{(2m)!\,(2x)^{2m}\,\Gamma\left(\nu+\frac{1}{2}\right)} \\ &\quad + \frac{(-1)^p\left(\frac{1}{2}-\nu\right)_{2p}}{(2p)!\,\Gamma\left(\nu+\frac{1}{2}\right)(2x)^{2p}}\int_0^{+\infty}\theta_0(u)e^{-u}u^{2p+\nu-\frac{1}{2}}\,du. \end{aligned}$$

Using the Gamma function property

$$\Gamma(x+1) = x\Gamma(x), \quad \forall x > 0,$$

The function $P(x,\nu)$ can be rewritten more explicitly as:

$$\begin{aligned} P(x,\nu) &= 1 + \sum_{m=1}^{p-1} \frac{(-1)^m\left(\frac{1}{2}-\nu\right)_{2m}\left(\nu+\frac{1}{2}\right)_{2m}}{(2m)!\,(2x)^{2m}} \\ &\quad + \frac{(-1)^p\left(\frac{1}{2}-\nu\right)_{2p}}{(2p)!\,\Gamma\left(\nu+\frac{1}{2}\right)(2x)^{2p}}\int_0^{+\infty}\theta_0(u)e^{-u}u^{2p+\nu-\frac{1}{2}}\,du. \end{aligned}$$

Attention is now focused on the study of the function $Q(x,\nu)$. By choosing an integer p sufficiently large so that $2p+1 \geq \nu - \frac{1}{2}$, and applying Taylor's formula with integral remainder again,

$$\left(1 \pm \frac{iu}{2x}\right)^{\nu-\frac{1}{2}} = \sum_{m=0}^{2p} \frac{\left(\frac{1}{2}-\nu\right)_m}{m!}\left(\frac{\pm u}{2ix}\right)^m + \frac{\theta'_{\pm}\left(\frac{1}{2}-\nu\right)_{2p+1}}{(2p+1)!}\left(\frac{\pm u}{2ix}\right)^{2p+1},$$

Licence CC BY-NC-ND 3.0

where

$$\theta'_{\pm}(u) = (2p+1)\int_0^1 (1-\epsilon)^{2p}\left(1 \pm \frac{i\epsilon u}{2x}\right)^{\nu-\frac{1}{2}-2p-1} d\epsilon,$$

which satisfies $|\theta'_{\pm}| \leq 1$ if $2p+1 > \nu - \frac{1}{2}$. This yields the following expression (setting $\theta'_0 := \frac{\theta'_+ + \theta'_-}{2}$):

$$\begin{aligned} B(x,u) &= \sum_{m=0}^{2p} \frac{(1-(-1)^m)\left(\frac{1}{2}-\nu\right)_m}{(m)!}\left(\frac{u}{2ix}\right)^m \\ &\quad + \frac{\left(\frac{1}{2}-\nu\right)_{2p+1}}{(2p+1)!}\left(\frac{u}{2ix}\right)^{2p+1}(\theta'_+(u)+\theta'_-(u)) \\ &= 2i\sum_{m=0}^{p-1} \frac{(-1)^{m+1}\left(\frac{1}{2}-\nu\right)_{2m+1}}{(2m+1)!}\left(\frac{u}{2x}\right)^{2m+1} \\ &\quad + \frac{2i(-1)^{p+1}\theta'_0(u)\left(\frac{1}{2}-\nu\right)_{2p+1}}{(2p+1)!}\left(\frac{u}{2x}\right)^{2p+1}. \end{aligned}$$

Proceeding analogously to the approach used for $P(x,\nu)$, the following expression is obtained:

$$\begin{aligned} Q(x,\nu) &= \sum_{m=0}^{p-1} \frac{(-1)^{m+1}\left(\frac{1}{2}-\nu\right)_{2m+1}\left(\frac{1}{2}+\nu\right)_{2m+1}}{(2m+1)!\,(2x)^{2m+1}} \\ &\quad + \frac{(-1)^{p+1}\left(\frac{1}{2}-\nu\right)_{2p+1}}{(2p+1)!\,\Gamma\left(\nu+\frac{1}{2}\right)(2x)^{2p+1}}\int_0^{+\infty}\theta'_0(u)e^{-u}u^{2p+\nu+\frac{1}{2}}\,du, \end{aligned}$$

A careful analysis of the preceding terms shows that the functions

$$R(x,\nu) := x\big(P(x,\nu)-1\big) \quad \text{and} \quad xQ(x,\nu)$$

remain bounded on the interval $[1,+\infty)$. Indeed, for $x > 1$

$$\begin{aligned} |R(x,\nu)| &\leq \sum_{m=1}^{p-1} \frac{\left(\frac{1}{2}-\nu\right)_{2m}\left(\nu+\frac{1}{2}\right)_{2m}}{(2m)!} + \frac{\left(\frac{1}{2}-\nu\right)_{2p}}{(2p)!\,\Gamma\left(\nu+\frac{1}{2}\right)}\int_0^{+\infty}|\theta_0(u)|\,e^{-u}u^{2p+\nu-\frac{1}{2}}\,du \\ &\leq \sum_{m=1}^{p-1} \frac{\left(\frac{1}{2}-\nu\right)_{2m}\left(\nu+\frac{1}{2}\right)_{2m}}{(2m)!} + \frac{\left(\frac{1}{2}-\nu\right)_{2p}}{(2p)!\,\Gamma\left(\nu+\frac{1}{2}\right)}\Gamma\left(2p+\nu+\frac{1}{2}\right), \end{aligned}$$

Licence CC BY-NC-ND 3.0

since $|\theta_0(u)| \leq 1$. And similarly,

$$\begin{aligned}|xQ(x,\nu)| &\leq \sum_{m=0}^{p-1} \frac{\left(\frac{1}{2}-\nu\right)_{2m+1}\left(\frac{1}{2}+\nu\right)_{2m+1}}{(2m+1)!} \\ &\quad + \frac{\left(\frac{1}{2}-\nu\right)_{2p+1}}{(2p+1)!\,\Gamma\left(\nu+\frac{1}{2}\right)} \int_0^{+\infty} |\theta_0'(u)|\, e^{-u} u^{2p+\nu+\frac{1}{2}}\, du \\ &\leq \sum_{m=0}^{p-1} \frac{\left(\frac{1}{2}-\nu\right)_{2m+1}\left(\frac{1}{2}+\nu\right)_{2m+1}}{(2m+1)!} + \frac{\left(\frac{1}{2}-\nu\right)_{2p+1}}{(2p+1)!\,\Gamma\left(\nu+\frac{1}{2}\right)} \Gamma\left(2p+\nu+\frac{3}{2}\right),\end{aligned}$$

since $|\theta_0'(u)| \leq 1$.
It follows that $Q(x,\nu)$ and $\frac{R(x,\nu)}{x}$ are of order $\mathcal{O}\left(\frac{1}{x}\right)$ for $x > 1$. Consequently, any linear combination of these functions with trigonometric functions, in particular sine and cosine, also remains of order $\mathcal{O}\left(\frac{1}{x}\right)$ when $x > 1$.
Based on (5.15) and (5.16), the Bessel functions can be expressed in terms of $R(x,\nu)$, as follows:

$$J_{\pm\nu}(x) = \sqrt{\frac{2}{\pi x}}\left[\cos\left(x \mp \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right) + \cos\left(x \mp \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right)\frac{R(x,\nu)}{x} - \sin\left(x \mp \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right) Q(x,\nu)\right],$$

$$Y_{\pm\nu}(x) = \sqrt{\frac{2}{\pi x}}\left[\sin\left(x \mp \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right) + \sin\left(x \mp \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right)\frac{R(x,\nu)}{x} + \cos\left(x \mp \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right) Q(x,\nu)\right].$$

These expansions lead to the asymptotic behavior of the Bessel functions as x tends to infinity:

$$J_{\pm\nu}(x) = \sqrt{\frac{2}{\pi x}}\left(\cos\left(x \mp \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right) + \mathcal{O}\left(\frac{1}{x}\right)\right),$$

$$Y_{\pm\nu}(x) = \sqrt{\frac{2}{\pi x}}\left(\sin\left(x \mp \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right) + \mathcal{O}\left(\frac{1}{x}\right)\right).$$

□

The final lemma is a result from real analysis that relates the asymptotic behavior of a positive function to that of its time average. It is applied in the argument establishing the optimality of the convergence rate.

Lemma 5.2.4. Let $g : [0, +\infty) \to \mathbb{R}$ be a continuous and positive function. Then

$$\limsup_{t\to\infty} g(t) \geq \limsup_{t\to\infty} \frac{1}{t}\int_0^t g(s)\, ds.$$

Proof. Let

$$A = \limsup_{t\to\infty} g(t).$$

Licence CC BY-NC-ND 3.0

If $A = +\infty$, the inequality is trivially satisfied. Suppose therefore that $A < +\infty$. Let $\varepsilon > 0$. By the definition of the limit superior, there exists $a_0 > 0$ such that

$$g(t) \leq A + \varepsilon \quad \text{for all } t \geq a_0.$$

For $t > a_0$, the integral is decomposed as follows:

$$\frac{1}{t}\int_0^t g(s)\,ds = \frac{1}{t}\int_0^{a_0} g(s)\,ds + \frac{1}{t}\int_{a_0}^t g(s)\,ds.$$

- The first term $\frac{1}{t}\int_0^{a_0} g(s)\,ds$ tends to 0 as $t \to \infty$ (since a_0 is fixed and g is continuous on $[0, a_0]$, so the integral is finite).
- For the second term, since $g(s) \leq A + \varepsilon$ for $s \geq a_0$

$$\frac{1}{t}\int_{a_0}^t g(s)\,ds \leq \frac{1}{t}\cdot(A+\varepsilon)(t-a_0) = (A+\varepsilon)\left(1 - \frac{a_0}{t}\right) \xrightarrow{t\to+\infty} (A+\varepsilon).$$

Taking the $\limsup_{t\to\infty}$:

$$\limsup_{t\to\infty} \frac{1}{t}\int_0^t g(s)\,ds \leq A + \varepsilon.$$

Since this holds for any $\varepsilon > 0$, then

$$\limsup_{t\to\infty} \frac{1}{t}\int_0^t g(s)\,ds \leq A = \limsup_{t\to\infty} g(t),$$

which is the desired inequality. □

Theorem 5.2.1 is now proved, based on the preliminary results established above.

Proof of the theorem 5.2.1. **Proof of (1):** Since $b \in \mathrm{Im}(\mathsf{Q})$, there exists $c \in \mathbb{R}^n$ such that $b = \mathsf{Q}c$. Thus the gradient of f is $\nabla f(x) = \mathsf{Q}x + b = \mathsf{Q}(x + c)$ and the set of minimizers of f is then given by:

$$S = \mathrm{argmin}(f) = \mathrm{Ker}(\mathsf{Q}) - c.$$

By the spectral theorem, the matrix Q admits the decomposition:

$$\mathsf{Q} = \mathsf{P}\mathsf{D}\mathsf{P}^\top \quad \text{with} \quad \mathsf{D} = \mathrm{diag}(\underbrace{\lambda_1, \ldots, \lambda_d}_{>0}, \underbrace{\lambda_{d+1}, \ldots, \lambda_n}_{=0})$$

where P is orthogonal ($\mathsf{P}^\top\mathsf{P} = \mathsf{P}\mathsf{P}^\top = \mathsf{I}_n$) and $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$. The change of variable:

$$X_\alpha(t) = \mathsf{P}^\top(x_\alpha(t) + c) \tag{5.17}$$

Licence CC BY-NC-ND 3.0

allows the system to be diagonalized. Substituting into (5.5) produces n independent scalar differential equations:

$$\ddot{X}_{\alpha,i}(t) + \frac{\alpha}{t}\dot{X}_{\alpha,i}(t) + \lambda_i X_{\alpha,i}(t) = 0, \quad i = 1, \dots, n \tag{5.18}$$

and the initial conditions transform into:

$$X_{\alpha,i}(t_0) = X_{0,i} = \langle \mathsf{P}_i, x_0 + c \rangle, \quad \dot{X}_{\alpha,i}(t_0) = X_{1,i} = \langle \mathsf{P}_i, x_1 \rangle$$

where P_i denotes the i^{th} column vector of the matrix P. According to Lemma 5.2.2,

$$X_{\alpha,j}(t) = \frac{1}{1-\alpha}\left[X_{1,j}t_0^{\alpha}t^{1-\alpha} - X_{1,j}t_0 + (1-\alpha)X_{0,j}\right], \qquad \text{if } \lambda_j = 0 \tag{5.19}$$

$$X_{\alpha,j}(t) = t^{\frac{1-\alpha}{2}}\left[A_{1,j}J_{\nu}(t\sqrt{\lambda_j}) + A_{2,j}Y_{\nu}(t\sqrt{\lambda_j})\right], \qquad \text{if } \lambda_j > 0 \tag{5.20}$$

where $\nu = \frac{\alpha-1}{2}$ is the order of the Bessel functions, and $A_{1,j}, A_{2,j}$ are determined by the initial conditions.
Let j be an integer between 1 and d. Using the asymptotic expansions of the Bessel functions (5.13-5.14) (Lemma 5.2.3), the following expressions are obtained for sufficiently large t:

$$J_{\nu}(\sqrt{\lambda_j}t) = \sqrt{\frac{2}{\pi t\sqrt{\lambda_j}}}\left[\cos\left(t\sqrt{\lambda_j} - \frac{\alpha\pi}{4}\right) + \mathcal{O}\left(\frac{1}{t}\right)\right]$$

$$Y_{\nu}(\sqrt{\lambda_j}t) = \sqrt{\frac{2}{\pi t\sqrt{\lambda_j}}}\left[\sin\left(t\sqrt{\lambda_j} - \frac{\alpha\pi}{4}\right) + \mathcal{O}\left(\frac{1}{t}\right)\right]$$

This gives the asymptotic expansion of the trajectory:

$$X_{\alpha,j}(t) = \sqrt{\frac{2}{\pi\sqrt{\lambda_j}}}t^{-\frac{\alpha}{2}}\left[A_{1,j}\cos\left(t\sqrt{\lambda_j} - \frac{\alpha\pi}{4}\right) + A_{2,j}\sin\left(t\sqrt{\lambda_j} - \frac{\alpha\pi}{4}\right) + \mathcal{O}\left(\frac{1}{t}\right)\right] \tag{5.21}$$

There exist constants $K > 0$ and $T_0 \geq 1$ such that for all $t > T_0$:

$$\begin{aligned} |X_{\alpha,j}(t)| &\leq \sqrt{\frac{2}{\pi\sqrt{\lambda_j}}}t^{-\frac{\alpha}{2}}\left(|A_{1,j}| + |A_{2,j}| + K\right) \\ &\leq C_j t^{-\frac{\alpha}{2}} \quad \text{where} \quad C_j = \sqrt{\frac{2}{\pi\sqrt{\lambda_j}}}(|A_{1,j}| + |A_{2,j}| + K) \end{aligned} \tag{5.22}$$

Estimation of the optimality gap: Let $x^* \in S$ (so that $x^* + c \in \mathsf{Ker(Q)}$). Starting from relation

Licence CC BY-NC-ND 3.0

(5.17), $x_\alpha(t) = \mathsf{P}X_\alpha(t) - c$, the development yields:

$$f(x_\alpha(t)) - f(x^*) = \frac{1}{2}\langle \mathsf{P}X_\alpha(t) - c, \mathsf{Q}(\mathsf{P}X_\alpha(t) - c)\rangle + \langle b, \mathsf{P}X_\alpha(t) - c\rangle - \frac{1}{2}\langle x^*, \mathsf{Q}x^*\rangle - \langle b, x^*\rangle$$

Using $b = \mathsf{Q}c$ and after simplification, the resulting expression is

$$\begin{aligned} f(x_\alpha(t)) - f(x^*) &= \frac{1}{2}\langle X_\alpha(t), \mathsf{D}X_\alpha(t)\rangle - \frac{1}{2}\langle x^* + c, \mathsf{Q}(x^* + c)\rangle \\ &= \frac{1}{2}\langle X_\alpha(t), \mathsf{D}X_\alpha(t)\rangle \quad (\text{since } \mathsf{Q}(x^* + c) = 0) \end{aligned} \tag{5.23}$$

Thus:

$$\begin{aligned} f(x_\alpha(t)) - \inf(f) &= \frac{1}{2}\sum_{j=1}^{d} \lambda_j X_{\alpha,j}^2(t) \quad \text{since for } i > d,\ \lambda_i = 0 \\ &\leq \frac{1}{2}\sum_{j=1}^{d} \lambda_j C_j^2 t^{-\alpha} \\ &= \frac{1}{t^\alpha}\left(\frac{1}{2}\sum_{j=1}^{d} \lambda_j C_j^2\right) \end{aligned} \tag{5.24}$$

which indeed demonstrates convergence at the announced rate $\mathcal{O}(t^{-\alpha})$.

Proof of (2): Consider an index $j \in \{1, \ldots, d\}$ corresponding to an eigenvalue $\lambda_j > 0$. From the asymptotic expansion (5.21), there exist $T_0 > 1$ and a function $M(t)$ bounded on $[T_0, +\infty)$ such that for all $t \geq T_0$:

$$X_{\alpha,j}(t) = \sqrt{\frac{2}{\pi\sqrt{\lambda_j}}}\, t^{-\frac{\alpha}{2}}\left(A_{1,j}\cos\left(t\sqrt{\lambda_j} - \frac{\alpha\pi}{4}\right) + A_{2,j}\sin\left(t\sqrt{\lambda_j} - \frac{\alpha\pi}{4}\right) + \frac{M(t)}{t}\right)$$

Squaring and multiplying by t^α yields:

$$t^\alpha X_{\alpha,j}^2(t) = \frac{2}{\pi\sqrt{\lambda_j}}\left[A_{1,j}^2\cos^2\left(\theta_j(t)\right) + A_{2,j}^2\sin^2\left(\theta_j(t)\right) + A_{1,j}A_{2,j}\sin\left(2\theta_j(t)\right) + \frac{M_1(t)}{t}\right]$$

where $\theta_j(t) = t\sqrt{\lambda_j} - \dfrac{\alpha\pi}{4}$ and $M_1(t)$ is a new bounded function defined by:

$$M_1(t) = 2A_{2,j}M(t)\sin\left(\theta_j(t)\right) + 2A_{1,j}M(t)\cos\left(\theta_j(t)\right) + \frac{M^2(t)}{t}$$

The previous expression permits a bound. Specifically:

Licence CC BY-NC-ND 3.0

- *Lower bound:* By bounding $M_1(t)/t$ from below by $-M_{1,\infty}/t$ with

$$M_{1,\infty} = \sup_{t \geq T_0} |M_1(t)|,$$

this yields :

$$t^\alpha X_{\alpha,j}^2(t) \geq \frac{2}{\pi\sqrt{\lambda_j}} \left(A_{1,j}^2 \cos^2\theta_j(t) + A_{2,j}^2 \sin^2\theta_j(t) + A_{1,j}A_{2,j}\sin(2\theta_j(t)) - \frac{M_{1,\infty}}{t} \right)$$

- *Upper bound:* Similarly, by bounding $M_1(t)/t$ from above by $+M_{1,\infty}/t$:

$$t^\alpha X_{\alpha,j}^2(t) \leq \frac{2}{\pi\sqrt{\lambda_j}} \left(A_{1,j}^2 \cos^2\theta_j(t) + A_{2,j}^2 \sin^2\theta_j(t) + A_{1,j}A_{2,j}\sin(2\theta_j(t)) + \frac{M_{1,\infty}}{t} \right)$$

Integrating these inequalities over $[T_0, t]$ and dividing by $t - T_0$:

$$\frac{1}{t-T_0}\int_{T_0}^t u^\alpha X_{\alpha,j}^2(u)\,du \geq \frac{2}{\pi\sqrt{\lambda_j}}\frac{1}{t-T_0}\int_{T_0}^t \left[A_{1,j}^2 \cos^2\left(\theta_j(u)\right) + A_{2,j}^2 \sin^2\left(\theta_j(u)\right) + A_{1,j}A_{2,j}\sin\left(2\theta_j(u)\right) - \frac{M_{1,\infty}}{u} \right] du$$

and

$$\frac{1}{t-T_0}\int_{T_0}^t u^\alpha X_{\alpha,j}^2(u)\,du \leq \frac{2}{\pi\sqrt{\lambda_j}}\frac{1}{t-T_0}\int_{T_0}^t \left[A_{1,j}^2 \cos^2\left(\theta_j(u)\right) + A_{2,j}^2 \sin^2\left(\theta_j(u)\right) + A_{1,j}A_{2,j}\sin\left(2\theta_j(u)\right) + \frac{M_{1,\infty}}{u} \right] du$$

Taking the limit as $t \to \infty$ for each term gives:

- Logarithmic term:

$$\lim_{t\to+\infty} \frac{1}{t-T_0}\int_{T_0}^t \pm\frac{M_{1,\infty}}{u}\,du = \pm M_{1,\infty} \lim_{t\to+\infty} \frac{\ln t - \ln T_0}{t-T_0} = 0$$

- Oscillatory term:

$$\lim_{t\to+\infty} \frac{1}{t-T_0}\int_{T_0}^t \sin(2\theta_j(u))\,du = \lim_{t\to+\infty} \frac{1}{t-T_0}\left[-\frac{1}{2\sqrt{\lambda_j}}\cos(2\theta_j(u)) \right]_{T_0}^t = 0$$

Licence CC BY-NC-ND 3.0

- Quadratic terms:

$$\lim_{t\to+\infty}\frac{1}{t-T_0}\int_{T_0}^{t}\cos^2(\theta_j(u))\,du=\lim_{t\to+\infty}\frac{1}{t-T_0}\left[\frac{1}{2}\left(\frac{1}{2\sqrt{\lambda_j}}\sin(2\theta_j(u))+u\right)\right]_{T_0}^{t}=\frac{1}{2}$$

$$\text{(and similarly)}\quad \lim_{t\to+\infty}\frac{1}{t-T_0}\int_{T_0}^{t}\sin^2\left(\theta_j(u)\right)\,du=\frac{1}{2}$$

It follows that:

$$\lim_{t\to+\infty}\frac{1}{t-T_0}\int_{T_0}^{t}u^{\alpha}X_{\alpha,j}^2(u)\,du=\frac{A_{1,j}^2+A_{2,j}^2}{\pi\sqrt{\lambda_j}} \tag{5.25}$$

Using relations (5.23) and (5.25), the following result holds:

$$\lim_{t\to+\infty}\frac{1}{t-T_0}\int_{T_0}^{t}u^{\alpha}(f(x_\alpha(u))-\inf(f))\,du=\frac{1}{2\pi}\sum_{j=1}^{d}\sqrt{\lambda_j}(A_{1,j}^2+A_{2,j}^2) \tag{5.26}$$

Furthermore, from Lemma 5.2.4:

$$\limsup_{t\to\infty}t^{\alpha}(f(x_\alpha(t))-\inf(f))\geq\lim_{t\to+\infty}\frac{1}{t-T_0}\int_{T_0}^{t}u^{\alpha}(f(x_\alpha(u))-\inf(f))du,$$

thus

$$\limsup_{t\to\infty}t^{\alpha}(f(x_\alpha(t))-\inf(f))\geq\frac{1}{2\pi}\sum_{j=1}^{d}\sqrt{\lambda_j}(A_{1,j}^2+A_{2,j}^2)>0 \tag{5.27}$$

Hence, if there existed $\beta>\alpha$ such that

$$f(x_\alpha(t))-\inf(f)\leq O\left(\frac{1}{t^{\beta}}\right),$$

then

$$\limsup_{t\to\infty}t^{\alpha}(f(x_\alpha(t))-\inf(f))=\lim_{t\to\infty}t^{\alpha}(f(x_\alpha(t))-\inf(f))=0,$$

which would contradict (5.27).

Proof of (3): From the previous relations (5.19) and (5.22), it follows that:

$$\lim_{t\to+\infty}X_\alpha(t)=X_\infty \tag{5.28}$$

Licence CC BY-NC-ND 3.0

where the components of X_∞ are given by:

$$X_{\infty,j} = \begin{cases} 0 & \text{if } \lambda_j > 0 \\ X_{0,j} - \frac{t_0}{1-\alpha} X_{1,j} & \text{if } \lambda_j = 0 \end{cases}$$

Identification of the minimizer: By the change of variable (5.17), one obtains:

$$x_{\alpha,\infty} := \lim_{t \to +\infty} x_\alpha(t) = \mathsf{P} X_\infty - c$$

Let us verify that $x_{\alpha,\infty} \in S = \mathsf{Ker}(\mathsf{Q}) - c$:

$$\mathsf{Q} x_{\alpha,\infty} = \mathsf{Q}(\mathsf{P} X_\infty - c) = \mathsf{P}\mathsf{D}\mathsf{P}^\top \mathsf{P} X_\infty - \mathsf{Q}c = \mathsf{P}\mathsf{D} X_\infty - b = -b$$

since $\mathsf{D} X_\infty = 0$ (by definition) and $b = \mathsf{Q}c$.

Convergence rate: Estimatation of the distance to the optimum

$$\begin{aligned} \|x_\alpha(t) - x_{\alpha,\infty}\|^2 &= \|\mathsf{P} X_\alpha(t) - c - \mathsf{P} X_\infty + c\|^2 \\ &\leq \|\mathsf{P}\|^2 \|X_\alpha(t) - X_\infty\|^2 \\ &= \|\mathsf{P}\|^2 \left[\sum_{j=1}^{d} X_{\alpha,j}^2(t) + \sum_{j=d+1}^{n} \left(\frac{X_{1,j} t_0^\alpha t^{1-\alpha}}{1-\alpha} \right)^2 \right] \end{aligned} \tag{5.29}$$

Using the bound (5.22) for non-zero eigenvalues, the following inequality holds for $t \geq T_0$:

$$\begin{aligned} \|x_\alpha(t) - x_{\alpha,\infty}\|^2 &\leq \frac{\|\mathsf{P}\|^2}{t^\alpha} \left(\sum_{j=1}^{d} \frac{2}{\pi \sqrt{\lambda_j}} (|A_{1,j}| + |A_{2,j}| + K)^2 \right. \\ &\qquad \left. + \frac{t_0^{2\alpha}}{(\alpha-1)^2} \sum_{j=d+1}^{n} X_{1,j}^2 t^{2-\alpha} \right) \\ &\leq \mathcal{O}\left(\frac{1}{t^\alpha} \right) \end{aligned} \tag{5.30}$$

This allows the convergence rate of the trajectories toward the minimizer to be deduced, completing the proof.

□

The following remarks summarize the main consequences of the preceding analysis.

Remark 5.2.1. This analysis identifies several properties of the accelerated dynamics for convex quadratic functions. The convergence rate $\mathcal{O}(t^{-\alpha})$ is directly determined by the parameter α,

Licence CC BY-NC-ND 3.0

providing a degree of freedom for optimization: larger values of α result in faster convergence. Compared to the general case of convex functions, where the rate is $o(1/t^2)$, a higher rate $\mathcal{O}(t^{-\alpha})$ is obtained for $\alpha > 3$. This improvement is explained by the algebraic structure of quadratic functions, which allows full use of the oscillatory properties of Bessel functions. The rate $\mathcal{O}(t^{-\alpha})$ represents a limit that cannot be uniformly improved.
The result also shows that the accelerated dynamics ensures convergence of the trajectories to a minimizer with the same order $\mathcal{O}(t^{-\alpha})$ as the objective function values. In many optimization algorithms, acceleration improves the rate of decrease of the objective value without providing precise control over the trajectories. In this case, the linear and oscillatory structure of the system, characterized by Bessel functions, allows both the optimal rate in objective values and control of the distance to the minimizer in the variable space. This trajectory convergence provides a framework for selecting the parameter α in order to achieve both speed and stability in the descent.

5.3 Numerical experiments

This section presents numerical simulations illustrating the theoretical results established in the previous sections, in particular Theorem 5.2.1, concerning the convergence rate of function values, its optimality, and the convergence rate of trajectories. The dynamics (AVD) are also compared with the Heavy Ball model (equation (5.1)) for various values of the parameter α. The simulations are designed to quantify the influence of the parameter α on the dynamics

$$(\mathrm{AVD}): \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \mathrm{Q}x(t) + b = 0, \quad x(t_0) = x_0, \quad \dot{x}(t_0) = x_1,$$

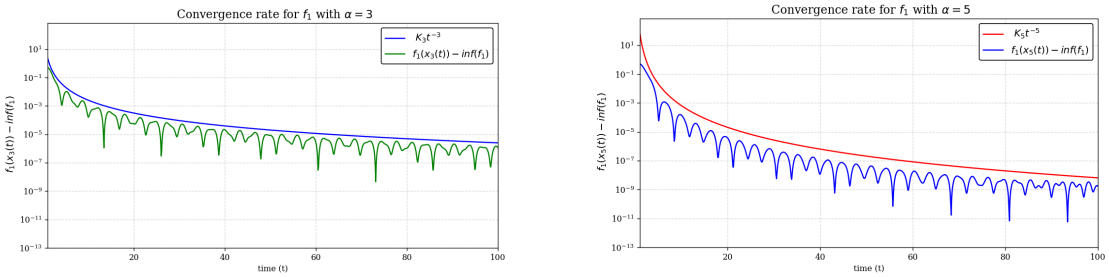
and to illustrate the consistency between analytical predictions and the observed behaviour of the system.
For all experiments, the differential equation (AVD) was solved numerically using a fourth–order Runge–Kutta scheme with adaptive step size, as implemented in the `SciPy` library [37]. This choice allows one to accurately track the dynamics for different values of α. The simulations were carried out on several convex quadratic functions, in order to cover a range of Hessian spectra and highlight the robustness of the method.
Example 5.3.1. The first test function $f_1 : \mathbb{R}^3 \to \mathbb{R}$ is defined by:

$$f_1(x) = \tfrac{1}{2}x^\top Q_1 x + b_1^\top x, \quad Q_1 = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad b_1 = (-2, 1, 1)^\top.$$

The spectrum of Q_1 is $\{0, 1, 3\}$, and a minimizer is $x^* = (1, 0, 0)^\top$ with $f_1(x^*) = -1$. The initial conditions are $t_0 = 1$, $x_0 = (1, 0, -1)^\top$, $x_1 = (0, 0, 0)^\top$.

Licence CC BY-NC-ND 3.0

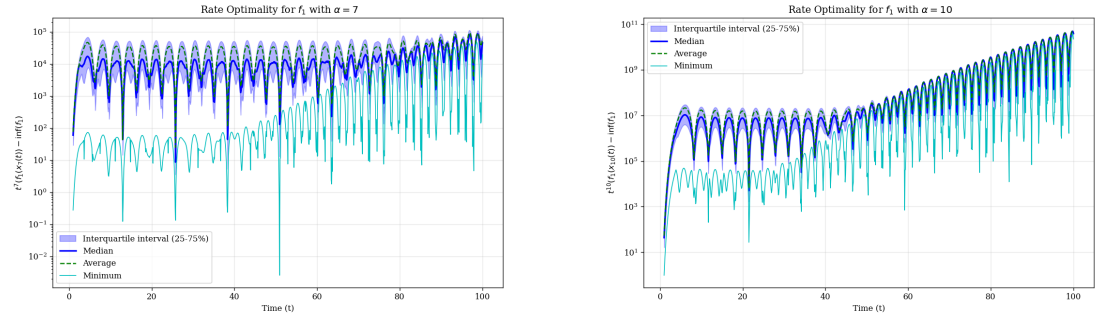
Figure 5.3 compares the time evolution of the suboptimality gap $f_1(x_\alpha(t)) - \inf(f_1)$ with the theoretical rate $1/t^\alpha$, while Figure 5.4 depicts the quantity $t^\alpha(f_1(x_\alpha(t)) - \inf(f_1))$, allowing an assessment of the rate's optimality.



(a) $\alpha = 3$. (b) $\alpha = 5$.

Figure 5.3: Convergence rate for f_1 (semi-logarithmic scale).

Figures 5.3(a) and 5.3(b) illustrate the convergence behaviour for the function f_1. For $\alpha = 3$, the gap $f_1(x_3(t)) - \inf(f_1)$ decreases in accordance with the theoretical reference $O(t^{-3})$, with good agreement between the numerical curve and the bound $K_3\, t^{-3}$ (where the K_i denote constants independent of t). When $\alpha = 5$, the decay is even faster and closely follows the rate $O(t^{-5})$. These results show that, for f_1, the (AVD) dynamics ensures accelerated convergence of order α, in agreement with the theoretical predictions.
Figure 5.4 illustrates the optimality of this rate. For 50 random initial conditions, the mean, median, minimum and interquartile range of the quantity $t^\alpha\big(f_1(x_\alpha(t)) - \inf(f_1)\big)$ are shown.



(a) $\alpha = 7$. (b) $\alpha = 10$.

Figure 5.4: Optimality of the rate for f_1.

Figures 5.4(a) and 5.4(b) confirm that the quantity $t^\alpha(f_1(x_\alpha(t)) - \inf(f_1))$ does not tend to zero, but stabilizes around a strictly positive value. Even the minimum of this quantity remains bounded away from zero, which experimentally confirms that the rate $O(1/t^\alpha)$ is optimal for f_1.

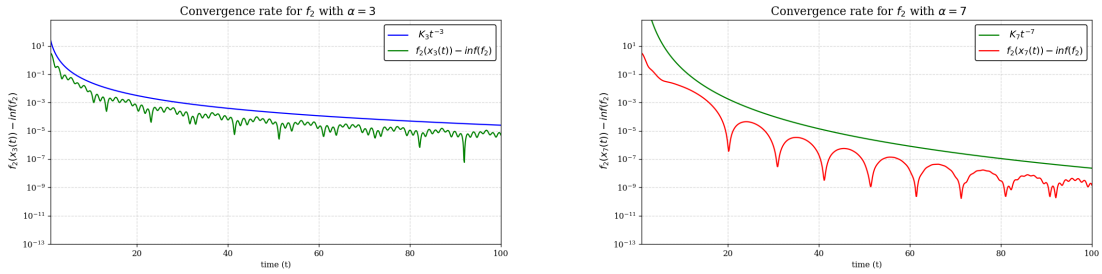
Licence CC BY-NC-ND 3.0

Example 5.3.2. The second test function $f_2 : \mathbb{R}^4 \to \mathbb{R}$ is defined by:

$$f_2(x) = \tfrac{1}{2}x^\top Q_2 x + b_2^\top x - 2, \quad b_2 = (-1.025, -1.525, 1.475, 0.975)^\top,$$

$$Q_2 = \begin{pmatrix} 1.525 & 1.025 & -0.975 & -1.475 \\ 1.025 & 1.525 & -1.475 & -0.975 \\ -0.975 & -1.475 & 1.525 & 1.025 \\ -1.475 & -0.975 & 1.025 & 1.525 \end{pmatrix}.$$

The spectrum of Q_2 is $\{0, 0.1, 1, 5\}$, more spread out than that of f_1, which makes it possible to see the influence of the spectral structure on transient oscillations. A minimizer is $x^* = (1, 0, -1, 1)^\top$ with $f_2(x^*) = -2.7625$. The initial conditions are $t_0 = 1$, $x_0 = (1, 0, -1, -1)^\top$, $x_1 = (0, 0, 0, 0)^\top$.



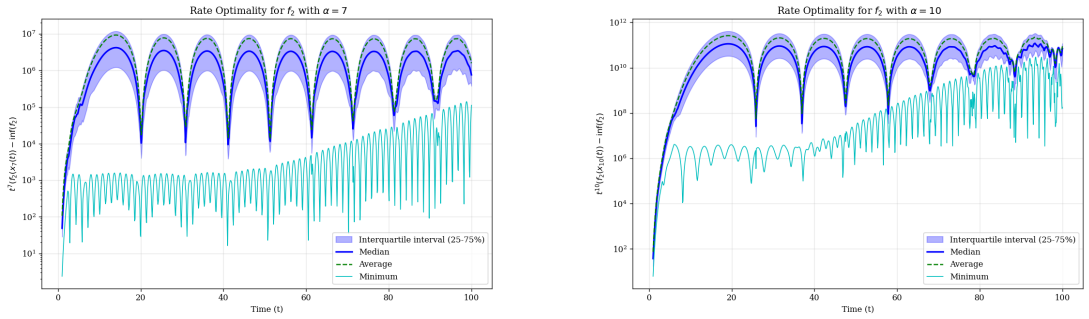
(a) $\alpha = 3$.

(b) $\alpha = 7$.

Figure 5.5: Convergence rate for f_2 (semi-logarithmic scale).

For the function f_2, Figures 5.5(a) and 5.5(b) compare the suboptimality gaps for $\alpha = 3$ and $\alpha = 7$. With $\alpha = 3$, the curve $f_2(x_3(t)) - \inf(f_2)$ decreases quickly, closely following the reference t^{-3}, and drops from 10^1 to 10^{-7} over the interval $[1, 100]$. For $\alpha = 7$, the convergence is even faster, confirming the robustness of the rate $O(1/t^7)$. Thus, increasing α improves the convergence speed, in accordance with the theoretical predictions.

Licence CC BY-NC-ND 3.0



(a) $\alpha = 7$. (b) $\alpha = 10$.

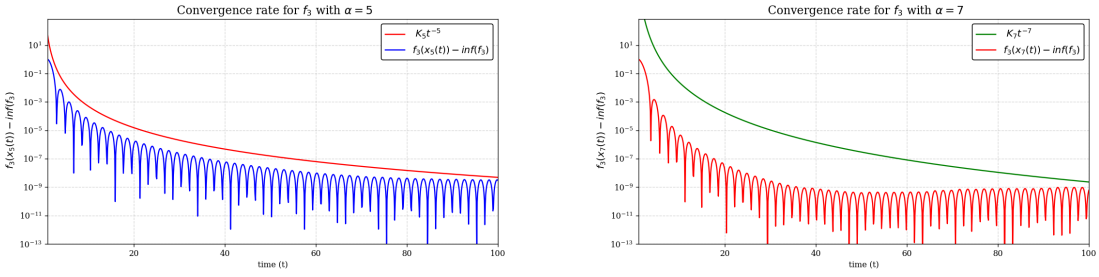
Figure 5.6: Optimality of the rate for f_2.

Figures 5.6(a) and 5.6(b) extend this optimality analysis for the function f_2. As for f_1, the minimum of the quantity $t^\alpha(f_2(x_\alpha(t)) - \inf(f_2))$ remains strictly positive, indicating that the convergence cannot be accelerated beyond the rate $O(1/t^\alpha)$. This similarity of behaviour between f_1 and f_2 confirms the robustness and generality of the optimality of the rate for the (AVD) dynamics.

Example 5.3.3. Finally, the last test function $f_3 : \mathbb{R}^4 \to \mathbb{R}$ is defined by:

$$f_3(x) = \tfrac{1}{2}x^\top Q_3 x + b_3^\top x, \quad Q_3 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}, \quad b_3 = (-1, -2, -1, 2)^\top.$$

The spectrum of Q_3 is $\{0, 1, 2, 3\}$, and a minimizer is $x^* = (0, 1, 0, -1)^\top$ with $f_3(x^*) = -2$. The initial conditions are $t_0 = 1$, $x_0 = (1, 1, 1, -1)^\top$, $x_1 = (0, 0, 0, 0)^\top$.



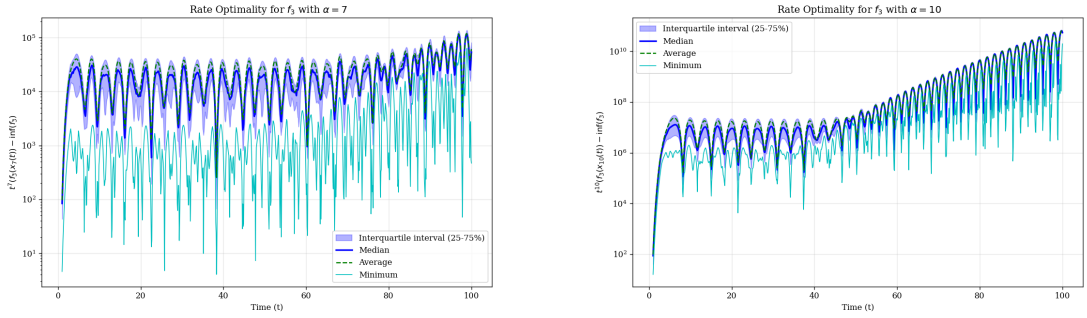
(a) $\alpha = 5$. (b) $\alpha = 7$.

Figure 5.7: Convergence rate for f_3 (semi-logarithmic scale).

Figures 5.7(a) and 5.7(b) extend the analysis to the function f_3. For $\alpha = 5$, the gap $f_3(x_5(t)) - \inf(f_3)$ decreases steadily, in accordance with the law $O(1/t^5)$. For $\alpha = 7$, the

Licence CC BY-NC-ND 3.0

decay is significantly faster, closely following the rate $O(1/t^7)$ from the very beginning. These results confirm that the (AVD) dynamics ensures convergence of order α for various convex quadratic functions, with significant acceleration as α increases.



(a) $\alpha = 7$. (b) $\alpha = 10$.

Figure 5.8: Optimality of the rate for f_3.

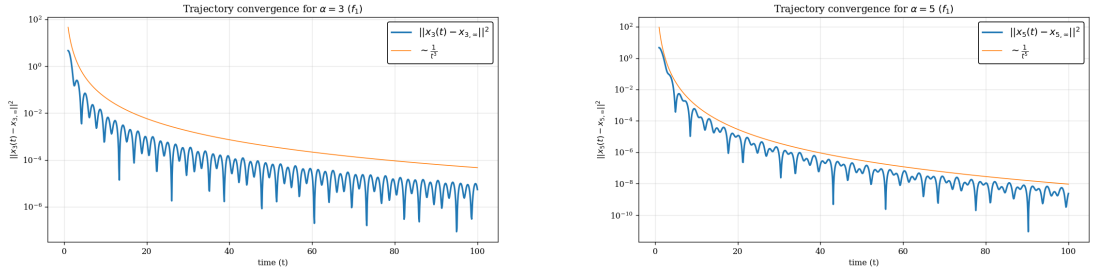
Figures 5.8(a) and 5.8(b) complete the optimality analysis for f_3. The study of the minimum of the quantity $t^\alpha\big(f_3(x_\alpha(t)) - \inf(f_3)\big)$ reveals a behaviour similar to that observed for f_1 and f_2. This minimum stabilizes at positive, bounded values, confirming that the rate $O(1/t^\alpha)$ constitutes an optimal convergence bound. The consistency of the results obtained for the three test functions robustly establishes the universal character of this optimality for the (AVD) dynamics. Thus, the identified theoretical bound appears non-improvable, independently of the objective function considered.

Illustration of the trajectory convergence rate: With respect to Convergence rate of the trajectories, the curves presented in Figures 5.9(a) and 5.9(b) show that the quantity

$$\|x_\alpha(t) - x_{\alpha,\infty}\|^2$$

decreases at a rate of order $\mathcal{O}(1/t^\alpha)$ for $\alpha = 3$ and $\alpha = 5$ in the case of the function f_1. In particular, the convergence is faster when the parameter α increases, as evidenced by the steeper slope of the curve associated with $\alpha = 5$. These results numerically validate the accelerating effect of the parameter α for a convex quadratic function.

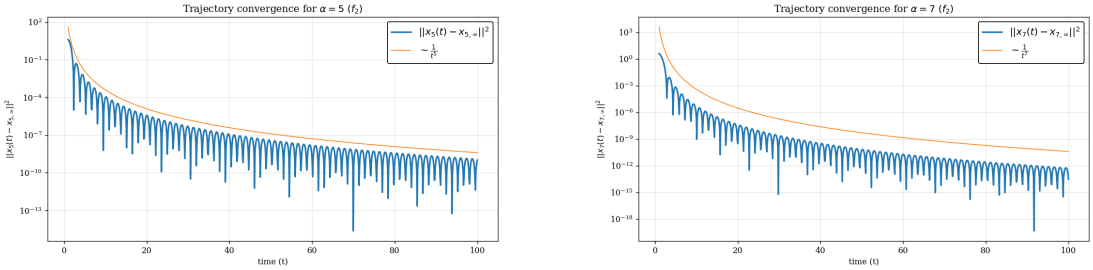
Licence CC BY-NC-ND 3.0



(a) $\alpha = 3$. (b) $\alpha = 5$.

Figure 5.9: Convergence rate of the trajectories on a semi-logarithmic scale (for f_1).

The results obtained for the function f_1 thus show that the convergence closely follows the theoretical rate $\mathcal{O}(1/t^{\alpha})$. For the function f_2, Figures 5.10(a) and 5.10(b) extend this observation to higher values of the parameter α.

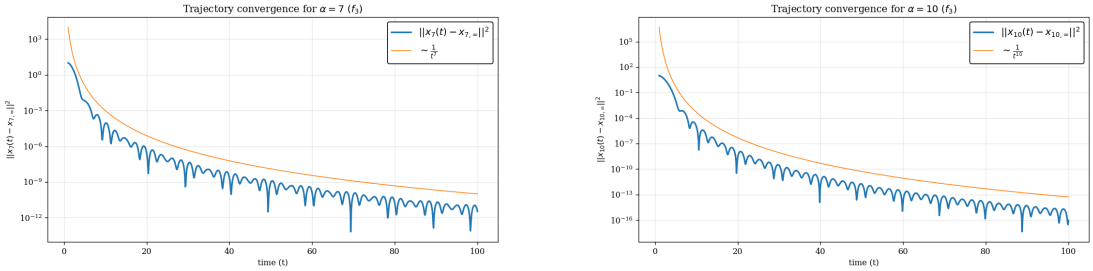


(a) $\alpha = 5$. (b) $\alpha = 7$.

Figure 5.10: Convergence rate of the trajectories on a semi-logarithmic scale (for f_2).

The gap $\|x_{\alpha}(t) - x_{\alpha,\infty}\|^2$ decreases according to a rate of order $1/t^5$ for $\alpha = 5$ and $1/t^7$ for $\alpha = 7$, which once again confirms the theoretical law. The convergence accelerates as α increases, clearly illustrating the effect of the parameter.
Finally, after analyzing the functions f_1 and f_2, Figures 5.11(a) and 5.11(b) present the results obtained for the function f_3, considering even higher values of α.

Licence CC BY-NC-ND 3.0

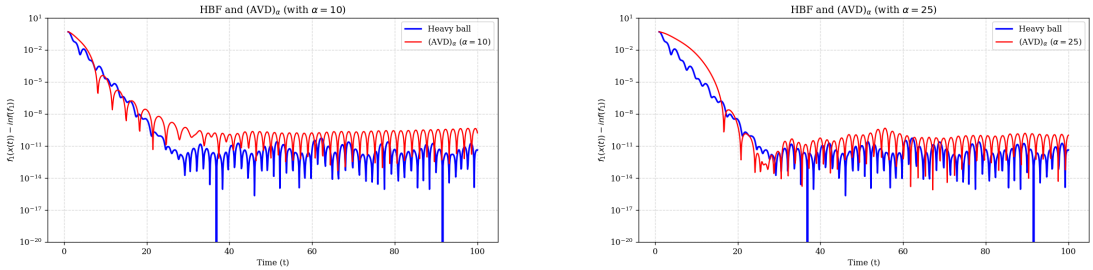


(a) $\alpha = 7$. (b) $\alpha = 10$.

Figure 5.11: Convergence rate of the trajectories on a semi-logarithmic scale (for f_3).

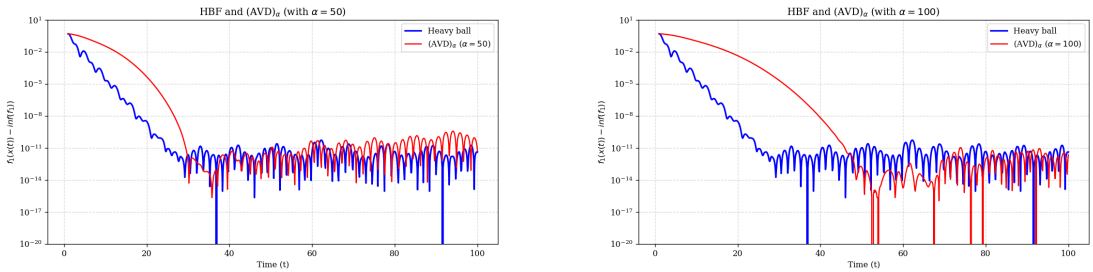
The results confirm the robustness of the theoretical convergence rate $\mathcal{O}(1/t^\alpha)$, with a decay of order $1/t^7$ for $\alpha = 7$ and $1/t^{10}$ for $\alpha = 10$. Increasing the parameter α induces a marked acceleration of the convergence, showing that this effect persists and is amplified for large values of α, independently of the convex function under consideration.

Comparison between the Heavy Ball and (AVD) models:



(a) For $\alpha = 10$. (b) For $\alpha = 25$.

Figure 5.12: Comparison between Heavy Ball and (AVD) for $\alpha = 10$ and $\alpha = 25$.



(a) For $\alpha = 50$. (b) For $\alpha = 100$.

Figure 5.13: Comparison between Heavy Ball and (AVD) for $\alpha = 50$ and $\alpha = 100$.

Licence CC BY-NC-ND 3.0

Across all four figures (5.12(a) to 5.13(b)), the blue curve corresponding to the Heavy Ball method consistently exhibits the same dynamics, which makes it possible to see the influence of the parameter value α on the behaviour of the (AVD) method shown in red. First, for moderate times, Heavy Ball clearly dominates: its exponential decay induces a very fast reduction of the error $f_1(x(t)) - \inf(f_1)$, much faster than that of (AVD), whose convergence is polynomial of order $1/t^\alpha$. The comparison of the four values of α reveals a clear phenomenon: when $\alpha = 10$, the red curve remains very far from the Heavy Ball curve over the entire time range. For $\alpha = 25$, it moves significantly closer but only meets the blue curve at very low error levels. For $\alpha = 50$, the polynomial decay is more pronounced, but (AVD) (for $\alpha = 50$) takes significantly more time to reach low error values, and it is only around $t \approx 40$ that the two methods exhibit comparable amplitudes in the oscillatory phase. For $\alpha = 100$, the dynamics show a different behaviour: the initial decay is very slow, but after some time the red curve eventually catches up with, and even slightly surpasses, the blue curve. It is also observed that the oscillations of (AVD) become more pronounced as α increases, reflecting increased dynamic inertia, while the Heavy Ball method maintains a more regular evolution. Thus, the graphical analysis confirms the superiority of Heavy Ball over moderate time horizons, while (AVD) with large α requires long-time regimes to express its asymptotic potential.

5.4 Conclusion and Perspectives

This chapter presents a detailed study of the vanishing damping dynamics (AVD), which serves as a continuous model for understanding Nesterov-type accelerated optimization methods.

The analysis focuses on the case of convex quadratic functions, which permits a detailed analytical investigation of the dynamics. In this setting, an explicit solution of (AVD) is derived using Bessel functions, revealing the oscillatory nature of the system. The convergence rate of the objective function values is of order $\mathcal{O}(1/t^\alpha)$, and this rate is optimal, representing an improvement over the known bounds in the general convex case. Moreover, the trajectories converge to a minimizer at the same rate, providing precise control over both function values and trajectories.

Numerical simulations confirm these theoretical findings and clearly illustrate the impact of the parameter α on the convergence speed. They also show that increasing α allows the (AVD) dynamics to approach the behavior of exponentially convergent methods, such as the Heavy Ball method, while maintaining a polynomial-type convergence. However, this study also emphasizes a point: the discrete case is not a straightforward translation of the continuous case, even within the quadratic framework. Phenomena observed in the continuous dynamics do not automatically carry over to discrete algorithms, and the question of the optimal choice of α in discrete schemes remains largely open.

Finally, this work opens several promising directions. Extending these results to the general convex case is a natural but delicate path, requiring new analytical approaches. Although a

Licence CC BY-NC-ND 3.0

complete theoretical solution seems challenging in the short term, numerical investigations could provide valuable guidance for the optimal tuning of the parameter α in discrete algorithms. These questions will be the subject of future research and are part of a broader effort to better understand the deep connections between continuous dynamics, discrete schemes, and acceleration in convex optimization.

Licence CC BY-NC-ND 3.0

6

Bibliography

Contents

References **152**

Licence CC BY-NC-ND 3.0

References

[1] R. E. Bruck, "Asymptotic convergence of nonlinear contraction semigroups in hilbert space," *Journal of Functional Analysis*, vol. 18, no. 1, pp. 15–26, 1975. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022123675900270

[2] B. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, pp. 1–17, 1964. [Online]. Available: https://doi.org/10.1016/0041-5553(64)90137-5

[3] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Dokl. Akad. Nauk SSSR*, vol. 269, pp. 543–547, 1983.

[4] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course* (Applied Optimization). Kluwer Academic Publishers, Boston, MA, 2004, vol. 87.

[5] H. Attouch and J. Peypouquet, "The rate of convergence of nesterov's accelerated forward-backward method is actually faster than $1/k^2$," *SIAM Journal on Optimization*, vol. 26, no. 3, pp. 1824–1834, 2016. [Online]. Available: https://doi.org/10.1137/15M1046095

[6] A. Chambolle and C. Dossal, "On the convergence of the iterates of the fast iterative shrinkage thresholding algorithm," *Journal of Optimization Theory and Applications*, vol. 166, pp. 968–982, 2015. [Online]. Available: https://doi.org/10.1007/s10957-015-0746-4

[7] V. Apidopoulos, J. F. Aujol, and C. Dossal, "The differential inclusion modeling fista algorithm and optimality of convergence rate in the case $b \leq 3$," *SIAM Journal on Optimization*, vol. 28, no. 1, pp. 551–574, 2018. [Online]. Available: https://doi.org/10.1137/17M1128642

[8] H. Attouch, Z. Chbani, and H. Riahi, "Rate of convergence of the nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$," *ESAIM: COCV*, vol. 25, no. 2, 2019. [Online]. Available: https://doi.org/10.1051/cocv/2017083

Licence CC BY-NC-ND 3.0

[9] W. J. Su, S. Boyd, and E. J. Candes, "A differential equation for modeling nesterov's accelerated gradient method: theory and insights," *Journal of Machine Learning Research*, vol. 17, pp. 1–43, 2016, Paper No. 153.

[10] H. Attouch and J. Fadili, *From the ravine method to the nesterov method and vice versa: a dynamical system perspective*, hal-03546187, 2022.

[11] F. Alvarez, H. Attouch, J. Bolte, and P. Redont, "A second-order gradient-like dissipative dynamical system with hessian-driven damping. application to optimization and mechanics," *Journal de Mathématiques Pures et Appliquées*, vol. 81, no. 8, pp. 747–779, 2002. [Online]. Available: https://doi.org/10.1016/S0021-7824(01)01253-3

[12] H. Attouch, J. Peypouquet, and P. Redont, "Fast convex minimization via inertial dynamics with hessian-driven damping," *Journal of Differential Equations*, vol. 261, pp. 5734–5783, 2016. [Online]. Available: https://doi.org/10.1016/j.jde.2016.08.020

[13] H. Attouch, A. Balhag, Z. Chbani, and H. Riahi, "Fast convex optimization via inertial dynamics combining viscous and hessian-driven damping with time rescaling," *Evolution Equations and Control Theory*, [Online]. Available: https://doi.org/10.3934/eect.2021010

[14] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su, "Understanding the acceleration phenomenon via high-resolution differential equations," *Mathematical Programming*, vol. 195, pp. 79–148, 2022. [Online]. Available: https://doi.org/10.1007/s10107-021-01681-8

[15] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi, "First-order algorithms via inertial systems with hessian-driven damping," *Mathematical Programming*, 2020. [Online]. Available: https://doi.org/10.1007/s10107-020-01591-1

[16] C. Alcesa, S. Laszlo, and T. Pinta, "An extension of the second order dynamical system that models nesterov's convex gradient method," *Applied Mathematics and Optimization*, vol. 84, no. 2, pp. 1687–1716, 2021. [Online]. Available: https://doi.org/10.1007/s00245-020-09692-1

[17] H. Brézis, *Analyse fonctionnelle: Théorie et applications*, 2nd. Paris, France: Masson, 2010.

[18] H. Brézis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert* (North-Holland Mathematical Studies 5). North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York, 1973, Notas Mat., No. 50 [Mathematical Notes].

[19] R. T. Rockafellar, *Convex Analysis* (Princeton Mathematical Series). Princeton, NJ: Princeton University Press, 1970, vol. 28.

Licence CC BY-NC-ND 3.0

[20] H. Bauschke and P. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* (CMS Books in Mathematics). New York: Springer, 2011.

[21] J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965. [Online]. Available: https://doi.org/10.24033/bsmf.1625

[22] J. F. Aujol, C. Dossal, and A. Rondepierre, *Optimal convergence rates for nesterov acceleration*, 2019. arXiv: 1805.05719 [math.OC]. [Online]. Available: https://arxiv.org/abs/1805.05719

[23] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd. Cambridge, UK: Cambridge University Press, 1944.

[24] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont, "Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity," *Mathematical Programming Series B*, vol. 168, pp. 123–175, 2018. [Online]. Available: https://doi.org/10.1007/s10107-016-0992-8

[25] R. May, "Asymptotic for a second-order evolution equation with convex potential and vanishing damping term," *Turkish Journal of Mathematics*, vol. 41, no. 3, pp. 681–685, 2017. [Online]. Available: https://doi.org/10.3906/mat-1512-28

[26] H. Attouch and A. Cabot, "Convergence rates of inertial forward-backward algorithms," *SIAM Journal on Optimization*, vol. 28, no. 1, pp. 849–874, 2018. [Online]. Available: https://doi.org/10.1137/17M1114739

[27] S. Adly and H. Attouch, "Complexity analysis based on tuning the viscosity parameter of the su-boyd-candes inertial gradient dynamics," *Set-Valued and Variational Analysis*, vol. 32, p. 17, 2024, 10.1007/s11228-024-00720-8. [Online]. Available: https://doi.org/10.1007/s11228-024-00720-8

[28] S. Adly and H. Attouch, "Accelerated optimization through time-scale analysis of inertial dynamics with asymptotic vanishing and hessian-driven dampings," *Optimization*, pp. 1–38, 2024, 10.1080/02331934.2024.2359540. [Online]. Available: https://doi.org/10.1080/02331934.2024.2359540

[29] L. F. Shampine and M. W. Reichelt, "The matlab ode suite," *SIAM Journal on Scientific Computing*, vol. 18, no. 1, pp. 1–22, 1997. [Online]. Available: https://doi.org/10.1137/S1064827594276424

[30] S. Adly, H. Attouch, and V. N. Vo, "Asymptotic behavior of newton-like inertial dynamics involving the sum of potential and nonpotential terms," *Fixed Point Theory Algorithms for Science and Engineering*, 2021. [Online]. Available: https://doi.org/10.1186/s13663-021-00702-7

Licence CC BY-NC-ND 3.0

[31] B. Polyak, *Introduction to Optimization* (Transl. Ser. Math. Engrg.). New York: Optimization Software, Inc., Publications Division, New York, 1987.

[32] H. Attouch, X. Goudou, and P. Redont, "The heavy ball with friction method. the continuous dynamical system, global exploration of the local minima of a real-valued function by asymptotical analysis of a dissipative dynamical system," *Communications in Contemporary Mathematics*, vol. 2, no. 1, pp. 1–34, 2000. [Online]. Available: https://doi.org/10.1142/S0219199700000025

[33] H. Attouch and A. Cabot, "Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity," *Journal of Differential Equations*, vol. 263, pp. 5412–5458, 2017. [Online]. Available: https://doi.org/10.1016/j.jde.2017.06.024

[34] H. Attouch, A. Cabot, Z. Chbani, and H. Riahi, "Rate of convergence of inertial gradient dynamics with time-dependent viscous damping coefficient," *Evolution Equations and Control Theory*, vol. 7, no. 3, pp. 353–371, 2018. [Online]. Available: https://www.aimsciences.org/article/id/94a05d99-4b60-48f8-ab35-bf67ea0690cc

[35] A. Cabot, H. Engler, and S. Gadat, "On the long time behavior of second order differential equations with asymptotically small dissipation," *Transactions of the American Mathematical Society*, vol. 361, pp. 5983–6017, 2009. [Online]. Available: https://doi.org/10.48550/arXiv.0710.1107

[36] M. Ghisi, M. Gobbino, and A. Haraux, "The remarkable effectiveness of time-dependent damping terms for second order evolution equations," *SIAM Journal on Control and Optimization*, vol. 54, no. 3, pp. 1266–1294, 2016. [Online]. Available: https://doi.org/10.1137/15M1029485

[37] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, et al., "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature Methods*, vol. 17, pp. 261–272, 2020. [Online]. Available: https://doi.org/10.1038/s41592-019-0686-2

Licence CC BY-NC-ND 3.0

Accélération inertielle en optimisation continue : ajustement du coefficient de viscosité α et convergence vers le flot proximal.

Résumé:
Cette thèse étudie l'accélération des méthodes d'optimisation du premier ordre à partir du lien entre algorithmes discrets et dynamiques continues. Elle analyse la dynamique accélérée $(\mathrm{AVD})_\alpha$ et montre que, pour les fonctions quadratiques convexes, l'écart à l'optimalité décroît en $\mathcal{O}(t^{-\alpha})$, avec des règles explicites pour le choix du paramètre α. Elle examine également le comportement asymptotique de la dynamique (ISIHD) lorsque $\alpha \to +\infty$, mettant en évidence une transition vers un flot proximal. Cette analyse est étendue au cadre des opérateurs cocoercifs pour les dynamiques (AVD), $(\mathrm{DIN\text{-}AVD})$ et (ISIHD). L'ensemble des résultats fournit un cadre d'analyse pour les mécanismes d'accélération en optimisation continue et pour l'étude des systèmes dynamiques inertiels.

Mots clés: opérateurs non potentiels; opérateur cocoercif; dynamique inertielle; amortissement hessien implicite; amortissement hessien explicite; viscosité évanescente; redimensionnement temporel; flot proximal; perturbation singulière; optimisation non convexe; flot de gradient.

Inertial acceleration in continuous optimization: adjusting the viscosity coefficient α and convergence to the proximal flow.

Abstract:
This thesis investigates the acceleration of first-order optimization methods through the connection between discrete algorithms and continuous-time dynamics. It studies the accelerated dynamic $(\mathrm{AVD})_\alpha$ and shows that, for convex quadratic functions, the optimality gap decays at the rate $\mathcal{O}(t^{-\alpha})$, with explicit rules for the choice of the parameter α. The thesis also analyzes the asymptotic behavior of the (ISIHD) dynamic as $\alpha \to +\infty$, showing a transition toward a proximal flow. This analysis is extended to the setting of cocoercive operators for the dynamics (AVD), $(\mathrm{DIN\text{-}AVD})$, and (ISIHD). The results as a whole provide a framework for analysing acceleration mechanisms in continuous optimisation and for studying inertial dynamic systems.

Keywords: non-potential operators; cocoercive operator; inertial dynamics; implicit hessian damping; explicit hessian damping; vanishing viscosity; time rescaling; proximal flow; singular perturbation; non-convex optimization; gradient flows.

Univ. de Limoges, CNRS, XLIM, UMR 7252, F-87000 Limoges, France