THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR
DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE MATHÉMATIQUES ET

INFORMATIQUE

MATHÉMATIQUES APPLIQUÉES ET CALCUL SCIENTIFIQUE

Par **Pierre-Jean BÉNARD**

Algorithmes non-convexes rapides pour la reconstruction de signaux parcimonieux sans grille

Sous la direction de : **Yann TRAONMILIN** et **Jean-François AUJOL**

Soutenue le 15 octobre 2024

Membres du jury :

M. Pierre WEISS	Directeur de Recherche	IRIT, CNRS Toulouse	Rapporteur
M. Jérôme IDIER	Directeur de Recherche	LS2N, CNRS Nantes	Rapporteur
Mme Caroline CHAUX	Directrice de Recherche	IRL, CNRS Aix-Marseille	Examinatrice
M. Baudoin Denis de SENNEVILLE	Directeur de Recherche	IMB, CNRS Bordeaux	Examinateur
Mme Irène WALDSPURGER	Chargée de Recherche	CEREMADE, CNRS Paris	Examinatrice
M. Yann TRAONMILIN	Chargé de Recherche	IMB, CNRS Bordeaux	Directeur de thèse
M. Jean-François AUJOL	Professeur	IMB, Université de Bordeaux	Co-directeur de thèse

Fast non-convex algorithms for off-the-grid sparse signal recovery

Pierre-Jean Bénard

Advisors : Yann Traonmilin and Jean-François Aujol

Algorithmes non-convexes rapides pour la reconstruction de signaux parcimonieux sans grille

Résumé en français

La reconstruction de signaux parcimonieux en dimension infinie est un problème classique dans le champs des problèmes inverses avec de nombreuses applications. Nous posons x_0 comme le signal inconnu à reconstruire que l'on modélise comme une somme de K mesures de Dirac sur $\mathbb{R}^d$. En se basant sur ce modèle, les seules informations nécessaires pour reconstruire x_0 sont ses amplitudes a et positions t associées. L'observation y correspond à l'acquisition composée de m mesures linéaires. Elle peut être modélisée dans le cas sans bruit par $y = Ax_0$ avec A un opérateur linéaire d'acquisition. L'une des manière de reconstruire x_0 à partir de y et A est de trouver le minimiseur d'un problème des moindres carrés non-convexe : $(\mathcal{P})$ $x^* \in \arg\min_{x \in \Sigma_{K,\epsilon}} \|y - Ax\|_2^2$ avec $\Sigma_{K,\epsilon}$ l'ensemble des signaux de K impulsions avec une contrainte de séparation entre les impulsions. Des garanties théoriques de reconstruction ont été données pour des opérateurs d'acquisition A ayant une propriété d'isométrie restreinte (RIP). Ceci avec d'autres résultats théoriques expliquent en partie le succès d'une méthode gourmande : l'algorithme Continuous Orthogonal Matching Pursuit (COMP) avec descente. Cependant, cette méthode reste coûteuse à cause de ses multiples descentes sur tous les paramètres. Nous introduisons alors l'algorithme COMP sur-paramétré avec descente de gradient projetée (OP-COMP + PGD) pour palier à ce problème.

Notre méthode OP-COMP + PGD est composée de deux parties distinctes. La première partie (OP-COMP) est une modification de COMP avec descente, nous supprimons les multiples descente sur tous les paramètres. En sur-paramétrisant le signal, nous compensons le manque de descente. En deuxième partie, PGD effectue une unique descente sur tous les paramètres. Elle réalise aussi une projection qui fusionne les impulsions trop proches les unes des autres. Cela impose au signal estimé de rester dans l'ensemble de solutions $\Sigma_{K,\epsilon}$.

Nous accompagnons ce nouvel algorithme avec une étude théorique sur OP-COMP indiquant que, pour un opérateur linéaire A suivant une RIP, les K premières itérations de OP-COMP initialisent des impulsions proches de celles de x_0. Avec des travaux sur la taille des bassins d'attraction de $\mathcal{P}$, nous assurons qu'avec une bonne initialisation du signal estimé, ce dernier se trouve dans le bassin d'attraction du minimum global et donc OP-COMP + PGD peut reconstruire x_0 exactement.

En termes de résultats expérimentaux, nous comparons OP-COMP + PGD à COMP avec descente sur plusieurs configurations. En somme, la qualité de reconstruction par OP-COMP + PGD est identique à celle de COMP avec descente. Au-delà de $K \geq 15$, OP-COMP devient moins coûteux avec un temps total de calcul d'ordre $\mathcal{O}(K)$ contre $\mathcal{O}(K^2)$ pour COMP avec descente.

En ajout de OP-COMP + PGD, nous développons deux autres techniques d'accélération.

Premièrement, la descente par blocs de coordonnées profite de la parcimonie de notre modèle. Cette méthode cherche à réduire les coûts de calculs lors de la descente de gradient projetée. Elle remplace PGD en effectuant de descente que sur les impulsions qui n'ont pas encore convergé. En comparant cette méthode à PGD, nous observons une amélioration du temps de calcul total de 35% sans perte de précision.

Deuxièmement, le sketching de l'opérateur linéaire consiste à compresser l'observation y pour réduire le nombre total de mesures et donc accélérer les calculs. En compressant avec la transformée de Fourier, nous résolvons une version approximée de $\mathcal{P}$. La qualité de la reconstruction obtenue dépend alors du nombre de mesures compressées. En comparant cette méthode avec OP-COMP + PGD, nous obtenons une précision comparable et une accélération du temps de calcul total de 80% avec un échantillonnage de 15%.

Mots-clés : problème inverse, non convexe, signal parcimonieux, dimension infinie

Fast non-convex algorithms for off-the-grid sparse signal recovery

Abstract in english

The recovery of sparse signals in infinite dimension is a classic problem in the field of inverse problems with many applications. Let x_0 be the unknown signal to recover, which we model as a sum of K Dirac measures on $\mathbb{R}^d$. Based on this model, the only information required to recover x_0 is its associated amplitudes a and positions t. The observation y corresponds to the acquisition of m linear measurements. It can be modeled in the noise-free case by $y = Ax_0$ with A a linear acquisition operator. One way of recovering x_0 from y and A is to find the minimizer of a non-convex least-squares problem : $(\mathcal{P})\quad x^* \in \arg\min_{x \in \Sigma_{K,\epsilon}} \|y - Ax\|_2^2$ with $\Sigma_{K,\epsilon}$ the set of signals from K spikes with a separation constraint between spikes. Theoretical recovery guarantees have been given for acquisition operators A with a restricted isometry property (RIP). This and other theoretical results partly explain the success of a greedy method : the Continuous Orthogonal Matching Pursuit (COMP) with descent algorithm. However, this method remains costly due to its multiple descents on all parameters. We then introduce the Over-Parametrized COMP algorithm with projected gradient descent (OP-COMP + PGD) to overcome this problem.

Our OP-COMP + PGD method consists of two distinct parts. The first part (OP-COMP) is a modification of COMP with descent, in which we remove the multiple descents on all parameters. By over-parametrizing the signal, we compensate for the lack of descent. In the second part, PGD performs a single descent on all parameters. It also enforces a projection that merges spikes that are too close to each other. This ensures that the estimated signal remains within the solution set $\Sigma_{K,\epsilon}$.

We accompany this new algorithm with a theoretical study on OP-COMP indicating that, for a linear operator A following a RIP, the first K iterations of OP-COMP initialize spikes close to those of x_0. With work on the size of the basins of attraction of $\mathcal{P}$, we ensure that with a good initialization of the estimated signal, the latter lies in the basin of attraction of the global minimum and so OP-COMP + PGD can recover x_0 exactly.

In terms of experimental results, we compare OP-COMP + PGD to COMP with descent on several configurations. In summary, the reconstruction quality of OP-COMP + PGD is identical to that of COMP with descent. Above $K \geq 15$, OP-COMP becomes less costly, with a total computation time in the order of $\mathcal{O}(K)$ compared to $\mathcal{O}(K^2)$ for COMP with descent.

In addition to OP-COMP + PGD, we develop two other acceleration techniques.

Firstly, block coordinate descent takes advantage of the sparsity of our model. This method seeks to reduce the computational cost of projected gradient descent. It replaces PGD by performing a descent only on spikes that have not yet converged. When comparing this method with PGD, we observe an improvement in total computation time of 35% with no loss of accuracy.

Secondly, sketching the linear operator involves compressing the y observation to reduce the total number of measurements and thus speed up computation. By sketching with the Fourier transform, we solve an approximation of $\mathcal{P}$. The quality of the recovery then depends on the number of compressed measurements. When comparing this method with OP-COMP + PGD, we obtain comparable accuracy and a total computation time speed-up of 80% with 15% sampling.

Keywords : inverse problem, non-convex, sparse signal, off-the-grid

Institut de Mathématiques de Bordeaux
IMB, UMR 5251 Université de Bordeaux, F 33 405 Talence, France.

Remerciements

C'est en dernière étape avant le rendu final que je me charge d'écrire cette partie. Un peu dans la panique et surtout par manque d'organisation, j'ai pas mal repoussé l'écriture de ces mots pour remercier toutes les personnes qui m'ont accompagné au cours de mes trois années en tant que doctorant. Je souhaite exprimer ma gratitude à tous les lecteurs de cette thèse. Je suis conscient que la rédaction peut comporter des imperfections et des maladresses, et je vous prie de bien vouloir m'en excuser. Votre compréhension et votre indulgence sont grandement appréciées. Merci de prendre le temps de lire mon travail et d'en apprécier le contenu malgré les éventuelles faiblesses dans la forme.

Je tiens à commencer en remerciant mes deux directeurs de thèse qui m'ont accompagné au cours de ces trois années : Yann et Jean-François. Merci à vous deux qui avez su créer un environnement de travail dans lequel j'ai pu m'épanouir tout en me formant au métier de chercheur. Merci à Yann pour sa bienveillance, sa patience et son aide précieuse. En m'aidant à rédiger mes articles, en me suggérant des pistes de recherche pertinentes, et par ta disponibilité constante, tu as grandement contribué à l'avancement de mon travail. Ton soutien indéfectible et tes conseils avisés ont été des atouts précieux tout au long de ce parcours. Merci à Jean-François pour ta rigueur et ta connaissance extensive des thèmes de recherche. Ton expertise approfondie, ta précision et ta disponibilité ont été inestimables pour guider mes travaux.

Je souhaite également remercier mes futurs rapporteurs Pierre Weiss et Jérôme Idier. Je suis reconnaissant pour le temps et l'énergie que vous consacrerez à l'évaluation de mes travaux de recherche. Je remercie par avance mes futurs examinateurs et examinatrices Caroline Chaux, Baudoin Denis de Senneville et Irène Waldspurger qui participerons à ma soutenance. Je suis à la fois honoré et anxieux à l'idée de soutenir devant des chercheurs de votre qualité que j'ai pu croiser tant de fois lors de mes recherches. Mention spéciale à Pierre Weiss qui, il y a maintenant cinq ans, m'a pris en stage de recherche dans ce même domaine sur lequel j'ai pu faire ma thèse. Jamais je n'aurai cru être là maintenant, mais c'est avec cette occasion que je me suis ouvert au monde de la recherche.

Je remercie de plus les doctorants, postdoctorants et stagiaires de l'IMB. Merci à Anne-Edgar, mon co-bureau qui, par sa présence sur une (très) large plage horaire, m'a poussé à être plus constant sur ma régularité. Merci aux doctorants, stagiaires de l'équipe IOP et rattachés : Antoine, Issa, Magalie, Gauthier, Jean, Samuel, Marien, Erell, Ali, Julien, Eloan, Jean-Frédéric, Paul, Théo, Martin, nos repas au Haut-Carré, aux Halles et nos sorties "extra-scolaires" m'ont énormément fait plaisir, permettant de passer/penser à chose, et sortir un peu de mon rythme habituel. Un merci tout particulier à Antoine, mon "frère" de thèse, toujours disponible pour une pause café/thé, avec qui j'ai pu beaucoup échanger sur mon sujet et le sien, mettant les choses à plat et parfois me débloquant dans mes travaux.

Merci également à tous les membres permanents de l'IMB avec qui j'ai pu échanger sur mes travaux, qui m'ont apporté des conseils et aidé administrativement. Notamment merci à la Cellule Informatique, qui de trop nombreuses fois, a réussi à me dépanner quand je bidouillais un peu trop mon ordinateur. Merci au pôle MCD pour leurs conseils techniques pour la mise en place de mes codes sur PlaFRIM et Laurent Facq pour ses formations pertinentes et éducatives. Merci aussi à l'ensemble du personnel administratif, qui m'a accompagné lors de l'organisation de mes déplacements.

Je tiens de même à exprimer ma profonde gratitude à mes proches, en commençant par ma famille. Papa, Maman, votre soutien inconditionnel, vos mots d'encouragement, votre disponibilité constante et votre accueil chaleureux à chaque visite m'ont été d'une aide

précieuse tout au long de ces trois années. Papa, merci pour ces plats maisons, rougails, samosas, bouchons. Maman, merci pour les sessions gaming à te coacher, ça m'a permis de passer du côté enseignant plutôt qu'élève ces fois-ci ! Merci aussi à mon frère préféré Ethan, ton humour et tes bons plans restau sur Paris m'enchante à chaque fois que je reviens !

Passons maintenant à mes amis, il faut bien que je les remercie aussi. Malgré mes huit années loin de la capitale, j'ai eu la chance de conserver de bons amis là-bas. Merci à Mattis, Samy, Luiggi, Benoit, Aaron, Ruben-R2, Antoine, Aline, Stan, Romane, Meyer, Geremy, Thomas, Franck, Maxime, Victor, ça fait toujours plaisir de vous revoir IRL-(quand mes nombreux jours de congés me l'ont permis) ou même sur la faille. Merci tout particulièrement à Mattis et Samy avec qui j'ai pu découvrir de nouveaux jeux, et passer de très bons moments en RP Viking. Avec le gang des Canadiens, j'ai également eu l'occasion de vivre sur deux fuseaux horaires différents en même temps, heureusement que les UUUUU sont éternels. Thanks to Ruben (oui je ne te compte plus dans la team Paris), Gaultier, Quentin, Pierre, Sarah, Maël, Caroline and Patricia. I greatly appreciated my time with you in Montreal and hope to have another occasion to be there again !

Pour finir, je veux remercier tout particulièrement Pauline. Cela fait bien plus que ces trois années que tu m'accompagnes, ton encouragement constant m'a permis de franchir les obstacles et de m'aventurer hors de ma zone de confort, explorant ainsi de nouveaux horizons. En me faisant sortir pour visiter un musée ou même tester le nouveau restaurant du coin, tu m'as aidé à me détendre et à me ressourcer. Grâce à ces moments de découverte et de détente, j'ai pu aborder mon travail avec une perspective fraîche et une énergie renouvelée. Grâce à toi, j'ai pu élargir mes perspectives et aborder mon travail avec un regard neuf. Ta patience, ta compréhension et ton amour ont été des piliers essentiels, me permettant de traverser cette période avec sérénité et assurance. Merci de croire en moi et de m'inspirer chaque jour à donner le meilleur de moi-même.

Chapitre 1

Introduction générale

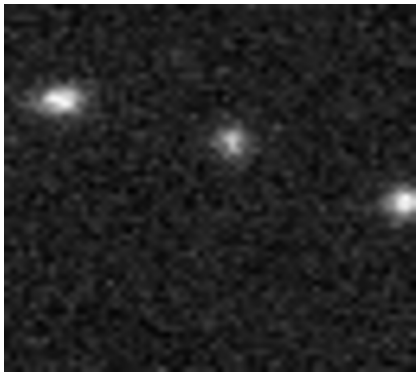
1.1 Un exemple d'application en microscopie

En microscopie, l'observation de micro-organismes, cellules et autres objets de petite taille est essentielle pour leur étude. C'est pourquoi il est important de pouvoir tirer un maximum d'information à partir de ces observations. Parmi ces informations, nous pouvons citer la taille ou la position d'un objet. Cependant, ces informations ne sont pas facilement accessibles à cause des déformations induites lors du processus d'acquisition. Par exemple, pour l'acquisition de cellules au microscope, ces déformations proviennent des limitations techniques et physique de l'appareil. Une cellule observée peut alors être floutée, ce qui rend plus difficile l'estimation de sa position et de sa taille. De plus, contrairement à une cellule qui évolue dans un espace en trois dimensions, son image (ou observation) au microscope est projetée sur un plan en deux dimensions, ce qui réduit d'autant plus l'information totale obtenue lors de l'acquisition. Ce problème d'estimation des positions et amplitudes de cellules en microscopie est appelé la "Single Molecule Localization Microscopy" (SMLM) qui sert de contexte applicatif principal dans nos travaux. Cependant, nos travaux et méthodes sont généralisables à de nombreux domaines tels qu'en astronomie (la localisation d'étoiles dans l'espace), l'écho-location (localisation de sources sonores) etc.

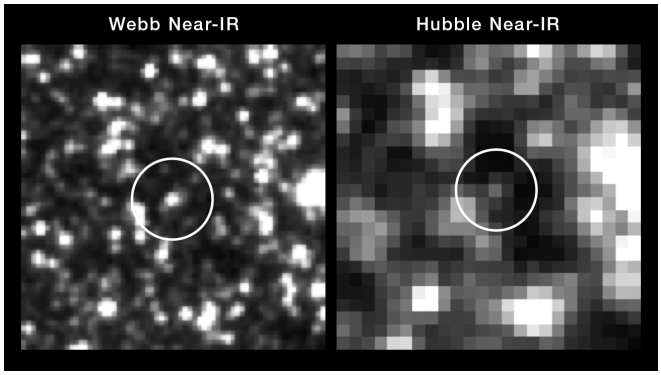
Dans un cadre plus formel, nous pouvons rattacher ce problème au champ du traitement du signal. Dans notre cas, un signal, noté x peut être une collection de cellules, d'étoiles ou de sources sonores définies par leurs positions et amplitudes. Ce signal est alors acquis par un opérateur d'acquisition A, un microscope, un télescope, une collection de microphones. Le résultat est une observation compressée y du signal. Cette catégorie de modélisations où l'on doit retrouver un signal à partir de son observation est appelée *problème inverse*. La notation générale pour ce type de problème est

$$y = Ax \tag{1.1}$$

En supposant que nous connaissons l'observation y et les propriétés de A (son type de déformation, transformation du signal original), nous devons alors reconstruire le signal inconnu x. À noter que l'opérateur A peut être linéaire ou non-linéaire en fonction du problème modélisé. Dans le cas de la SMLM avec le microscope, la linéarité de l'opérateur A s'explique par le fait que l'observation de l'ensemble des cellules correspond à la somme des observations des cellules. Cette propriété de linéarité est élément important dans notre étude et c'est pourquoi nous nous restreindrons uniquement à la résolution de problèmes inverses linéaires en traitement du signal.



(a) Microscope, Challenge SMLM 2016 [74]



(b) Télescopes Webb et Hubble [63]

FIGURE 1.1 – Observations de (a) cellules et (b) d'étoiles à travers différentes méthodes d'acquisition, respectivement un microscope et les télescopes Webb et Hubble.

Cette définition générale de problème inverse linéaire est applicable à une grande variété de modélisations. Bien que nous nous intéressons uniquement au problème d'estimation des positions et amplitudes d'objets, il existe d'autres problèmes qui peuvent s'écrire de la même manière. Nous pouvons citer ceux liés à la reconstruction d'un signal image tels que le débruitage, le défloutage, l'inpainting, la super-résolution etc. Le débruitage cherche à enlever le bruit associé à une observation (par exemple lorsqu'un appareil photo capture une scène où le capteur reçoit des perturbations non désirées). Le défloutage consiste à enlever le flou d'une image lorsque qu'un sujet de l'image ou la caméra bouge lors de la capture. Le flou peut être modélisé comme la convolution entre l'image nette et un noyau de flou pour chaque pixel de l'image. Si le noyau de flou est inconnu, on parle alors de défloutage aveugle. L'inpainting cherche à reconstruire une image alors qu'une partie de ses pixels sont manquants lors de l'acquisition. Et, la super-résolution consiste (comme son nom l'indique) à augmenter la résolution d'une image. Tout ces problèmes sont similaires et reconstruire une image à partir d'une observation dégradée peut parfois recourir à plusieurs de ces méthodes.

1.2 Application du modèle de problème inverse à la reconstruction de signaux parcimonieux

Dans notre cas, pour l'exemple des cellules, seules leurs positions et amplitudes importent. Une idée est donc de considérer l'ensemble des cellules à reconstruire uniquement comme un signal défini par ces paramètres. Ce type de signaux rentre dans la catégorie des signaux parcimonieux. Il s'agit d'un type de signal qui ne dépend que d'un petit nombre de paramètres bien choisis. Pour le comparer à un signal plus classique, par exemple une petite image en noir et blanc composée de $64 \times 64 = 4\,096$ pixels (et autant de paramètres). Un signal parcimonieux composé de 10 cellules en trois dimensions est quant à lui composé de seulement $10 \times (3+1) = 40$ paramètres (trois pour chaque coordonnée de sa position et une pour son amplitude). Pour un autre exemple de signal parcimonieux, une sinusoïde $\sin(c \cdot x)$ qui dans l'espace des fréquence peut être représentée avec une seule fréquence c d'intensité 1.

Pour revenir à notre modèle de signal parcimonieux, nous modélisons nos cellules par leur centre (un unique point contrairement à une vraie cellule qui a un volume et une surface non nuls). Nous souhaitons donc retrouver la position du centre associé à chaque cellule. Un modèle couramment utilisé est la mesure de Dirac. Cet objet fait parti de la catégorie des distributions, une généralisation des fonctions qui permet, par exemple, d'étendre certaines notions de dérivabilité. Cet objet peut être vu comme une impulsion qui vaut 0 partout sauf en un point. Notre signal est alors composé d'une ou plusieurs impulsions appartenant à l'espace des mesures signées $\mathcal{M}$ sur $\mathbb{R}^d$ avec $d \in \mathbb{N}^*$ le nombre de dimensions. Une impulsion est notée δ_t (ou $\delta(t)$) avec $t \in \mathbb{R}^d$ sa position. Elle est généralement associée à une amplitude $a \in \mathbb{R}$. Nous pouvons alors représenter x comme une somme d'impulsions,

$$x = \sum_{i=1}^{K} a_i \delta_{t_i} \tag{1.2}$$

avec $t \in \mathbb{R}^{K \times d}$ les positions et $a \in \mathbb{R}^K$ les amplitudes associées aux différentes impulsions. Ce signal est en effet parcimonieux car il suffit de connaitre uniquement l'ensemble de ses positions et de ses amplitudes pour pouvoir le reconstruire.

Pour la suite des travaux, nous posons x_0 comme le signal parcimonieux inconnu à reconstruire et x un signal parcimonieux quelconque.

Nous pouvons alors définir l'ensemble des paramètres associés à un signal parcimonieux composé de K impulsions,

$$\Theta_K := \left\{ (a_1, \dots, a_K, t_1, \dots, t_K) \in \mathbb{R}^{K \times (d+1)} \right\}. \tag{1.3}$$

Enfin, l'ensemble des signaux parcimonieux peut être modélisé à partir de Θ_K,

$$\Sigma_K := \left\{ x = \sum_{i=1}^{K} a_i \delta_{t_i}, (a_1, \dots, a_K, t_1, \dots, t_K) \in \Theta_K \right\}. \tag{1.4}$$

Notons que les paramètres que nous cherchons à estimer appartiennent au domaine continu. Contrairement au domaine discret où, pour estimer les positions de K impulsions parmi N positions possibles, il faudrait juste tester $\binom{N}{K}$ combinaisons différentes, il est impossible de faire de même dans le domaine continu. À cela s'ajoute le fait que reconstruire exactement un signal dans le domaine continu implique d'estimer parfaitement l'ensemble de ses positions et amplitudes malgré une acquisition dégradée.

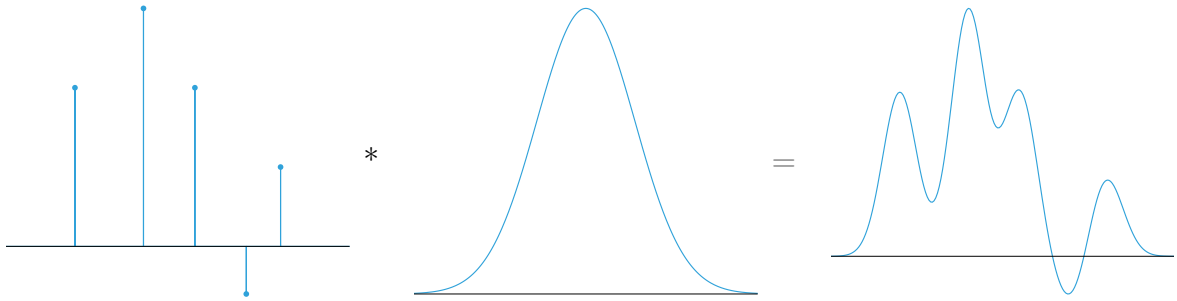
1.2.1 Définition de l'observation acquise par application d'un opérateur linéaire

Concernant l'observation y, il s'agit d'une collection de m mesures appartenant à $\mathbb{C}^m$. Sachant que A est linéaire, nous pouvons réécrire (1.1) par linéarité,

$$y = Ax = A \sum_{i=1}^{K} a_i \delta_{t_i} \tag{1.5}$$

$$y = \sum_{i=1} a_i A \delta_{t_i} \tag{1.6}$$

$$y_l = \sum_{i=1} a_i (A \delta_{t_i})_l. \tag{1.7}$$



(a) x (b) α_l (c) y

FIGURE 1.2 – Représentation d'un signal parcimonieux x composé de 5 mesures de Dirac ou impulsions (gauche), un noyau gaussien d'observation (centre) et le résultat de l'observation du signal parcimonieux x par un opérateur linéaire de convolution gaussien (droite).

Nous devons alors définir $(A\delta_t)_l$ pour tout $l \in \{1, \dots, m\}$. Cette application de l'opérateur linéaire A envoie de l'espace des mesures $\mathcal{M}$ sur $\mathbb{R}^d$ dans $\mathbb{C}^m$. C'est à dire que $Im(A) \subset \mathbb{C}^m$ est définie par

$$\forall l \in \{1, \dots, m\}, \quad (A\delta_t)_l = \langle \delta_t, \alpha_l \rangle \tag{1.8}$$

$$= \int_{\mathbb{R}^d} \alpha_l(t) \mathrm{d}\delta(t) \tag{1.9}$$

$$= \alpha_l(t) \tag{1.10}$$

avec $(\alpha_l)_{l=1,\dots,m}$ une collection de fonctions, $\langle \delta_t, \alpha_l \rangle$ le crochet de dualité entre une mesure finie signée δ_t et d'une fonction qui évalue l'impact de cette mesure selon un paramètre l de l'opérateur d'observation (voir la sous-Section 1.2.2 plusieurs exemples d'opérateurs linéaires dérivés de différents modèles). En revenant à l'équation (1.7), nous obtenons

$$y_l = \sum_{i=1}^{K} a_i \alpha_l(t_i). \tag{1.11}$$

Cette formulation explicite sera utile par la suite pour tous calculs utilisés dans les différentes méthodes de reconstruction de x. Qualitativement, nous pouvons alors en déduire que l'observation y est la combinaison linéaire des réponses pour chaque impulsion dans le signal.

Dans la Figure 1.2, nous montrons un exemple d'un signal parcimonieux x composé de 5 mesures de Dirac (ou impulsions dans la suite de nos travaux) observé par un opérateur linéaire à noyau gaussien. L'application de cet opérateur sur notre signal peut être simplement décrit comme la convolution du signal par le noyau gaussien associé à l'opérateur de mesure. Le résultat est alors l'observation $y = Ax$.

1.2.2 Opérateurs linéaires

Parmi les nombreux opérateurs linéaires qui existent et qui sont utilisés pour modéliser des problèmes inverses de traitement du signal, nous nous concentrons sur deux d'entre eux. Nous avons tout d'abord l'échantillonnage de Fourier qui correspond à l'application de

la transformée de Fourier sur un nombre fini de fréquences. D'un autre côté en microscopie, les opérateurs linéaires sont souvent associés à des fonctions dites "point spread functions" (PSF). Comme les signaux parcimonieux sont composés de sommes de sources ponctuelles, les PSFs décrivent alors la réponse impulsionnelle de ces sources. L'usage des PSFs a une portée plus grande que la microscopie et peut également être utilisé dans d'autres domaines de l'optique ou même de l'imagerie en général.

L'échantillonnage de Fourier

L'échantillonnage de Fourier renvoie une collection de mesures $y \in \mathbb{C}^m$ liées à des fréquences $(\omega_l)_{l=1,\dots,m}$ avec $\omega_l \in \mathbb{R}^d$. Nous obtenons alors

$$(A\delta_t)_l = \alpha_l(t) = \frac{1}{\sqrt{m}} \mathrm{e}^{-j\langle \omega_l, t\rangle}. \tag{1.12}$$

Le choix des fréquences $(\omega_l)_l$ fait l'objet d'autres travaux et ne sera pas abordé ici. Néanmoins, nous expliciterons dans la Section 3 un ordre de grandeur pour déterminer le nombre de mesures adéquat à la reconstruction exacte du signal x.

Nous notons qu'il existe une multitude d'opérateurs linéaires. Nous verrons un autre exemple d'opérateur dans le Chapitre 5 avec le MA-TIRF. Nous avons choisi ce modèle car il représente un modèle général d'illumination réaliste dans le contexte de la microscopie.

1.3 Méthodes pour la reconstruction de signaux parcimonieux

Il existe de nombreuses méthodes pour la reconstruction d'un signal parcimonieux à partir de son observation dégradée. On peut citer trois méthodes différentes. Tout d'abord, il y a les méthodes dites "convexes" qui cherchent à minimiser une fonctionnelle en lien avec le problème qui a subit une relaxation convexe, garantissant un minimum global. À l'inverse, il existe également les méthodes dites "non-convexes". Elles cherchent à minimiser une fonctionnelle non relaxée. Elle ne garantissent alors pas de minimum global unique mais ont un potentiel de reconstruction plus grand. Ces deux types de méthodes sont robustes car accompagnées de beaucoup de travaux théoriques garantissant la reconstruction du signal inconnu x_0. En troisième type de méthode, nous avons les méthodes basées sur l'apprentissage d'un modèle à partir d'un jeu de données. Ces méthodes, assez récentes grâce aux avancées technologiques sur les réseaux de neurones profonds, sont particulièrement efficaces lors de la reconstruction mais manquent des certaines garanties théoriques pour être robustes.

1.3.1 Méthodes convexes

Ces méthodes cherchent à minimiser une fonctionnelle convexe afin d'en tirer une solution à l'équation (1.1). Cette fonctionnelle convexe est composée d'un terme d'attache aux données et d'un terme de régularisation. Elle s'écrit alors de la manière suivante,

$$\min_{x\in\Sigma} \|y - Ax\|_2^2 + \mathcal{R}(x) \tag{1.13}$$

avec $\|y - Ax\|_2^2$ le terme d'attache aux données qui cherche à minimiser la distance entre une estimation Ax^* et l'observation y et $\mathcal{R}(x)$ le terme de régularisation qui impose

un a priori sur le signal à reconstruire. À noter qu'il existe de nombreux choix pour la régularisation $\mathcal{R}$. Il est en général choisi de prendre une norme pour la régularisation. Cette norme impose une structure au signal que l'on cherche à reconstruire et peut également convexifier le problème. On peut tout à fait chercher à réduire le nombre d'impulsions de notre estimation avec la régularisation $\|\cdot\|_0$, c.-à-d., $\|x\|_0 \equiv \#\{i, a_i \neq 0\}$. Nous notons cependant que le choix de la pseudo-norme 0 pour la régularisation ne rend pas le problème convexe, le choix d'une autre norme est donc plus adapté. On peut chercher à minimiser la somme totale des amplitudes de notre estimation en utilisant la norme 1, c'est à dire, $\|x\|_1 \equiv \sum_{i=1}^{K} |a_i|$.

En dimension finie

En dimension finie, l'équation (1.1) peut être réécrite sous la forme $y = Ax$ avec $x \in \mathbb{R}^N$ un vecteur parcimonieux (c'est à dire un vecteur dont peu de ses composantes sont non nulles et donc que $K \ll N$), $A \in \mathcal{M}_{m,N}(\mathbb{C})$ une matrice représentant un opérateur linéaire et $y \in \mathbb{C}^m$ le vecteur des observations.

Comme le vecteur à reconstruire x_0 est de taille finie, il suffit uniquement de reconstruire les composantes non-nulles de ce vecteur. L'algorithme de Frank-Wolfe [50] cherche à résoudre un problème similaire à (1.13) avec un problème quadratique avec des contraintes de positivité pour le vecteur à reconstruire.

Pour revenir à des méthodes plus récentes, nous retrouvons les méthodes de LASSO [83] et de Basis Pursuit (BP) [29]. Ces deux méthodes sont similaires car elles cherchent à résoudre le problème (1.13) avec $\mathcal{R}$ étant la norme 1 des amplitudes du vecteur estimé. Plusieurs méthodes découlent du LASSO et de BP. Nous pouvons citer par exemple le "proximal forward backward splitting" [32] et le Least Angle Regression (LARS) [43].

Des résultats théoriques de reconstruction exacte du vecteur inconnu existent également pour accompagner les résultats empiriques dans [38, 39].

De nouvelles conditions sont données pour la reconstruction exacte avec la régularisation en norme ℓ_1 [23, 40]. Parmi les applications, nous notons celle de la reconstruction des données liées à l'étude sismique apportant de nouvelles conditions [41, 91] et l'introduction d'une méthode basée sur la formulation duale du problème (1.13) [70].

En dimension infinie

En passant de la dimension finie à infinie, nous cherchons également à retrouver les positions exactes des impulsions sur le domaine continu. Ce problème en dimension infinie est aussi connu sous le nom de super-résolution où l'on cherche à retrouver des informations sur les positions et amplitudes à une précision "sous-pixelique" (c'est à dire avec une meilleure précision que celle présente dans l'observation y).

Le BLASSO [34] est une extension du LASSO avec comme fonction de régularisation la norme variation totale pour les mesures qui est une généralisation de la norme ℓ_1 et garantissant la reconstruction en fonction de la distance entre les impulsions à reconstruire. Dans le cas de la reconstruction d'impulsions à amplitudes positives, la stabilité de l'algorithme dépend de la distance entre les impulsions.

La version continue de Frank-Wolfe, aussi appelée Sliding Frank-Wolfe (SFW) ou Frank-Wolfe avec descente est une extension au domaine continu [37] et minimise BLASSO. Des méthodes telles que le Space-Scale Hybrid SFW [60], cherchent à raffiner et améliorer les résultats de ce dernier en se concentrant sur la stabilisation numérique dans le cas d'un changement d'échelle du noyau. Il a également été montré avec l'algorithme SFW [37] que

le signal estimé est garanti de converger vers x^* si il existe un certificat dual non-dégénéré. Cette condition est difficile à vérifier en pratique (au-delà du cas du filtrage passe-bas).

Parmi les autres types d'algorithmes, nous pouvons citer les algorithmes d'affinage adaptatifs [46] qui estiment la position des impulsions en procédant par recherche dichotomique d'une grille qui se raffine au fil des itérations.

Également, il existe les méthodes qui cherchent directement à résoudre le problème (1.13) dans l'espace des mesures. Par exemple, [35] propose une formulation exacte du problème avec une régularisation ℓ_1 directement dans l'espace des mesures de Radon.

Limitations

Comme vu dans cette section, il existe de nombreuses façons d'aborder la résolution du problème convexe 1.13. Cependant, il est clair qu'en dimension finie (support discret), la reconstruction du signal n'est pas suffisamment précise.

De plus, parmi les algorithmes mis en place pour résoudre le problème convexe, la plupart sont gourmands. C'est à dire qu'ils convergent de manière itérative vers la solution en ajoutant une impulsion qui maximise la descente vers le minimum global à chaque itération. La vitesse de convergence de ce processus est alors ralentie à chaque itération, ce qui limite la capacité de reconstruction des signaux avec un grand nombre d'impulsions.

Actuellement, en dimension infinie, l'algorithme de Sliding Frank-Wolfe [37] correspond à l'état de l'art dans son domaine.

Nous verrons avec l'étude des méthodes non-convexes, que si l'on se place suffisamment proche de la solution x^*, le problème non-convexe est convexe localement.

1.3.2 Méthodes non-convexes

En plus des méthodes convexes pour la résolution du problème inverse (1.1), il existe également des méthodes qui se basent sur une formulation non-convexe. Elle s'écrit de la manière suivante,

$$\min_{x \in \Sigma} \|y - Ax\|_2^2 \tag{1.14}$$

En dimension finie

En dimension finie, plusieurs méthodes ont été introduites pour résoudre (1.14).

La méthode de Matching Pursuit (MP) se base sur la décomposition d'un signal à travers une transformation en temps-fréquence. Cet algorithme cherche à reconstruire le signal inconnu en ajoutant une impulsion à chaque itération de manière à diminuer le résidu. Une amélioration directe à MP est Orthogonal Matching Pursuit (OMP) [69]. Cette méthode ajoute une nouvelle impulsion de manière à ce que la nouvelle estimation soit orthogonale au résidu actuel. De cette façon, la nouvelle impulsion ajoutée est la meilleure estimation possible à chaque itération. La preuve de ce résultat de reconstruction exacte d'un signal x_0 par OMP est donnée dans [88, 89].

Une autre méthode non dérivée d'OMP est l'estimation par modèle Poisson-Gaussien [42]. De même, à chaque itération, ce modèle ajoute une impulsion de manière à maximiser la vraisemblance à posteriori. Cette approche d'abord étudiée d'un point de vue statistique cherche à estimer les composants sinusoïdaux d'un signal observé. Un première étape consiste en la détection du nombre de paramètre à estimer, la deuxième étape localise leurs positions et la dernière étape estime leurs amplitudes.

Par ailleurs, outre les méthodes dérivées de OMP, l'algorithme itératif de seuillage fort (Iterative Hard Thresholding) [19, 20] cherche à résoudre le problème (1.13) avec une régularisation ℓ_0 ce qui rend ce problème non-convexe. Il effectue une projection du signal estimé à chaque itération sur l'ensemble de composantes non nulles du signal à reconstruire.

Une autre méthode consiste à appliquer des pénalisations non-convexes. Pour le problème (1.13) avec une régularisation ℓ_q avec $q < 1$, en découpant cette formulation non-convexe en une différence de problèmes convexes et en utilisant un LASSO, il est possible de reconstruire x_0 [51, 28]. En prenant une pénalisation non-convexe ℓ_0 au problème (1.14), l'algorithme CEℓ_0 [81, 82] est la généralisation des méthodes avec seuillages durs.

En dimension infinie

Le passage de la dimension finie à infinie se fait de la même manière que pour le cas convexe, la plupart des algorithmes de cette section sont des extensions des algorithmes sur grilles qui résolvent le problème non-convexe (1.14).

Des garanties théoriques pour la reconstruction de x_0 avec (1.14) ont été données par Gribonval et coll. [53] : par exemple lorsque les fréquences sont tirées avec une distribution gaussienne bien choisie et $m \gtrsim O(K^2 d \operatorname{polylog}(K, d))$, nous avons que $x^* = x_0$. En généralisant à l'estimation des raies spectrales [21], dans le cas d'une observation continue des fréquences jusqu'à une valeur de coupure, il a été montré dans [24] que l'on peut retrouver x_0 s'il obéit à une contrainte de séparation. En effet, en utilisant des mesures de Fourier jusqu'à une fréquence f_c, les signaux parcimonieux avec une distance minimale entre les impulsions de $2/f_c$ peuvent être récupérés exactement. En pratique, le Continuous Orthogonal Matching Pursuit [59] (COMP) a permis de minimiser (1.14). L'OMP en continu s'inspire de l'OMP sur la grille. Des succès dans le cas continu ont été démontrés dans des cas à partir de mesures de Fourier gaussiennes aléatoires [44, 45]. OMP en continu permet alors d'étendre le problème de reconstruction de signaux parcimonieux à d'autres problèmes (tel que le K-means [59]).

En algorithme dérivé de COMP, il existe COMP avec descente et remplacement [57] qui ajoute de nouvelles impulsions et peut éventuellement remplacer celles qui n'apportent plus assez d'information au signal. Cela permet une meilleure reconstruction de x_0 au prix d'un coût de calcul plus élevé.

D'un autre côté, certaines méthodes comme la descente de gradient projetée (PGD) [86] appliquent une contrainte de séparation entre les impulsions initialisées sur une grille en fusionnant les impulsions trop proches les unes des autres.

En combinant les idées d'une initialisation par COMP avec descente et une estimation par PGD, [10] effectue une unique descente sur les paramètres afin d'éviter l'étape de descente coûteuse à chaque itération de COMP avec descente.

Concernant les travaux sur les projections approximatives, [18, 52] introduisent des algorithmes de projection itératifs qui transforment un signal sur-paramétré pour le faire appartenir à l'ensemble souhaité tout en gardant une vitesse de calcul raisonnablement faible. Nous retrouvons des applications plus orientées autour de l'estimation de matrices de rangs faibles [30] et de l'acquisition compressée [22] pour la descente de gradient projetée.

Limitations

Comme énoncé dans cette section, la non-convexité du problème (1.14) ne permet pas de converger vers le minimum global en partant de n'importe quelle initialisation. Par

ailleurs, le manque d'études théoriques sur cette initialisation des méthodes non-convexes ne permet pas d'évaluer la robustesse de ces méthodes. Quelques résultats en lien avec une initialisation sur grille [86] donnent alors des pistes de recherche.

De même que pour le modèle convexe, les algorithmes du modèle non-convexe sont pour la plupart gourmands et donc ralentissent à chaque itération, ne les rendant pas adaptés pour la reconstruction de signaux parcimonieux avec de nombreuses impulsions.

Concernant l'étude théorique sur la convergence des algorithmes, si nous travaillons suffisamment près du minimum global, alors nous sommes dans un domaine convexe localement [87, 85].

1.4 Contributions

Maintenant que nous avons pu voir un panel de l'ensemble des méthodes de reconstruction de signaux parcimonieux et leurs limites, nous allons décrire l'ensemble des contributions que nous avons pu apporter à ce domaine.

1.4.1 Chapitre 2 : Continuous Orthogonal Matching Pursuit sur-paramétré avec descente de gradient projetée

L'algorithme OP-COMP avec PGD (pour Over-Parametrized Continuous Orthogonal Matching Pursuit with Projected Gradient Descent) est notre contribution centrale. Il s'agit d'une méthode rapide sans grille pour reconstruire des signaux parcimonieux sur des domaines avec n'importe quel nombre de dimensions. Nous proposons d'utiliser une initialisation gourmande hors grille sur-paramétrée basée sur la méthode COMP sans étape de descente. Nous effectuons ensuite une descente de gradient projetée qui fournit l'estimation finale. Afin d'accélérer encore plus cette étape de descente, nous implémentons une version accélérée de FISTA (Fast Iterative Shrinkage Algorithm) avec relance. Nous justifions la suppression de l'étape de descente à chaque itération par l'ajout d'une étape de sur-paramétrisation ainsi qu'une étape de projection dans la descente de gradient finale. Nous comparons la complexité algorithmique des deux méthodes OP-COMP avec PGD et COMP avec descente pour mettre en avant l'avantage principal de notre méthode : elle devient plus efficace par rapport à COMP avec descente pour des signaux avec beaucoup d'impulsions. Nous détaillons également les choix des différentes implémentation de notre algorithme.

1.4.2 Chapitre 3 : Garanties théoriques pour l'initialisation d'une solution

Nous donnons des garanties théoriques pour COMP sans descente sous une condition de propriété d'isométrie restreinte (RIP) sur l'opérateur de mesure A. En général, COMP sans descente ne donne pas les vraies positions de x_0. Cependant, nous montrons que la sortie des K-ièmes premières itérations de COMP sans descente approxime les vraies positions avec une précision qui dépend de la qualité de la constante RIP. Qualitativement, cela montre que l'on peut supprimer l'étape de descente de COMP avec descente et rester dans un bassin d'attraction de x_0 si l'on dispose de suffisamment de mesures. Nous ajoutons à ce résultat, une autre garantie théoriques pour des signaux avec amplitudes fixes dans le but d'affaiblir nos hypothèses. En pratique, l'ajout d'une sur-paramétrisation permet d'aller au-delà de la qualité d'approximation donnée par notre résultat à la K-ième

itération. Nous mettons en évidence un lien entre le nombre de mesures d'un opérateur linéaire et sa constante RIP associée. C'est alors avec le nombre de mesures que nous pouvons contrôler l'initialisation d'un signal par COMP sans descente.

1.4.3 Chapitre 4 : Analyses expérimentales de la fonctionnelle non-convexe

Bien que la sur-paramétrisation permette de mieux localiser les impulsions, nous montrons expérimentalement que la Hessienne de la fonctionnelle sur-paramétrée considérée devient très mal conditionnée et même négative autour de la vraie solution. Cela montre la nécessité d'une étape de projection pour réduire le nombre d'impulsions afin d'estimer les positions avec précision d'une manière efficace sur le plan calculatoire. De plus, nous accompagnons cette analyse par une étude quantitative sur la taille du bassin d'attraction de la fonctionnelle non-convexe. Cette partie est en lien direct avec les travaux théoriques de [85].

1.4.4 Chapitre 5 : Expériences et applications à la microscopie

Nous fournissons des expériences en 2D et 3D qui montrent le succès de l'algorithme et jusqu'à une amélioration de cinq fois les temps de calcul par rapport à COMP avec descente. Pour ces expériences nous utilisons deux opérateurs linéaires différents, l'échantillonnage de Fourier et MA-TIRF. Avec l'échantillonnage de Fourier, nous comparons OP-COMP avec PGD contre COMP avec descente sur plusieurs signaux parcimonieux composé de nombreuses impulsions en 2D et 3D. Avec MA-TIRF, nous comparons OP-COMP avec PGD contre COMP avec descente sur deux problèmes : l'estimation des impulsions pour l'étalonnage des microscopes et la localisation de molécules avec des données provenant du défi SMLM [74]. Nous montrons que OP-COMP avec PGD estime les positions avec une précision au moins aussi bonne que COMP avec descente avec des temps de calcul améliorés. L'amélioration des calculs augmente lorsque le nombre K de molécules illuminées augmente, montrant que nous gagnons un ordre de grandeur par rapport à K.

1.4.5 Chapitre 6 : Accélération avec la descente par blocs de coordonnées

Nous introduisons une méthode accélérée de récupération de signaux parcimonieux basée sur OP-COMP avec PGD. Nous proposons une méthode de descente par blocs de coordonnées (BCD, *block coordinate descente*) projetée qui effectue l'étape de descente sur les impulsions qui n'ont pas encore convergé. L'ingrédient principal de cette BCD est une stratégie de sélection de blocs basée sur la structure spécifique du modèle d'impulsions parcimonieuses afin de décomposer le problème en petits blocs plus simples à résoudre. Nous définissons les trois ingrédients définissant le BCD dans le cadre de la méthode OP-COMP avec PGD et justifions qualitativement nos choix. Nous fournissons des expériences synthétiques dans le contexte de la microscopie qui montrent une amélioration jusqu'à 35% du temps de calcul par rapport à PGD.

1.4.6 Chapitre 7 : Accélération par méthode de "*sketching*"

Nous introduisons une méthode accélérée pour la reconstruction de signaux parcimonieux à l'aide de COMP sur-paramétré avec descente de gradient projetée, applicable à des algorithmes similaires tels que COMP avec descente et Sliding Frank-Wolfe. Nous proposons de compresser les mesures de y en approximant un opérateur de compression idéal dans l'espace des mesures, ce qui conduit à une réduction effective du nombre de mesures m. Cette méthode améliore considérablement les temps de calcul globaux OP-COMP avec PGD. En effet, nous présentons des expériences synthétiques dans le contexte de la microscopie qui montrent une amélioration jusqu'à 80% du temps de calcul par rapport à l'opérateur linéaire non compressé avec peu ou pas de perte de précision.

1.4.7 Liste des publications

Nous listons ici l'ensemble des publications associées à ce manuscrit,

- P.-J. Bénard, Y. Traonmilin, and J.-F. Aujol. Fast off-the-grid sparse recovery with over-parametrized projected gradient descent. In *2022 30th EUSIPCO*, pages 2206–2210. IEEE, 2022
- P.-J. Bénard, Y. Traonmilin, J.-F. Aujol, and E. Soubies. Estimation of off-the grid sparse spikes with over-parametrized projected gradient descent: theory and application. *Inverse Problems*, 40(5):055010, 2024
- Y. Traonmilin, J.-F. Aujol, P.-J. Bénard, and A. Leclaire. On strong basins of attractions for non-convex sparse spike estimation: upper and lower bounds. *Journal of Mathematical Imaging and Vision*, 66(1):57–74, 2024
- P.-J. Bénard, Y. Traonmilin, and J. F. Aujol. Projected Block Coordinate Descent for sparse spike estimation. In *2024 32nd EUSIPCO*. IEEE, Feb. 2024. https://hal.science/hal-04462779
- P.-J. Bénard, Y. Traonmilin, and J.-F. Aujol. Sketched over-parametrized projected gradient descent for sparse spike estimation. preprint, https://hal.science/hal-04584951/, Apr. 2024

1.4.8 Recherche reproductible

L'ensemble des codes pour les expériences dans les chapitres suivant sont disponibles aux adresses : [11, 12, 13].

1.4.9 Remerciements pour fonds et ressources

L'ensemble de ces travaux a été subventionné par l'Agence Nationale de la Recherche (ANR) sous les références ANR-20-CE40-0001 (Projet EFFIREG) et ANR-18-CE92-0050 (Projet SUPREMATIM). Les expériences présentées dans ces travaux ont été réalisées à l'aide du centre de calculs expérimental PlaFRIM (Inria, CNRS, Université de Bordeaux, Bordeaux INP et Conseil Régional d'Aquitaine).

Table des matières

Résumé en Français i
Abstract in English ii
Remerciements iii

1 Introduction générale **1**
1.1 Un exemple d'application en microscopie 1
1.2 Application du modèle de problème inverse à la reconstruction de signaux parcimonieux 2
1.2.1 Définition de l'observation acquise par application d'un opérateur linéaire 3
1.2.2 Opérateurs linéaires 4
1.3 Méthodes pour la reconstruction de signaux parcimonieux 5
1.3.1 Méthodes convexes 5
1.3.2 Méthodes non-convexes 7
1.4 Contributions 9
1.4.1 Chapitre 2 : Continuous Orthogonal Matching Pursuit sur-paramétré avec descente de gradient projetée 9
1.4.2 Chapitre 3 : Garanties théoriques pour l'initialisation d'une solution 9
1.4.3 Chapitre 4 : Analyses expérimentales de la fonctionnelle non-convexe 10
1.4.4 Chapitre 5 : Expériences et applications à la microscopie 10
1.4.5 Chapitre 6 : Accélération avec la descente par blocs de coordonnées 10
1.4.6 Chapitre 7 : Accélération par méthode de "*sketching*" 11
1.4.7 Liste des publications 11
1.4.8 Recherche reproductible 11
1.4.9 Remerciements pour fonds et ressources 11

2 Un algorithme rapide : Continuous Orthogonal Matching Pursuit sur-paramétré avec une descente de gradient projetée **15**
2.1 Une initialisation avec COMP sur-paramétré 16
2.1.1 Méthode précédente : l'initialisation sur grille 16
2.1.2 Présentation des algorithmes COMP avec descente et COMP sur-paramétré 17
2.1.3 Le choix de la sur-paramétrisation 18
2.2 Une estimation avec la Descente de Gradient Projetée 19
2.3 Complexité en temps 20
2.4 Implémentation de COMP et des algorithmes de descente avec FISTA avec relance 21
2.5 Résumé du chapitre et conclusions 23

3 L'analyse théorique de COMP sans descente **24**
3.1 Définitions . . . 24
3.2 Théorème général . . . 26
3.3 Théorème pour signaux avec des amplitudes fixes . . . 27
3.4 Schéma de preuve . . . 28
3.5 Relation entre l'hypothèse RIP et le nombre de mesures . . . 29
3.6 Résumé du chapitre et conclusions . . . 29

4 Analyse expérimentale des bassins d'attraction de la fonctionnelle non-convexe **31**
4.1 Le problème mal posé de la sur-paramétrisation . . . 31
4.1.1 Présentation des expériences . . . 31
4.1.2 Résultats et remarques . . . 33
4.1.3 Les méthodes de projection . . . 37
4.2 Estimation de la taille du bassin d'attraction . . . 37
4.2.1 Méthode et paramètres des expériences . . . 38
4.2.2 Expériences pour l'estimation de la taille du bassin d'attraction . . . 38
4.3 Résumé du chapitre et conclusions . . . 42

5 Application à la super-résolution en microscopie et comparaison avec les méthodes état de l'art **43**
5.1 La réflexion fluorescente interne fluorescente total à plusieurs angles (MA-TIRF) . . . 43
5.2 Métriques pour évaluer la performance des algorithmes . . . 44
5.3 Analyse comparative des algorithmes et initialisation de paramètres . . . 45
5.4 Étalonnage du microscope à partir d'acquisitions de billes fluorescentes . . . 46
5.5 Le problème de localisation de SMLM . . . 50
5.6 Vers une amélioration de la précision de la méthode SMLM, en reconstruisant des signaux avec un grand nombre d'impulsions . . . 55
5.7 Résumé du chapitre et conclusions . . . 57

6 Méthodes de descente par blocs appliquées à la descente de gradient projetée **58**
6.1 Introduction à la descente par blocs . . . 58
6.2 L'algorithme de descente de gradient projetée par blocs . . . 59
6.2.1 Un résultat pour guider la sélection des blocs . . . 59
6.2.2 Description de l'algorithme . . . 61
6.3 Expériences . . . 62
6.3.1 Comparaison entre BCD et PGD avec MA-TIRF . . . 64
6.3.2 Comparaison entre BCD et PGD avec la transformée de Fourier . . . 67
6.4 Résumé du chapitre et conclusions . . . 69

7 Méthodes de "sketching" pour la super-résolution **70**
7.1 Introduction au sketching . . . 70
7.2 L'application du sketching à la super-résolution pour le SMLM . . . 71
7.2.1 Compression de mesures gaussiennes dans l'espace des mesures . . . 71
7.2.2 Application à la SMLM . . . 73
7.3 Expériences . . . 75
7.4 Résumé du chapitre et conclusions . . . 78

8 Conclusions **82**
8.1 Récapitulatif des travaux 82
8.2 Limitations associées et suggestions pour de futurs travaux 83

A Glossaire et notations **85**
A.1 Acronymes 85
A.2 Variables 85
A.3 Variables dans la preuve 86

B Preuves des théorèmes du Chapitre 3 **87**
B.1 Preuve des résultats techniques 87
B.2 Preuve des résultats intermédiaires 90
B.3 Preuve de l'hérédité avec amplitudes non fixées (Théorème 3.3) 98
B.4 Preuve principale des Théorèmes 3.3 et 3.4 104
B.5 Preuve de l'hérédité avec amplitudes fixées (Théorème 3.4) 105

Chapitre 2

Un algorithme rapide : Continuous Orthogonal Matching Pursuit sur-paramétré avec une descente de gradient projetée

Dans ce chapitre, nous décrivons notre méthode Over-Parametrized Continuous Orthogonal Matching Pursuit (OP-COMP) avec Projected Gradient Descent (PGD) qui consiste en deux parties : une initialisation gourmande sur-paramétrée hors grille (OP-COMP) et une descente de gradient projetée (PGD). Pour COMP sur-paramétrisé (OP-COMP), nous explicitons l'algorithme et justifions les différences avec la méthode état de l'art COMP avec descente. Pour la descente de gradient projetée (PGD), nous détaillons le choix d'une projection heuristique (à l'opposé de la projection orthogonale) et de son apport dans notre méthode. Nous comparons également la complexité algorithmique de notre méthode avec l'état de l'art COMP avec descente pour justifier une accélération pour la reconstruction de signaux avec un grand nombre d'impulsions. Enfin, nous nous attardons sur quelques implémentations techniques dans notre méthode, avec par exemple l'ajout d'une descente de gradient accélérée (qui ne modifie pas l'idée générale de notre méthode).

Soit $x_0 = \sum_{i=1}^{K} a_i \delta_{t_i}$, le signal à reconstruire composé de K impulsions. Pour retrouver x_0, nous cherchons à minimiser le problème des moindres-carrés non-convexe

$$x_* = \underset{x \in \Sigma_{K,\epsilon}}{\arg\min} \|Ax - y\|_2^2 \tag{2.1}$$

avec $\Sigma_{K,\epsilon}$ une contrainte de séparation sur les positions des impulsions. Sachant que retrouver les amplitudes et positions du signal permettent de le décrire complètement, nous considérons l'ensemble des paramètres $\Theta_{K,\epsilon}$ dérivé de (1.3) où

$$\Theta_{K,\epsilon} \coloneqq \left\{\theta = (a_1, \dots, a_K, t_1, \dots, t_K) \in \mathbb{R}^{K(d+1)}, \forall i, j \in \{1, \dots, K\}, i \neq j, \|t_i - t_j\|_2 > \epsilon\right\}. \tag{2.2}$$

Cet ensemble $\Theta_{K,\epsilon}$ désigne l'ensemble des paramètres (amplitudes et positions) d'un signal composé de K impulsions avec une contrainte de séparation entre chacune des ses impulsions. Nous pouvons alors réécrire tout signal x comme

$$x = \phi(\theta) \coloneqq \sum_{i=1}^{K} b_i \delta_{s_i}. \tag{2.3}$$

Nous posons que x_0 peut également s'écrire de la forme $x_0 = \phi(\theta_0)$ et qu'il appartient donc au modèle de faible dimension

$$\Sigma_{K,\epsilon} \coloneqq \phi(\Theta_{K,\epsilon}) = \left\{ \sum_{i=1}^{K} a_i \delta_{t_i}, (a_i, t_i) \in \Theta_{K,\epsilon} \right\}. \tag{2.4}$$

Le problème des moindres carrés non-convexe (1.14) peut être écrit de manière équivalente :

$$\theta^* \in \underset{\theta \in \Theta_{K,\epsilon}}{\arg\min}\, g(\theta) \quad \text{avec } g(\theta) \coloneqq \|A\phi(\theta) - y\|_2^2. \tag{2.5}$$

Nous avons donc que la méthode OP-COMP avec PGD cherche à estimer les paramètres θ^* (amplitudes et positions) du signal à reconstruire qui minimisent la fonctionnelle g (2.5) afin de reconstruire x_0. La première étape (OP-COMP) agit comme une initialisation de nos paramètres θ_{init}. La deuxième étape (PGD) part de cette initialisation pour estimer les paramètres θ_{esti}. Dans les bonnes conditions, notre méthode retourne alors $\theta_{\text{esti}} = \theta^*$.

Nous procédons avec un algorithme en deux parties distinctes (initialisation avec OP-COMP et estimation avec PGD) car l'algorithme de descente utilisé dans PGD converge vers le minimiseur global de g, si il est initialisé dans son bassin d'attraction. L'initialisation par OP-COMP cherche donc à créer un signal qui soit dans suffisamment proche de x^* pour que l'estimation par PGD converge bien vers la solution de (2.5). Nous détaillons plus en détail ces choix dans le Chapitre 3.

2.1 Une initialisation avec COMP sur-paramétré

2.1.1 Méthode précédente : l'initialisation sur grille

Avant l'introduction de COMP comme étape d'initialisation d'un signal x_{init} pour PGD, une méthode gourmande d'initialisation sur-paramétrée sur grille était utilisée [86]. Cette méthode est basée sur l'emploi de la rétro-projection de l'observation y sur une grille. En effet, l'observation y n'est pas toujours composée de mesures en espace (c'est à dire appartenant à une grille Γ discrétisant l'espace des positions t du signal à reconstruire). Par exemple, pour l'échantillonnage de Fourier, les mesures y_l représentent le coefficient du signal x_0 associé à la fréquence ω_l. Pour construire cette rétro-projection, nous utilisons une grille Γ et des observations y. Nous avons alors, pour l'échantillonnage de Fourier,

$$z_\Gamma = \sum_{s_j \in \Gamma} z_{\Gamma,j} \delta_{s_i} \quad \text{avec} \quad z_{\Gamma,j} = \sum_{l=1}^{m} y_l e^{j\langle \omega_l, s_j \rangle}. \tag{2.6}$$

avec z_Γ la rétro-projection des observations y sur la grille Γ. Il est alors possible d'extraire de manière gourmande un θ_{init} initialisé à partir des plus grandes amplitudes dans z_Γ [86]. Bien sûr, l'étape critique pour l'utilisation d'une seule passe de PGD est que l'initialisation sur-paramétrée θ_{init} soit suffisamment proche de θ^*. Plus la séparation ϵ entre les impulsions de x_0 sera proche de 0, plus il faudra une grille Γ fine pour une rétro-projection fidèle du signal à reconstruire. Bien que cette méthode sur grille soit efficace en 2D pour l'initialisation de PGD (et utile pour visualiser facilement la qualité de l'échantillonnage), le fléau de la dimension limite la possibilité d'étendre une telle méthode sur grille à des domaines de dimensions supérieures. C'est pourquoi nous introduisons une méthode d'initialisation sans grille COMP sur-paramétré.

2.1.2 Présentation des algorithmes COMP avec descente et COMP sur-paramétré

Introduction de OMP en continu

Les deux algorithmes COMP avec descente et COMP sur-paramétré sont sont similaires et sont décrit dans l'Algorithme 1.

Algorithm 1 Continuous Orthogonal Matching Pursuit algorithm (COMP). Le COMP avec descente effectue une reconstruction des impulsions selon l'état de l'art. Nous utilisons une version sur-paramétrée de COMP sans descente comme initialisation de notre descente de gradient projetée.

procedure COMP($A, y, K,$ `is_sliding`)
  $r^{(0)} \leftarrow y$
  $t^{(0)} \leftarrow \{\}$
  for $k = 1 \rightarrow K$ **do**
    $T \leftarrow \arg\max_t \langle A\delta_t, r^{(k-1)} \rangle$
    $t^{(k)} \leftarrow t^{(k-1)} \cup \{T\}$
    $a^{(k)} \leftarrow \arg\min_a \|A \sum_{i=1}^{k} a_i \delta_{t_i^{(k)}} - y\|_2^2$ ▷ Mise à jour des amplitudes
    if `is_sliding` **then**
      $a^{(k)}, t^{(k)} \leftarrow$ `descente`$(g, a^{(k)}, t^{(k)})$
    end if
    $r^{(k)} \leftarrow y - A \sum_{i=1}^{k} a_i^{(k)} \delta_{t_i^{(k)}}$ ▷ Mise à jour du résidu
  end for
  retourne $a^{(K)}, t^{(K)}$
end procedure

Nous notons dans cet algorithme l'introduction du résidu $r^{(k)}$ dans la résolution de (2.5) avec COMP. Il s'agit de la différence entre l'observation y du signal à reconstruire et de l'observation $Ax^{(k)}$ du signal estimé. Ce résidu nous apporte l'information non estimée avec $x^{(k)}$. Par ailleurs, nous remarquons que la fonctionnelle g (2.5) à l'itération k peut s'écrire $g(\theta^k) = \|r^{(k)}\|_2^2$. Donc en minimisant la norme du résidu (c'est à dire en reconstruisant toute l'information de y), nous résolvons le problème de minimisation de g.

Description de COMP avec descente

Comme décrit dans l'Algorithme 1, COMP avec descente itère sur le nombre total d'impulsions à ajouter. Il trouve un emplacement où l'impulsion maximise la corrélation avec le résidu des mesures. Il effectue ensuite une mise à jour des amplitudes et actualise le résidu.

Dans le cas spécifique de COMP avec descente, après la mise à jour des amplitudes, une descente sur tous les paramètres déjà initialisés est effectuée afin d'obtenir une meilleure initialisation à chaque itération de la boucle.

Finalement, le résidu est mis à jour avec les nouveaux paramètres obtenus à cette itération. La boucle se finit alors au bout de K itérations et retourne les paramètres estimés $a^{(K)}$ et $t^{(K)}$.

Description de COMP sans descente sur-paramétré

Pour COMP sur-paramétré, l'idée est de supprimer la partie de descente sur les paramètres déjà initialisés. Nous supposons qu'une initialisation d'un signal x_{init} sans descente à chaque itération de COMP est suffisamment proche de x^* pour pouvoir converger vers cette solution avec la descente de gradient projetée.

La méthode COMP sur-paramétré est une modification de COMP avec descente. Il s'agit de la simple exécution de COMP sans descente avec $K_{\text{op}} > K$ itérations. Son fonctionnement reste identique à celui de COMP avec descente (voir paragraphe précédent pour la description de l'Algorithme 1)

Cependant, comme le nombre de d'impulsions initialisées n'est pas fixé, il faut déterminer un critère d'arrêt pour OP-COMP. En effet, il ne s'agit que d'une première initialisation brute avec l'étape de raffinement avec PGD. Pour estimer si nous pouvons arrêter les itérations de OP-COMP, nous regardons deux critères. Si la norme du résidu $r^{(k)}$ diminue en dessous d'un seuil ou si le nombre d'impulsions à ajouter dépasse une certaine quantité, nous arrêtons les itérations.

Différence entre COMP sur-paramétré et COMP avec descente

Nous pouvons alors choisir d'effectuer l'étape de descente ou pas (avec le paramètre `is_sliding`), et même contrôler la sur-paramétrisation du signal initialisé (en modifiant le paramètre K) si nous le souhaitons. Le passage de COMP avec descente à OP-COMP est donc très simple et contrôlable avec deux paramètres K et `is_sliding` dans l'Algorithme 1.

Le rôle de la sur-paramétrisation est de compenser les imprécisions induites à chaque itération par la suppression de l'étape de descente. Nous pouvons considérer qu'il s'agit d'une généralisation hors grille de l'initialisation gourmande sur grille proposée précédemment [86]. Nous allons maintenant faire une étude qualitative sur le contrôle de la sur-paramétrisation dans OP-COMP.

2.1.3 Le choix de la sur-paramétrisation

Une question légitime est de se demander quand stopper la sur-paramétrisation du signal x_{init} que nous allons noter K_{op}. Dans des travaux précédents, il a été proposé de choisir un multiple fixe du nombre réel estimé d'impulsions [86]. Dans notre nouvelle méthode, nous ajoutons des impulsions jusqu'à ce que le résidu atteigne un seuil ou qu'il cesse de diminuer. Cette méthode permet de s'assurer que l'ajout d'impulsions supplémentaires avec COMP n'apporterait pas plus d'informations dans l'initialisation de la descente du gradient projeté (puisque le résidu ne diminuerait pas davantage avec OP-COMP).

Nous définissons alors un critère d'arrêt relatif à l'observation du signal à reconstruire $y = Ax_0$ et à l'observation de notre estimation Ax,

$$\frac{\|Ax - Ax_0\|_2}{\|Ax_0\|_2} \leq \epsilon. \tag{2.7}$$

Nous pouvons alors, à partir de ce critère, estimer la précision relative de notre estimation x avec le signal à reconstruire x_0. Grâce à la propriété d'isométrie restreinte de constante γ sur notre opérateur linéaire A par respect de la norme à noyau $\|\cdot\|_h$ (voir Définition 4) et en supposant que $x \in \Sigma_{K,\epsilon}$, nous avons alors que le critère (2.7) implique

$$\sqrt{1-\gamma}\|x - x_0\|_h \leq \epsilon\sqrt{1+\gamma}\|x_0\|_h \quad \text{c-à-d} \quad \sqrt{\tfrac{1-\gamma}{1+\gamma}}\tfrac{\|x-x_0\|_h}{\|x_0\|_h} \leq \epsilon. \tag{2.8}$$

Cela signifie que si A a une constante RIP γ suffisamment proche de 0 (c'est à dire $\sqrt{(1-\gamma)/(1+\gamma)}$ est proche de 1), nous pouvons alors avoir un bon contrôle sur l'erreur relative de l'estimation des impulsions uniquement à partir de leurs observations. Dans nos expériences (voir le Chapitre 5), nous observons qu'un seuil associé au critère relatif de $\epsilon = 5\%$ est suffisant pour stopper les itérations de OP-COMP et obtenir de bonnes initialisations. Ce critère nous donne une bonne estimation du moment où il faut arrêter la sur-paramétrisation, car il garantit que la solution sur-paramétrée atteint la même valeur de la fonctionnelle qu'une estimation de K impulsions qui estime bien les vraies positions de x_0.

Il convient de noter que l'algorithme plus robuste COMP avec descente et avec remplacement a été proposé dans [57]. Dans la version avec remplacement, $2K$ impulsions sont estimés pour produire une estimation avec K impulsions (ce qui rend les dernières étapes de descente plus coûteuses). Nous choisissons de nous comparer à la version plus rapide sans remplacement car l'objectif principal de notre algorithme est de fournir une estimation rapide (c'est-à-dire que nous nous plaçons dans le cas le moins favorable pour la comparaison des temps d'exécution).

2.2 Une estimation avec la Descente de Gradient Projetée

En deuxième partie de notre algorithme OP-COMP avec PGD, nous effectuons la descente de gradient projetée (PGD). Nous retrouvons plusieurs apparitions de la PGD, comme introduite dans [19], pour l'acquisition spectrale compressée dans [22] ou dans ce cas spécifique de la super-résolution et la reconstruction de signaux parcimonieux dans [86]. Il s'agit d'une manière d'effectuer une simple descente de gradient sur la fonction g tout en garantissant que nos paramètres estimés θ appartiennent à notre ensemble d'arrivée $\Theta_{K,\epsilon}$, c'est à dire que notre signal estimé x_0 appartient bien à $\Sigma_{k,\epsilon}$.

Nous avons observé dans [86, 10] que la sur-paramétrisation (c'est-à-dire OP-COMP) permet d'assurer la localisation approximative de toutes les impulsions dans le signal même sans utiliser l'étape de descente (voir également la Section 4 et les travaux connexes pour une discussion sur la sur-paramétrisation avec un grand nombre d'impulsions). C'est un moyen d'aller au-delà des résultats théoriques de la Section 3. Malheureusement, la Hessienne de la fonction considérée devient mal conditionnée et non positive, ce qui entraîne une convergence très lente de la descente de gradient classique vers les vraies positions de x_0, comme nous le verrons dans la Section 4. Ceci motive alors le fait d'ajouter une étape de projection dans la descente.

Pour réduire le nombre d'impulsions, nous effectuons une projection heuristique P_ϵ basée sur la contrainte de séparation à chaque étape de la descente. Si deux impulsions sont trop proches l'une de l'autre, leurs positions et leurs amplitudes sont fusionnées, c'est-à-dire que la descente projetée est décrite par

$$(a^{(n+1)}, t^{(n+1)}) = P_\epsilon \left(\texttt{descent}(a^{(n)}, t^{(n)}) \right) \tag{PGD}$$

où notre implémentation rapide de $\texttt{descent}(a^{(n)}, t^{(n)})$ est décrite dans la Section 2.4.

La projection heuristique P_ϵ est décrite dans l'Algorithme 2. Une étape importante de notre projection heuristique P_ϵ est que les positions fusionnées sont les barycentres (en utilisant la valeur absolue des amplitudes) des positions précédentes. De cette manière, nous nous assurons que l'étape de projection produit une impulsion située dans l'enveloppe

convexe des positions que l'on fusionne. Comme nous supposons que pour un ϵ suffisamment petit, les positions de toutes les impulsions fusionnées appartiennent au même bassin d'attraction, la sortie de P_ϵ appartient également à ce bassin d'attraction. Notons que P_ϵ effectue également un seuillage des très petites amplitudes car les impulsions de petites amplitudes n'ajoutent pas beaucoup d'informations et ralentissent le calcul. Le seuil est fixé manuellement. En pratique, nous choisissons un seuil $\texttt{threshold} = 10\%$ de l'amplitude minimale dans la vérité terrain, car cela permet d'éliminer certaines impulsions inutiles sans perdre en précision pour la reconstruction de x_0.

Il est possible d'envisager une "projection" théorique sur la contrainte de séparation, par exemple, en considérant pour une norme donnée sur $\mathcal{M}$, le problème suivant

$$\inf_{x \in \Sigma_{K,\epsilon}} \left\| x - \sum_{i=1}^{K} a_i \delta_{t_i} \right\|. \tag{2.9}$$

Cependant, le choix de la norme (norme à noyau, distance de Wasserstein, etc.) et la résolution de ce problème ne sont pas simples. Pour montrer la convergence de cette méthode, il faudrait garantir que la projection conserve les positions dans les bassins d'attraction de la fonctionnelle, ce qui n'entre pas dans le cadre de notre étude.

Algorithm 2 Projection sur la contrainte de séparation P_ϵ.

procedure P$(a, t, \epsilon, \texttt{threshold})$
  $\texttt{sort}(a_i, t_i)_{i=1,\dots,k}$ such that $|a_1| \geq \cdots \geq |a_k|$
  for $i = 1, \dots, k$ **do**
    if $a_i = 0$ **then**
      Skip iteration
    else if $a_i \leq \texttt{threshold}$ **then**
      $a_i \leftarrow 0$
    else
      Get $J = \{j, j \geq i, \|t_i - t_j\|_2 \leq \epsilon\}$
      $t_i \leftarrow \frac{1}{\sum_{j \in J} |a_j|} \sum_{j \in J} |a_j| t_j$
      $a_i \leftarrow \sum_{j \in J} a_j$
      $a_j \leftarrow 0$ for all $j \in J, j \neq i$
    end if
  end for
  return $(a_i, t_i)_{i \in I}$ with $I = \{i, a_i \neq 0\}$
end procedure

2.3 Complexité en temps

Pour estimer le temps gagné avec notre méthode OP-COMP avec PGD par rapport à la méthode COMP avec descente, nous analysons qualitativement le nombre d'itérations dans l'étape de descente (qui est la plus coûteuse en temps de calcul) dans chaque méthode (c'est à dire PGD pour notre méthode et la descente de gradient à chaque itération pour COMP avec descente). Pour la méthode COMP avec descente, le temps d'exécution est de l'ordre de

$$\mathtt{T}_{\texttt{SCOMP}} = \mathcal{O}(N \times (1 + \cdots + K)) = \mathcal{O}(NK^2) \tag{2.10}$$

avec N le nombre d'itérations à chaque descente de gradient et K le nombre d'impulsions. Pour notre descente de gradient projetée (PGD) avec un signal sur-paramétré, le temps

d'exécution est de

$$\mathtt{T}_{\mathtt{PGD}} = \mathcal{O}(K \sum_i N_i C_i) \tag{2.11}$$

avec $N_0 = \sum_i N_i$ le nombre d'itérations dans la méthode de descente de gradient projetée, N_i le nombre d'itérations entre chaque projection et C_i la constante symbolisant la sur-paramétrisation entre chaque projection, c'est-à-dire qu'après la i-ième projection, il reste $K_i = C_i K$ impulsions.

Comme les facteurs de sur-paramétrisation C_i sont d'ordre $\mathcal{O}(1)$, nous avons alors que $K_i = C_i K$ (le nombre d'impulsion après chaque projection) sont d'ordre $\mathcal{O}(K)$. Nous observons que lorsque le nombre d'impulsions K augmente, la descente de gradient projetée PGD devient plus intéressante que COMP avec descente. Qualitativement, nous gagnons un facteur K par rapport à COMP avec descente. Nous remarquons également que PGD est plus rapide si les projections ont lieu au début de la descente. De plus, par rapport à l'initialisation basée sur la grille de [86], la complexité de COMP sur-paramétrée n'est plus exponentielle dans la dimension du support des impulsions.

D'un point de vue informatique, cela montre l'intérêt d'éviter l'étape de descente pour un grand nombre d'impulsions si possible.

2.4 Implémentation de COMP et des algorithmes de descente avec FISTA avec relance

Nous décrivons ici les choix spécifiques faits pour chaque partie de COMP (Algorithme 1).

- Calcul de $\arg\max_t \langle A\delta_t, r^{(k-1)} \rangle$: nous effectuons une descente sur $-\langle A\delta_t, r^{(k-1)} \rangle$ en utilisant FISTA avec relance (voir Algorithme 3) jusqu'à convergence vers une estimation $t_{\mathtt{conv}}$. Pour les mesures spatiales sur une grille avec une réponse impulsionnelle concentrée autour des impulsions, nous initialisons de manière gourmande la descente avec une position sur la grille avec la plus forte amplitude (comme cela est fait dans la partie expérimentale de nos travaux avec un opérateur linéaire A tiré de la microscopie). Pour les mesures de Fourier et autre mesures ne vivant pas sur une grille, nous avons observé qu'une initialisation aléatoire sur $\mathcal{D}$ était efficace [10]. Afin de limiter la sur-paramétrisation totale du signal avec des positions initialisées approximatives, il est possible effectuer cette étape plusieurs fois (avec une éventuelle parallélisation du processus) avec d'obtenir la meilleure estimation $t_{\mathtt{conv}}$ à chaque étape de OP-COMP.
- Calcul de $\arg\min_a \|A \sum_{i=1}^k a_i \delta_{t_i^{(k)}} - y\|_2^2$: Nous pouvons résoudre ceci en utilisant une méthode traditionnelle des moindres-carrés à dimension finie. En effet, nous avons que

$$A \sum_{i=1}^k a_i \delta_{t_i} = \sum_{i=1}^k a_i (A\delta_{t_i}) = Ba \tag{2.12}$$

  où B est la matrice donc les colonnes sont définies par $B_i = A\delta_{t_i}$.
- Calcul de $\mathtt{descent}(a^{(k)}, t^{(k)})$: comme il est crucial d'effectuer cette étape rapidement étant donné la dimension des objets, nous utilisons l'algorithme de FISTA avec relance [68, 1] sur la fonction

$$g(a,t) = \left\| A \sum_{i=1}^k a_i \delta_{t_i} - y \right\|_2^2 . \tag{2.13}$$

En effet, l'inertie introduite par FISTA peut causer problème sur la vitesse de convergence. Au lieu de diminuer vers le minimum du bassin d'attraction, il peut tourner en spirale autour de ce dernier. En regardant la distance entre le point estimé et le minimum du bassin d'attraction lors de cette descente en spirale (de manière générale $\|x - x_0\|_2$ ou dans notre cas des signaux parcimonieux $\|Ax - y\|_2$), on remarque que cette distance augmente et diminue au fur et à mesure des itérations, entraînant une convergence lente (bien que plus rapide qu'avec une descente de gradient classique). C'est ce qu'on appelle le phénomène de rebond. Cet effet de rebond ralentit la convergence du système. Pour contrer cet effet, la méthode de FISTA avec relance, qui consiste à réinitialiser l'inertie au stade le plus bas du rebond, a été introduite. Alors que certains articles travaillent sur une estimation du moment où il faut redémarrer l'inertie [3], nous pouvons simplement évaluer notre critère `restart_cond` et redémarrer l'inertie à ce point. Aucun travail théorique ne prouve l'accélération de cette méthode par rapport à FISTA, mais elle est très efficace en pratique et simple à mettre en œuvre, [6, 7, 68]. Il existe plusieurs critères `restart_cond` faciles à implémenter et qui effectuent un relance au plus bas du rebond. Nous avons le "critère de fonction" : $f(x^{(k+1)}) > f(x^{(k)})$ et le "critère de gradient" : $\nabla f(y^{(k)})^T(x^{(k+1)} - x^{(k)})$. En pratique, nous choisissons le critère de gradient car tous sont composants sont déjà calculés lors de la boucle de FISTA et qu'il est numériquement plus stable lorsque notre estimation est très proche du minimum du bassin d'attraction [68].

Algorithm 3 FISTA avec relance pour une fonction $f(z)$.

procedure FISTA RESTART($z, N, \tau,$ `restart_cond`)
  $x^{(0)} \leftarrow z$
  $y^{(0)} \leftarrow z$
  $\theta_0 \leftarrow 1$
  for $k = 0, 1, \ldots, N$ **do**
    $x^{(k+1)} \leftarrow y^{(k)} - \tau \nabla f(y^{(k)})$
    $\theta_{k+1} \leftarrow \left(1 + \sqrt{1 + 4\theta_k^2}\right)/2$
    $\beta_{k+1} \leftarrow (\theta_k - 1)/\theta_{k+1}$
    $y^{(k)} \leftarrow x^{(k)} + \beta_{k+1}(x^{(k+1)} - x^{(k)})$
    if `restart_cond`$(y^{(k)}, x^{(k)}, x^{(k+1)})$ **then**
      $\theta_{k+1} \leftarrow 1$
    end if
  end for
  return $x^{(N)}$
end procedure

Nous notons que pour tous les algorithmes de descente utilisés dans COMP avec descente et OP-COMP avec PGD, nous utilisons FISTA avec relance. Dans PGD, après chaque projection, si au moins deux impulsions sont fusionnées, nous réinitialisons l'inertie du système. En effet, la différence $x^{\text{après}} - x^{\text{avant}}$ n'est plus possible car ces deux vecteurs ne sont pas de la même taille.

2.5 Résumé du chapitre et conclusions

Dans ce chapitre, nous avons pu voir en détail l'implémentation de notre méthode OP-COMP avec PGD. Cette méthode est notre fil conducteur tout le long de ce nos travaux. Nous avons justifié qualitativement l'utilisation des idées de sur-paramétrisation et de projection de notre estimation pour une reconstruction exacte du signal x_0. Cette sur-paramétrisation devient plus intéressante avec les signaux à reconstruire composé d'un grand nombre K d'impulsions. En effet, nous gagnons en vitesse de calcul par rapport aux autres méthodes état de l'art COMP avec descente et Frank-Wolfe avec descente (SFW) d'un facteur d'ordre K. En outre, notre méthode peut également être accélérée en fonction de nos implémentations pratiques. Par exemple, l'utilisation de FISTA avec relance pour tout descente de gradient (dans PGD ou le calcul de $\arg\max_t \langle A\delta_t, r\rangle$ dans OP-COMP) permet à notre méthode de résoudre le problème initial (2.1).

Dans le prochain chapitre, nous cherchons à accompagner notre nouvelle méthode d'une analyse théorique concernant son étape d'initialisation (OP-COMP). En effet, peut-on contrôler l'erreur d'estimation venant de la suppression de l'étape de descente à chaque itération ?

Chapitre 3

L'analyse théorique de COMP sans descente

Le but de OP-COMP est d'initialiser un signal x_{init} suffisamment proche de la solution x^* de manière à ce que les itérés obtenus par la descente de gradient projetée (PGD) convergent vers la solution. Pour converger, il faut qu'au départ, x_{init} soit dans le bassin d'attraction de x^*. Nous montrons dans ce chapitre qu'avec quelques hypothèses supplémentaires, en plus de celles nécessaires pour le bon fonctionnement de COMP avec descente, il n'est pas nécessaire d'effectuer l'étape de descente avec COMP pour obtenir une approximation contrôlée des positions des impulsions par rapport à la solution x^*. Dans [44, 45], il est montré que COMP sans descente récupère exactement K impulsions en K étapes pour une réponse impulsionnelle exponentielle, mais ces résultats ne peuvent pas être appliqués pour les filtres passe-bas tels que les filtres gaussiens qui sont plus courants dans le traitement du signal et de l'image. En particulier, le cas gaussien est proche de l'exemple de la microscopie.

Nous commençons par introduire quelques définitions et hypothèses sur l'opérateur linéaire A et sur les mesures de Dirac ainsi que nos deux hypothèses fondamentales sur le noyau (la cohérence mutuelle) et l'opérateur linéaire associé A (la propriété d'isométrie restreinte). Nous allons continuer avec l'énonciation de notre résultat principal ainsi que d'un deuxième résultat avec des hypothèses moins fortes sur le contrôle des signaux avec une amplitude fixe. Nous détaillerons dans les grandes ligne un schéma de preuve pour nos deux résultats. Et enfin, nous discuterons du lien qu'entretient la propriété d'isométrie restreinte avec le nombre de mesures m de notre observation y. Les parties techniques des preuves sont renvoyées en Annexe B.

3.1 Définitions

Nous introduisons tout d'abord la notion de norme à noyau qui permet d'utiliser la propriété d'isométrie restreinte (RIP) dans l'estimation des impulsions sur un domaine continu (voir par exemple [84]),

Définition 1 (Noyau, produit scalaire et norme)**.** *Pour des mesures finies et signées sur $\mathbb{R}^d$, l'espace de Hilbert construit avec le noyau h (une fonction dérivable sur $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$) est défini par le produit scalaire suivant entre 2 mesures ν_1 et ν_2,*

$$\langle \nu_1, \nu_2 \rangle_h = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(t_1, t_2) \mathrm{d}\nu_1(t_1) \mathrm{d}\nu_2(t_2). \tag{3.1}$$

Nous définissons alors

$$\|\nu_1\|_h^2 = \langle \nu_1, \nu_1 \rangle_h. \tag{3.2}$$

Nous obtenons la relation

$$\|\nu_1 + \nu_2\|_h^2 = \|\nu_1\|_h^2 + 2\langle \nu_1, \nu_2 \rangle_h + \|\nu_2\|_h^2. \tag{3.3}$$

Un objet fondamental pour la suite de notre étude est le concept de dipôle. Il nous permet d'étudier plus finement des propriétés sur les mesures de Dirac en évaluant leurs différences deux à deux.

Définition 2 ((ξ-)Dipôle et séparation)**.** *Un ξ-dipôle (noté dipôle par simplicité) est une mesure $\pi_1 = a_1\delta_{t_1} - b_1\delta_{s_1}$ où $\|t_1 - s_1\|_2 \leq \xi$, avec $a_1, b_1 \in \mathbb{R}$. Deux dipôles $\pi_1 = a_1\delta_{t_1} - b_1\delta_{s_1}$ et $\pi_2 = a_2\delta_{t_2} - b_2\delta_{s_2}$ sont ϵ-séparés si leurs supports sont strictement ϵ-séparés (par rapport à la norme ℓ^2 sur $\mathbb{R}^d$), c'est-à-dire, si $\|t_1 - t_2\|_2 > \epsilon$, $\|t_1 - s_2\|_2 > \epsilon$, $\|s_1 - t_2\|_2 > \epsilon$ et $\|s_1 - s_2\|_2 > \epsilon$.*

Nous utilisons pour le reste de ce chapitre (et le reste du manuscrit) le noyau Gaussien défini par $h(t, s) = \mathrm{e}^{-\|t-s\|^2/(2\sigma^2)}$ utilisé dans [53]. Nous avons alors que notre noyau h respecte l'Hypothèse 1. Une hypothèse type est que le noyau rend les dipôles séparés incohérents, c'est à dire que leur produit scalaire retourne un petit résidu contrôlé par le noyau.

Hypothèse 1. *Un noyau Gaussien h suit cette hypothèse si il existe une constante μ_h telle que, pour tout dipôles π_1, π_2 $\frac{\epsilon}{3}$-séparés, on a $\langle \pi_1, \pi_2 \rangle_h \leq \mu_h \|\pi_1\|_h \|\pi_2\|_h$ (cohérence mutuelle). Sachant que l'on fixe le noyau h dans ce manuscrit, par souci de simplification des notations, on note $\mu := \mu_h$.*

Le choix de la constante $\frac{\epsilon}{3}$ pour la séparation des dipôle devient pertinent plus tard dans la preuve de notre théorème principal. Pour détailler brièvement, si nous supposons deux $\frac{\epsilon}{3}$-dipôles $\pi_1 = a_1\delta_{t_1} - b_1\delta_{s_1}, \pi_2 = a_2\delta_{t_2} - b_2\delta_{s_2}$ avec $\delta_{t_1}, \delta_{t_2}$ qui sont ϵ-séparés (c'est à dire que $\|t_1 - t_2\|_2 > \epsilon$), alors les deux dipôles π_1 et π_2 sont $\frac{\epsilon}{3}$-séparés (pour les détails calculatoires, voir la Preuve B.4).

Avec un noyau Gaussien h, nous avons les propriétés suivantes.

Lemme 3.1 (Conséquences d'avoir choisi un noyau Gaussien)**.** *Soit $h(t, s) = \mathrm{e}^{-\|t-s\|_2^2/(2\sigma^2)}$ un noyau Gaussien de variance $2\sigma^2$.*
Alors,

1. *le noyau h est symétrique par rapport à 0 et est invariant par translation ;*
2. *nous avons que $h(t, t) = 1 = \max_{t \in \mathbb{R}^d, s \in \mathbb{R}^d} |h(t, s)|$;*
3. *si h vérifie l'Hypothèse 1, alors pour toutes impulsions δ_t, δ_s $\frac{\epsilon}{3}$-séparées,*

$$\langle \delta_t, \delta_s \rangle_h = \mathrm{e}^{-\frac{\|t-s\|_2^2}{2\sigma^2}} \leq \mu \|\delta_t\|_h \|\delta_s\|_h = \mu. \tag{3.4}$$

Le lemme suivant est une approximation d'une identité de Pythagore où la norme au carré de la somme des dipôles est approximativement égale à la somme des normes au carré des dipôles.

Lemme 3.2. *(de [84]) Supposons que pour tous dipôles π_1, π_2 $\frac{\epsilon}{3}$-séparés, $\langle \pi_1, \pi_2 \rangle_h \leq \mu \|\pi_1\|_h \|\pi_2\|_h$ (cohérence mutuelle). Alors pour k dipôles $\frac{\epsilon}{3}$-séparés $\pi_1, \ldots, \pi_k$ tels que $\max_i \|\pi_i\|_h > 0$, nous avons*

$$1 - (k-1)\mu \leq \frac{\|\sum_{i=1,k} \pi_i\|_h^2}{\sum_{i=1,k} \|\pi_i\|_h^2} \leq 1 + (k-1)\mu. \tag{3.5}$$

Pour définir la propriété d'isométrie restreinte (RIP), nous avons besoin de la notion d'ensemble sécant.

Définition 3 (Ensemble sécant)**.** *L'ensemble sécant* $\mathcal{S}(\Sigma_{K,\epsilon})$ *de l'ensemble modèle* $\Sigma_{K,\epsilon}$ *est défini par*

$$\mathcal{S}(\Sigma_{K,\epsilon}) := \{x - y, x \in \Sigma_{K,\epsilon}, y \in \Sigma_{K,\epsilon}\}. \tag{3.6}$$

Notons que $\mathcal{S}(\Sigma_{K,\epsilon})$ *peut être écrit comme* $\Sigma_{K,\epsilon} - \Sigma_{K,\epsilon}$.

Nous avons maintenant les outils nécessaires pour présenter la propriété d'isométrie restreinte (voir [53]).

Définition 4 (Propriété d'isométrie restreinte (RIP))**.** *L'opérateur linéaire* A *a une* γ*-RIP sur* $\mathcal{S}(\Sigma_{K,\epsilon})$ *de constante* $0 < \gamma < 1$ *si pour tout* $x \in \mathcal{S}(\Sigma_{K,\epsilon})$ *:*

$$(1-\gamma)\|x\|_h^2 \leq \|Ax\|_2^2 \leq (1+\gamma)\|x\|_h^2. \tag{3.7}$$

Cette propriété nous permet de borner les distances entre deux éléments de $\mathcal{S}(\Sigma_{K,\epsilon})$ après l'application de l'opérateur linéaire A en fonction de leur distance avant application de A. Pour $x_1, x_2 \in \mathcal{S}(\Sigma_{K,\epsilon})$, nous avons alors que $\langle Ax_1, Ax_2\rangle \approx \langle x_1, x_2\rangle_h$. Qualitativement, la RIP implique que l'opérateur linéaire A est une transformation quasiment orthogonale.

Supposons que l'opérateur A a une γ-RIP sur $\mathcal{S}(\Sigma_{K,\epsilon})$ et considérons des dipôles π_i deux à deux ϵ-séparés. Alors nous avons

$$(1-\gamma)\left\|\sum_{i=1}^{k}\pi_i\right\|_h^2 \leq \left\|A\sum_{i=1}^{k}\pi_i\right\|_2^2 \leq (1+\gamma)\left\|\sum_{i=1}^{k}\pi_i\right\|_h^2. \tag{3.8}$$

3.2 Théorème général

Énonçons notre résultat principal.

Théorème 3.3. *Soit* $K \in \mathbb{N}$ *avec* $K > 0$*. Soit* $\epsilon > 0$*. Soit* $x_{K,\epsilon} = \sum_{i=1}^{K} a_i\delta_{t_i} \in \Sigma_{K,\epsilon}$ *où* $\|a\|_\infty = |a_1| \geq |a_2| \geq \ldots \geq |a_K|$*. Notons* $\alpha = |a_1|/|a_K|$*. Supposons que l'opérateur linéaire* A *a une* γ*-RIP sur* $\mathcal{S}(\Sigma_{K,\frac{\epsilon}{3}})$ *avec un noyau Gaussien* h *de variance* σ^2 *respectant l'Hypothèse 1. Nous définissons* $s_1^*, \ldots, s_K^*$ *comme étant les* K *premières sorties ordonnées de COMP sans descente (Algorithme 1) appliqué à* $y = Ax_{K,\epsilon}$*. Soit*

$$\xi := \sqrt{2\sigma^2 \ln\left(\frac{1}{1-(4K\alpha-1)(\mu+\gamma)}\right)}. \tag{3.9}$$

Supposons que

$$\mu + \gamma \leq \frac{1}{5K(4K\alpha-1)} \tag{3.10}$$

et

$$\epsilon^2 > 18\ln\left(\frac{10}{9}\right)\sigma^2. \tag{3.11}$$

Alors pour tout $l \in \{0, \ldots, K\}$*, il existe un* $i \in \{0, \ldots, K\}$*, tel que* $\|s_l^* - t_i\|_2 \leq \xi$.

Ce résultat indique que si le signal inconnu suit une hypothèse de séparation et l'opérateur A a une γ-RIP sur de tels signaux séparés par rapport à une norme à noyau suivant l'hypothèse de cohérence mutuelle, alors nous pouvons limiter l'erreur d'estimation des positions après COMP sans descente. L'hypothèse principale de ce théorème est la condition que $\mu + \gamma = \mathcal{O}\left(\frac{1}{K^2}\right)$ qui semble restrictive ; elle dépend aussi du rapport des amplitudes α. Nous exprimons dans la suite un second théorème où les amplitudes sont connues et fixées à 1, ce qui donne des conditions plus favorables. Nous attribuons ces conditions restrictives au fait que nous n'avons pas d'orthogonalité exacte entre les impulsions dans ce cadre continu par rapport au cas discret de dimension finie ([49, Théorème 6.24]).

L'hypothèse $\epsilon^2 > 18 \ln\left(\frac{10}{9}\right) \sigma^2$ est garantie si

$$\epsilon > 1.37\sigma. \tag{3.12}$$

Cette inégalité signifie que la séparation minimale nécessaire pour une reconstruction complète des positions des impulsions est plus grande que l'écart-type du noyau h. Ceci est en accord avec l'idée que si deux impulsions sont trop proches l'une de l'autre par rapport à la largeur du noyau, leurs observations ne sont pas suffisamment séparées pour être récupérées indépendamment. Notons que nous n'avons pas cherché à obtenir les meilleures constantes afin de garder la preuve aussi simple que possible.

L'erreur de localisation des impulsions est bornée par ξ qui converge vers 0 lorsque $\mu+\gamma$ est suffisamment petit. Ainsi, pour des opérateurs A très bien conditionnés par rapport au modèle Σ, l'algorithme COMP sans descente en K étapes a de bonnes garanties de reconstruction. Notons que pour qu'une descente finale sur les paramètres converge, la borne ξ doit être plus petite que la taille du bassin d'attraction de g.

En rappelant la définition d'un bassin d'attraction Λ de g ([84, Définition 1.3]), les résultats sur les tailles des bassins d'attraction de [84, Corollaire 3.1], [87, Corollaire 2.1] et [85, Théorème 3.1] nous permettent alors de déduire qu'une borne inférieure pour la taille de Λ est contrôlée par la RIP sur A, la cohérence mutuelle μ_h du noyau h.

En conséquence, en utilisant ces résultats, nous venons de prouver qu'il existe un opérateur linéaire (avec éventuellement un très grand nombre de mesures) pour lequel COMP avec une descente de gradient finale converge vers x_0. Par exemple, avec un opérateur linéaire composé d'un grand nombre de mesures de Fourier aléatoires suivant la RIP, ce résultat est vrai (voir [85] pour une étude précise des bassins d'attraction de (2.1)).

La preuve de ce théorème est une récurrence sur les étapes de COMP où la reconstruction approximative à une étape garantit la reconstruction approximative aux étapes suivantes. La principale difficulté est de contrôler tous les résidus générés par le fait que nous n'avons qu'une orthogonalité approximative entre les atomes (par le biais de la norme à noyau).

Bien que nous effectuons notre analyse de COMP sans descente dans le cas gaussien uniquement pour simplifier les preuves, il est possible d'étendre notre résultat à des noyaux plus généraux (tels que les noyaux exponentiels, voir par exemple [44]) puisque l'hypothèse principale est la cohérence mutuelle du noyau et sa monotonicité.

3.3 Théorème pour signaux avec des amplitudes fixes

Le Théorème précédent 3.3 est un cas général qui s'applique aux signaux d'amplitude variable. Pour les signaux parcimonieux de la forme $x_0 = \sum_{i=1}^{K} \delta_{t_i}$, nous donnons un théorème avec des conditions plus faibles.

Théorème 3.4. *Soit $K \in \mathbb{N}$ avec $K > 0$. Soit $\epsilon > 0$. Soit $x_{K,\epsilon} = \sum_{i=1}^{K} \delta_{t_i} \in \Sigma_{K,\epsilon}$. Supposons que l'opérateur linéaire A a une γ-RIP sur $\mathcal{S}(\Sigma_{K,\frac{\epsilon}{3}})$ avec un noyau Gaussien h de variance σ^2. Nous définissons $s_1^*, \ldots, s_K^*$ comme étant les K premières sorties ordonnées de COMP sans descente (Algorithme 1) appliqué à $y = Ax_{K,\epsilon}$. Soit*

$$\xi := \sqrt{2\sigma^2 \ln\left(\frac{1}{1-(4K-1)(\mu+\gamma)}\right)}. \tag{3.13}$$

Supposons que

$$\mu + \gamma \leq \frac{1}{10\,(4K-1)} \tag{3.14}$$

et

$$\epsilon^2 > 18 \ln\left(\frac{10}{9}\right)\sigma^2. \tag{3.15}$$

Alors pour tout $l \in \{0, \ldots, K\}$, il existe un $i \in \{0, \ldots, K\}$, tel que $\|s_l^* - t_i\|_2 \leq \xi$.

3.4 Schéma de preuve

Sans entrer dans les détails de la preuve, faisons une ébauche de cette dernière pour en expliquer son déroulement. La preuve complète pour les théorèmes avec et sans amplitudes se trouve dans l'Annexe B. Les deux Théorèmes 3.3 et 3.4 sont similaires. Il s'agit d'une preuve par récurrence. Nous voulons montrer qu'à chaque étape de COMP sans descente, la différence entre le signal à reconstruire et son estimation partielle peut être décomposés en deux signaux dont on peut contrôler les normes avec des bornes bien choisies. On note $\mathcal{X}_{K,l}^{\epsilon,\xi}$ l'ensemble auquel appartient la différence $X_{K,l} = x_0 - x_{\text{esti},l}$ avec x_0 le signal à reconstruire et $x_{\text{esti},l}$ l'estimation partielle construite à partir des l premières étapes de COMP sans descente. Nous définissons alors $\mathcal{X}_{K,l}^{\epsilon,\xi}$ par

$$\mathcal{X}_{K,l}^{\epsilon,\xi} = \Sigma_{K-l,\epsilon} + Z_{l,\xi,\epsilon/3} \tag{3.16}$$

avec $\Sigma_{K-l,\epsilon}$, l'ensemble des impulsions qu'il reste à estimer et $Z_{l,\xi,\epsilon/3}$ l'ensemble de la différence entre les impulsions déjà estimées de x_0 et $x_{\text{esti,l}}$. Par récurrence, en arrivant à montrer que si pour un l, en supposant que $X_{K,l} \in \mathcal{X}_{K,l}^{\epsilon,\xi}$, alors $X_{K,l+1} \in \mathcal{X}_{K,l+1}^{\epsilon,\xi}$, nous pouvons montrer que la $l+1$ ème impulsion estimée avec COMP sans descente estime dans un rayon ξ une impulsion pas encore estimée. Par récurrence, en le montrant pour tout $l < K$, nous avons alors que pour toute impulsion de x_0, il existe une impulsion de $x_{\text{esti,K}}$ telle que leur distance est inférieure à ξ.

La partie la plus technique de cette récurrence vient du fait que l'on doit assurer que l'impulsion estimée δ_{s^*} par COMP sans descente est forcément initialisée près d'une impulsion du signal à reconstruire x_0. Nous savons que l'impulsion estimée δ_{s^*} par COMP sans descente obéit au fait que $s^* \in \arg\max_{\bar{s}\in\mathbb{R}^d} |\langle A\delta_{\bar{s}}, AX_{K,l}\rangle|$. Nous créons alors trois espaces E_1, E_2 et E_3, deux à deux distincts et dont l'union est l'espace tout entier avec

- $E_1 = \{s \in \mathbb{R}^d : \exists i_0 \in \{1, \ldots, k\} \,/\, \|s - t_{i_0}\|_2 < \xi\}$;
- $E_2 = \{s \in \mathbb{R}^d : \exists i_0 \in \{1, \ldots, k\} \,/\, \xi \leq \|s - t_{i_0}\|_2 < \epsilon/3\}$;
- $E_3 = \{s \in \mathbb{R}^d : \exists i_0 \in \{1, \ldots, k\} \,/\, \|s - t_{i_0}\|_2 \geq \epsilon/3\}$.

En prenant $s_1 \in E_1$ et $s_2 \in E_2$, nous déduisons qu'il existe au moins un élément $s_1 \in E_1$ tel que pour tout élément $s_2 \in E_2$, nous avons $|\langle A\delta_{s_1}, AX_{K,l}\rangle| > |\langle A\delta_{s_2}, AX_{K,l}\rangle|$. En

prouvant également cette inégalité avec tous les éléments de l'ensemble E_3 à la place de E_2, nous obtenons alors que $\arg\max_{\bar{s}\in\mathbb{R}^d} |\langle A\delta_{\bar{s}}, AX_{K,l}\rangle|$ est forcément compris dans E_1, ce qui conclut notre récurrence. Pour plus de détails techniques ainsi que les calculs et résultats intermédiaires, nous renvoyons le lecteur à l'Annexe B.

3.5 Relation entre l'hypothèse RIP et le nombre de mesures

La différence principale entre les deux théorèmes est la borne supérieure moins restrictive sur $\mu + \gamma$ dans le Théorème 3.4 (avec amplitudes fixées). Au lieu d'avoir une borne supérieure de l'ordre de $\frac{1}{K^2}$, elle est simplement de l'ordre de $\frac{1}{K}$ sans dépendre du rapport des amplitudes α. Cela signifie que le noyau gaussien peut être plus grand car il implique une condition moins stricte sur l'observation pour avoir une récupération complète du vrai signal.

Dans le contexte des mesures gaussiennes aléatoires, nous pouvons garantir que la condition sur γ est presque surement vérifiée si

$$m \geq C\gamma^{-2}\left[K^2 d \cdot \left(1 + \log Kd + \log\frac{R}{\epsilon} - \log\zeta\right) - K\log\zeta\right] \cdot \log K\mathrm{e} \cdot \min(\log \mathrm{e}K, d) \tag{3.17}$$

avec C, ζ des constantes, R une borne sur la norme des impulsions et ϵ la distance minimale de séparation entre ces dernières. En simplifiant (3.17), cela veut dire que le nombre de mesures m est de l'ordre de $\gamma^{-2}K^2 d$ multiplié par certains termes logarithmiques, grâce à [54, Théorème 3.1]. Dans notre cas, cela donne un nombre de mesures m nécessaire pour garantir une limite supérieure de γ d'ordre $K^6 d$. Pour les signaux d'amplitude fixe, m doit être d'ordre $K^4 d$. Ces deux limites supérieures sont obtenues en négligeant les termes logarithmiques

Contrairement au cas des mesures gaussiennes aléatoires pour un échantillonnage de Fourier, il existe peu de résultat sur le lien entre la RIP et les opérateurs linéaires utilisés en microscopie. Sachant que certains se basent sur un noyau gaussien, il serait alors possible d'en déduire les conditions suffisantes pour avoir une RIP.

Pour le cas de la reconstruction d'un signal parcimonieux discret, le succès d'OMP est garanti (en négligeant les termes logarithmiques) pour $m = \mathcal{O}(K)$ dans le cas où γ est borné par une constante [25]. Une question ouverte est donc de déterminer si nos conditions sur γ peuvent être assouplies pour obtenir de meilleurs résultats de reconstruction en ce qui concerne le nombre d'impulsions K.

3.6 Résumé du chapitre et conclusions

Dans ce chapitre, nous avons introduit plusieurs notions importantes dans l'étude des problèmes inverses et des signaux parcimonieux. Tout d'abord, notre Hypothèse 1 de cohérence mutuelle nous permet de contrôler le produit scalaire de deux mesures de Dirac suffisamment séparées. D'un autre côté, nous avons la propriété d'isométrie restreinte (RIP, Définition 4) qui contrôle la norme de l'application par un opérateur linéaire d'un élément x de $\Sigma_{K,\epsilon}$ par sa norme. Ces deux hypothèses (sur le noyau h et l'opérateur linéaire A) sont fondamentales dans notre théorème principal sur le contrôle de l'initialisation des K premières itérations de OP-COMP.

Nos résultats énoncent alors un lien fort entre la distance d'initialisation des impulsions, la cohérence mutuelle du noyau μ_h et la γ-RIP sur A. Si notre fonction à reconstruire x_0 a des amplitudes fixes, l'hypothèse sur $\mu + \gamma$ devient moins forte d'un facteur K. Le lien alors établi entre le nombre de mesures m et la γ-RIP de A s'en voit alors diminué d'un facteur K^2.

En conclusion, nos deux résultats (Théorèmes 3.3 et 3.4) imposent un contrôle sur la position des impulsions initialisées par COMP sans descente. Bien que les bornes dans nos hypothèses dépendent du noyau gaussien, seules la cohérence mutuelle et la RIP restent nécessaires. Nos résultats sont alors généralisables à d'autres noyaux et opérateurs linéaires associés tant qu'ils respectent ces hypothèses.

Bien que dans le cas idéal de nos théorèmes, nous observons que seules K impulsions initialisées sont suffisantes, en pratique, nous sommes rarement dans ces conditions d'où l'ajout de l'étape de sur-paramétrisation. Une question ouverte reste alors de savoir si il est possible de généraliser nos résultats pour un nombre d'impulsions initialisées quelconque. Ce nouveau résultat serait possible en affaiblissant nos hypothèses de cohérence mutuelle et de RIP.

Chapitre 4

Analyse expérimentale des bassins d'attraction de la fonctionnelle non-convexe

Dans ce chapitre, nous étudions de manière quantitative les effets de la sur-paramétrisation sur la trajectoire de notre estimation dans notre problème à résoudre (2.1). En effet, la sur-paramétrisation d'une estimation peut entrainer des problèmes de convergence en terme de temps de calcul. Sachant que notre fonction $g(\theta) := \|A\phi(\theta) - y\|_2^2$ avec $\theta \in \Theta_{K,\epsilon}$ possède $K \times (d+1)$ paramètres en entrée, il est alors difficile d'en estimer correctement la géométrie globale et encore plus de la représenter graphiquement. De plus, cette fonction n'est pas convexe, trouver le bassin d'attraction du minimum global est alors nécessaire pour minimiser g avec des méthodes de descente de gradient pour fonctions convexes.

Dans un premier lieu, nous décrivons le problème mal-posé de la sur-paramétrisation et les méthodes pour contourner ou supprimer ce problème. Nous effectuons une étude expérimentale sur la sur-paramétrisation en comparant la vitesse de convergence entre plusieurs signaux dont certains sur-paramétrés. Nous évoquerons également les différentes méthodes de projection pour réduire le nombre total d'impulsions.

Deuxièmement, nous étudions l'effet du nombre de mesures sur la taille des bassins d'attraction du minimum global. Nous fournissons, de plus, une étude quantitative qui illustre la taille du bassin d'attraction du problème (2.5).

4.1 Le problème mal posé de la sur-paramétrisation

Dans cette section, nous montrons par des expériences simples que, même si la sur-paramétrisation permet de minimiser la quantité (2.5), elle conduit à un problème mal conditionné qui entraîne une convergence lente de la descente de gradient. Ceci illustre le fait qu'effectuer une projection, même heuristique dans la descente est une partie clé de notre algorithme OP-COMP avec PGD pour une convergence rapide vers la solution.

4.1.1 Présentation des expériences

Nous nous plaçons dans le cas de la reconstruction d'une impulsion (notons que ce cas est souvent très instructif pour l'étude des limites des algorithmes de super-résolution [36, 85]). Des résultats sur les bassins d'attractions montrent que si la descente est initialisée suffisamment près d'une impulsion observée, alors sous la condition d'une propriété

d'isométrie restreinte (RIP, voir Définition 4) sur l'opérateur linéaire A, il y aura convergence vers la position désirée (à un taux linéaire).

Nous nous plaçons en 2D, c'est-à-dire que $x_0 = \delta_t$ avec $t = (0, 0)$. Nous observons ce signal par des mesures de Fourier aléatoires (où la RIP avec un noyau gaussien est garanti théoriquement pour un suffisamment grand nombre de mesures). Pour cet exemple, nous prenons un nombre de mesures $m = 40$. Nous initialisons ensuite notre signal x_{init} comme une somme d'impulsions positionnées près de notre impulsion à reconstruire δ_t. Ceci a pour but de recréer la sortie de OP-COMP avec x_{init} suffisamment proche de la solution pour être dans le bassin d'attraction du minimum global.

Pour nos expériences, nous comparons quatre cas.

- Cas 1 : nous posons $x_{\text{init}} = \delta_s$ où s est proche de t, $\|t - s\|_2 = 0.2$. Dans ce cas, le nombre de vraies impulsions est égal au nombre d'impulsions initialisées et il n'y a pas besoin de projection.
- Cas 2 : nous posons $x_{\text{init}} = \frac{1}{2}\sum_{i=1}^{2} \delta_{s_i}$ où $s_1 = (0.2, 0)$ et $s_2 = (-0.2, 0)$.
- Cas 3 : nous posons $x_{\text{init}} = \frac{1}{5}\sum_{i=1}^{5} \delta_{s_i}$ où pour tout $i = 1, \dots, 5, \|t - s_i\|_2 = 0.2$ et chaque s_i est à égale distance des autre $s_j, i \neq j$.
- Cas 4 : nous posons $x_{\text{init}} = \frac{1}{5}\sum_{i=1}^{5} \delta_{s_i}$ où pour tout $i = 1, \dots, 5, \|t - s_i\|_2 = 10^{-4}$ et chaque s_i est à égale distance des autres $s_j, i \neq j$. Les impulsions initialisées sont très proches de la vraie impulsion δ_t (comme dans le cas 3 mais initialisées très près de la solution).

Pour chaque cas, nous effectuons une simple descente de gradient sans projection pour un grand nombre d'itérations et avec un pas τ sélectionné par une recherche linéaire.

Une manière classique d'étudier la bonne formulation des algorithmes du premier ordre est de calculer la Hessienne H de la fonction g dans (2.13) (calculée dans [84, Proposition 2.2]). Pour rappel, nous avons que $H \in \mathcal{M}_{k(d+1),k(d+1)}(\mathbb{R})$ vaut pour tout $\theta \in \mathbb{R}^{k\times(d+1)}$

$$H = \begin{bmatrix} H_1 & H_{12} \\ H_{21} & H_2 \end{bmatrix} \tag{4.1}$$

avec

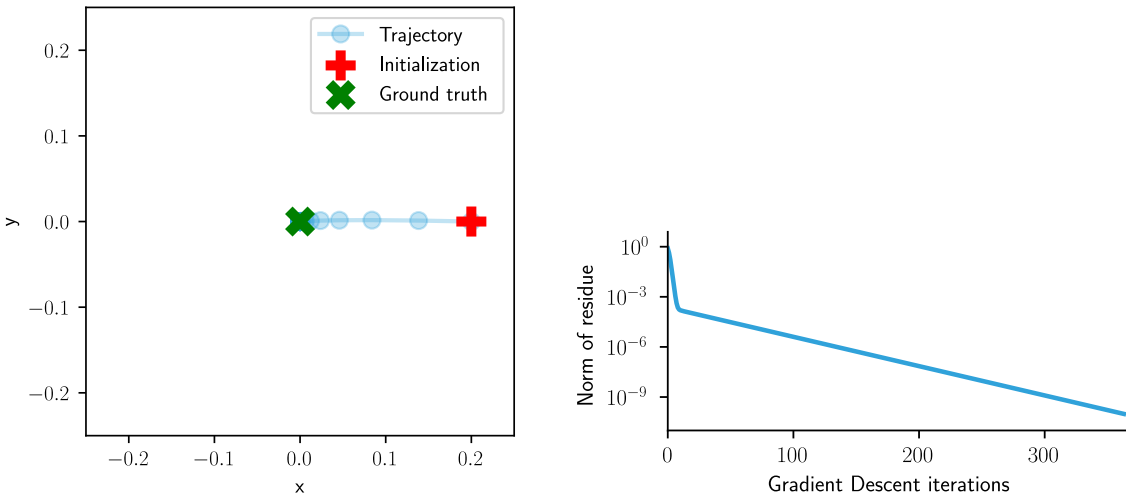
$$\begin{aligned}
H_1 \in \mathcal{M}_{k,k}(\mathbb{R}) \qquad & H_{1,r,s} = 2\Re\langle A\delta_{t_r}, A\delta_{t_s}\rangle \\
H_2 \in \mathcal{M}_{kd,kd}(\mathbb{R}) \qquad & H_{1,r,j_1,s,j_2} = 2a_r a_s \Re\langle A\delta'_{t_r,j_1}, A\delta'_{t_s,j_2}\rangle + \mathbf{1}_{r=s} 2a_r \Re\langle A\delta''_{t_r,j_1,j_2}, A\phi(\theta) - y\rangle \\
H_{12} \in \mathcal{M}_{k,kd}(\mathbb{R}) \qquad & H_{1,r,s,j} = -2a_s \Re\langle A\delta_{t_r}, A\delta'_{t_s,j}\rangle - \mathbf{1}_{r=s} 2\Re\langle A\delta'_{t_s,j}, A\phi(\theta) - y\rangle \\
H_{21} \in \mathcal{M}_{kd,k}(\mathbb{R}) \qquad & H_{21} = H_{12}^T
\end{aligned}$$

avec $r, s \in \{1, \dots, k\}$ les indices des impulsions, $j, j_1, j_2 \in \{1, \dots, d\}$ les indices des dimensions et $\delta'_{t,j}, \delta''_{t,j_1,j_2}$, respectivement les dérivées directionnelles du premier et deuxième ordre de l'impulsion δ_t (plus de détails et d'informations dans [84]).

Le conditionnement et la positivité de H permettent de montrer la convexité locale et donnent des informations sur la vitesse de convergence de la descente de gradient. Nous rappelons l'expression du conditionnement de H, noté $\kappa(H)$, avec H une matrice carrée complexe,

$$\kappa(H) = \frac{|\lambda_{max}(H)|}{|\lambda_{min}(H)|} \tag{4.2}$$

où $|\lambda_{min}(H)|$ et $|\lambda_{max}(H)|$ sont respectivement les modules des valeurs propres minimales et maximales de H.



(a) (b)

FIGURE 4.1 – (gauche) Les itérations des impulsions intermédiaires x_{temp} et la trajectoire de $x_{\text{init}} = \delta_s$ vers x_0 en utilisant une descente de gradient. (droite) Norme du résidu $\|y - Ax_{\text{temp}}\|_2$ à chaque itération de la descente de gradient.

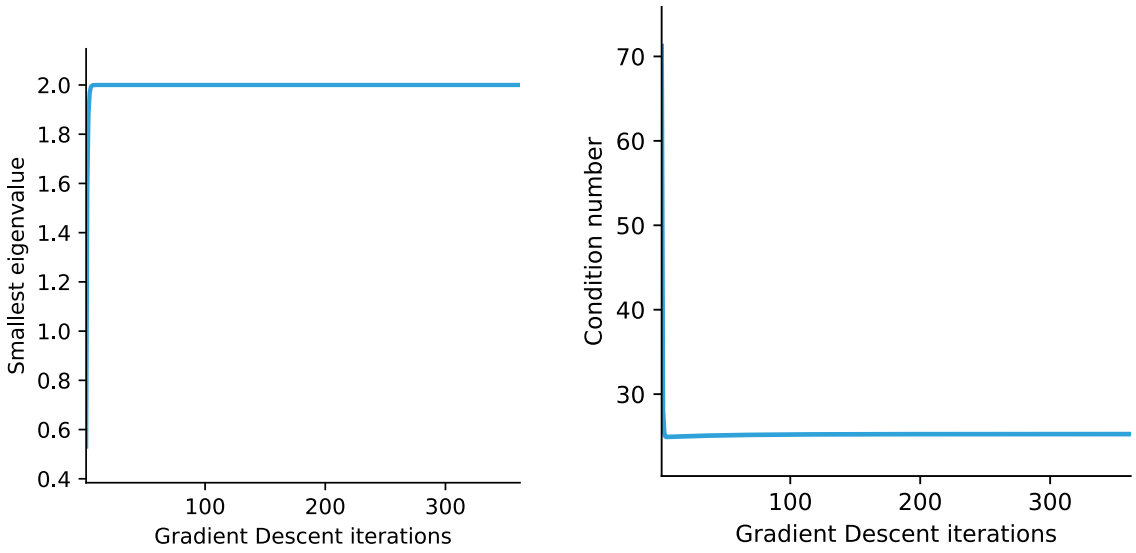
4.1.2 Résultats et remarques

Dans le Cas 1, l'impulsion estimée converge vers l'impulsion réelle, voir Figure 4.1. La norme du résidu diminue jusqu'à 10^{-14}, cette valeur en partie due au nombre limité de mesures utilisées pour l'échantillonnage de Fourier. Pour la plus petite valeur propre de H, elle n'est négative que pour la première itération (nous avons réglé manuellement l'initialisation). Mais après sa première itération de descente de gradient, elle redevient positive et se stabilise à $\lambda_{min} \approx 2$, voir Figure 4.2. Pour le conditionnement de H, il se stabilise également à $\kappa(H) \approx 25$. Cette valeur est assez faible par rapport à nos tests suivants, le système est numériquement stable. La positivité de λ_{min} indique que H est définie positive pour notre signal estimé, c'est-à-dire qu'il se trouve dans un bassin d'attraction.

Dans le Cas 2, les trajectoires des impulsions pour le signal sur-paramétré semblent converger vers l'impulsion de signal à reconstruire, voir Figure 4.3. Cependant, lorsque nous observons la norme du résidu du système, elle diminue très lentement et n'a pas le même taux de convergence linéaire que dans le cas 1. L'erreur d'estimation des positions est d'environ 10^{-3} après 10 000 itérations de la descente de gradient. En examinant la plus petite valeur propre et le conditionnement du système, nous observons qu'au moins une valeur propre de la hessienne est négative, voir Figure 4.4. Cela indique que la fonctionnelle sur-paramétrée estimée n'est pas convexe autour du minimiseur global. Lorsque nous examinons le conditionnement de la hessienne, bien qu'il soit suffisamment raisonnablement loin l'optimum global, il continue d'augmenter au cours des itérations et de la convergence. Cela implique que le système est numériquement instable à mesure qu'il se rapproche du minimiseur global.

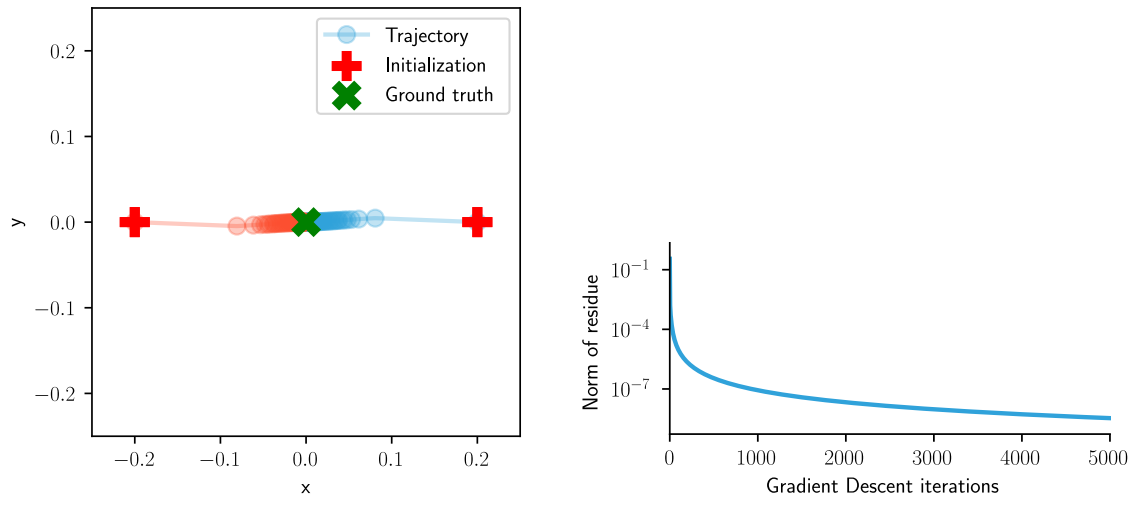
Dans le Cas 3, nous observons les mêmes phénomènes que dans le Cas 2. Les cinq impulsions du signal sur-paramétré convergent vers la solution. Cependant, cette convergence est très lente, voir Figure 4.5. Pour la plus petite valeur propre de la hessienne, elle est encore négative mais se rapproche de 0 au fil des itérations de descente de gradient. Quant au conditionnement de H, il augmente et conduit à une instabilité de calcul (voir Figure 4.6). L'erreur d'estimation des positions est d'environ 10^{-3} après 10 000 itérations.

L'idée pour le Cas 4 est de voir ce qui se passe lorsque nous initialisons les impulsions



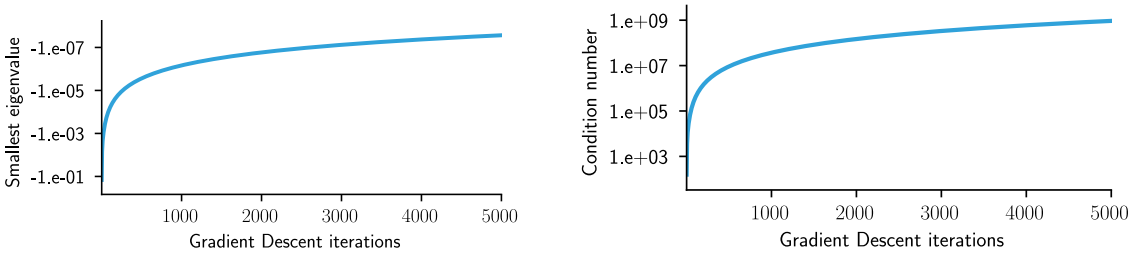
(a) (b)

FIGURE 4.2 – (gauche) Plus petite valeur propre et (droite) conditionnement κ de H à chaque itération de la descente de gradient pour un signal composé d'une impulsion.



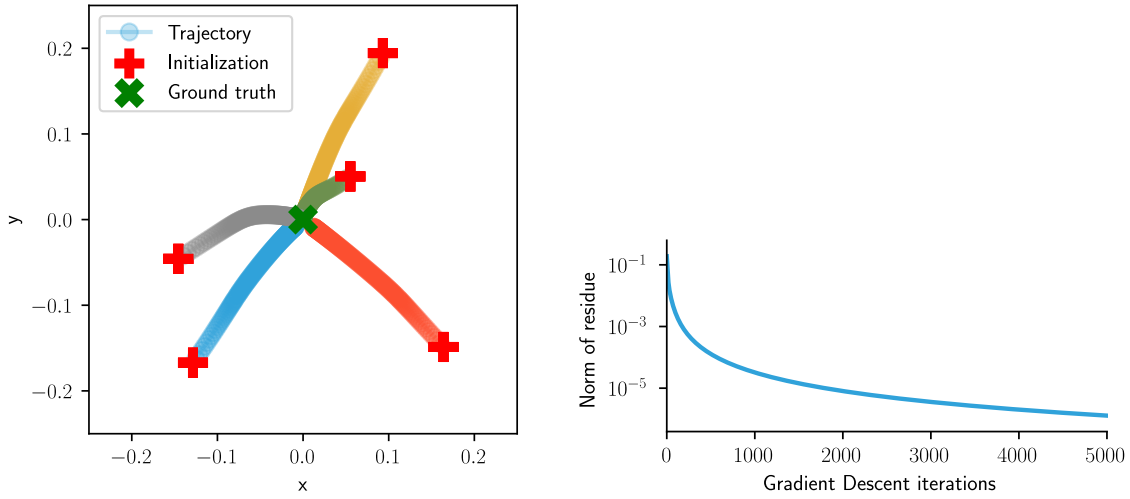
(a) (b)

FIGURE 4.3 – (gauche) Les itérations des impulsions intermédiaires x_{temp} et la trajectoire de $x_{\text{init}} = \frac{1}{2}\delta_{s_1} + \frac{1}{2}\delta_{s_2}$ vers x_0 en utilisant une descente de gradient. (droite) Norme du résidu $\|y - Ax_{\text{temp}}\|_2$ à chaque itération de la descente de gradient.



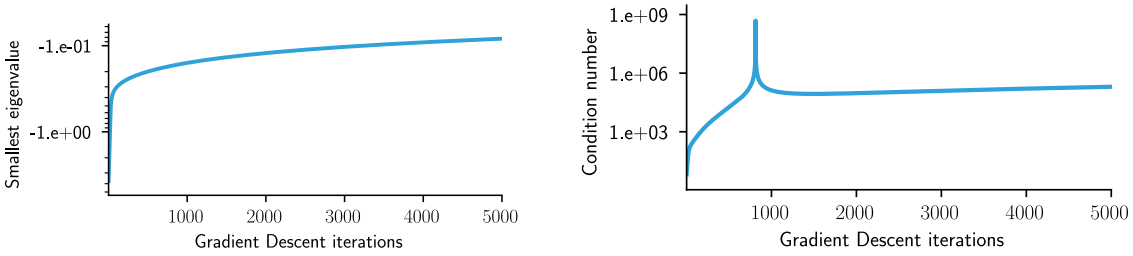
(a) (b)

FIGURE 4.4 – (gauche) Plus petite valeur propre et (droite) conditionnement κ de H à chaque itération de la descente de gradient pour un signal sur-paramétré composé de deux impulsions.



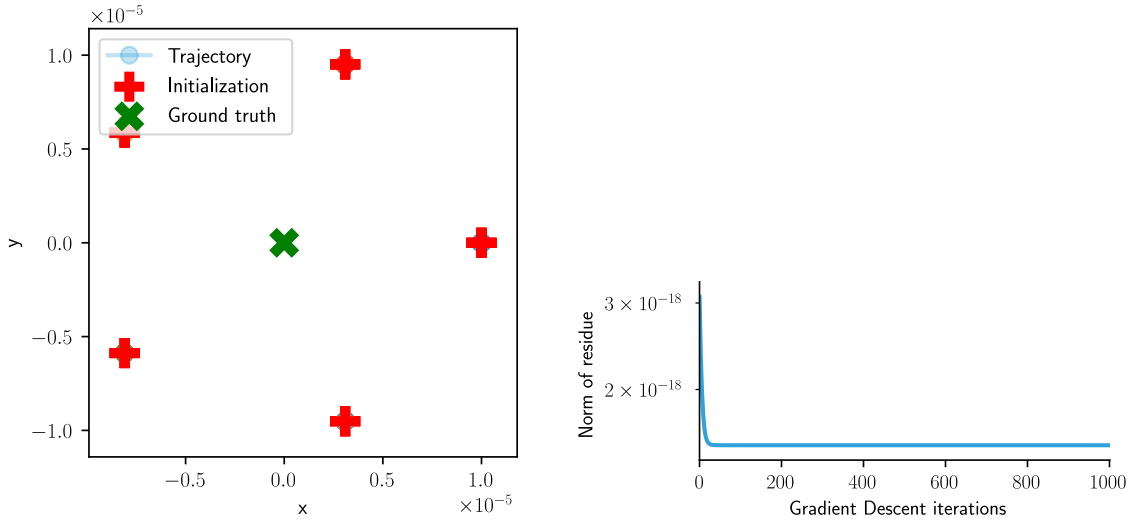
(a) (b)

FIGURE 4.5 – (gauche) Les itérations des impulsions intermédiaires x_{temp} et la trajectoire de $x_{\text{init}} = \frac{1}{5}\sum_{i=1}^{5}\delta_{s_i}$ vers x_0 en utilisant une descente de gradient. (droite) Norme du résidu $\|y - Ax_{\text{temp}}\|_2$ à chaque itération de la descente de gradient.



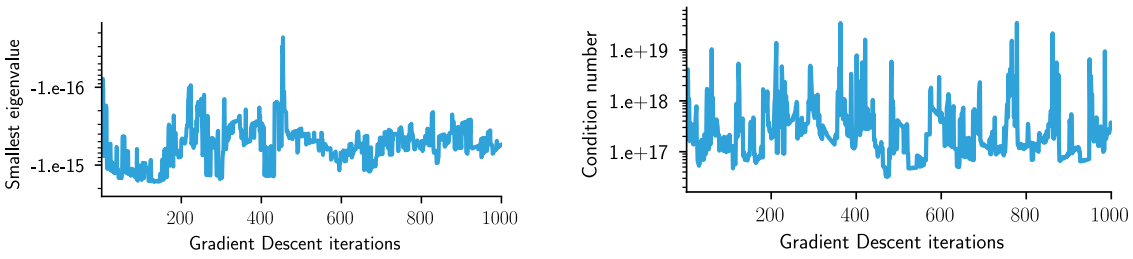
(a) (b)

FIGURE 4.6 – (gauche) Plus petite valeur propre et (droite) conditionnement κ de H à chaque itération de la descente de gradient pour un signal sur-paramétré composé de cinq impulsions.



(a) (b)

FIGURE 4.7 – (gauche) Les itérations des impulsions intermédiaires x_{temp} et la trajectoire de $x_{\text{init}} = \frac{1}{5}\sum_{i=1}^{5}\delta_{s_i}$ vers x_0 en utilisant une descente de gradient. (droite) Norme du résidu $\|y - Ax_{\text{temp}}\|_2$ à chaque itération de la descente de gradient.



(a) (b)

FIGURE 4.8 – (gauche) Plus petite valeur propre et (droite) conditionnement κ de H à chaque itération de la descente de gradient pour un signal sur-paramétré composé de cinq impulsions.

plus près de la solution (distance inférieure à 10^{-5}) que l'estimation obtenue après toutes les itérations d'une descente de gradient du Cas 3. Dans le cas 4, la descente ne semble pas rapprocher nos impulsions initialisées de x_0, voir Figure 4.7. Entre la première et la dernière itération de la descente de gradient, la différence de résidu est de l'ordre de 10^{-15} et le résidu est de l'ordre de 10^{-10}. Il y a une convergence des itérés vers x_0 mais elle est très lente. A cette échelle, nous attribuons ce comportement au mauvais conditionnement et à la non-convexité de la fonctionnelle : la plus petite valeur propre est très petite ($\lambda_{min} \approx -10^{-15}$) et ne devient positive pendant aucune étape de la descente, voir Figure 4.8.

Nous constatons que lorsque le nombre d'impulsions initialisées est supérieur au nombre d'impulsions à reconstruire, le signal estimé ne converge pas vers la vérité terrain avec une bonne précision et en un temps fini. Les itérations de gradient n'améliorent pas la solution en raison du caractère mal posé de cette situation. Une solution à ce problème consiste à réduire le nombre d'impulsions pendant la descente. En pratique, la solution que nous proposons consiste à fusionner les impulsions proches les unes des autres (voir

Algorithme 2) afin d'éviter de traiter des problèmes mal conditionnés et de projeter les itérés dans un bassin d'attraction convexe du minimum global. Nous devons remarquer que le choix du critère de séparation pour fusionner les impulsions est un compromis entre la précision de l'algorithme (fusionner les impulsions proches les unes des autres) et la vitesse de calcul (projeter aussi vite que possible pendant la descente).

4.1.3 Les méthodes de projection

Malgré les travaux précédemment cités, lorsque nous nous retrouvons avec un signal initialisé x_{init} qui est sur-paramétré en sortie de OP-COMP, nous souhaitons tout de même réduire le nombre total d'impulsion pour pourvoir associer exactement une impulsion du signal estimée x_{esti} avec une impulsion du signal à reconstruire x_0. De surcroît, en pratique l'accélération liée au conditionnement du gradient n'apporte pas de résultats satisfaisants en terme de temps de calculs. L'utilisation d'une méthode de projection pour réduire le nombre d'impulsion semble pertinent de ce point de vue là.

Comme précédemment citées dans le Chapitre 2, il existe plusieurs types de projection. Il y a la projection $\mathcal{P}_{\text{theo}}$ dite "théorique" qui projette orthogonalement un élément $\tilde{x}$ dans $\Sigma_{K,\epsilon}$ de telle manière à ce que $\mathcal{P}_{\text{theo}} = \arg\min_{x\in\Sigma_\epsilon} \|x - \tilde{x}\|_h^2$. Cependant, résoudre ce problème n'est pas trivial. En effet, il est non-convexe et peut être approximé par $\arg\min_{x\in\Sigma_{K,\epsilon}} \|Ax - A\tilde{x}\|_2^2$ en utilisant la propriété d'isométrie restreinte(RIP). Or c'est un problème comme cette approximation que nous cherchons à résoudre dans un premier temps. Il n'est donc pas envisageable, avec nos outils actuels, de trouver la formulation explicite de cette projection orthogonale.

D'un autre côté, la projection "heuristique" (Algorithme 2) introduite dans [86] à l'avantage d'être simple à implémenter. De plus, cette projection envoie bien $\tilde{x}$ dans Σ_ϵ tout en réduisant le nombre total d'impulsions car elle fusionne les impulsions qui ont une distance entre elles inférieure à ϵ. Cette propriété importante de réduction du nombre d'impulsions se combine bien au fait que nous avons en sortie de OP-COMP un signal sur-paramétré.

4.2 Estimation de la taille du bassin d'attraction

Nous cherchons à estimer la taille du bassin d'attraction de g dans (2.5). Pour cela, nous partons des travaux théoriques de [85] et plus précisément du Théorème 4.1. Ce théorème nous annonce qu'avec l'opérateur de mesures de Fourier gaussiennes aléatoires, alors nous avons que la taille du bassin d'attraction h est borné par

$$h \leq \sigma\sqrt{1 - \mathcal{O}\left(\frac{1}{\sqrt{m}}\right)} \tag{4.3}$$

avec σ l'écart-type associé au tirage aléatoire des mesures de Fourier gaussiennes et par ϵ.

Dans cette section, en utilisant des mesures de Fourier aléatoires comme opérateur linéaire, nous fournissons des expériences sur la taille des bassins d'attraction en fonction des différents paramètres du problème : le nombre de mesures, la dimension d du support, le choix des fréquences (et du noyau induit) et la séparation minimale entre les impulsions. Ces expériences illustrent les résultats théoriques de [85].

4.2.1 Méthode et paramètres des expériences

Avant de procéder à la description de nos expériences, nous rappelons quelques définitions de [84] adaptés à notre problème de minimisation de g (2.5),

Définition 5 (Descente de gradient)**.** *Considérons un point initial $\theta_0 \in \Theta$, un pas de descente $\tau > 0$ et g (2.5) une fonction lisse et différentiable. Nous considérons alors la séquence des θ_n tels que*

$$\theta_{n+1} = \theta_n - \tau \nabla g(\theta_n) \tag{4.4}$$

Définition 6 (Bassin d'attraction)**.** *Considérons l'ensemble $\Lambda \subset \Theta$ un bassin d'attraction de g (2.5) si il existe un $\theta^* \in \Lambda$, tel que si $\theta_0 \in \Lambda$, alors la séquence des θ_n obtenue par une descente de gradient converge vers θ^*.*

Ce bassin d'attraction est l'ensemble des signaux qui convergent avec une descente de gradient vers le minimum de g.

Étant donné un signal de base $x_0 = \phi(\theta^*)$ à reconstruire et un ensemble de paramètres, nous répétons l'expérience suivante 10 000 fois : nous effectuons une descente de gradient avec recherche linéaire (pour une convergence pratique plus rapide) et avec une initialisation aléatoire θ_{init} et nous notons si avec la descente, notre ensemble de paramètres a convergé vers θ^* ou non avec un seuil de 0.01 sur l'erreur de localisation $\|\theta_{\text{nit}_{\max}} - \theta^*\|_2$ après un nombre fixe d'itérations égale à 500. Nous calculons la taille du bassin d'attraction comme étant la plus petite distance β telle que 99.8% des expériences qui n'ont pas convergé vers θ^* ont été initialisées en dehors de du bassin d'attraction Λ_β, c'est-à-dire que l'initialisation dans Λ_β produit une convergence avec une forte probabilité.

Les fréquences des mesures gaussiennes suivent une distribution normale $\mathcal{N}(0, \frac{1}{\lambda^2} I)$. Pour illustrer à quel point les mesures se rapprochent du noyau gaussien moyen lorsque leur nombre augmente, nous montrons la rétro-projection des mesures (voir Équation 2.6) d'une impulsion sur une grille dans la Figure 4.9.

4.2.2 Expériences pour l'estimation de la taille du bassin d'attraction

L'effet de la dimension

Pour étudier l'effet de la dimension d, nous utilisons le signal $x_0 = \delta_0$ et prenons $d \in \{1, 2, 3\}$. Nous initialisations de manière aléatoire θ_{init} avec

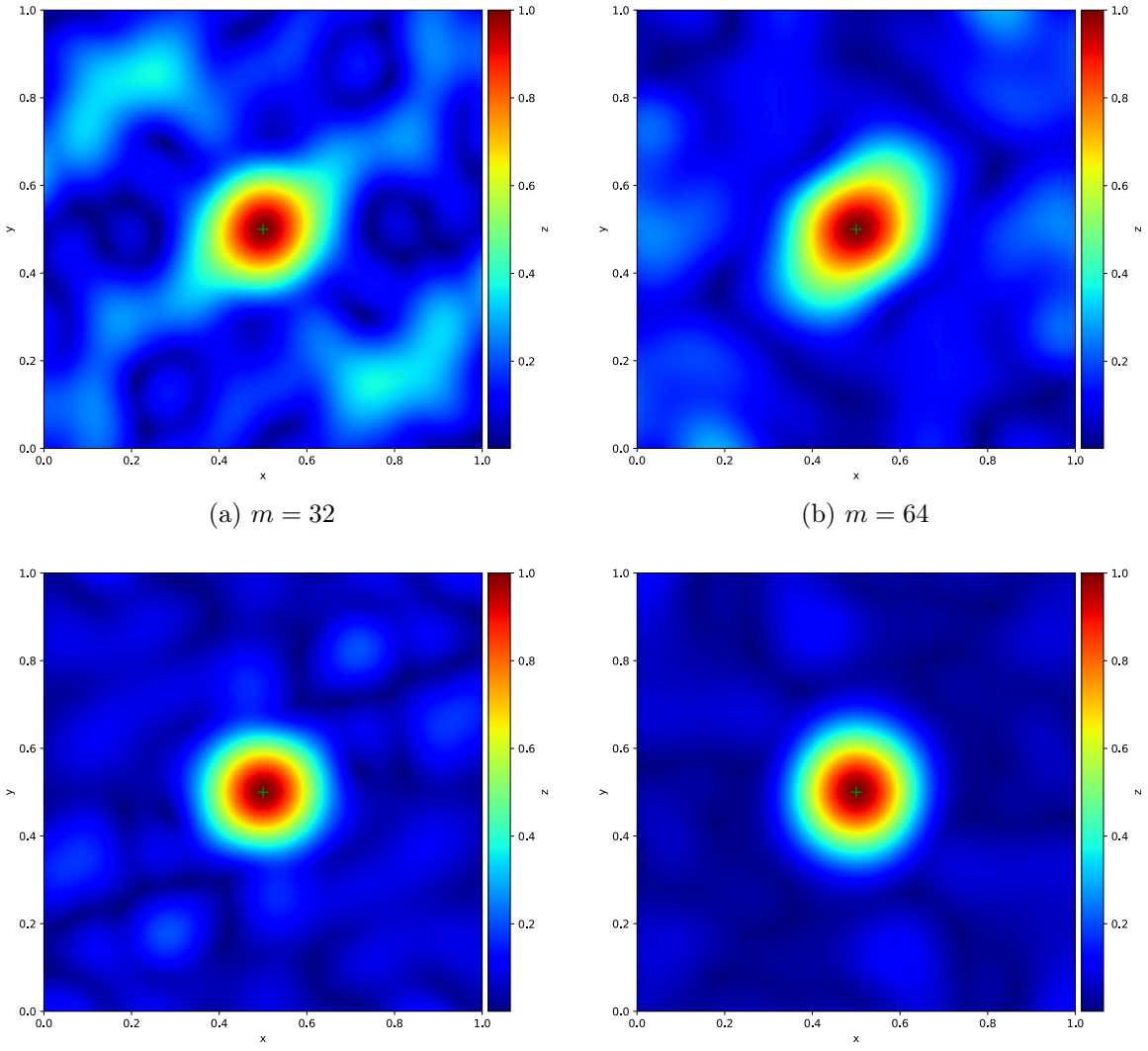
$$\theta_{\text{init}} \in \{t \in \mathbb{R}^d, 0.01 \leq \|t\|_2 \leq 0.5\}. \tag{4.5}$$

Nous obtenons alors les résultats montrés dans la Figure 4.10.

Nous observons que la taille du bassin d'attraction ne dépend pas de d comme montré dans [85].

L'effet du nombre d'impulsions

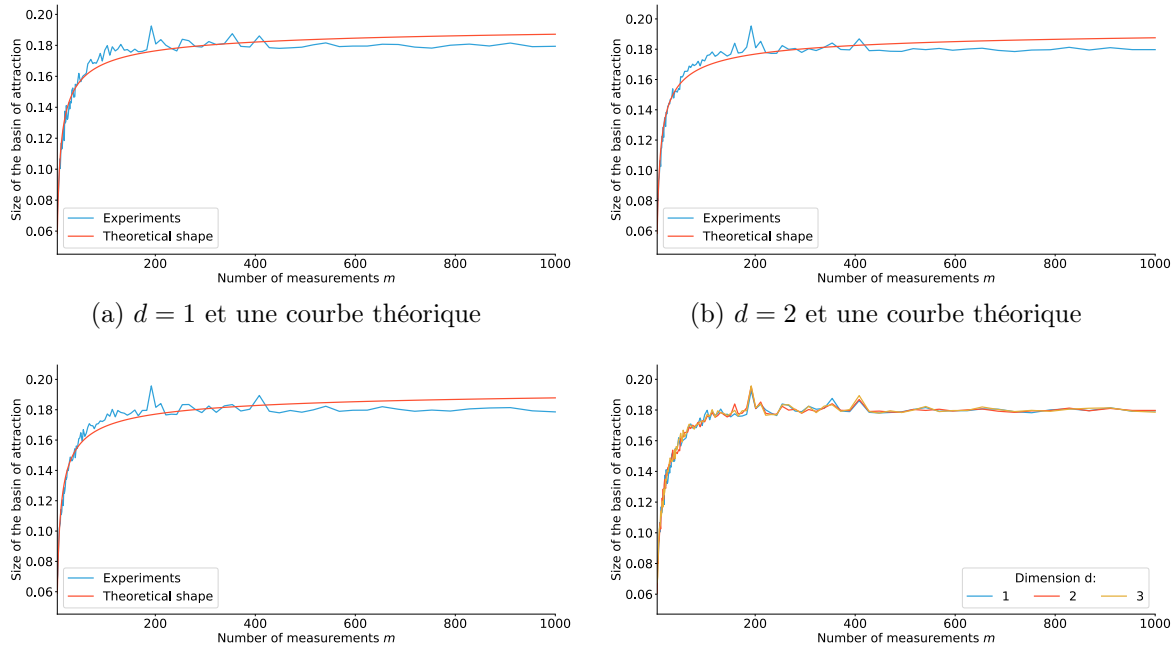
Nous vérifions dans cette section que le comportement par rapport à m ne change pas en fonction du nombre d'impulsions K. Nous utilisons le signal $x_0 = \sum_{i=1}^K \delta_{t_i}$ avec $\{t_1, \ldots, t_K\}$ les K-ièmes racines du cercle unitaire de rayon $r = 1$ afin de fixer une distance constante entre les impulsions. Nous réalisons l'expérience pour $K \in \{2, 3, 4, 5\}$. Nous initialisons un signal $x_{\text{init}} = \sum_{i=1}^K \delta_{s_i}$ tel que pour tous $i = 1, \ldots, K$, $0.01 \leq \|t_i - s_i\|_2 \leq 0.5$. Pour évaluer la distance entre x_0 et l'estimation x^*, nous prenons $\max_{i=1,\ldots,k} \|t_i^* - t_i\|_2$.



(a) $m = 32$ (b) $m = 64$

(c) $m = 128$ (d) $m = 512$

FIGURE 4.9 – Rétro-projection du signal $x_0 = \delta_t$ observé à l'aide de m mesures de Fourier aléatoire. Nous remarquons que l'opérateur de mesure converge vers un noyau Gaussien quand m augmente.



(a) $d = 1$ et une courbe théorique

(b) $d = 2$ et une courbe théorique

(c) $d = 3$ et une courbe théorique

(d) $d = 1, 2, 3$

FIGURE 4.10 – Taille du bassin d'attraction en fonction du nombre de mesures m pour différentes dimensions d. La courbe théorique est obtenue par une estimation par les moindres carrés de la fonction $C_1\sqrt{1 - C_2/\sqrt{m}}$ paramétrée par C_1 et C_2.

Nous obtenons les résultats présentés dans la Figure 4.11. Nous observons que la taille des bassins se comporte de manière similaire pour différents k. Il y a un léger décalage pour les petits m, comme on peut s'y attendre étant donné la nature asymptotique du résultat dans [85].

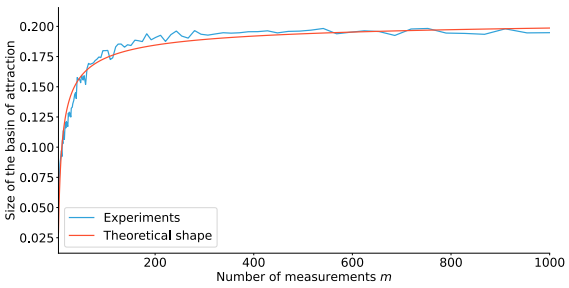
L'effet de la séparation minimale et le choix des mesures

Nous montrons dans la Figure 4.12 l'effet de la séparation minimale entre les impulsions sur la taille des bassins d'attraction, pour 5 impulsions avec une configuration donnée. Nous changeons le paramètre de séparation en modifiant simplement les positions des impulsions.

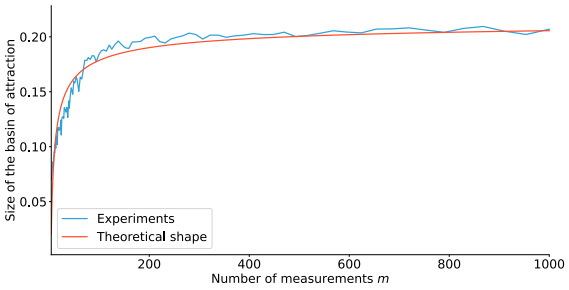
Comme prévu, la taille des bassins d'attraction augmente en fonction de la séparation minimale, mais seulement jusqu'à une certaine limite : cette expérience montre l'impact du choix de σ (le paramètre qui définit le choix de la distribution des fréquences aléatoires). Afin d'atteindre une taille de bassin d'attraction proche de la séparation minimale, il doit être réglé en conséquence.

Ceci est confirmé par la Figure 4.13, où nous répétons les expériences avec différents σ avec une séparation minimale fixe $\epsilon = 0.5$. Cela confirme les résultats théoriques dans [85] : la taille des bassins augmente en fonction de σ, jusqu'à une limite correspondant à la séparation minimale.

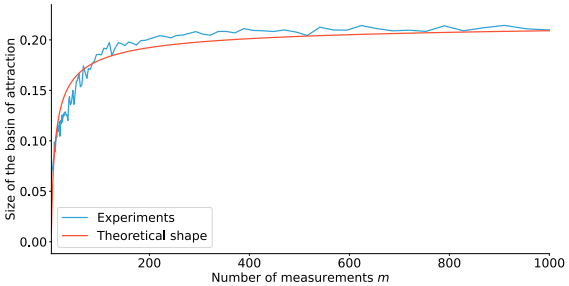
Pour conclure cette partie, nous remarquons que la taille du bassin d'attraction ne semble pas dépendre de la dimension d de notre problème ni du nombre d'impulsions K à reconstruire. Cependant, comme attendu avec la borne 4.3, la taille du bassin d'attraction dépend bien de l'écart-type de notre noyau gaussien et donc de la séparation minimale relative entre les éléments du support. Nous parlons de séparation relative car deux réponses impulsionnelles seront plus ou moins distinctes l'une de l'autre en fonction



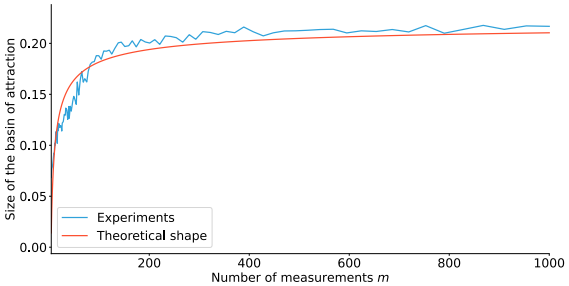
(a) $K = 2$ et une courbe théorique



(b) $K = 3$ et une courbe théorique

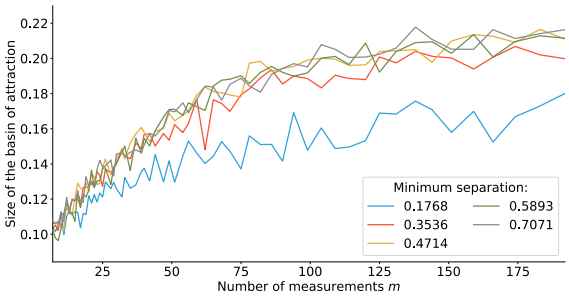


(c) $K = 4$ et une courbe théorique

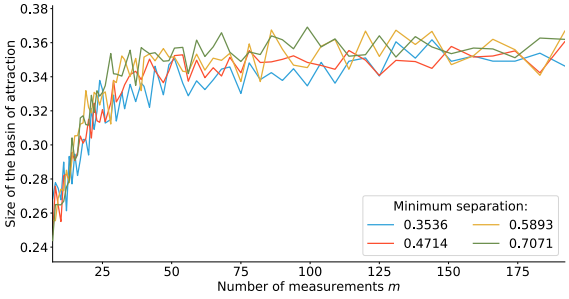


(d) $K = 5$ et une courbe théorique

FIGURE 4.11 – Taille du bassin d'attraction en fonction du nombre de mesures m avec différents nombres d'impulsions K. La courbe théorique est obtenue par une estimation par les moindres carrés de la fonction $C_1\sqrt{1 - C_2/\sqrt{m}}$ paramétrée par C_1 et C_2.



(a) $\sigma = 0.1$



(b) $\sigma = 0.2$

FIGURE 4.12 – Taille du bassin d'attraction en fonction de la séparation minimale ϵ.

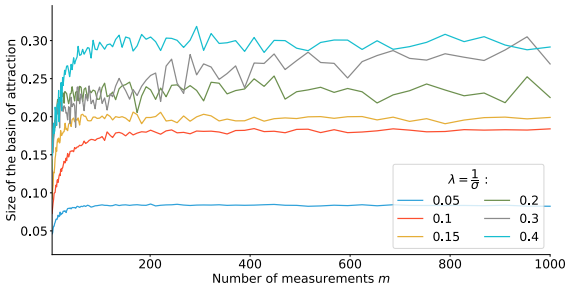


FIGURE 4.13 – Taille du bassin d'attraction en fonction du paramètre σ. Séparation minimale à $\epsilon = 0.5$.

de l'écart-type noyau gaussien associé.

De plus, nous notons qu'un nombre minimum de mesures est nécessaire pour obtenir une taille de bassin d'attraction suffisamment grande. Cependant, choisir un trop grand nombre de mesures peut également nuire à l'aspect pratique de la reconstruction du signal car la taille du bassin d'attraction n'augmente plus de manière significative et le temps de calcul peut être impacté par m.

4.3 Résumé du chapitre et conclusions

Dans ce chapitre, nous avons pu quantifier l'effet de la sur-paramétrisation d'un signal estimé dans la résolution du problème non-convexe 2.1. Dans un premier lieu, nous avons observé que si notre estimation est sur-paramétrée, la vitesse de convergence vers x_0 diminue. Le nombre d'itérations pour atteindre la même valeur pour la norme d'un résidu qu'avec un signal sans sur-paramétrisation est bien supérieur. En utilisant la Hessienne de la fonction non-convexe, au moins l'une des valeurs propre associée au système est négative pour les signaux sur-paramétrés, ce qui implique que nous ne sommes pas dans le bassin d'attraction du minimum global (au sens de la convexité). De plus, avec le conditionnement de la Hessienne qui augmente considérablement, l'instabilité numérique peut également causer d'autres imprécisions dans la reconstruction du signal x_0. Ces résultats appuient alors l'utilisation d'une méthode de projection afin de réduire le nombre d'impulsions pour accélérer la convergence.

Quant aux travaux sur la taille des bassins d'attraction, nous observons un lien clair avec le nombre de mesures d'une observation. Quantitativement, nous remarquons que la variance associée au noyau gaussien occupe un rôle important dans la taille du bassin d'attraction tout comme le nombre de mesures de notre observation. Une possibilité par la suite peut être d'étendre ces travaux de calculs sur les estimations de taille de bassins d'attraction pour des opérateurs linéaires autres que Fourier aléatoire.

Chapitre 5

Application à la super-résolution en microscopie et comparaison avec les méthodes état de l'art

Dans ce chapitre, nous appliquons notre algorithme OP-COMP avec PGD à deux problèmes de microscopie et comparons ses performances à la méthode état de l'art COMP avec descente. Nous rappelons même si nous ne nous comparons pas directement à Sliding Frank-Wolfe, les algorithmes COMP avec descente et SFW sont très similaires. Donc les temps de calculs et précision des signaux estimés pour ces deux algorithmes sont équivalents. Plus précisément, nous considérons les problèmes de calibration de microscope à partir d'acquisitions de micro-billes fluorescentes (voir Section 5.4) et de super-résolution avec la reconstruction de signaux (position et amplitudes des cellules observées) à partir de données SMLM (Section 5.5). Bien que les deux problèmes soient liés à la reconstruction de signaux composés d'impulsions parcimonieuses, dans le premier cas, les micro-billes suivent généralement une distance de séparation minimale conforme aux hypothèses de nos résultats théoriques. Nous comparons à nouveau ces deux algorithmes sur un dernier exemple, celui où le signal à reconstruire est composé de beaucoup d'impulsions sans structure particulière ni séparation minimale. Nous cherchons à illustrer avec ce dernier exemple l'avantage que procure notre méthode à la fois sur le plan calculatoire et sur la précision de l'estimation selon différentes métriques. Le code des expériences suivantes peut être téléchargé à l'adresse suivante : [11].

Tout d'abord, explicitons l'un des opérateurs utilisé en microscopie.

5.1 La réflexion fluorescente interne fluorescente total à plusieurs angles (MA-TIRF)

Le principe de la TIRF est d'éclairer l'échantillon avec un angle d'incidence dans le régime de la réflexion totale [5]. Dans ce régime, bien que toute la lumière incidente soit réfléchie, une onde évanescente est créée au-dessus de la lamelle de verre. Elle permet d'exciter les fluorophores dans une fine couche de quelques centaines de nanomètres à la surface basale des cellules. Le MA-TIRF consiste alors simplement à considérer un ensemble d'acquisitions pour des angles TIRF multiples. Comme la décroissance axiale de l'onde évanescente est liée à l'angle d'incidence, les données MA-TIRF fournissent des informations sur la position axiale des fluorophores. Soit $y \in \mathbb{R}^{N_1 \times N_2 \times K_{\text{angle}}}$ notre observation avec N_1 et N_2 le nombre de pixels pour les deux premières dimensions et

K_{angle} le nombre d'angles d'incidence. Bien que nous n'ayons pas d'information directes sur les positions des cellules pour le troisième axe, c'est en combinant les observations avec les multiples angles d'incidence que nous pouvons retrouver ces informations.

Chaque entrée de y est exprimée comme une somme pondérée de l'application d'une réponse impulsionnelle $\alpha_{i,k}$, avec $i \in \{1, \dots, N_1 N_2\}$, $k \in \{1, \dots, K_{\text{angle}}\}$, sur chaque impulsion de $x_0 = \sum_{j=1}^{K} a_j \delta_{t_j}$ avec $t_j \in \mathbb{R}^3$, c'est-à-dire $y_{i,k} = \sum_{j=1}^{K} a_j \alpha_{i,k}(t_j)$. L'expression de $\alpha_{i,k}(t_j)$ pour le modèle MA-TIRF est donnée par

$$\alpha_{i,k}(t_j) := \frac{\xi(t_{j,3}) \mathrm{e}^{-s_k t_{j,3}}}{2\pi\sigma_1\sigma_2} \int_{\Omega_i} \mathrm{e}^{-\left(\frac{(t_{j,1}-s_1)^2}{2\sigma_1^2} + \frac{(t_{j,2}-s_2)^2}{2\sigma_2^2}\right)} \mathrm{d}s_1 s_2 \tag{5.1}$$

où $\xi(t_{j,3}) = \left(\sum_{k=1}^{K_{\text{angle}}} \mathrm{e}^{-2s_k t_{j,3}}\right)^{-1/2}$ et Ω_i la région du i-ième pixel de la caméra. Elle correspond à la combinaison de l'excitation évanescente dans la dimension axiale (la troisième dimension) et de la convolution latérale (deux premières dimensions) avec la réponse impulsionnelle, que nous approximons par un noyau gaussien. Pour plus de détails sur ce modèle, nous renvoyons le lecteur vers [37]. Afin d'implémenter cette formulation de la forme continue à sa forme discrète, nous obtenons

$$\alpha_{i,k}(t_j) = \frac{\xi(t_{j,3}) \mathrm{e}^{-s_k t_{j,3}}}{4} \Big(\operatorname{erf}(v_x^+) - \operatorname{erf}(v_x^-) \Big) \Big(\operatorname{erf}(v_y^+) - \operatorname{erf}(v_y^-) \Big) \tag{5.2}$$

avec $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2} \mathrm{d}t$ la fonction d'erreur et

$$v_x^- = \frac{t_{j,1} - \Delta_{\text{px}}/2 - x_i}{\sqrt{2}\sigma_1} \qquad v_x^+ = \frac{t_{j,1} + \Delta_{\text{px}}/2 - x_i}{\sqrt{2}\sigma_1} \tag{5.3}$$

$$v_y^- = \frac{t_{j,2} - \Delta_{\text{px}}/2 - y_i}{\sqrt{2}\sigma_2} \qquad v_y^+ = \frac{t_{j,2} + \Delta_{\text{px}}/2 - y_i}{\sqrt{2}\sigma_2} \tag{5.4}$$

avec (x_i, y_i) les coordonnées du centre du i-ième pixel et Δ_{px} sa largeur (il s'agit également de sa longueur sachant que nous considérons le cas du pixel carré). Cette déduction de la formulation continue est évidente en considérant $\Omega_i = [x_i - \Delta_{\text{px}}/2, x_i + \Delta_{\text{px}}/2] \times [y_i - \Delta_{\text{px}}/2, y_i + \Delta_{\text{px}}/2]$.

Nous remarquons finalement que la formulation discrète est largement plus complexe que pour le modèle de l'échantillonnage de Fourier.

5.2 Métriques pour évaluer la performance des algorithmes

Pour évaluer les performances de chaque algorithme, nous prenons en compte le temps de calcul et la norme du résidu de l'estimation $g(\theta_{\text{esti}})$. Cependant, ces deux métriques à elles seules ne nous donnent pas assez d'information pour quantifier la proximité de notre estimation au signal original c'est à dire estimer le nombre d'impulsions correctement estimées.

Nous introduisons alors les notations de vrai positif (TP, true positive), faux positif (FP, false positive) et faux négatif (FN, false negative). Elles étiquettent chaque impulsions estimées en fonction de leurs position par rapport aux impulsions de la vérité terrain (GT, ground truth).

- TP : une impulsion de GT et une impulsion estimée sont proches l'une de l'autre. Dans ce cas là, on estime correctement une impulsion de x_0 par une impulsion de x_{esti}.
- FN : une impulsion de GT est seule. Dans ce cas, nous n'estimons pas d'impulsion alors qu'il existe bel et bien une impulsion du signal x_0 à reconstruire.
- FP : une impulsion estimée est seule. Dans ce cas, nous estimons une impulsion alors qu'il n'y a pas d'impulsion du signal x_0 à reconstruire.

Notons également que l'introduction d'un "vrai négatif" (TN, true negative) n'a pas de sens dans notre cas d'un signal parcimonieux en dimension infinie car il s'agirait de lister l'ensemble des positions de $\mathbb{R}^d$ où il n'y a ni impulsion de c_0 ni impulsion de x_{esti}. Ces notations sont utilisées pour calculer les métriques suivantes [74] : l'indice de Jaccard (Jac), le rappel (Recall) et la précision (Pre),

$$\text{Jac} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FP} + \#\text{FN}} \qquad \text{Rec} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FN}} \qquad \text{Pre} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FP}}. \tag{5.5}$$

Ces indices et métriques nous permettent d'évaluer la capacité d'un algorithme à fournir une bonne estimation. Les scores sont compris entre 0 et 1, allant de la plus mauvaise à la meilleure performance. Nous notons que pour identifier une impulsion estimée comme TP, nous vérifions sa distance à l'impulsion de GT la plus proche notée $d_{\texttt{TP,GT}}$. Dans notre cas en microscopie, l'espace dans dans lequel les impulsions à reconstruire sont situées est un cuboïde de $[0, 6.4] \times [0, 6.4] \times [0, 0.8]$ µm. Nous distinguons deux cas en fonction de la plus petite distance entre deux impulsions de la vérité terrain notée $d_{\texttt{min}}$. Soit les impulsions de la vérité terrain sont bien séparées les unes des autres, c'est-à-dire que $d_{\texttt{min}} \geq 40$nm, alors $d_{\texttt{TP,GT}} = 20$nm. Soit $d_{\texttt{min}} < 40$nm, alors $d_{\texttt{TP,GT}} = \frac{1}{2} d_{\texttt{min}}$.

Nous introduisons également l'erreur quadratique moyenne (RMSE) entre les paires d'impulsions GT et estimées (uniquement à partir de l'ensemble des impulsions de TP) le long de chaque dimension,

$$\text{RMSE}_{x_i} = \sqrt{\frac{1}{\#\text{TP}} \sum_{j \in \text{TP}} \left([x_j]_i - [x_{0,j}]_i\right)^2}. \tag{5.6}$$

Cette métrique nous donne une estimation de la proximité du signal estimé par rapport à la vérité terrain dans chaque dimension pour chaque paire d'impulsions de GT et TP.

5.3 Analyse comparative des algorithmes et initialisation de paramètres

Comme indiqué précédemment, il existe deux différences principales entre COMP avec descente et OP-COMP avec PGD. La première est l'absence de l'étape de descente dans OP-COMP à chaque itération, la seconde est l'ajout d'une étape de sur-paramétrisation dans OP-COMP et d'une unique descente sur tous les paramètres avec des projections effectuées dans PGD. Pour vérifier si les deux différences entre les algorithmes apportent une valeur ajoutée, nous introduisons dans nos expériences l'algorithme COMP avec descente modifié : COMP + GD. Il s'agit d'un algorithme hybride entre COMP avec descente et OP-COMP + PGD. Pour l'initialisation, il se comporte comme OP-COMP (c'est à dire qu'il n'effectue pas de descente sur les paramètres à chaque itération) à la différence qu'il s'arrête à exactement K itérations pour un signal de K impulsions. Il effectue ensuite

une descente de gradient sur tous les paramètres mais ne réalise aucune projection. Cet algorithme permet de s'assurer que la sur-paramétrisation avec projection est un élément clé dans l'amélioration de la reconstruction des positions des impulsions.

En ce qui concerne l'étape de projection dans OP-COMP avec PGD, pour définir la distance de projection `eps_dist`, nous utilisons un oracle : nous calculons à l'avance la plus petite distance entre tous les impulsions dans le vrai signal `min_dist` et définissons la distance de projection `eps_dist` comme étant $0.5 \times$ `min_dist`. Toutefois, ce paramètre défini par l'utilisateur peut être facilement estimé à partir de l'observation de l'ensemble de données. En pratique, s'il est trop grand, on risque de projeter deux impulsions estimées qui ne devraient pas l'être. Dans le cas contraire, s'il est trop petit, PGD ne fusionne pas les impulsions qui devraient l'être et ralentit le calcul en raison du grand nombre d'impulsions et de la convergence lente d'un problème mal posé (voir le Chapitre 4).

5.4 Étalonnage du microscope à partir d'acquisitions de billes fluorescentes

Une pratique standard pour l'étalonnage en microscopie à fluorescence consiste à prendre des images de micro-billes fluorescentes. Elles se comportent comme des sources ponctuelles et les images associées constituent donc des mesures directes de la réponse impulsionnelle du système. Le processus d'étalonnage consiste alors en une estimation conjointe de la réponse impulsionnelle et de la position des micro-billes à partir de ces données. Ce problème est généralement traité par des méthodes alternatives et nous nous concentrons ici sur le sous-problème de localisation des micro-billes. Habituellement, ces échantillons d'étalonnage sont préparés de manière à ce que les micro-billes suivent une distance de séparation minimale, conformément aux hypothèses des résultats théoriques de nos travaux.

Dans ce contexte, nous avons synthétisé des billes fluorescentes uniformément dans le volume d'acquisition. Nous avons d'abord établi une grille uniforme de points dans notre volume. Ensuite, à partir de chaque point $x = (x_1, x_2, x_3)$ de la grille, nous avons généré aléatoirement une impulsion aux coordonnées $t = (t_1, t_2, t_3)$ à l'intérieur d'une ellipse centrée sur x, c'est à dire,

$$\frac{(x_1 - t_1)^2}{(s_1/2)^2} + \frac{(x_2 - t_2)^2}{(s_2/2)^2} + \frac{(x_3 - t_3)^2}{(s_3/2)^2} \leq 1. \tag{5.7}$$

avec s_1, s_2, s_3 les longueurs de chaque axe de l'ellipse. Quant aux amplitudes des impulsions, elles sont échantillonnées uniformément entre 1 et 1.5. Nous montrons dans la Figure 5.1 une coupe de l'observation y selon deux angles d'incidence différents. Nous observons qu'il y a une séparation entre les billes, certaines réponses se recoupant encore (c'est-à-dire que nous ne sommes pas dans le cas le plus favorable en termes de séparation).

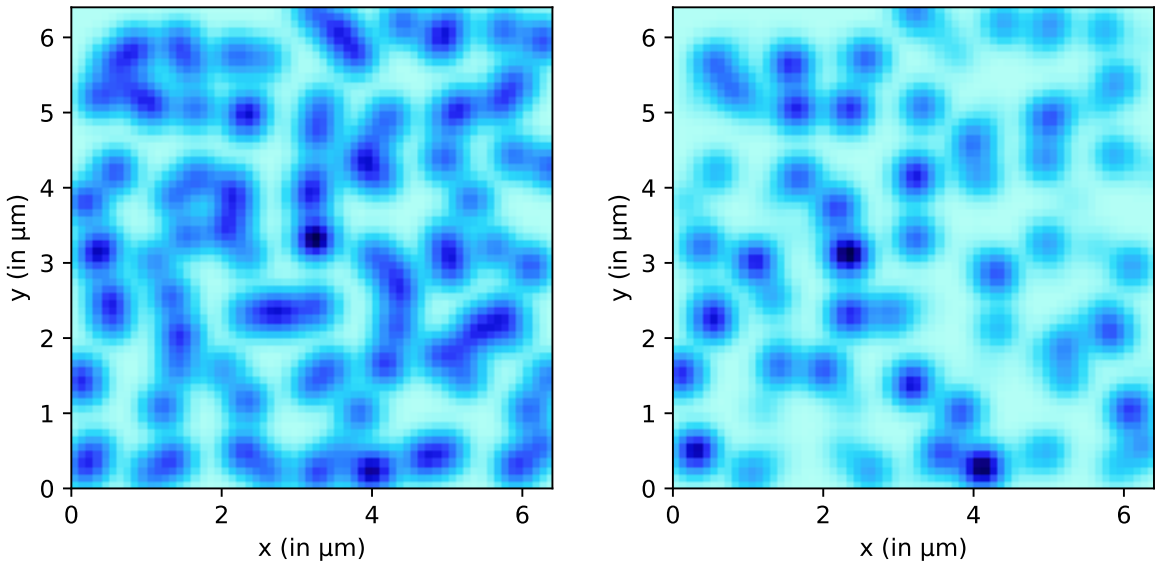


FIGURE 5.1 – Observation d'un signal composé de 98 impulsions le long de deux angles d'incidence différents avec le modèle MA-TIRF.

Nous comparons OP-COMP + PGD, COMP + GD et COMP avec descente sur un signal composé de 98 impulsions. La localisation 3D est présentée dans la Figure 5.2. La plupart des impulsions ont été estimées par les trois méthodes,

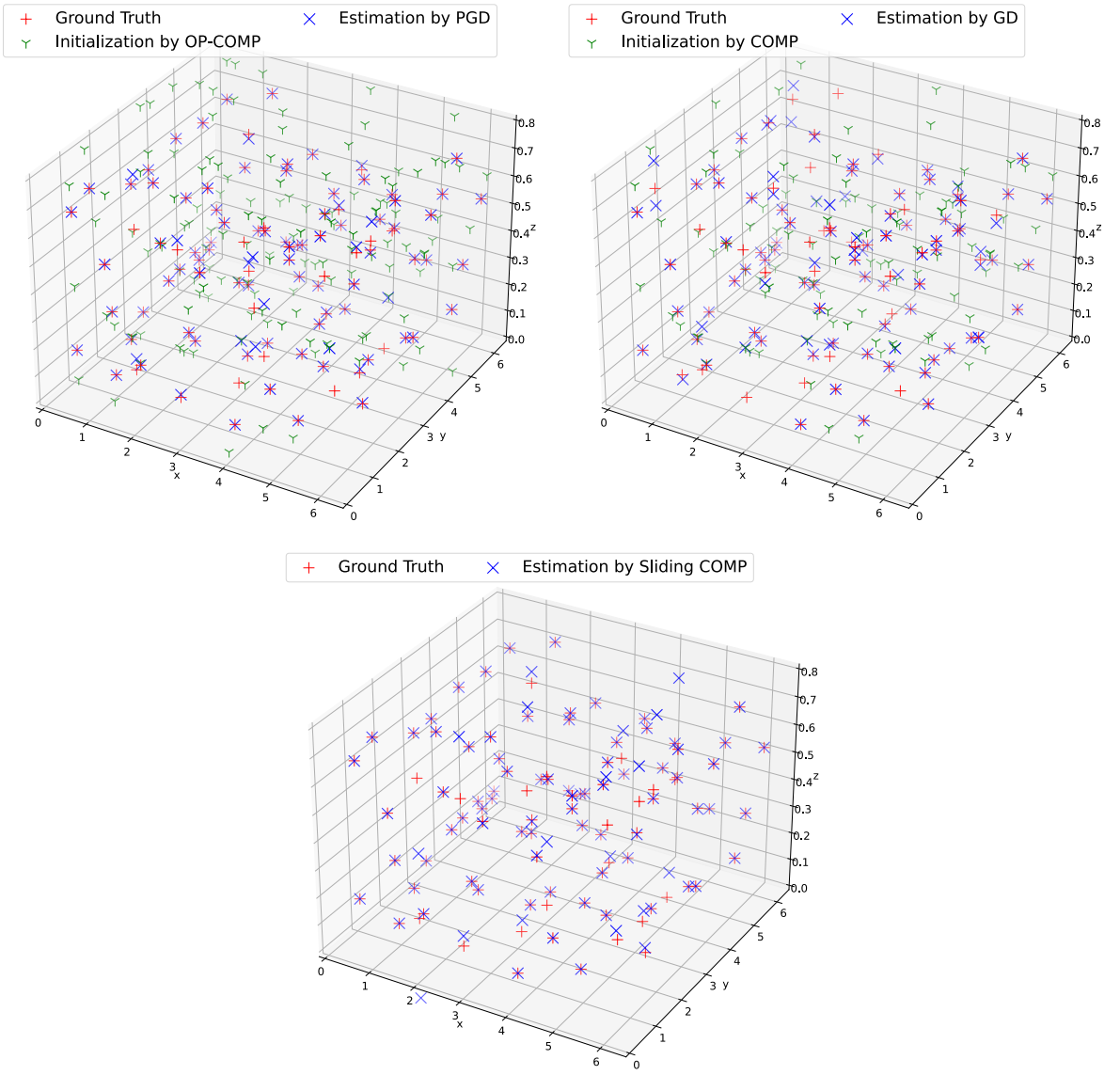


FIGURE 5.2 – Représentation en 3D d'un signal de $K = 98$ impulsions et de son estimations par OP-COMP (en vert) puis PGD (en bleu) (en haut à gauche), de son estimation par COMP (en vert) puis GD (en bleu) (en haut à droite) et de son estimation par COMP avec descente (en bleu) (en bas au centre).

Nous traçons les résidus $\|y - Ax\|_2$ à chaque étape des trois algorithmes dans les Figures 5.3, 5.4 et 5.5. À la fin de OP-COMP, le signal initialisé est composé de 135 impulsions. Après 10 000 itérations et environ 35 projections avec PGD, la norme du résidu cesse de décroître et nous obtenons un signal composé de 95 impulsions. Nous observons que la norme du résidu diminue beaucoup moins avec GD sans projection. Par rapport au COMP avec descente, la norme du résidu de OP-COMP avec PGD est similaire.

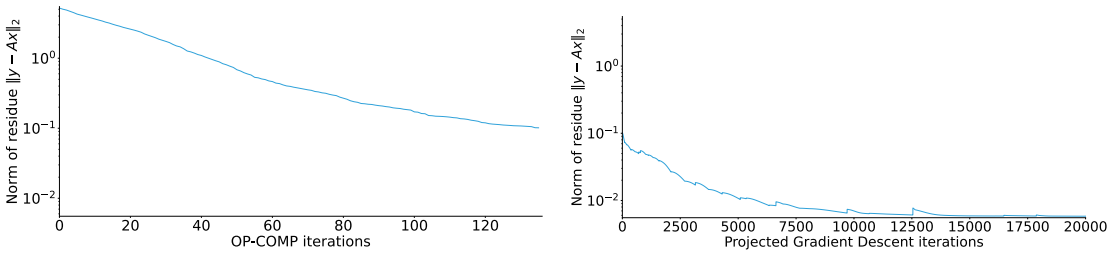


FIGURE 5.3 – Norme des résidus $\|y - Ax\|_2$ à chaque itération de OP-COMP (gauche), de la Descente de Gradient Projetée (droite) pour un signal de 98 impulsions.

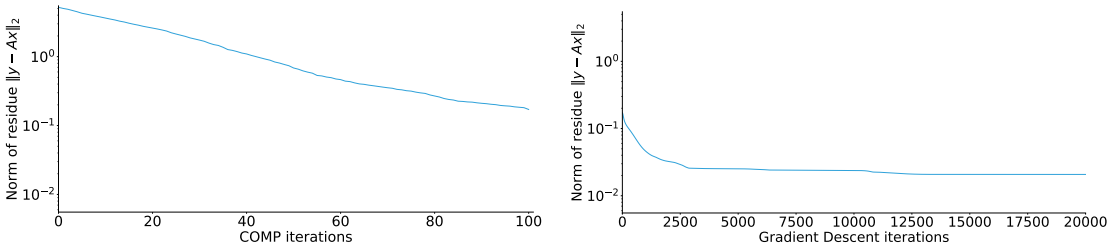


FIGURE 5.4 – Norme des résidus $\|y - Ax\|_2$ à chaque itération de COMP (gauche), de la Descente de Gradient (droite) pour un signal de 98 impulsions.

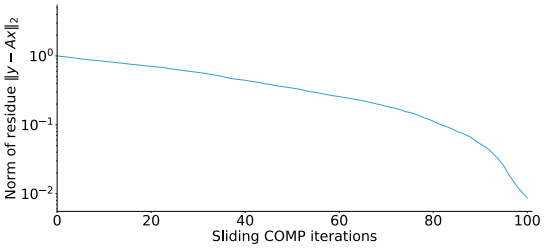


FIGURE 5.5 – Norme des résidus $\|y - Ax\|_2$ à chaque itération de COMP avec descente pour un signal de 98 impulsions.

Pour comparer ces estimations en termes de localisation, nous utilisons les métriques que nous avons introduites précédemment (voir la Section 5.2) et la norme du résidu de chaque estimation. Les résultats et les temps de calcul sont présentés dans le Tableau 5.1. Nous concluons que pour la résolution de notre problème non-convexe (2.5), OP-COMP + PGD est plus performant que son homologue sans sur-paramétrisation ni projection [59].

	OP-COMP + PGD	COMP + GD	COMP avec descente
Temps de calcul	129 minutes	115 minutes	272 minutes
Norme du residu de l'estimation	5.84×10^{-3}	2.08×10^{-2}	8.69×10^{-3}
Index Jaccard	0.771	0.567	0.664
Rappel	0.857	0.727	0.806
Précision	0.884	0.720	0.790
RMSE (x_1, x_2, x_3) en nm	$(2.70, 1.68, 2.05)$	$(3.39, 3.56, 4.83)$	$(1.43, 1.58, 1.55)$

TABLE 5.1 – Tableau de comparaison entre OP-COMP + PGD, COMP + GD et COMP avec descente sur le temps de calcul, la norme des résidus, l'index de Jaccard, les métriques de rappel et de précision, et sur la RMSE pour chaque dimension.

Nous observons qu'en termes de temps de calcul, COMP avec descente est largement surclassée par OP-COMP avec PGD et COMP avec GD. Pour les trois métriques du Jaccard, Rappel et Précision, nous constatons que OP-COMP + PGD est plus performant que COMP + GD et COMP avec descente.

Nous concluons que dans le cas où nous devons estimer un grand nombre d'impulsions dans un signal avec une (relativement) bonne séparation entre impulsions, OP-COMP avec PGD est un algorithme bien adapté pour reconstruire un tel signal car il offre un temps de calcul plus faible et de meilleures performances par rapport à l'état de l'art COMP avec descente.

5.5 Le problème de localisation de SMLM

La méthode de "Single molecule localization microscopy" (SMLM) est actuellement l'une des techniques de super-résolution les plus puissantes en microscopie de fluorescence, avec la possibilité d'atteindre une résolution allant jusqu'à dix nanomètres [74]. Contrairement à la microscopie à fluorescence classique, où toutes les molécules fluorescentes sont activées et imagées en même temps, elle procède par activation séquentielle et localisation de sous-ensembles parcimonieux de ces molécules. La densité des molécules activées sur chaque image SMLM est un compromis entre la difficulté du problème de localisation (plus facile pour les faibles densités) et la résolution temporelle de l'imagerie en direct (meilleure pour les fortes densités).

La SMLM tridimensionnelle peut être obtenue en utilisant une réponse impulsionnelle modifiée dont la forme varie axialement (troisième dimension). Il s'agit par exemple des réponses populaires que sont l'astigmatisme [56] et la double hélice [72]. Pour coder la profondeur des molécules par le biais de variations axiales, ces réponses impulsionnelles exotiques sont accompagnées d'un support latéral plus important que les réponses conventionnelles, ce qui accroît la difficulté du problème de localisation. Une alternative consiste à exploiter les excitations MA-TIRF pour coder la profondeur des molécules tout en conservant une réponse conventionnelle plus étroite. Cette approche a été étudiée numériquement et comparée à l'utilisation de l'astigmatisme et des réponses de la double

hélice dans [37].

Dans nos travaux, nous considérons les structures de micro-tubules simulées dans [74]. À partir des positions disponibles des molécules simulées $K_{\texttt{tot}}$, nous générons des images SMLM (avec MA-TIRF), chacune contenant un sous-ensemble de $K \ll K_{\texttt{tot}}$ molécules. Dans la référence de l'état de l'art [37], K varie entre 5 et 15. Lorsque K augmente, la probabilité que deux molécules soient proches l'une de l'autre augmente et le problème de localisation devient plus difficile. Si cela se produit, il devient plus difficile de récupérer leurs positions et amplitudes réelles avec les méthodes d'illumination standard. En effet, dans ce cadre là, nous n'imposons aucune contrainte de séparation entre les molécules. Dans nos expériences, nous évaluons les performances de OP-COMP avec PGD et de COMP avec descente sur des lots de 100 images SMLM, où un lot correspond à une observation d'un signal composé de $K \in \{5, 10, 15, 30\}$ impulsions (c'est à dire une image SMLM).

En ce qui concerne les paramètres du modèle (5.1), nous considérons le cadre décrit dans [37]. En particulier, nous utilisons le nombre de pixels du détecteur à grille $N_1 = N_2 = 64$, le nombre d'angles différents $K_{\texttt{angle}} = 4$ et la longueur d'onde d'excitation $\lambda_l = 660$nm. Cette dernière est directement liée à la décroissance de l'onde évanescente de l'illumination TIRF. En outre, conformément à [37], les variances σ_1 et σ_2 du filtre gaussien utilisé dans le modèle (5.1) ont été fixées à $\sigma_1 = \sigma_2 = 0.42\lambda_l/\mathrm{NA}$ avec $\mathrm{NA} = 1.49$ l'ouverture numérique du système. Cela correspond à un ajustement du modèle de filtre gaussien au filtre mesuré expérimentalement dans le cadre du défi SMLM [74]. Ainsi, λ_l est également lié à l'étalement de la réponse impulsionnelle. Une longueur d'onde plus petite conduit à une réponse plus étroite et permet une meilleure séparabilité des observations des molécules.

Dans ce contexte, contrairement au problème d'étalonnage présenté dans la section 5.4, l'hypothèse de séparation concernant la forme des mesures n'est pas complètement vérifiée : certaines impulsions de la vérité de terrain sont très proches les unes des autres et l'opérateur de mesures linéaires A ne donne pas suffisamment d'informations pour les distinguer. Compte tenu de ces lacunes, une application directe de OP-COMP sans descente + PGD ne fonctionne pas aussi bien que dans la section précédente. Dans ces expériences, nous ajoutons une petite étape de descente à chaque itération pour OP-COMP et COMP. Par rapport à COMP avec descente, où plusieurs centaines d'itérations sont effectuées à chaque étape de descente, nous n'effectuons que 10 à 20 itérations de descente. Ce nombre est suffisamment faible pour ne pas ralentir l'ensemble des algorithmes tout en améliorant la convergence globale.

Enfin, pour effectuer une comparaison adéquate, le nombre d'itérations dans la descente de gradient finale (projetée) est fixé de manière à fournir des performances similaires en termes de reconstruction des positions lorsque cela est possible. En comparant les temps d'exécution de OP-COMP avec PGD, COMP avec GD et COMP avec descente dans la Figure 5.6 (en échelle logarithmique) et leurs différences relatives dans le Tableau 5.2, nous observons que pour $K \in \{5, 10\}$ (c'est à dire un petit nombre d'impulsions), notre algorithme OP-COMP avec PGD et COMP avec GD sont plus lents que COMP glissant. Mais à mesure que le nombre d'impulsions par signaux augmente, nous constatons un gain de temps croissant pour OP-COMP avec PGD par rapport à COMP avec descente, comme le prévoit la complexité de calcul théorique des algorithmes. Nous gagnons 10% de temps de calcul pour $K = 15$ et jusqu'à 75% pour $K = 30$. En effet, à mesure que le nombre d'impulsions augmente, l'étape de descente à chaque itération du COMP avec descente est plus longue à calculer. Nous rappelons de la Section 2.3 que les temps d'exécution pour OP-COMP avec PGD et COMP avec descente sont respectivement de l'ordre

de $\mathcal{O}(K)$ contre $\mathcal{O}(K^2)$.

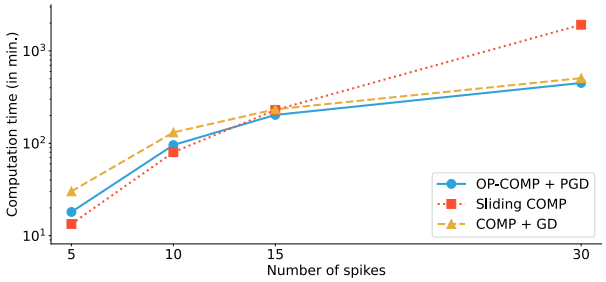


FIGURE 5.6 – Temps de calcul en *échelle logarithmique* pour OP-COMP avec PGD, COMP avec GD et COMP avec descente sur un lot de 100 signaux de taille $K = 5,\ 10,\ 15$ et 30.

	Nombre d'impulsions K			
Différence relative entre COMP avec descente et	5	10	15	30
OP-COMP avec PGD (en %)	34.8	19.6	−11.6	−76.6
COMP avec GD (en %)	125.5	63.9	1.7	−73.6

TABLE 5.2 – Différences relatives de temps de calcul en pourcentage entre COMP avec descente et les deux méthodes OP-COMP avec PGD et COMP avec GD sur 100 signaux de taille $K = 5,\ 10,\ 15$ et 30.

Les résultats de la reconstruction en termes d'index de Jaccard, de Rappel et de Précision sont présentés dans la Figure 5.7.

Nous observons que les différents index et métriques sont très similaires pour les trois méthodes pour $K = 5,\ 10$ et 15. Cela signifie que même si nos méthodes OP-COMP avec PGD et COMP avec GD sont plus rapides que le COMP avec descente, il n'y a pas de dégradation significative pour la qualité de l'estimation.

Pour le cas de $K = 30$, nous remarquons une baisse de performance pour chaque index et métrique pour toutes les méthodes. Ce résultat est prévisible en raison de la taille de notre signal par rapport à notre observation y, car il y a une plus grande probabilité d'avoir deux impulsions très proches l'une de l'autre dans un même signal. La méthode OP-COMP avec PGD et la méthode COMP avec descente ont des métriques similaires dans ce contexte pour les trois mesures de performance. Cependant, pour COMP avec GD, nous constatons une dégradation des performances par rapport à OP-COMP avec PGD. Ceci valide l'avantage d'avoir une initialisation sur-paramétrée et une projection dans la descente de gradient finale.

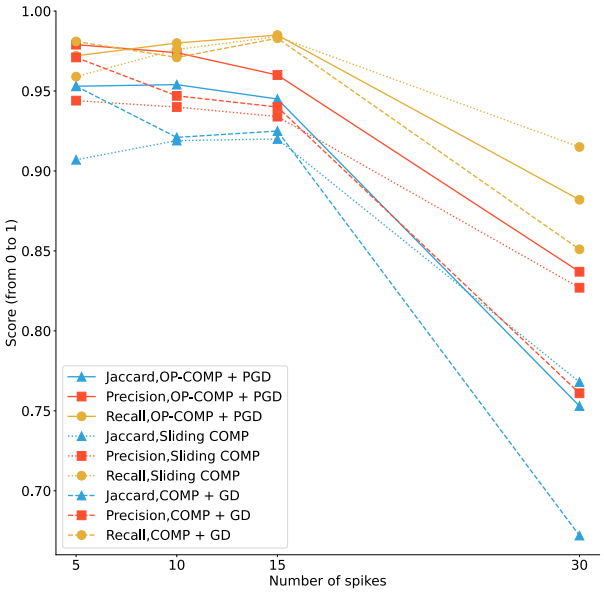


FIGURE 5.7 – Scores de l'index de Jaccard, des métriques du Rappel et de la Précision pour OP-COMP avec PGD, COMP avec GD, et COMP avec descente sur 100 signaux de taille $K = 5$, 10, 15 et 30.

En ce qui concerne la métrique RMSE pour chaque dimension, nous constatons que la distance entre les impulsions GT et TP est faible pour les trois algorithmes (voir Figure 5.8). Nous notons que le RMSE dans la troisième dimension (celle associée aux différents angles d'incidence dans MA-TIRF) est plus faible que la RMSE pour les deux premières dimensions. Ceci peut être dû à une profondeur de la zone d'acquisition plus faible 800nm comparé à 6400nm pour la largeur et longueur de la zone. Nous émettons tout de même l'hypothèse que $K_{\texttt{angle}} = 4$ pourrait être abaissé, et donc réduire le nombre total de mesures.

Comme prévu, nous observons également que le RMSE augmente avec le nombre d'impulsions K car l'observation y limite notre précision et la probabilité d'avoir deux impulsions de vérité terrain proches l'une de l'autre augmente avec plus d'impulsions par signaux.

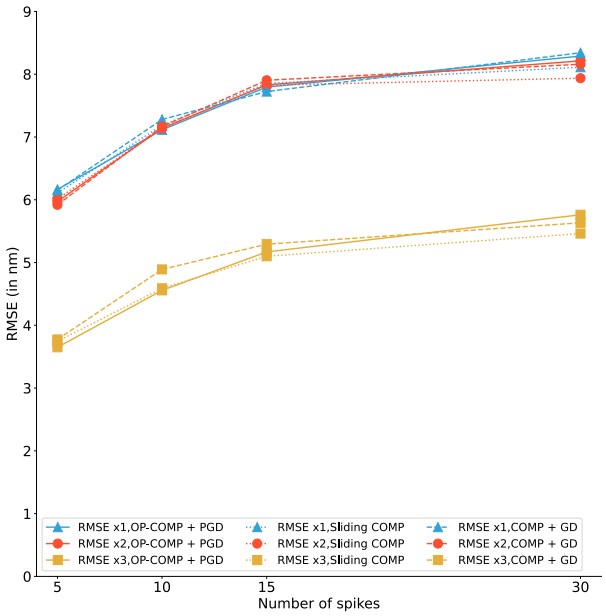


FIGURE 5.8 – Le RMSE dans chaque dimension pour OP-COMP avec PGD, COMP avec GD, et COMP avec descente pour 100 signaux de taille $K = 5$, 10, 15 et 30.

Nous combinons également tous les signaux de nos tests avec $K = 15$ dans un seul graphique dans le même style que dans [37] pour obtenir la Figure 5.9. Nous constatons que les trois estimations sont proches de nos micro-tubules initiaux. Cela confirme les performances du RMSE présentées précédemment, car pour les trois algorithmes, leurs RMSE sont très similaires et proches de 0 dans chaque dimension.

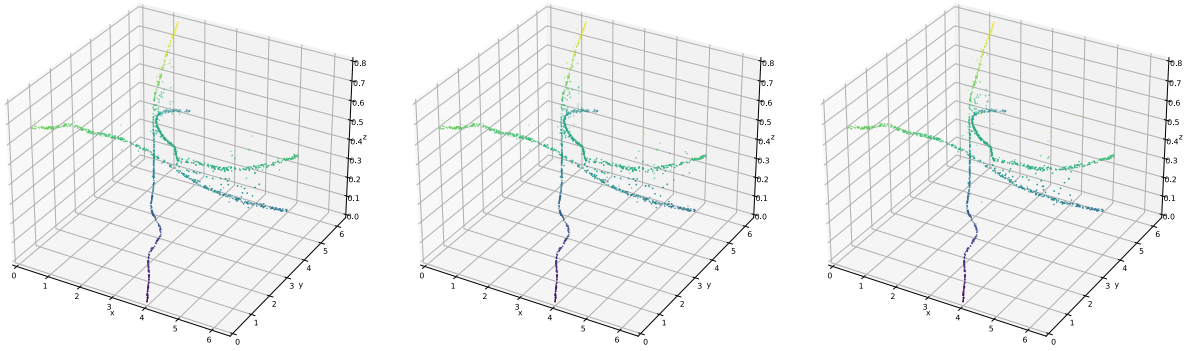


FIGURE 5.9 – Représentation en 3D d'une estimation d'un signal de taille $K = 15 \times 100$ par OP-COMP avec PGD (gauche), par COMP avec GD (centre) et par COMP avec descente (droite) avec un gradient de couleur variant avec la troisième dimension des positions et une taille associée à l'amplitude de chaque impulsion.

En conclusion de cette sous-section, même si les signaux reconstruits ne correspondent pas parfaitement à notre hypothèse de séparation (distances très faibles entre les impulsions pour une observation de faible qualité), OP-COMP avec PGD est aussi performant que l'état de l'art COMP avec descente en comparant avec les métriques classiques. Lorsque le nombre d'impulsions dans un signal augmente, OP-COMP avec PGD devient plus rapide que COMP avec descente. Nous présentons dans la Section 5.6 une expérience SMLM avec des signaux de 50 impulsions avec une longueur d'onde beaucoup plus petite

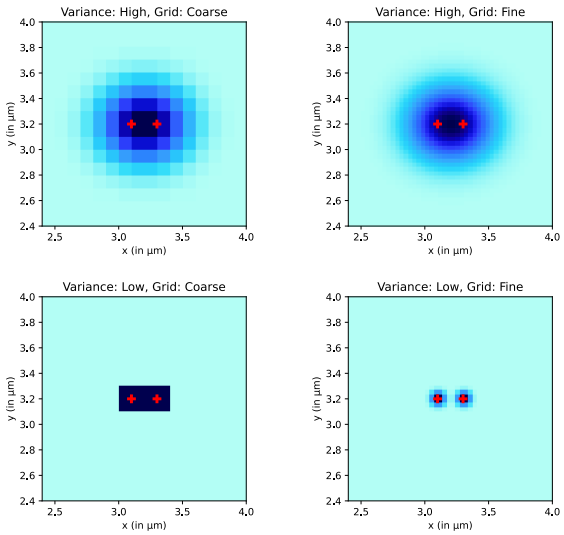


FIGURE 5.10 – Réponses impulsionnelles d'un signal par un opérateur de mesures gaussiennes avec une grande variance sur une grille grossière (en haut à gauche), avec une grande variance sur une grille fine (en haut à droite), avec une petite variance sur une grille grossière (en bas à gauche), avec une petite variance sur une grille fine (en bas à droite).

(qui ne sont malheureusement pas encore disponibles dans le cadre réel). Cette expérience montre que, si la résolution des instruments s'améliore, notre algorithme fournira une amélioration conséquente en temps de calcul.

5.6 Vers une amélioration de la précision de la méthode SMLM, en reconstruisant des signaux avec un grand nombre d'impulsions

Nous comparons les trois algorithmes OP-COMP avec PGD, COMP avec GD et COMP avec descente sur un signal avec $K = 50$ impulsions. Comme notre méthode s'adapte bien à un plus grand nombre d'impulsions, notre objectif est de réduire le nombre de signaux dans un lot en augmentant le nombre d'impulsions K dans chaque signal du lot. Cependant, dans le cadre classique, nous notons que la reconstruction des impulsions à partir de leur observation est limitée par la qualité de cette observation.

L'idée est de diminuer la largeur du noyau d'observation. Pour le modèle MA-TIRF, cela peut être réalisé en diminuant la longueur d'onde d'excitation λ_l. Nous avons choisi de remplacer $\lambda_l = 660$nm dans le cadre classique par $\lambda_l = 110$nm, la grille adaptée est plus fine avec 192×192 points au lieu de 64×64 sur le même domaine.

Dans la Figure 5.10, nous remarquons qu'utiliser uniquement une longueur d'onde d'excitation λ_l plus petite ou une grille plus fine ne permet pas de gagner en qualité sur l'observation à elles seules. La séparation entre les deux impulsions n'est pas suffisante par rapport à l'écart-type du noyau gaussien associé qui est trop grande. C'est en combinant les deux améliorations (diminuer λ_l et avoir une grille plus fine) que la séparation des deux réponses impulsionnelles est visible.

Même si ces paramètres (faible variance et grille fine) sont peu courants et très difficiles à mettre en place dans le cadre pratique, nous anticipons ces progrès dans les méthodes

d'acquisition et les améliorations potentielles apportées par l'algorithme OP-COMP avec PGD par rapport à l'algorithme COMP avec descente.

En comparant le signal estimé par OP-COMP avec PGD, COMP avec GD et COMP avec descente, nous représentons à la fois le signal à reconstruire x_0 et le signal estimé x_{esti}, (voir Figure 5.11). Nous observons que les signaux estimés sont proches du signal à reconstruire avec les trois méthodes.

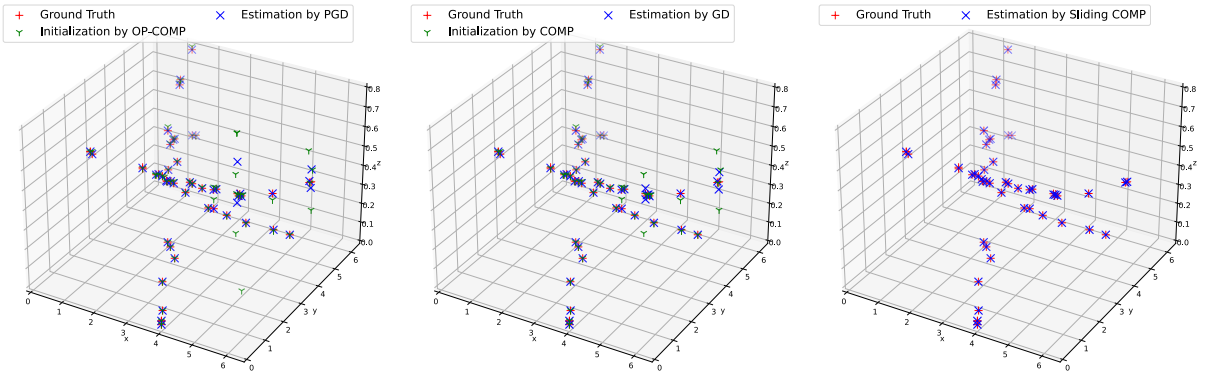


FIGURE 5.11 – Représentation en 3D d'un signal composé de $K = 50$ impulsions et son estimation par OP-COMP (en vert) et ensuite par PGD (en bleu) (gauche), son estimation par COMP (en vert) et puis par GD (en bleu) (centre) et son estimation par COMP avec descente (en bleu) (droite).

Pour comparer ces estimations, nous utilisons les mêmes métriques que nous avons introduites précédemment (temps de calcul, Index Jaccard, Rappel, Précision, RMSE) et la norme du résidu de chaque estimation. Nous obtenons le Tableau 5.3 suivant,

	OP-COMP avec PGD	COMP avec GD	COMP avec descente
Temps de calcul	333 minutes	327 minutes	537 minutes
Norme du résidu de l'estimation	3.58×10^{-4}	1.09×10^{-3}	2.46×10^{-2}
Index de Jaccard	0.839	0.776	0.741
Rappel	0.940	0.882	0.860
Précision	0.887	0.843	0.865
RMSE (x_1, x_2, x_3) en nm	$(3.07, 2.70, 1.50)$	$(3.17, 2.65, 2.47)$	$(4.36, 2.80, 2.07)$

TABLE 5.3 – Tableau des comparaison entre OP-COMP avec PGD, COMP avec GD et COMP avec descente sur le temps de calcul, la norme du résidu, les métriques Jaccard, Rappel et Précision et sur le RMSE dans chaque dimensions.

Nous observons dans le Tableau 5.3 que toutes les métriques de OP-COMP avec PGD sont meilleures que celles de COMP avec descente avec une amélioration de 40% du temps de calcul. Nous obtenons plus rapidement un résultat plus proche de la vérité terrain en termes de RMSE et d'index Jaccard, Rappel et Précision. Lorsque l'on compare OP-COMP avec PGD et COMP avec GD, la sur-paramétrisation avec projection est plus

efficace car nous obtenons des gains significatifs dans les trois métriques (Jaccard, Rappel et Précision).

Nous notons que dans ce signal de $K = 50$ impulsions, notre méthode a initialisé son signal avec $K_{\texttt{init}} = 50$ impulsions et a convergé après PGD vers un signal estimé avec $K_{\texttt{esti}} = 48$ impulsions. Cela explique le temps de calcul plus rapide par rapport à COMP + GD, et signifie que la sur-paramétrisation peut être facultative car elle n'améliorerait pas beaucoup plus le résultat dans ce cas là.

5.7 Résumé du chapitre et conclusions

Dans ce chapitre, nous avons pu comparer notre méthode OP-COMP + PGD à l'état de l'art COMP avec descente et également à la troisième méthode introduite COMP + GD. Dans l'ensemble des expériences réalisées, nous observons que plus le nombre d'impulsions à reconstruire est grand, plus notre méthode OP-COMP + PGD est rapide en termes de temps de calcul face à COMP avec descente. Pour ce qui est de la performance des autres métriques, OP-COMP + PGD est équivalente voire meilleur que COMP avec descente. Les signaux estimés sont au moins aussi proche du signal à reconstruire x_0 que les signaux estimés par COMP avec descente. Face à COMP + GD, nous observons que la combinaison de la sur-paramétrisation et de la projection des impulsions améliore la qualité du signal estimé (impulsions très proches de celles de la vérité terrain) sans endommager les performances de temps de calcul. Nous remarquons que plus un signal à reconstruire est bien séparé (une grande distance entre chaque pair d'impulsions), moins grande est la sur-paramétrisation du signal initialisé. Avec la variété d'expériences introduites, nous notons que OP-COMP + PGD produit de bonnes estimations dans n'importe quel cas. Avec une contrainte de séparation élevée ou faible, avec un petit ou grand nombre d'impulsions, avec une structure explicite des positions des impulsions (micro-tubules), les positions et amplitudes estimées sont très proches de celles de x_0.

Cependant, nous notons que notre nouvelle méthode OP-COMP avec PGD introduit de nouveaux paramètres supplémentaires qu'il faut soit estimer, soit définir à l'avance manuellement. Nous pouvons citer la distance de projection utilisée pour fusionner les impulsions trop proches les unes des autres ou alors le facteur de sur-paramétrisation maximal. Comme ouverture, nous suggérons de travailler sur des méthodes d'estimation automatique des ces paramètres afin de simplifier les démarches et formaliser ces méthodes.

Dans les chapitres suivants, nous nous attarderons sur des méthodes d'accélérations qui utilisent les spécificités de notre modèle afin réduire le temps de calcul pour la reconstruction de signaux parcimonieux.

Chapitre 6

Méthodes de descente par blocs appliquées à la descente de gradient projetée

Dans ce chapitre, nous explorons les méthodes de descente par blocs et leur application au problème de reconstruction de signaux parcimonieux. Il s'agit d'une catégorie d'algorithmes et méthodes qui se base sur la structure d'un problème pour procéder à une découpe de ce problème en blocs afin de traiter chaque bloc séparément et accélérer la résolution du problème global. Dans un premier temps, nous introduisons de manière générale la descente par blocs. Puis nous développons cette idée pour l'intégrer à la reconstruction de signaux parcimonieux en exploitant leur structure de petite dimension (petit nombre de paramètres avec uniquement les amplitudes et les positions des impulsions). Nous énonçons également un Lemme afin de justifier ce découpage en blocs et ainsi la réduction de notre problème de reconstruction de nombreuses impulsions en plusieurs problèmes de reconstruction d'une seule impulsion. Enfin, nous procédons à quelques expériences avec différents opérateurs linéaires afin de tester l'efficacité de cette nouvelle méthode à OP-COMP + PGD.

6.1 Introduction à la descente par blocs

Pour les méthodes de descente de gradient, une idée pour accélérer le temps de calcul est la descente par blocs ou *block coordinate descente* (BCD). Elle consiste à n'effectuer des itérations de la descente que sur un sous-ensemble des coordonnées des paramètres que nous estimons.

L'idée d'une structure sous-jacente pour séparer le signal en blocs est courante dans de nombreuses applications (voir par exemple [80, 92]). Parmi les nombreuses applications, nous pouvons citer la factorisation de matrices non-négatives [94], la régression logistique parcimonieuse [93]. Le problème du LASSO a également été résolu avec des méthodes de descente par blocs [61, 92].

La définition de la descente par blocs repose sur la stratégie de création des blocs, la règle de sélection des blocs et la règle de mise à jour des blocs [66].

Stratégie de création des blocs. Les blocs peuvent être soit fixés à l'avance, soit variables et recalculés à chaque itération. Les blocs fixes permettent un meilleur contrôle et une meilleure séparation des sous-problèmes [75] lorsque la structure du problème est connue avant les calculs [62]. Les blocs variables permettent une plus grande liberté, par exemple en créant des blocs basés sur des méthodes gourmandes [90] telles que Gauss-Southwell.

Règle de sélection des blocs. Comme le choix du bloc à mettre à jour peut conduire à différents minima, il s'agit d'un domaine d'intérêt majeur pour les méthodes de descente par blocs [80]. Les idées de base telles que les variantes cycliques pour le choix de l'indice sont courantes car il s'agit de méthodes déterministes dont la convergence est prouvée [8]. Du point de vue de la mise en œuvre, elles sont simples et ne nécessitent pas de surcharge de calcul. Au lieu de parcourir tous les blocs dans un ordre cyclique, on peut utiliser des méthodes aléatoires en choisissant les blocs en fonction d'une distribution uniforme [65], d'une distribution pondérée [2] ou simplement à partir de règles arbitraires [73]. Les méthodes gourmandes sont également utilisées dans le processus de sélection des blocs. Elles choisissent le bloc le plus approprié à chaque itération en résolvant un problème d'optimisation [66]. Une méthode gourmande bien connue, pour son efficacité théorique et pratique et sa simplicité, est la règle de Gauss-Southwell (GS) qui sélectionne un bloc sur la base de la norme de son gradient,

$$b_k \in \arg\max_{b \in \mathcal{B}} \| \left[\nabla g(\theta^k)\right]_b \|_2. \tag{6.1}$$

où $\mathcal{B}$ est l'ensemble de tous les blocs. D'autres méthodes gourmandes telles que Maximum Improvement [27] ou GS-Lipschitz [67] sont également des choix valables comme règles de sélection des blocs. Les méthodes gourmandes sont plus performantes que les méthodes cycliques ou aléatoires, mais au prix d'un coût de calcul plus élevé [67].
Règle de mise à jour des blocs Le rôle de cette règle est de faire en sorte que le bloc choisi diminue au maximum la fonction objective (comme (1.13)). Il est important d'utiliser une règle de mise à jour des blocs adaptée au problème. En commençant par la descente de gradient classique, nous pouvons définir la taille du pas de différentes manières (fixe, recherche linéaire approximative ou taille de pas optimale) pour améliorer les performances de la descente [17]. De nombreuses autres méthodes de descente utilisées dans l'optimisation convexe et non-convexe sont utilisées dans la descente par blocs. Parmi les nombreuses méthodes introduites, nous pouvons utiliser descente proximale [4], le gradient stochastique [64] en fonction, par exemple, de la géométrie de la fonction objectif. Dans notre cas, nous souhaitons améliorer la descente du gradient projetée à partir de OP-COMP avec PGD en utilisant une méthode de FISTA avec relance pour la règle de mise à jour des blocs. Une première utilisation de FISTA pour la méthode de descente par blocs se trouve dans [71]. Bien qu'il n'y ait pas de garanties théoriques pour la convergence de FISTA dans les méthodes de descente par blocs, elle fonctionne bien dans divers cas [26].

6.2 L'algorithme de descente de gradient projetée par blocs

Dans cette section, nous motivons l'idée d'utiliser un système de blocs pour diviser notre problème en une somme de problèmes plus petits et plus simples. En outre, nous décrivons notre méthode : la descente de gradient projetée par blocs. Cette descente n'opère que sur les paramètres des impulsions du signal estimé qui n'ont pas encore convergé.

6.2.1 Un résultat pour guider la sélection des blocs

La descente par blocs est un moyen d'accélérer le temps de calcul des algorithmes de descente en utilisant la structure sous-jacente du problème. Dans notre cas, plus il y

a d'impulsions dans le signal à reconstruire, plus il y a de paramètres à calculer dans le processus de descente. Comme notre modèle contient une contrainte de séparation, chaque impulsion à récupérer est estimée par une ou plusieurs impulsions initialisées indépendamment des autres vraies impulsions. Cela conduit à deux cas : soit il y a une impulsion initialisée estimant une impulsion du signal à reconstruire, soit il y a un groupe d'impulsions initialisées. Pour les estimations d'une seule impulsion, la convergence vers l'impulsion à reconstruire associée est rapide. Au contraire, pour les groupes d'impulsions initialisées, leur convergence est beaucoup plus lente vers une vraie impulsion (voir Chapitre 4, [16]). Nous proposons de ne calculer que les gradients des impulsions qui n'ont pas encore convergé. Ensuite, nous montrons qualitativement pourquoi nous pouvons utiliser notre modèle pour séparer la reconstruction de chaque impulsion en utilisant notre contrainte de séparation avec un signal initialisé idéalisé.

Soit $x_0 = \sum_{i=1}^K a_i \delta_{t_i} \in \Sigma_{K,\epsilon}$ le signal à reconstruire et $x = \sum_{j=1}^K b_j \delta_{s_j}$ le signal initialisé obtenu avec COMP. Supposons que x est bien initialisé, c'est à dire que pour tout $i \in \{1, \dots, K\}, \|t_i - s_i\|_2 < \frac{\epsilon}{3}$. En calculant la différence des deux signaux, nous obtenons

$$x_0 - x = \sum_{i=1}^K (a_i \delta_{t_i} - b_i \delta_{s_i}) = \sum_{i=1}^K \nu_i \tag{6.2}$$

avec $\nu_i = a_i \delta_{t_i} - b_i \delta_{s_i}$ un $\frac{\epsilon}{3}$-dipôle (voir Définition 2).
Le Lemme suivant nous donne des bornes pour mieux interpréter la structure en blocs de notre modèle avec notre initialisation x.

Lemme 6.1. *Soit $x_0 = \sum_{i=1}^K a_i \delta_{t_i} \in \Sigma_{K,\epsilon}$ et $x = \sum_{i=1}^K b_i \delta_{s_i}$ de telle manière à ce que $x_0 - x = \sum_{i=1}^K \nu_i$ soit une somme de $\frac{\epsilon}{3}$-dipôles. Soit A un opérateur linéaire qui envoie de l'espace des mesures sur $\mathbb{R}^d$ vers $\mathbb{C}^m$. Posons $y = Ax_0$ l'observation de x_0 par A. Supposons que l'opérateur A suive l'hypothèse suivante,*

$$\langle A\nu_1, A\nu_2 \rangle \leq \mu_A \|A\nu_1\|_2^2 \|A\nu_2\|_2^2 . \tag{6.3}$$

Alors, nous obtenons

$$\|y - Ax\|_2^2 \leq \sum_{i=1}^K \|A\nu_i\|_2^2 + \mu_A \sum_{\substack{i,j=1 \\ i \neq j}}^K \|A\nu_i\|_2^2 \|A\nu_j\|_2^2 . \tag{6.4}$$

De plus, considérons $g(\theta)$, voir (2.5), la fonctionnelle à minimiser avec Θ l'ensemble des paramètres de x et $\theta \in \Theta$. Soit $\partial_{i,r} g(\theta)$ la dérivée partielle de $g(\theta)$ par rapport à la r-ième coordonnée de s_i. Alors, nous obtenons

$$\left| \partial_{i,r} \|y - Ax\|_2^2 - \partial_{i,r} \|A\nu_i\|_2^2 \right| \leq 2|b_i|(K-1)\mu_A \|\partial_{i,r} A\nu_i\|_2^2 \max_{j \in \{1,\dots,K\}} \|A\nu_j\|_2^2. \tag{6.5}$$

Preuve. Nous commençons par écrire $x_0 - x$ comme une somme de $\frac{\epsilon}{3}$-dipôles $\sum_{i=1}^K \nu_i$. Pour obtenir (6.4), nous développons $\|y - Ax\|_2^2$,

$$\|y - Ax\|_2^2 = \|A(x_0 - x)\|_2^2 = \left\| A \sum_{i=1}^K \nu_i \right\|_2^2 \tag{6.6}$$

$$= \sum_{i=1}^K \|A\nu_i\|_2^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^K \langle A\nu_i, A\nu_j \rangle \tag{6.7}$$

Nous utilisons notre hypothèse (6.3) de manière à pouvoir borner $\langle A\nu_i, A\nu_j \rangle$,

$$\|y - Ax\|_2^2 \leq \sum_{i=1}^{K} \|A\nu_i\|_2^2 + \mu_A \sum_{\substack{i,j=1 \\ i \neq j}}^{K} \|A\nu_i\|_2^2 \|A\nu_j\|_2^2 . \tag{6.8}$$

Ceci termine la démonstration pour la première partie du Lemme. Maintenant, nous cherchons à prouver (6.5). Grâce au résultat de [84, Proposition 2.1], nous avons

$$\partial_{i,r}\|y - Ax\|_2^2 = 2b_i \Re \left\langle A\partial_{i,r}\delta_{s_i}, A\sum_{j=1}^{K} \nu_j \right\rangle \tag{6.9}$$

$$= 2b_i \Re \langle A\partial_{i,r}\delta_{s_i}, A\nu_i \rangle + 2b_i \Re \left\langle A\partial_{i,r}\delta_{s_i}, A\sum_{\substack{j=1 \\ j \neq i}}^{K} \nu_j \right\rangle . \tag{6.10}$$

En utilisant notre hypothèse (6.3), nous obtenons (en utilisant le fait que les dérivées d'une mesure de Dirac sont des dipôles généralisés [84]) :

$$\partial_{i,r}\|y - Ax\|_2^2 \leq \; 2b_i \Re \langle A\partial_{i,r}\delta_{s_i}, A\nu_i \rangle + 2b_i\mu_A \|\partial_{i,r}A\delta_{s_i}\|_2^2 \sum_{\substack{j=1 \\ j \neq i}}^{K} \|A\nu_j\|_2^2. \tag{6.11}$$

Nous bornons chaque $\|A\nu_j\|_2^2$ par leur maximum $\max_{j\in\{1,\dots,K\}} \|A\nu_j\|_2^2$. En outre, nous avons que $2b_i \Re \langle A\partial_{i,r}\delta_{s_i}, A\nu_i \rangle = \partial_{i,r}\|A\nu_i\|_2^2$. Finalement, nous obtenons que

$$\left|\partial_{i,r}\|y - Ax\|_2^2 - \partial_{i,r}\|A\nu_i\|_2^2\right| \leq 2|b_i|\mu_A(K-1)\|\partial_{i,r}A\nu_i\|_2^2 \max_{j\in\{1,\dots,K\}} \|A\nu_j\|_2^2. \tag{6.12}$$

□

En utilisant le résultat (6.4) du Lemme 6.1, nous obtenons que pour minimiser approximativement $\|y - Ax\|_2^2$, une condition suffisante est de minimiser individuellement chaque $\|A\nu_i\|_2^2$, c'est-à-dire de récupérer K signaux distincts chacun composé d'une impulsion.

Pour distinguer les impulsions qui ont convergé des impulsions qui n'ont pas encore convergé, nous comparerons leurs gradients en utilisant le résultat (6.5) du Lemme 6.1. En effet, la différence absolue entre $\partial_{i,r}\|y - Ax\|_2^2$ et $\partial_{i,r}\|A\nu_i\|_2^2$ est petite : Pour les opérateurs d'observation courants (par exemple, le filtre gaussien), le paramètre de cohérence mutuelle μ_A est petit et les termes $\|\partial_{i,r}A\nu_i\|_2^2$ et $\sup_{j\in\{1,\dots,K\}} \|A\nu_j\|2^2$ sont bornés. Dans la pratique, la sur-paramétrisation du signal initialisé produit plus d'impulsions que de dipôles, mais le cas plus simple donné par le Lemme 6.1 suggère que des blocs d'impulsions avec de grands gradients devraient être pris en compte pour la stratégie de sélection.

6.2.2 Description de l'algorithme

Notre nouvel algorithme : la descente de gradient projetée par blocs (Algorithme 4) est une amélioration de la descente de gradient projeté (PGD) [86], qui effectue une descente sur tous les paramètres tout en imposant une contrainte de séparation sur les positions des impulsions estimées. Nous conservons cette caractéristique principale et nous intégrons notre accélération de descente par blocs. Notre méthode est dérivée de la règle de Gauss-Southwell. Nous choisissons nos blocs dans l'ensemble des blocs $\mathcal{B}_{\texttt{super-res}}$, où chaque bloc représente une impulsion, c'est à dire $\mathcal{B}_{\texttt{super-res}} \coloneqq \{b_i = (a_i, t_i), i = 1, \dots, K\}$. À

chaque itération de la boucle principale, nous calculons la norme des gradients ($\|\nabla[g(\theta)]_i\|$) associés à chaque impulsion du signal estimé avec $\nabla[g(\theta)]_i := \nabla g((a_i, t_i))$, c'est-à-dire les gradients pour l'amplitude et la position de la i-ième impulsion. Ensuite, nous créons deux ensembles I et J. L'ensemble $I = \{i, \|\nabla[g(\theta)]_i\|_2 \geq \texttt{threshold} \times \max_j \|\nabla[g(\theta)]_j\|_2\}$ représente toutes les impulsions qui n'ont pas encore convergé. En pratique, nous fixons manuellement la valeur de `threshold` et nous discutons de ce choix plus en détail dans la section suivante. Nous fixons $J = \{1, \ldots, K\} \backslash I$. Nous notons $x_I = \sum_{i \in I} a_i \delta_{s_i}$ et $x_J = \sum_{j \in J} a_j \delta_{s_j}$.

Avec seulement les impulsions sélectionnées de I, nous effectuons une descente de gradient projetée sur x_I pour un nombre fixe d'itérations $\texttt{m}_{\texttt{it}}$ (ou jusqu'à convergence). Notez que l'observation dans cette descente partielle projetée y_I est $y - Ax_J$, c'est-à-dire l'observation originale mise à jour avec les impulsions non sélectionnées de cette itération. Si une projection se produit au cours de cette descente, nous arrêtons la descente du gradient projeté. Nous fusionnons les deux signaux x_I et x_J et nous calculons à nouveau I et J. Nous poursuivons ce processus jusqu'à la convergence de notre signal estimé x. Pour effectuer la descente, nous choisissons l'algorithme de FISTA avec relance [68] car, dans la pratique, il est plus performant qu'une descente de gradient classique et offre une convergence rapide.

Algorithm 4 Descente de Gradient Projetée par Blocs.

procedure BCD($y, A, \theta^{(0)}, \texttt{n}_{\texttt{it}}, \texttt{m}_{\texttt{it}}\texttt{threshold}$)
  $n \leftarrow 0$
  $k \leftarrow \#\texttt{spikes}$
  while $n < n_{it}$ **do**
    $I = \{i, \left\|\nabla[g(\theta^{(n)})]_i\right\|_2 \geq \texttt{threshold} \times \max_j \left\|\nabla[g(\theta^{(n)})]_j\right\|_2\}$
    $J = \{1, \ldots, k\} \setminus I$
    $\tilde{\theta}^{(n)} \leftarrow \left\{\theta_i^{(n)}, i \in I\right\}, \quad \bar{\theta}^{(n)} \leftarrow \left\{\theta_j^{(n)}, j \in J\right\}$
    $\tilde{y} \leftarrow y - A\phi(\bar{\theta}^{(n)})$
    $\theta^{*(n)} \leftarrow \texttt{projected_descent}(\tilde{y}, \tilde{\theta}^{(n)}, \texttt{m}_{\texttt{it}})$
    $\theta^{(n+1)} \leftarrow \left\{\theta_i^{*(n)} \text{ if } i \in I \text{ else } \bar{\theta}_i^{(n)}\right\}_{i=1,\ldots,k}$
    $k \leftarrow \#\texttt{spikes}$
  end while
  return $\theta^{(n_{\texttt{it}})}$
end procedure

Le Diagramme 6.1 représente une itération de la descente de gradient projetée par blocs. En l'observant plus en détails, nous remarquons que la nouvelle observation $\tilde{y}$ ne prend pas en compte l'information liée aux impulsions déjà estimées par $\bar{\theta}$. Visuellement, pour un opérateur linéaire en microscopie, c'est comme si l'on supprimait la réponse impulsionnelles de certaines impulsions pour diminuer le nombre restant d'impulsions à reconstruire.

6.3 Expériences

Dans cette section, nous comparons notre nouvelle descente de gradient projetée par blocs avec l'état de l'art : la descente de gradient projeté (PGD). Nous comparons nos

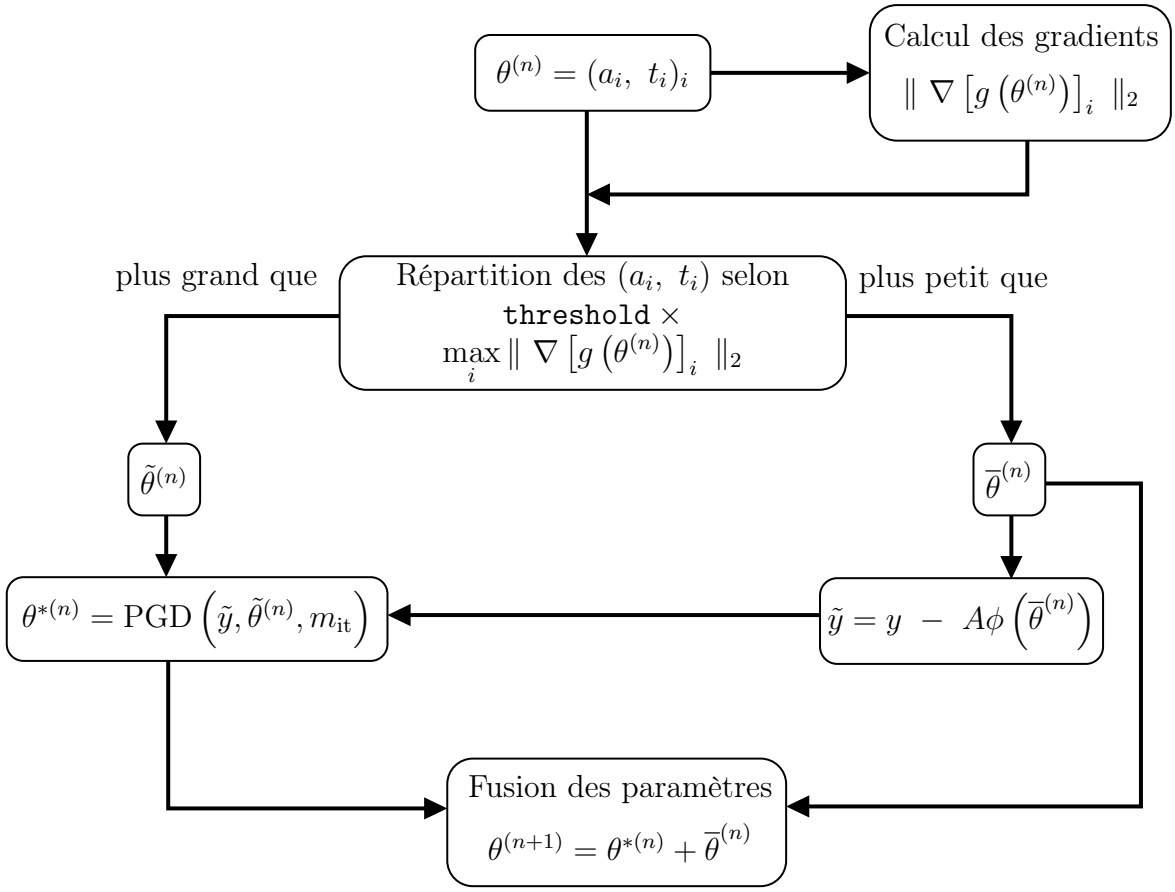
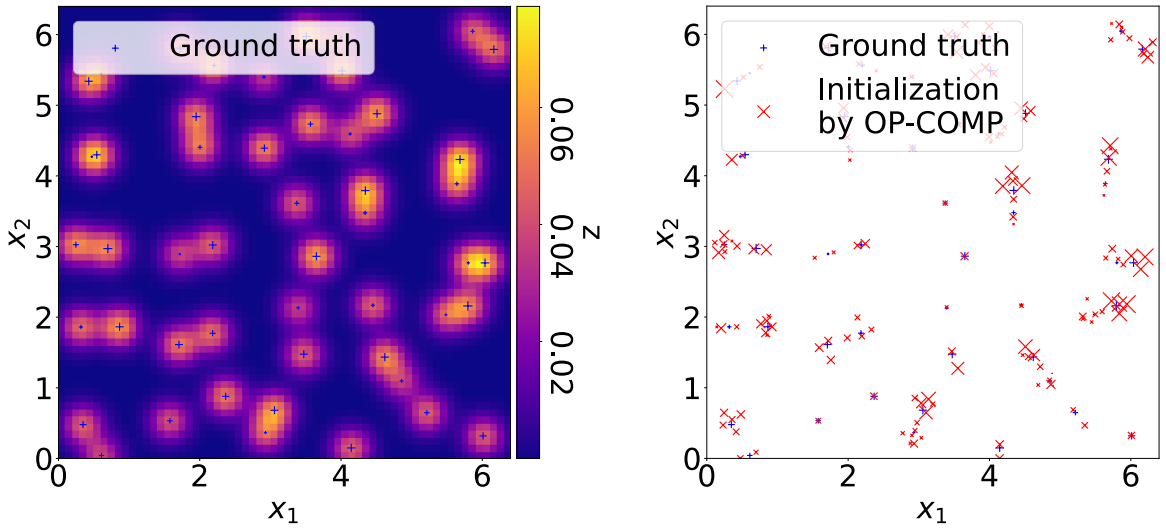


FIGURE 6.1 – Diagramme de fonctionnement d'une itération de la Descente de Coordonnées par Blocs



(a) (b)

FIGURE 6.2 – (a) Vérité terrain x_0 et un plan de son observation y par MA-TIRF, (b) Vérité terrain x_0 et le signal initialisé x_{init} par OP-COMP.

algorithmes avec deux opérateurs linéaires différents : MA-TIRF et l'échantillonnage de Fourier.

6.3.1 Comparaison entre BCD et PGD avec MA-TIRF

Le signal à reconstruire dans chaque expérience est composé de 50 mesures de Dirac pour MA-TIRF en 3 dimensions [37]. Comme ces méthodes visent à récupérer un signal à partir d'une estimation initiale, nous utilisons notre méthode COMP sur-paramétrée (OP-COMP) pour l'initialisation du signal.

L'opérateur linéaire utilisé pour les observations est l'opérateur MA-TIRF [37]. Il est utilisé en microscopie pour récupérer des cellules dans des images 3D. Les positions t sont dans un domaine de taille 6.4 µm×6.4 µm×0.8 µm. L'observation est composée de 4 plans qui représentent différents angles de focalisation. Chaque plan est composé de 64×64 pixels. Au total, $\dim(y) = 64 \times 64 \times 4 = 16\,384$. Les amplitudes suivent une distribution uniforme $U([1, 2])$. Le code de ces expériences peut être téléchargé à l'adresse [12].

Dans la Figure 6.2a, nous présentons un plan focal de l'observation. Il correspond à une image prise sous un certain angle. Bien qu'il s'agisse d'une image en 2 dimensions, la combinaison de plusieurs angles est utile pour reconstruire les impulsions en 3D. Nous voyons sur cette figure que la plupart des impulsions sont bien séparés. Dans la Figure 6.2b, nous avons le signal à reconstruire x_0 et le signal initialisé par OP-COMP. Il est à noter que chaque impulsion de x_0 a au moins une impulsion initialisée proche d'elle. De plus, la sur-paramétrisation est clairement visible car le nombre d'impulsions initialisées est supérieur au nombre d'impulsions à reconstruire.

Dans la Figure 6.3a, nous observons que les impulsions sont bien reconstruites par PGD comme prévu. Dans la Figure 6.3c, nous montrons la norme du résidu $\|y - Ax\|_2^2$ à travers les itérations pour PGD (en bleu) et BCD (en rouge). Nous arrêtons la descente lorsque la norme du résidu devient inférieure à 2×10^{-8}. Pour PGD, la norme du résidu diminue et atteint son minimum à 2×10^{-8}. Les sauts soudains dans la norme du résidu sont dus à la projection utilisée.

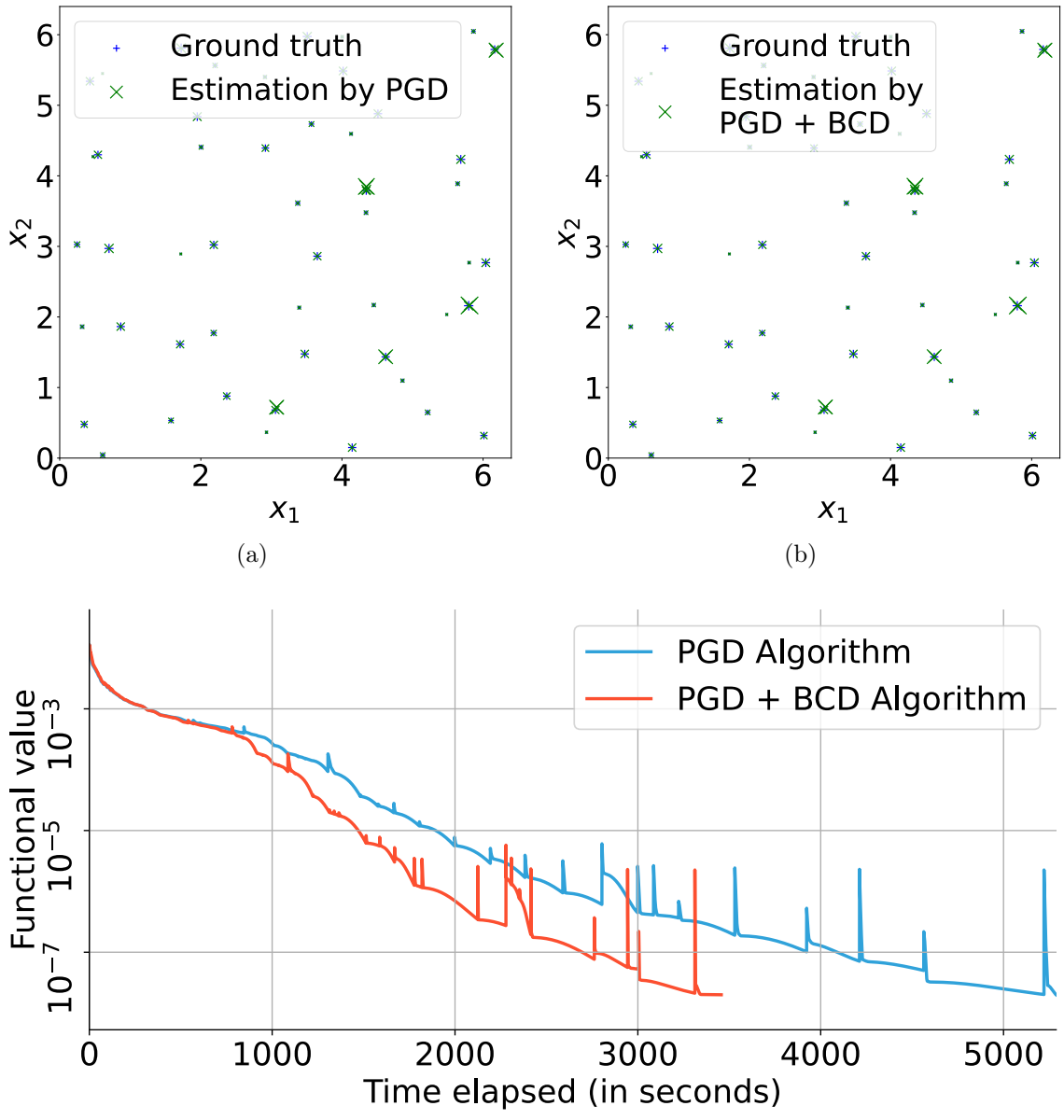
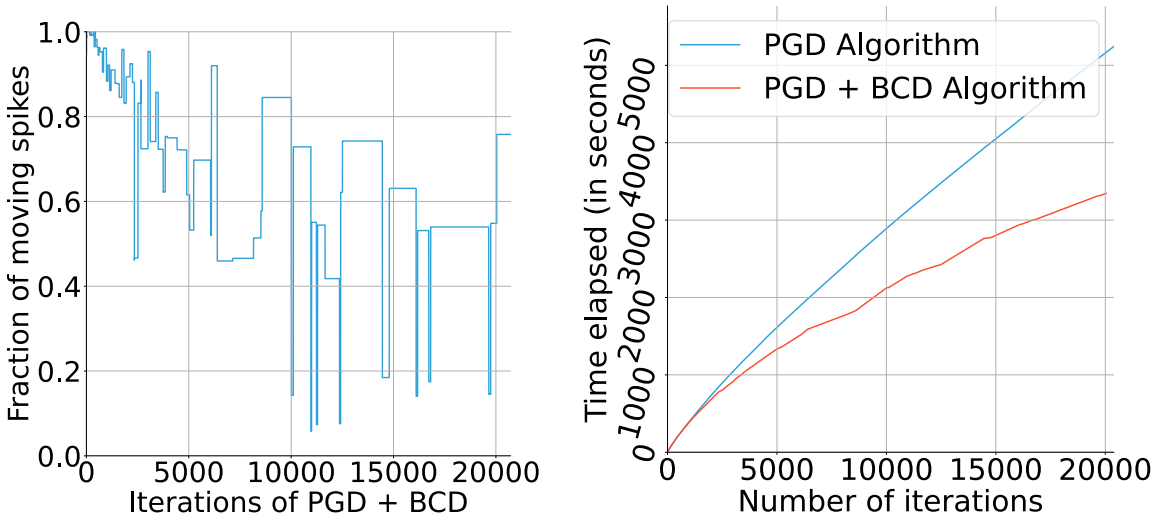


FIGURE 6.3 – (a) Vérité terrain x_0 et le signal estimé x_{esti} par PGD avec MA-TIRF, (b) Vérité terrain x_0 et le signal estimé x_{esti} par BCD avec MA-TIRF, (c) Comparaison des normes des résidus entre PGD et BCD en fonction du temps de calcul.



(a) (b)

FIGURE 6.4 – (a) Fraction des impulsions dont le gradient est calculé pendant BCD, (b) Temps écoulé pour PGD et BCD en fonction du nombre d'itérations.

Dans la Figure 6.3b, nous obtenons un signal estimé par BCD similaire à celui obtenu avec PGD. Dans la Figure 6.3c (en rouge), comme pour la PGD, nous observons la norme du résidu à travers les itérations de la BCD. Nous remarquons que pour la même norme du résidu, BCD converge beaucoup plus rapidement avec une accélération de près de 35% en temps de calcul.

Alors que le démarrage des deux méthodes prend le même temps pour environ 2 000 itérations, nous observons une nette accélération au fur et à mesure que le temps passe. En effet, comme le montre la Figure 6.4a, la fraction des gradients calculés à chaque itération de BCD diminue. Une bonne valeur empirique pour le choix de `threshold` dans l'Algorithme 4 est 10^{-3}. Cette valeur est utilisée pour toutes nos expériences. Elle offre un bon compromis en termes de minimisation de g (voir l'équation (2.5)) et de temps de calcul. En effet, ce seuil limite le nombre de blocs utilisés pour chaque `projected_descent` en ne sélectionnant que les impulsions non convergées. Au début de la descente, presque tous les gradients sont encore calculés à chaque itération, mais au fil des itérations, la fraction des gradients calculés diminue, d'où un gain de temps de calcul par rapport à la PGD. Les Figures 6.4a et 6.4b révèlent qu'en calculant moins de gradients, nous pouvons gagner du temps sur le calcul total.

Après avoir été testés sur 10 signaux similaires composés de 50 impulsions avec une grande distance entre les impulsions (pour respecter la contrainte de séparation), les résultats restent les mêmes et nous constatons qu'en général, BCD est 35% plus rapide que PGD pour MA-TIRF. Les autres mesures utilisées pour comparer les deux estimations sont le Jaccard, la Précision et le Rappel. Pour ces mesures, il n'y a pas de différence particulière entre les deux méthodes PGD et BCD.

Nous notons que pour d'autres types d'opérateurs linéaires (tels que la l'échantillonnage de la transformée de Fourier), les performances de la BCD peuvent varier, mais elle ne dégrade pas la vitesse de calcul.

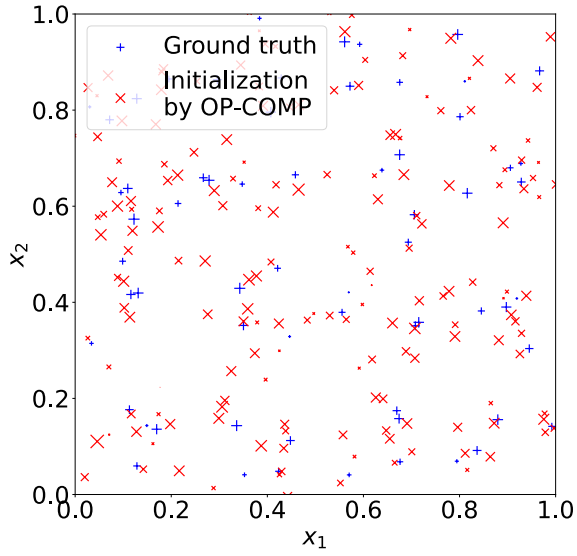


FIGURE 6.5 – Vérité terrain x_0 et le signal initialisé x_{init} par OP-COMP

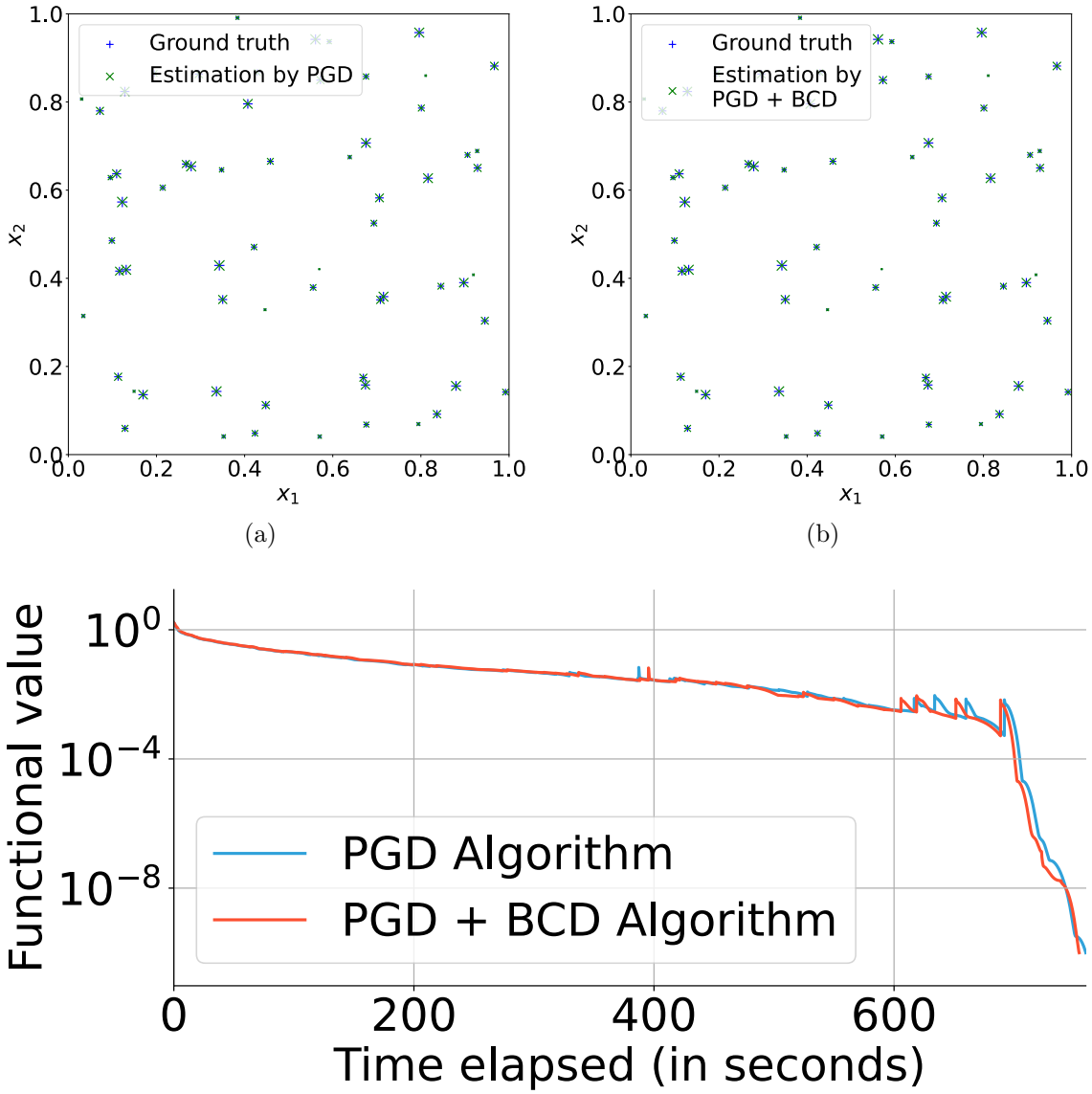
6.3.2 Comparaison entre BCD et PGD avec la transformée de Fourier

Notre but est de reconstruire un signal composé de $K = 64$ impulsions vivant dans $[0, 1]^3$. Dans ce cas, l'opérateur linéaire utilisé pour les observations est la transformée de Fourier. Il produit m mesures de Fourier à partir de fréquences choisies au hasard sur une distribution gaussienne (ou fixées à l'avance). Une mesure obtenue avec la transformée de Fourier et associée à la fréquence ω_l peut être décrite comme $(Ax)_l = \sum_{j=1}^{K} a_j \int_{\mathbb{R}^3} e^{-i\langle \omega_l, t_j \rangle} \mathrm{d}\delta_{t_j}$. Plus il y a de fréquences, meilleure est l'observation. Pour ces expériences, nous prenons $m = 20 \times K = 1\,280$. Les amplitudes suivent une distribution uniforme $U([1, 2])$.

Dans la Figure 6.5, nous observons que l'initialisation (en rouge) par OP-COMP donne une bonne première estimation du signal à reconstruire (en bleu). Le facteur de sur-paramétrisation dans cet exemple est d'environ 3 ($K_{\text{op}} \approx 3 \times K$), ce qui signifie qu'il y a environ 190 impulsions initialisées pour 64 de vraies impulsions.

Dans la Figure 6.6a, nous observons que toutes les impulsions ont été reconstruites. Nous le confirmons dans la Figure 6.6c, où la norme du résidu pour PGD (en bleu) a diminué en dessous de 10^{-10}. De même pour BCD, dans les Figures 6.6b et 6.6c (en rouge), nous observons que tous les impulsions ont été reconstruites. Cependant, contrairement aux expériences avec MA-TIRF, les deux algorithmes convergent rapidement (en moins de 800 secondes contre plus de 5000 secondes pour PGD avec MA-TIRF). Dans le cas de la transformée de Fourier en tant qu'opérateur linéaire, il n'y a pratiquement aucune différence entre BCD et PGD. Il semble que si le signal est bien initialisé et que la descente est rapide, notre méthode n'est pas nécessaire et n'améliore pas le temps de calcul.

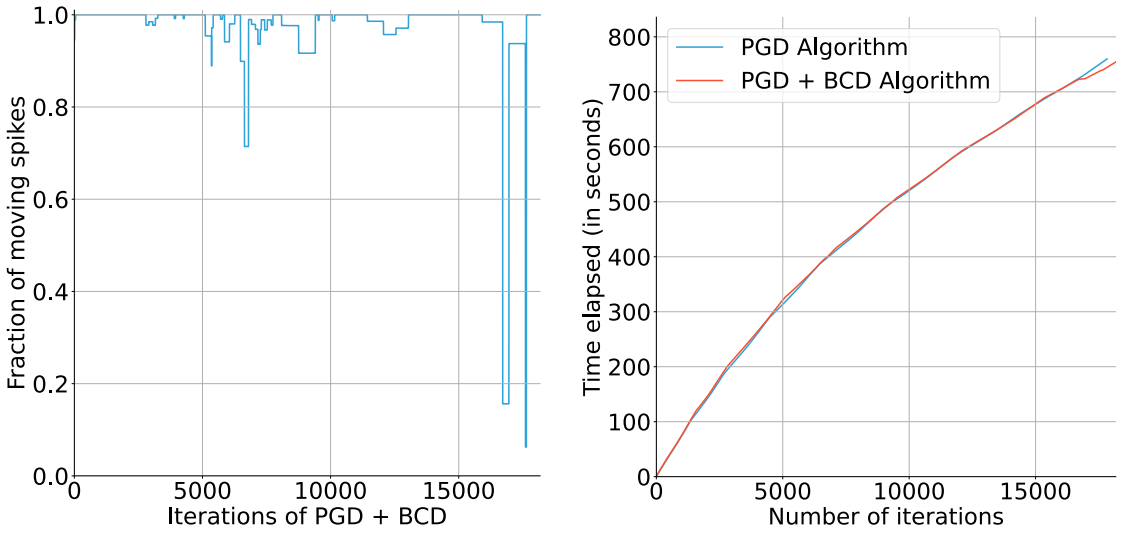
Dans la Figure 6.7b, nous observons que le temps de calcul des deux méthodes est identique. Comme dans la Figure 6.7a, pour la majorité des itérations, tous les gradients sont calculés pendant la descente. Pour certaines itérations, un faible pourcentage des gradients n'est pas calculé. D'une part, le temps gagné en ne calculant pas ces gradients est perdu avec le calcul de la sélection des blocs. D'une autre part, le calcul du gradient avec la transformée Fourier est beaucoup plus rapide qu'avec MA-TIRF. Cela est dû au plus petit nombre de mesures dans Fourier et à l'opérateur linéaire moins complexe.



(a) (b)

(c)

FIGURE 6.6 – (a) Vérité terrain x_0 et le signal estimé x_{esti} par PGD avec la transformée de Fourier, (b) Vérité terrain x_0 et le signal estimé x_{esti} par BCD avec la transformée de Fourier, (c) Comparaison des normes des résidus entre PGD et BCD en fonction du temps de calcul.



(a) (b)

FIGURE 6.7 – (a) Fraction des impulsions dont le gradient est calculé pendant BCD, (b) Temps écoulé pour PGD et BCD en fonction du nombre d'itérations.

6.4 Résumé du chapitre et conclusions

En conclusion de ce chapitre, nous observons que la structure en blocs pour la reconstruction des signaux parcimonieux permet d'améliorer le temps de calcul des algorithmes de descente. En effet, dans un contexte réaliste de microscopie, le temps gagné par BCD est de 35% par rapport aux méthodes précédentes. Pour les travaux futurs, nous pouvons essayer d'autres méthodes gourmandes pour une meilleure sélection de blocs. En outre, il est possible de paralléliser la descente sur les différents blocs, ce qui constitue un avantage majeur des méthodes de descente par blocs.

Contrairement à notre nouvelle méthode d'accélération décrite dans ce chapitre qui ne peut s'appliquer qu'à la reconstruction de signaux parcimonieux (de part leur modélisation en impulsions avec amplitudes et positions distinctes et séparables), il existe d'autres méthodes d'accélérations qui peuvent s'ajouter de manière indépendante à la méthode de descente par blocs.

Dans le chapitre suivant, nous nous attardons sur une deuxième méthode d'accélération de OP-COMP avec PGD qui cherche cette fois-ci à simplifier les calculs en lien avec l'opérateur linéaire utilisé.

Chapitre 7

Méthodes de "sketching" pour la super-résolution

Dans ce chapitre, nous étudions les méthodes de "sketching". Ces méthodes cherchent à créer un "sketch" (une compression) de l'observation y pour diminuer la charge de calculs des algorithmes. Cette diminution vient avec un certain coût, en effet ce sketching de l'opérateur linéaire peut dégrader l'opérateur et donc ajouter une imprécision supplémentaire dans la résolution du problème.

Nous verrons alors dans un premier lieu l'état de l'art des méthodes de sketching dans la littérature. Puis nous chercherons à appliquer ces méthodes pour le cas de la super-résolution avec un opérateur linéaire gaussien. Nous accompagnons notre nouvelle méthode de sketching avec un Lemme appliquant un contrôle sur la solution théorique donnée par une méthode de sketching. Ensuite, nous proposons un exemple de sketching d'un opérateur linéaire dérivé du cas de la microscopie. Nous utilisons ce nouvel opérateur linéaire compressé afin de comparer notre méthode de sketching avec un opérateur linéaire classique dans la méthode OP-COMP avec PGD.

7.1 Introduction au sketching

Avec nos méthodes d'estimation d'impulsions, le coût de calcul dépend du nombre de mesures m. En particulier, cela pose un problème dans le contexte de la SMLM où, pour certains opérateurs linéaires, les mesures sont une collection d'images de microscope échantillonnées sous différents angles. Le nombre de mesures est alors beaucoup plus important que le nombre de paramètres que nous devons estimer. Une idée consiste à compresser les mesures en un "sketch" afin d'accélérer le temps de calcul tout en récupérant le signal de base avec une bonne précision.

Issues de l'acquisition compressée, les méthodes de sketching permettent des opérations rapides sur de grandes quantités de données. Récemment, Gribonval et coll. ont proposé un cadre de sketching sous la forme d'un problème linéaire inverse dans l'espace des mesures [53, 54, 55]. Il a été appliqué avec succès pour estimer les sommes des mesures de Dirac [58], les paramètres des mélanges de Gaussiennes [59, 78, 79, 76], les paramètres des réseaux neuronaux définissant la densité sous-jacente d'une base de données [77, 79] ou la localisation de sources [47, 48].

Nous proposons d'adapter le cadre du sketching à un contexte SMLM où l'opérateur de mesure dans l'espace des mesures est déjà conçu.

Dans ce chapitre, notre principale contribution est une méthode accélérée pour la reconstruction de signaux parcimonieux avec COMP sur-paramétré (OP-COMP) avec une

descente de gradient projetée (PGD) et applicable à des algorithmes similaires comme COMP avec descente et Frank-Wolfe avec descente. Nous proposons de compresser les mesures de l'observation y en approximant un opérateur de sketching idéal dans l'espace des mesures, ce qui conduit à une réduction effective du nombre de mesures m. Cette méthode améliore considérablement les temps de calcul globaux de OP-COMP avec PGD. En effet, nous fournissons des expériences synthétiques dans le contexte de la microscopie qui montrent une amélioration jusqu'à 80% du temps de calcul par rapport à l'opérateur linéaire sans compression sans perte de précision.

7.2 L'application du sketching à la super-résolution pour le SMLM

Comme évoqués quand le Chapitre 5, nous appliquons nos algorithmes dans des problèmes en microscopie tels que le SMLM ou l'étalonnage à partir de micro-billes fluorescentes. Pour ce faire, nous utilisons le modèle MA-TIRF (5.1) pour simuler les réponses impulsionnelles de cellules observées avec un microscope. Il existe d'autres modèles tels que l'astigmatisme ou le modèle à double hélices.

Tous ces modèles, dans le cas de la reconstruction d'impulsions en trois dimensions peuvent être approximés avec l'opérateur dont les réponses impulsionnelles à chaque pixel sont modélisés par

$$\alpha_{i,k}(t) = \frac{f_k(z)}{2\pi\sigma_x\sigma_y} \times \mathrm{e}^{-\left(\frac{(x_i-x)^2}{2\sigma_x^2}+\frac{(y_i-y)^2}{2\sigma_y^2}\right)} \tag{7.1}$$

avec $t = (x, y, z) \in \mathbb{R}^3$, $\sigma_x, \sigma_y \in \mathbb{R}^+$. Les couples $(x_i, y_i) \in \Gamma \subset \mathbb{R}^2$ sont les positions du i-ième pixel sur une grille finie Γ à un certain angle d'incidence donné k, c'est à dire que nous considérons une approximation gaussienne de la réponse impulsionnelle du microscope. Les fonctions $f_k : \mathbb{R} \mapsto \mathbb{R}$ prennent en entré la profondeur (la composante z des positions t) et modélisent l'impact de l'angle d'incidence. Nous pouvons citer l'opérateur linéaire MA-TIRF (multi-angle total internal reflection fluorescence) [37] qui peut être approximé par (7.1).

7.2.1 Compression de mesures gaussiennes dans l'espace des mesures

L'objectif de l'opérateur de sketching est de compresser les mesures tout en conservant l'information pertinente de l'observation originale y. Grâce au cadre de Gribonval et coll. [53], il peut être défini sur l'espace de mesures de dimension infinie $\mathcal{M}$. Considérons un processus de mesure gaussien 2D idéal A^{id} défini par

$$\alpha_i^{id}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y}\mathrm{e}^{-\left(\frac{(x_i-x)^2}{2\sigma_x^2}+\frac{(y_i-y)^2}{2\sigma_y^2}\right)} \tag{7.2}$$

pour tout $(x_i, y_i) \in \mathbb{R}^2$, c'est à dire que nous avons accès à la convolution complète d'une mesure de Dirac avec une réponse impulsionnelle gaussienne. Nous pouvons alors considérer un opérateur de *sketching* $\mathcal{S} : L^2(\mathbb{R}^2) \to \mathbb{C}^{m_\mathcal{S}}$ de l'espace des fonctions de carré intégrable vers l'espace $\mathbb{C}^{m_\mathcal{S}}$ de dimension réduite $m_\mathcal{S}$. Nous appelons $\mathcal{S}A^{id}$ l'opérateur idéal *compressé* correspondant. Une manière de compresser un opérateur linéaire A^{id} est de prendre des mesures de Fourier aléatoires avec un nombre fini de fréquences $m_\mathcal{S}$

bien choisies $\omega_l \in \mathbb{R}^2$. Cette opération, pour une fréquence donnée ω_l et pour tout $x = \sum_{r=1}^{K} b_r \delta_{s_r}$ s'écrit comme

$$(\mathcal{S}(A^{id}x))_l = [\mathcal{F}(A^{id}x)](w_l) = \sum_{r=1}^{K} b_r [\mathcal{F}(A^{id}\delta_{s_r})](w_l) \tag{7.3}$$

où $\mathcal{F}$ est la transformée de Fourier continue, c'est à dire

$$[\mathcal{F}(A^{id}\delta_{s_r})](w_l) = \int_{(x_i,y_i)\in\mathbb{R}^2} \alpha_i(s_r) e^{-j\langle \omega_l,(x_i,y_i)\rangle} dx_i dy_i. \tag{7.4}$$

Il a été montré dans [54] qu'il est possible de récupérer les positions et les amplitudes de gaussiennes suffisamment séparées avec un choix de fréquences gaussiennes aléatoires avec $m_{\mathcal{S}} \gtrapprox \mathcal{O}(K^2 d \operatorname{polylog}(K,d))$ fréquences en minimisant, à partir du problème compressé

$$g_{\mathcal{S}}(\theta) := \|\mathcal{S}A^{id}\phi(\theta) - \mathcal{S}A^{id}x_0\|_2^2 \tag{7.5}$$

sous une contrainte de séparation. Malheureusement, nous n'avons pas accès à des opérateurs de mesures gaussiennes idéaux A^{id} dans la pratique : ils sont échantillonnés sur une grille. Cependant, dans la méthode SMLM, cette grille est généralement choisie pour satisfaire les limites de Shannon-Nyquist des instruments de mesure, c'est-à-dire qu'il y a peu de perte d'information par le processus de discrétisation.

Bien que nous puissions calculer pour toute estimation $x = \phi(\theta)$ les mesures idéales compressées $\mathcal{S}A^{id}\phi(\theta)$, nous ne pouvons pas calculer exactement $\mathcal{S}A^{id}x_0$ car seul Ax_0 est disponible. Au lieu de cela, nous approximons $\mathcal{S}A^{id}x_0$ avec une version discrétisée appelée $\mathcal{S}_d y = \mathcal{S}_d A x_0$ directement obtenue à partir des mesures de y. Nous proposons de minimiser la fonction approchée

$$g_{\mathcal{S}_d}(\theta) := \|\mathcal{S}A^{id}\phi(\theta) - \mathcal{S}_d y\|_2^2. \tag{7.6}$$

L'opérateur de mesure approximé $\mathcal{S}_d$ effectue une *transformée de Fourier à temps discret tronquée* de y sur la grille Γ :

$$(\mathcal{S}_d y)_l = \frac{1}{|\Gamma|} \sum_{i=1}^{|\Gamma|} y_i \mathrm{e}^{-j\langle \omega_l,(x_i,y_i)\rangle} \tag{7.7}$$

avec (x_i, y_i) la position du i-ième pixel de la grille Γ, y_i sa mesure associée et $|\Gamma|$ le nombre total de pixels dans Γ. Notons que pour calculer $\mathcal{S}_d y$, nous itérons sur la grille Γ, ce que nous voulions éviter à l'origine. Cependant, ce calcul de $\mathcal{S}_d y$ ne doit être effectué qu'une seule fois. Cela a donc un faible impact sur le temps de calcul total par rapport à l'utilisation de l'opérateur linéaire classique A dans le processus d'optimisation.

L'introduction d'une approximation $\mathcal{S}_d$ de l'opérateur de sketching $\mathcal{S}$ pose le problème que $\arg\min_{\theta\in\Theta} g_{\mathcal{S}_d} \neq \arg\min_{\theta\in\Theta} g_{\mathcal{S}}$. Grâce au Lemme 7.1 suivant, nous garantissons que le problème compressé idéal (où $A^{id}x_0$ serait disponible) est approximativement minimisée par la solution de la fonctionnelle du problème compression approximée (7.6).

Lemme 7.1. *Soit T_Γ l'opérateur d'échantillonnage linéaire sur une grille Γ. Soit A^{id} un opérateur de mesure idéal et $A = T_\Gamma A^{id}$ sa discrétisation sur une grille Γ. Soit $y = A\phi(\theta_0)$ une observation d'un signal $\phi(\theta_0)$ sur Γ. Soit $\mathcal{S}$ et $\mathcal{S}_d$ un opérateur de sketching et son approximation discrète sur la grille Γ.*

Supposons que

$$\|\mathcal{S} - \mathcal{S}_d T_\Gamma\|_{\mathrm{op}}^{A^{id},\Sigma} := \sup_{x\in\Sigma} \frac{\|\mathcal{S}A^{id}x - \mathcal{S}_d T_\Gamma A^{id}x\|_2}{\|A^{id}x\|_2} \leq \eta \tag{7.8}$$

et soit $\theta^* \in \arg\min_{\theta\in\Theta} g_{\mathcal{S}_d}(\theta)$. *Alors nous avons que pour tout* $\theta \in \Theta$,

$$\sqrt{g_{\mathcal{S}}(\theta^*)} \leq \sqrt{\min_{\theta\in\Theta} g_{\mathcal{S}}(\theta)} + 2\eta\|y\|_2. \tag{7.9}$$

Preuve. Soit $\theta^* \in \arg\min_{\theta\in\Theta} g_{\mathcal{S}_d}(\theta)$. Nous avons que pour tout $\theta \in \Theta$

$$\begin{aligned}\sqrt{g_{\mathcal{S}}(\theta^*)} &= \|\mathcal{S}A^{id}\phi(\theta^*) - \mathcal{S}A^{id}\phi(\theta_0)\|_2 &(7.10)\\ &= \|\mathcal{S}A^{id}\phi(\theta^*) - \mathcal{S}_dA\phi(\theta_0) + \mathcal{S}_dA\phi(\theta_0) - \mathcal{S}A^{id}\phi(\theta_0)\|_2 &(7.11)\end{aligned}$$

Avec l'inégalité triangulaire, nous obtenons

$$\begin{aligned}\sqrt{g_{\mathcal{S}}(\theta^*)} &\leq \|\mathcal{S}A^{id}\phi(\theta^*) - \mathcal{S}_dA\phi(\theta_0)\|_2 + \|\mathcal{S}_dA\phi(\theta_0) - \mathcal{S}A^{id}\phi(\theta_0)\|_2 &(7.12)\\ &= \sqrt{g_{\mathcal{S}_d}(\theta^*)} + \|(\mathcal{S} - \mathcal{S}_dT_\Gamma)A^{id}\phi(\theta_0)\|_2 &(7.13)\\ &\leq \sqrt{g_{\mathcal{S}_d}(\theta^*)} + \|\mathcal{S} - \mathcal{S}_dT_\Gamma\|_{\text{op}}^{A^{id},\Sigma}\|y\|_2 &(7.14)\end{aligned}$$

Soit $\tilde{\theta} \in \arg\min_{\theta\in\Theta} g_{\mathcal{S}}(\theta)$, nous avons de façon similaire

$$\begin{aligned}\sqrt{g_{\mathcal{S}_d}(\theta^*)} \leq \sqrt{g_{\mathcal{S}_d}(\tilde{\theta})} &= \|\mathcal{S}A^{id}\phi(\tilde{\theta}) - \mathcal{S}_dA\phi(\theta_0)\|_2 &(7.15)\\ &\leq \|\mathcal{S}A^{id}\phi(\tilde{\theta}) - \mathcal{S}A^{id}\phi(\theta_0)\|_2 + \|\mathcal{S}A^{id}\phi(\theta_0) - \mathcal{S}_dA\phi(\theta_0)\|_2 &(7.16)\\ &\leq \sqrt{g_{\mathcal{S}}(\tilde{\theta})} + \|\mathcal{S} - \mathcal{S}_dT_\Gamma\|_{\text{op}}^{A^{id},\Sigma}\|y\|_2 &(7.17)\end{aligned}$$

Nous en déduisons alors des inégalités (7.14) et (7.17) que

$$\sqrt{g_{\mathcal{S}}(\theta^*)} \leq \sqrt{g_{\mathcal{S}}(\tilde{\theta})} + 2\|\mathcal{S} - \mathcal{S}_dT_\Gamma\|_{\text{op}}^{A^{id},\Sigma}\|y\|_2 \tag{7.18}$$

□

L'hypothèse principale du Lemme 7.1 utilise la norme de l'opérateur pour quantifier la différence entre l'opérateur de compression $\mathcal{S}$ et son approximation $\mathcal{S}_d$. La borne η existe si la discrétisation respecte la condition de Shannon-Nyquist qui peut être approximativement vraie avec un petit η pour des observations gaussiennes, et lorsque $\mathcal{S}A^{id}$ et A^{id} garantissent la reconstruction de x_0 (par exemple, si l'opérateur vérifie une propriété d'isométrie restreinte, voir [53]).

7.2.2 Application à la SMLM

Comme le montre l'expression des fonctions de mesure dans (7.1), plusieurs images 2D représentant différents angles d'incidence sont capturées. Nous proposons d'appliquer l'opération de sketching indépendamment sur chaque angle. Nous donnons une expression explicite de l'opération de sketching en 2D, angle par angle, avec le Lemme 7.2 suivant. Appelons A_k^{id} les mesures gaussiennes idéales correspondant à l'angle d'incidence k. Nous avons juste besoin de la transformée de Fourier continue de $A_k^{id}x_0$ pour implémenter la méthode de sketching en pratique.

Lemme 7.2. *Soit* $\alpha_{i,k}$ *défini par* (7.1), *avec* $t = (x, y, z) \in \mathbb{R}^3$, $\sigma = (\sigma_x, \sigma_y) \in \mathbb{R}^2$. *Alors la Transformée de Fourier de l'opérateur de mesure idéal gaussien est*

$$[\mathcal{F}(A_k^{id}\delta_t)](w_l) = f_k(z)\mathrm{e}^{-\frac{\langle\omega_l^2,\sigma^2\rangle}{2} - j\langle\omega_l,(x,y)\rangle} \tag{7.19}$$

avec $\omega = (\omega_x, \omega_y) \in \mathbb{R}^2$.

Preuve. Ce Lemme est une conséquence directe du fait que la transformée de Fourier d'une gaussienne est une gaussienne. Soit

$$\alpha_{i,k}(t) = \frac{f_k(z)}{2\pi\sigma_x\sigma_y} \times \mathrm{e}^{-\left(\frac{(x_i-x)^2}{2\sigma_x^2}+\frac{(y_i-y)^2}{2\sigma_y^2}\right)} \tag{7.20}$$

l'opérateur linéaire défini plus tôt dans ce chapitre (7.1). Prenons la transformée de Fourier de (7.1),

$$[\mathcal{F}(A_k^{id}\delta_t)](w_l) = \int_{\mathbb{R}^2} \alpha_{i,k}(t)\mathrm{e}^{-j\langle\omega_l,s_i\rangle}\mathrm{d}s_i \tag{7.21}$$

avec $t = (x, y, z) \in \mathbb{R}^3$, $\omega_l = (\omega_{l,x}, \omega_{l,y}) \in \mathbb{R}^2$ et $s_i = (x_i, y_i) \in \mathbb{R}^2$.

$$[\mathcal{F}(A_k^{id}\delta_t)](w_l) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{f_k(z)}{2\pi\sigma_x\sigma_y}\mathrm{e}^{-\left(\frac{(x_i-x)^2}{2\sigma_x^2}+\frac{(y_i-y)^2}{2\sigma_y^2}\right)}\mathrm{e}^{-j\langle\omega_l,s_i\rangle}\mathrm{d}s_i \tag{7.22}$$

$$[\mathcal{F}(A_k^{id}\delta_t)](w_l) = \frac{f_k(z)}{2\pi\sigma_x\sigma_y}\left(\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{(x_i-x)^2}{2\sigma_x^2}}\mathrm{e}^{-j\omega_{l,x}x_i}\mathrm{d}x_i\right)\left(\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{(y_i-y)^2}{2\sigma_y^2}}\mathrm{e}^{-j\omega_{l,y}y_i}\mathrm{d}y_i\right) \tag{7.23}$$

Soit I la première intégrale,

$$I := \int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{(x_i-x)^2}{2\sigma_x^2}}\mathrm{e}^{-j\omega_{l,x}x_i}\mathrm{d}x_i = \int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{(x_i-x)^2+2j\omega_{l,x}x_i\sigma_x^2}{2\sigma_x^2}}\mathrm{d}x_i \tag{7.24}$$

Nous isolons x_i du reste des autres termes dans l'exposant,

$$(x_i-x)^2 + 2j\omega_{l,x}x_i\sigma_x^2 = x_i^2 + 2x_i(j\omega_{l,x}\sigma_x^2 - x) + x^2 \tag{7.25}$$

$$= (x_i + j\omega_{l,x}\sigma_x^2 - x)^2 - (j\omega_{l,x}\sigma_x^2 - x)^2 + x^2 \tag{7.26}$$

Nous injectons ce résultat dans notre intégrale pour obtenir

$$I = \int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{(x_i+j\omega_{l,x}\sigma_x^2-x)^2-(j\omega_{l,x}\sigma_x^2-x)^2+x^2}{2\sigma_x^2}}\mathrm{d}x_i \tag{7.27}$$

$$= \mathrm{e}^{\frac{(j\omega_{l,x}\sigma_x^2-x)^2-x^2}{2\sigma_x^2}}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{(x_i+j\omega_{l,x}\sigma_x^2-x)^2}{2\sigma_x^2}}\mathrm{d}x_i \tag{7.28}$$

$$\tag{7.29}$$

En procédant par substitution dans cette intégrale et en utilisant $x_i' := \frac{x_i+j\omega_{l,x}\sigma_x^2-x}{\sqrt{2}\sigma_x}$, nous obtenons que $\mathrm{d}x_i' = \frac{1}{\sqrt{2}\sigma_x}\mathrm{d}x_i \leftrightarrow \mathrm{d}x_i = \sqrt{2}\sigma_x\mathrm{d}x_i'$. Nous avons alors que

$$I = \sqrt{2}\sigma_x\mathrm{e}^{\frac{(j\omega_{l,x}\sigma_x^2-x)^2-x^2}{2\sigma_x^2}}\int_{-\infty}^{+\infty}\mathrm{e}^{-(x_i')^2}\mathrm{d}x_i' \tag{7.30}$$

En reconnaissant l'intégrale d'une Gaussienne en une dimension qui est égale à $\sqrt{\pi}$,

$$I = \sqrt{2\pi}\sigma_x\mathrm{e}^{\frac{(j\omega_{l,x}\sigma_x^2-x)^2-x^2}{2\sigma_x^2}} \tag{7.31}$$

$$= \sqrt{2\pi}\sigma_x\mathrm{e}^{-\frac{\omega_{l,x}^2\sigma_x^2}{2}-j\omega_{l,x}x} \tag{7.32}$$

Nous notons que nous obtenons un résultat similaire avec l'autre intégrale dans (7.23). En injectant notre nouveau résultat (7.32), nous obtenons,

$$[\mathcal{F}(A_k^{id}\delta_t)](w_l) = \frac{f_k(z)}{2\pi\sigma_x\sigma_y}\left(\sqrt{2\pi}\sigma_x \mathrm{e}^{-\frac{\omega_{l,x}^2\sigma_x^2}{2}-j\omega_{l,x}x}\right)\left(\sqrt{2\pi}\sigma_y \mathrm{e}^{-\frac{\omega_{l,y}^2\sigma_y^2}{2}-j\omega_{l,y}y}\right) \tag{7.33}$$

En simplifiant, nous avons finalement que

$$[\mathcal{F}(A_k^{id}\delta_t)](w_l) = f_k(z)\mathrm{e}^{-\frac{\omega_{l,x}^2\sigma_x^2}{2}-j\omega_{l,x}x}\mathrm{e}^{-\frac{\omega_{l,y}^2\sigma_y^2}{2}-j\omega_{l,y}y} \tag{7.34}$$

$$= f_k(z)\mathrm{e}^{-\frac{\langle\omega_l^2,\sigma^2\rangle}{2}-j\langle\omega_l,(x,y)\rangle}. \tag{7.35}$$

□

Avec le Lemme 7.2 et avec (7.7), nous disposons de tous les outils nécessaires pour récupérer un signal à partir d'observations compressées. En particulier, les gradients de la fonctionnelle compressée sont facilement calculés,

$$\frac{\partial}{\partial x}[\mathcal{F}(A_k^{id}\delta_t)](w_l) = -j\omega_x[\mathcal{F}(A_k^{id}\delta_t)](w_l) \tag{7.36}$$

$$\frac{\partial}{\partial y}[\mathcal{F}(A_k^{id}\delta_t)](w_l) = -j\omega_y[\mathcal{F}(A_k^{id}\delta_t)](w_l) \tag{7.37}$$

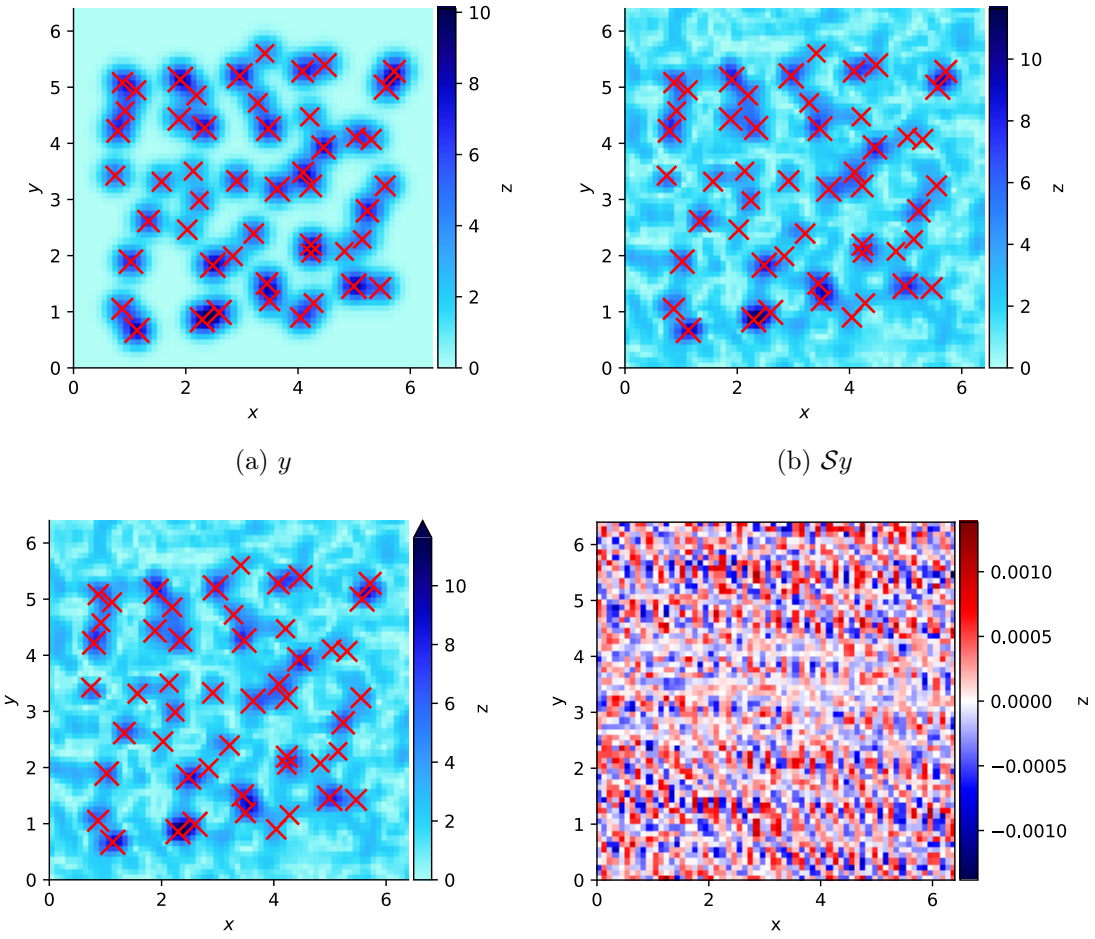
$$\frac{\partial}{\partial z}[\mathcal{F}(A_k^{id}\delta_t)](w_l) = f_k'(z)\mathrm{e}^{-\frac{\langle\omega_l^2,\sigma^2\rangle}{2}-j\langle\omega_l,(x,y)\rangle}. \tag{7.38}$$

7.3 Expériences

Dans cette section, nous appliquons notre méthode de sketching à la méthode référence OP-COMP avec PGD [16]. Nous comparons l'algorithme original à la version avec sketching où l'opérateur de mesure (respectivement les mesures) est remplacé par l'opérateur de mesure avec sketching (respectivement les mesures compressées), c'est-à-dire que nous étudions l'impact de la réduction de m mesures à $m_{\mathcal{S}} < m$ mesures.

Nous estimons des signaux composés de 50 impulsions en 3 dimensions. Les opérateurs linéaires A utilisés pour ces expériences suivent l'approximation gaussienne du modèle MA-TIRF. Les positions d'entrée sont dans un cuboïde ($[0, 6.4] \times [0, 6.4] \times [0, 0.8]$ µm) et les amplitudes sont uniformément distribuées entre 1 et 2. La sortie (le vecteur de mesure y) est constituée de 4 différentes grilles 2D composées de 64×64 pixels pour un total de $m = 4 \times 64 \times 64 = 16\,384$ mesures. Comme nous réalisons le sketching de A séparément sur chaque grille 2D et pour une meilleure lisibilité, nous notons $m_{\mathcal{S}} = 4 \times K \times m_{\text{spike}}$ avec $K = 50$ dans nos expériences et m_{spike} le nombre de mesures par impulsion. Nous étudions d'abord un exemple de reconstruction du signal avec A et $\mathcal{S}A^{id}$ avec $m_{\text{spike}} = 12$. Nous testons ensuite $\mathcal{S}A^{id}$ avec différentes valeurs de $m_{\text{spike}} \in \{3, 6, \ldots, 45, 48\}$. De plus, comme les fréquences de l'opérateur compressé sont tirées au hasard, nous répétons le processus pour chaque valeur de m_{spike} 16 fois avec différentes fréquences et signaux x_0. Le code de ces expériences peut être téléchargé à l'adresse [13].

Nous notons que pour le processus d'initialisation avec OP-COMP (détaillé dans le Chapitre 2), il existe différentes techniques d'initialisation d'une impulsion en fonction du type d'observations. D'une part, si elles représentent l'intensité de chaque pixel sur une grille, on peut créer une première estimation simple en prenant la position du pixel dont l'indice correspond à l'$\arg\max_i y_i$. C'est le cas de MA-TIRF et d'autres opérateurs linéaires de microscopie. D'une autre part, si les observations ne représentent pas les

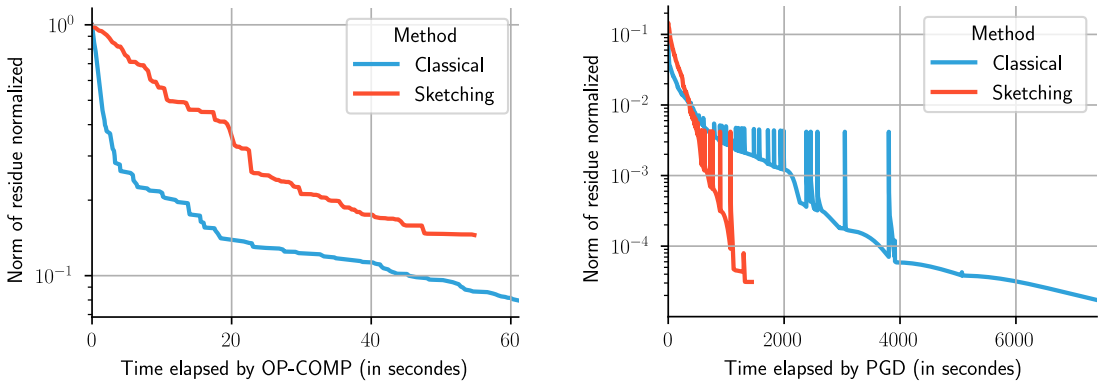


(a) y (b) $\mathcal{S}y$

(c) $\mathcal{S}_d y$ (d) $\mathcal{S}y - \mathcal{S}_d y$

FIGURE 7.1 – Vérité terrain x_0 (en rouge) et la grille en 2D d'un des angles d'incidence obtenu à partir de (a) l'observation $y = Ax_0$, (b) la rétro-projection de l'observation du compression idéal $\mathcal{S}y = \mathcal{S}A^{id}x_0$, (c) la rétro-projection de l'observation du compression approximé $\mathcal{S}_d y = \mathcal{S}_d A x_0$, (d) la différence $\mathcal{S}y - \mathcal{S}_d y$. La petite différence indique que $\mathcal{S}_d y$ est une bonne approximation de $\mathcal{S}y$.

intensités des pixels de la grille, il faut donc prendre une première estimation au hasard dans le cuboïde. C'est le cas pour l'échantillonnage de Fourier et l'opérateur linéaire MA-TIRF compressé. Cette raison met en évidence une différence essentielle entre les opérateurs linéaires compressés et les opérateurs dont l'observation est structurée par une grille, tels que MA-TIRF.



(a) (b)

FIGURE 7.2 – (a) La norme du résidu r au cours des itérations de OP-COMP avec A (en bleu) et $\mathcal{S}A^{id}$ (en rouge) par rapport au temps de calcul écoulé, (b) La norme du résidu r au cours des itérations de PGD avec A (en bleu) et $\mathcal{S}A^{id}$ (en rouge) par rapport au temps de calcul écoulé. Le temps de calcul total avec le sketching est $\approx 80\%$ plus court que sans sketching.

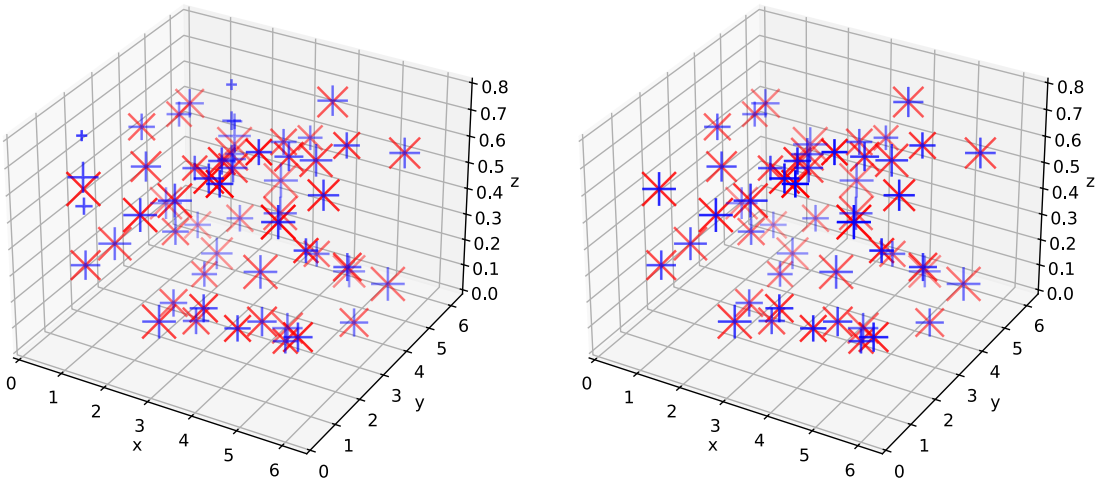
Comme les fonctionnelles pour les méthodes compressées et non-compressées sont différentes, nous utilisons les versions normalisées de ces fonctionnelles :

$$g_{\text{norm}} = \frac{\|A\phi(\theta) - y\|_2}{\|y\|_2} \quad \text{et} \quad g_{\mathcal{S}_d,\text{norm}} = \frac{\|\mathcal{S}A^{id}\phi(\theta) - \mathcal{S}_d y\|_2}{\|\mathcal{S}_d y\|_2}. \tag{7.39}$$

Détails avec $m_{\text{spike}} = 12$: Tout d'abord, nous comparons, dans la Figure 7.1, les mesures originales avec la rétro-projection des mesures idéales et réelles avec sketching sur la grille Γ. Nous remarquons que la différence entre les mesures idéales et les mesures réelles avec sketching est très faible. Dans cet exemple, nous avons $\|y\|_2 \approx 200$ et $\|\mathcal{S}y - \mathcal{S}_d y\|_2 \approx 6 \times 10^{-3}$. Cela correspondrait à $\eta \approx 3 \times 10^{-5}$ dans le Lemme 7.1. Par conséquent, cet exemple (ainsi que les expériences suivantes) vérifie qualitativement le Lemme 7.1.

Nous comparons dans la Figure 7.2 et la Figure 7.3 le résultat d'OP-COMP + PGD avec et sans sketching. Avec $m_{\text{spike}} = 12$ mesures par impulsion représentant environ 15% du nombre total de mesures ($m = 16\,384$) obtenues avec A, la méthode OP-COMP + PGD avec sketching converge avec environ 38 000 itérations en 26 minutes contre 75 000 itérations en 124 minutes pour la méthode sans sketching. Cela représente un gain de près de 80% en temps de calcul, sans perte de performance de l'estimation (voir l'expérience suivante).

Performance de l'estimation avec $m_{\text{spike}} \in \{3, 6, \ldots, 45, 48\}$: Dans la Figure 7.4, nous comparons les performances de reconstruction avec la métrique suivante : Nous désignons le RMSE dans chaque dimension spatiale par $RMSE_i = \sqrt{\frac{1}{|TP|}\sum_{j\in TP}\left([t_j]_i - [t_{0,j}]_i\right)^2}$ avec i la dimension sur laquelle le RMSE est calculée, t nos positions estimées de notre signal x_{esti}, t_0 nos positions des impulsions de x_0 et TP l'ensemble $|TP|$ de paires d'impulsions de la vérité terrain et de impulsions estimées. Nous construisons cet ensemble TP en



(a) x_0 and x^* (b) x_0 and x^* (sketched)

FIGURE 7.3 – Vérité terrain x_0 (en rouge) et le signal estimé x_{init} avec OP-COMP + PGD. (a) Méthode classique OP-COMP + PGD, (b) Méthode sketched OP-COMP + PGD. Le sketched OP-COMP + PGD reconstruit les impulsions de x_0 avec la même précision que OP-COMP + PGD.

appariant les impulsions si elles sont suffisamment proches les unes des autres. Le rayon choisi est de 0.02 µm (qui est plus petit que $\epsilon/2$, avec ϵ la séparation entre les impulsions de la vérité terrain).

Nous observons dans la Figure 7.4 que le RMSE médian le long de chaque axe diminue à mesure que le nombre de mesures sketched $m_{\mathcal{S}}$ augmente. À partir d'environ 15%, nous remarquons que les RMSE médians obtenus avec notre opérateur sketched $\mathcal{S}A^{id}$ sont inférieurs aux RMSE obtenus avec l'opérateur classique A (notez que les performances devraient être légèrement dégradées, mais le sketching peut introduire une opération de lissage qui profite au processus d'optimisation).

En comparant le temps de calcul pour chaque valeur de $m_{\mathcal{S}}$ dans la Figure 7.5, nous remarquons que le temps écoulé pour la reconstruction de x_0 avec le sketching est linéaire par rapport à $m_{\mathcal{S}}$. Avec $m_{\mathcal{S}}/m \approx 15\%$, nous avons une accélération par rapport à la méthode sans sketching jusqu'à 80%. Au-dessus de $m_{\mathcal{S}}/m \approx 60\%$, il n'est pas utile d'utiliser l'opérateur linéaire de sketching par rapport à la méthode originale (en raison des calculs supplémentaires pour le sketching).

7.4 Résumé du chapitre et conclusions

Nous avons montré que l'utilisation d'un opérateur linéaire compressé au lieu d'un opérateur linéaire classique lorsque le nombre de mesures est élevé peut améliorer considérablement les performances sans perdre en précision sur les positions et les amplitudes reconstruites des impulsions. Nos résultats sont étayés par un solide contexte théorique et des expériences. Nous avons également mis en évidence que le compromis entre le temps de calcul et la précision peut être explicitement choisi. Nous avons proposé d'utiliser une méthode de sketching indépendamment des mesures 2D. Un travail futur possible serait d'étudier si une méthode de sketching en 3D complète améliorerait nos résultats. Il serait également intéressant d'étudier la légère amélioration surprenante des performances que

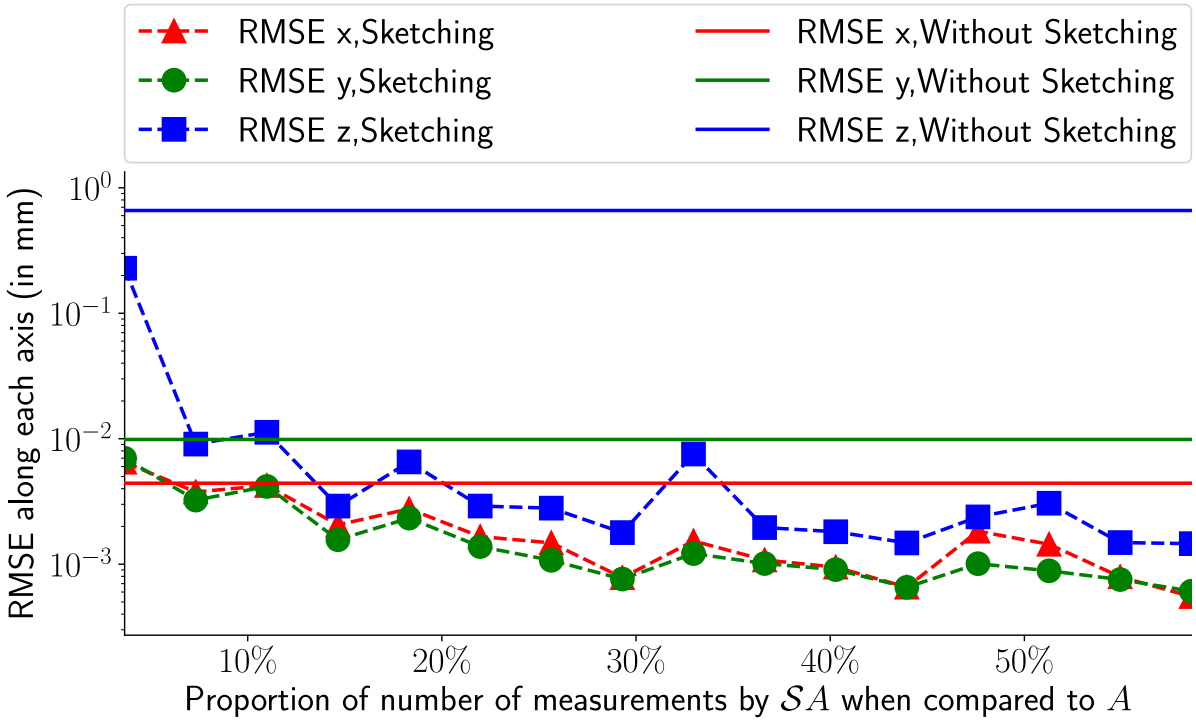


FIGURE 7.4 – Le médian du RMSE pour chacune des dimensions $\{x, y, z\}$ obtenu avec OP-COMP + PGD sur des signaux composés de 50 impulsions par rapport au taux de compression de mesures du sketching. Le RMSE pour chaque dimension du signal estimé x_{esti} avec la méthode de sketching décroit quand le nombre de mesures m_{spike} augmente et atteint les performances de la méthode originale avec un taux de compression de 15%.

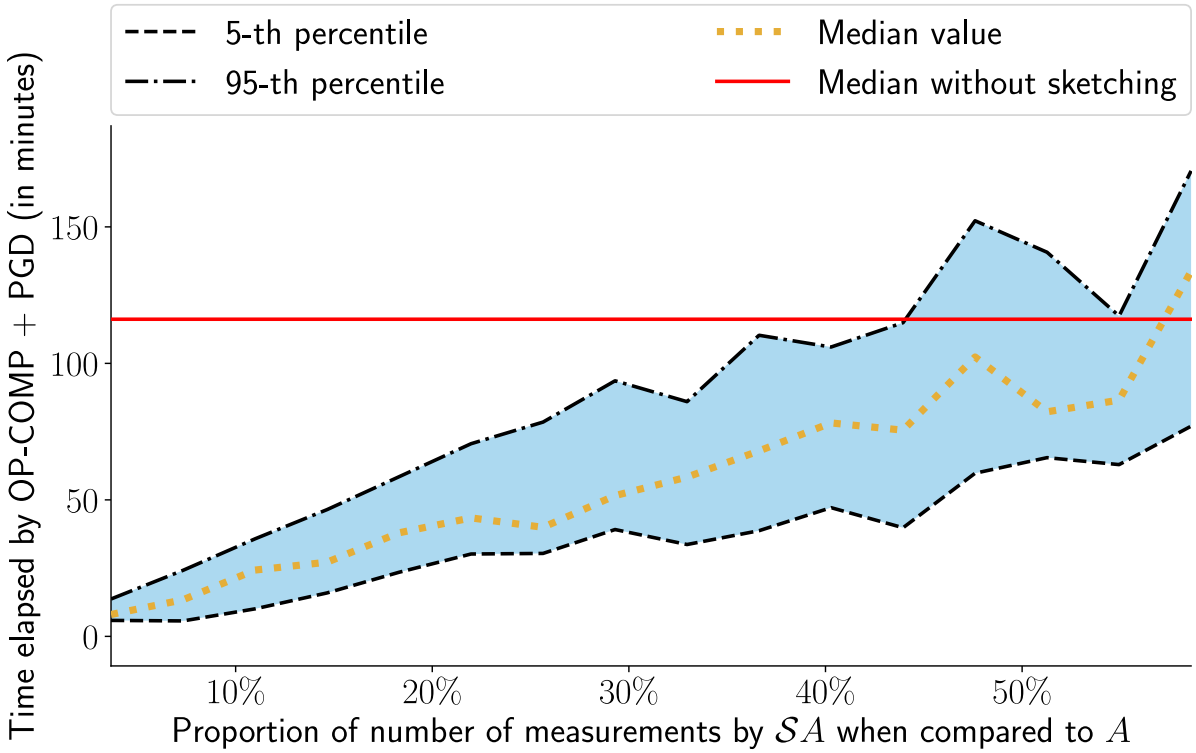


FIGURE 7.5 – Le temps de calcul de OP-COMP + PGD pour des signaux de 50 impulsions en fonction du taux de compression $m_{\mathcal{S}}/m$. En bleu, l'étendue des résultats pour l'ensemble des expériences avec la médiane (en jaune), le 5-ième et 95-ième centile. En rouge, la médiane des résultats sans sketching. Le temps de calcul est proportionnel à $m_{\mathcal{S}}$. Avec $m_{\mathcal{S}}/m \lesssim 60\%$, la méthode avec sketching est plus rapide que la méthode originale.

nous avons observée avec la méthode de sketching. Comme vu dans le Chapitre 6 sur la descente par blocs de coordonnées, BCD n'apporte pas d'accélération significative dans le cas de la transformée de Fourier comme opérateur linéaire A. Cela montre que la méthode de sketching rend la descente par blocs inutile dans ce contexte.

Chapitre 8

Conclusions

Pour finir, nous allons maintenant faire un récapitulatif de l'ensemble des travaux présentés et de leurs implications. Nous commencerons par un rapide résumé de chaque méthode avec leurs avantages et nouveautés par rapport à de précédents travaux. Ensuite, nous discuterons de leurs défauts et limitations. Et enfin, nous évoquerons les pistes de recherches issues des idées précédemment évoquées.

8.1 Récapitulatif des travaux

En définitif, l'ensemble de nos méthodes présentées dans ces travaux concernent l'algorithme COMP sur-paramétré avec descente de gradient projetée. Cette méthode, elle même issue de l'état de l'art COMP avec descente, apporte des modifications qui la rendent plus rapide lors de la reconstruction de signaux parcimonieux avec beaucoup d'impulsions. En effet, la suppression de la descente à chaque itération de l'initialisation du signal élimine l'étape la plus couteuse de COMP avec descente. Ce ralentissement est dû au nombre croissant de paramètres à calculer. Cette suppression de la descente s'accompagnant d'une perte de précision pour l'estimation du signal, nous ajoutons alors une étape de sur-paramétrisation pour compenser ce défaut. En complément de cette sur-paramétrisation, afin de faire converger les impulsions estimées vers le signal inconnu et de supprimer les impulsions superflues, nous ajoutons une étape de descente sur les paramètres avec une projection à chaque itération.

L'ensemble de ces modifications de COMP sur-paramétré avec PGD par rapport à COMP avec descente lui confèrent un avantage non négligeable plus le signal à reconstruire possède d'impulsions. En effet, nous avons que $\mathtt{T}_{\mathtt{SCOMP}} = \mathcal{O}(K^2)$ alors que $\mathtt{T}_{\mathtt{PGD}} = \mathcal{O}(K)$. Notons tout de même que bien que nous passons de K descentes à une seule, pour la reconstruction de signaux avec un petit nombre d'impulsions à estimer, COMP avec descente reste plus rapide. De plus, cette accélération ne s'accompagne pas d'une perte de précision sur l'estimation finale.

En plus de fournir des résultats expérimentaux, nous apportons des résultats théoriques sur notre méthode COMP sur-paramétré. En effet, le manque de descente à chaque itération ajoute une imprécision sur le signal initialisé. Nous montrons alors que dans le cas d'un opérateur comportant suffisamment de mesures, c'est à dire qu'il suit a une propriété d'isométrie restreinte, alors l'ensemble des K premières impulsions de COMP sans descente est arbitrairement proche des impulsions du signal à reconstruire. Ce résultat très important valide notre méthode car la descente à chaque itération de COMP avec descente n'est pas justifiée. Cependant, les conditions nécessaires pour utiliser ce résultat sont difficilement réalisables en l'état et nous devons alors sur-paramétriser le signal

pour obtenir des résultats satisfaisants. Une autre implication en lien avec ce résultat est l'estimation du nombre de mesures nécessaires afin de garantir une propriété d'isométrie restreinte sur certains opérateurs linéaires.

En plus de l'accélération procurée par OP-COMP avec PGD, nous développons également d'autres méthodes d'accélération. La première tiens compte du fait que les paramètres que nous estimons (positions et amplitudes des impulsions) peuvent être regroupées par blocs. Chaque bloc représentant une impulsion, nous pouvons alors séparer le problème en effectuant la descente finale uniquement sur les impulsions qui n'ont pas encore convergé. Une autre méthode d'accélération consiste alors à faire une compression de l'opérateur linéaire afin de compresser le nombre initial de mesures pour réduire le coût de calcul de la descente. Une méthode de compression est d'appliquer la transformée de Fourier sur l'opérateur linéaire. Ces deux méthodes d'accélérations apportent toutes deux une réduction du coût de calcul total et sont compatibles entre elles.

Au final nous avons vu plusieurs méthodes d'accélération qui permettent de résoudre le problème de reconstruction de signaux parcimonieux plus rapidement et sans perte de précision.

8.2 Limitations associées et suggestions pour de futurs travaux

Bien que nos nouvelles méthodes apportent une accélération significative par rapport à l'état de l'art, elles sont en proie à quelques défauts qu'ils nous faut décrire.

L'état de sur-paramétrisation apporté par OP-COMP permet de palier à l'absence de l'étape de descente mais engendre lui aussi un ralentissement pendant la descente de gradient finale. Ce ralentissement est principalement dû au mauvais conditionnement du système. En effet, le signal estimé converge très lentement vers le minimum global de g. Ceci limite les capacités de reconstruction et nous force à supprimer suffisamment d'impulsions pour être de nouveau bien conditionné. Un meilleur pré-conditionnement du problème [31, 33] est alors une piste à étudier pour limiter le ralentissement de la sur-paramétrisation.

Par ailleurs, l'heuristique utilisée dans PGD pour projeter le signal dans son espace d'arrivée $\Sigma_{K,\epsilon}$ n'est pas optimale. En effet, elle ajoute un nouveau paramètre à contrôler qui détermine la distance à partir de laquelle fusionner deux impulsions. Une idée future pourrait être d'apporter un estimateur pour cette distance de projection en fonction de la vitesse de convergence du signal. Finalement pour OP-COMP, la résolution de $\arg\max_t \langle A\delta_t, r^{(k-1)} \rangle$ pour initialiser la position de la prochaine impulsion est difficile à résoudre en pratique de part la non-convexité du problème.

Concernant le résultat pour l'initialisation des K premières impulsions, il ne tient pas en compte des approximations faites sur $\arg\max_t \langle A\delta_t, r^{(k-1)} \rangle$, ce qui diffère du cas pratique. Un axe de recherche peut être de s'orienter sur une version sur-paramétrée du résultat en tenant compte des approximations dans le calcul de $\arg\max_t \langle A\delta_t, r^{(k-1)} \rangle$. De plus, elle pourrait également permettre de relaxer les hypothèses de départ liées à la propriété d'isométrie restreinte γ de l'opérateur linéaire ainsi que la cohérence mutuelle μ du noyau. En effet, elles dépendent toutes deux du nombre d'impulsions du signal à reconstruire avec $\mu + \gamma \leq \mathcal{O}\left(\frac{1}{K^2}\right)$.

Pour ce qui est des accélérations avec les méthodes de descente par blocs de coordonnées et de compression de l'opérateur linéaire, elles rajoutent des paramètres en plus à gérer et qui impactent de manière significative l'efficacité des deux méthodes. Il est alors

possible d'explorer des techniques d'estimations de paramètres pour ces deux méthodes. Plus spécifiquement, pour la descente par blocs, un avantage de la séparation du signal en plusieurs signaux est l'intégration d'une parallélisation possible pour chaque bloc.

Finalement, avec l'arrivée ces dernières années des algorithmes basés sur des réseaux de neurones, nous constatons une nouvelle piste pour la reconstruction de signaux parcimonieux. En effet, ces méthodes basées sur l'apprentissage profond donnent de bons résultats rapidement. Cependant, leur manque de garanties théoriques et parfois de robustesse les différencies des méthodes formelles que nous avons évoquées au cours de nos travaux. Une approche hybride pourrait s'avérer être un bon compromis entre les deux types de méthodes. En effet, nous pouvons imaginer un algorithme qui initialise un signal grâce à un réseau de neurones profonds et qui par la suite effectue une descente sur les paramètres pour une meilleure convergence vers la solution.

Annexe A

Glossaire et notations

A.1 Acronymes

— SMLM : Single Molecule Localization Microscopy
— COMP : Continuous Orthogonal Matching Pursuit
— OP-COMP : COMP sur-paramétré
— PGD : Descente de gradient projetée
— BCD : Descente par blocs de coordonnées
— RIP : Propriété d'isométrie restreinte
— P_ϵ : Projection qui envoi un élément sur l'ensemble respectant la contrainte de séparation minimale ϵ
— FISTA : Fast Iterative Shrinkage-Thresholding Algorithm
— MA-TIRF : Multi-Angle Total Internal Reflection Fluorescence

A.2 Variables

— $K \in \mathbb{N}^*$: nombre d'impulsions
— $d \in \mathbb{N}^*$: dimension du support des impulsions
— $a \in \mathbb{R}^K$: amplitudes des impulsions
— $t \in \mathbb{R}^{K \times d}$: positions des impulsions
— $x_0 = \sum_{i=1}^K a_i \delta_{t_i}$: signal parcimonieux à reconstruire
— A : opérateur linéaire de mesure
— A^{id} : opérateur linéaire de mesure idéal
— $m \in \mathbb{N}^*$: nombre de mesures
— $y \in \mathbb{C}^m$: observation
— $\epsilon \in \mathbb{R}^*$: séparation minimal entre les impulsions
— $\Sigma_{K,\epsilon}$: ensemble des signaux de K impulsions avec une contrainte de séparation entre les impulsions
— $\mathcal{S}(\Sigma_{K,\epsilon})$: espace sécant de $\Sigma_{K,\epsilon}$. Voir Définition 3
— $\gamma \in]0,1[$: Constante RIP. Voir Définition 4
— h : noyau associé à A
— μ : Constante de cohérence mutuelle du noyau. Voir Définition 1
— ν : mesure finie et signée sur $\mathbb{R}^d$
— π : Dipôle de mesures. Voir Définition 2
— ξ : Séparation maximale entre des impulsions et leurs estimations par COMP sans descente, $\xi = \max_i \min_l \|s_l^* - t_i\|_2$. Voir Théorème 3.3
— σ^2 : variance du noyau Gaussien

A.3 Variables dans la preuve

— $Z_{l,\xi,\frac{\epsilon}{3}}$: l'ensemble des dipôles ξ-concentrés, deux à deux ϵ-séparés. Voir Équation (B.45)

— $\mathcal{X}_{K,l}^{\epsilon,\xi}$: l'ensemble $\Sigma_{K-l,\epsilon} + Z_{l,\xi,\frac{\epsilon}{3}}$. Voir Équation (B.162)

Annexe B

Preuves des théorèmes du Chapitre 3

Dans ce chapitre, nous détaillons les preuves des deux Théorèmes 3.3 et 3.4 du chapitre précédente. La stratégie de la preuve est décrite dans la Section 3.4. Avant de débuter les preuves principales, nous commençons d'abord par introduire des Lemmes intermédiaires.

B.1 Preuve des résultats techniques

Dans cette section, nous détaillons les preuves de résultats techniques notamment en rapport avec les bornes sur produits scalaires et normes des dipôles et mesures de Dirac. Nous commençons avec la preuve du Lemme 3.1,

Preuve du Lemme 3.1. Nous avons que $h(t,s) = \mathrm{e}^{-\|t-s\|_2^2/(2\sigma^2)}$,

1. Nous pouvons écrire $h(t,s) = \rho(\|t-s\|_2)$ où $\rho(x) = \mathrm{e}^{-x^2/(2\sigma^2)}$ pour $t, s \in \mathbb{R}^d$, c'est à dire que h est symétrique par rapport à 0 et est invariant par translation.
2. Nous avons que $h(t,t) = \rho(0) = 1$. Cependant, $-x^2/(2\sigma^2) \leq 0$ pour $x \in \mathbb{R}$ et devient une égalité pour $x = 0$, c'est à dire que $h(t,t) = \max_{t\in\mathbb{R}^d, s\in\mathbb{R}^d} |h(t,s)|$.
3. Pour n'importe quels impulsions $\delta_{t_0}, \delta_{s_0}$ telles que $\|t_0 - s_0\|_2 = \epsilon/3$, d'après la définition du produit scalaire associé au noyau h, nous avons que $\langle \delta_{t_0}, \delta_{s_0}\rangle_h = \mathrm{e}^{-\|t_0-s_0\|_2^2/(2\sigma^2)} = \mathrm{e}^{-(\epsilon/3)^2/(2\sigma^2)}$. En outre, à partir de notre Hypothèse 1 de cohérence mutuelle sur le noyau h et de (2.), nous obtenons $\langle \delta_{t_0}, \delta_{s_0}\rangle_h \leq \mu \|\delta_{t_0}\|_h \|\delta_{s_0}\|_h = \mu$. En combinant ces deux résultats, nous montrons que

$$\langle \delta_{t_0}, \delta_{s_0}\rangle_h = \mathrm{e}^{-\frac{(\epsilon/3)^2}{2\sigma^2}} \leq \mu \|\delta_t\|_h \|\delta_s\|_h = \mu. \tag{B.1}$$

   En prenant t, s tels que $\|t-s\|_2 \leq \epsilon/3 = \|t_0 - s_0\|_2$, nous avons

$$\langle \delta_t, \delta_s\rangle_h \leq \mathrm{e}^{-\frac{(\epsilon/3)^2}{2\sigma^2}} \leq \mu. \tag{B.2}$$

□

Lemme B.1 (Borne supérieure de la norme d'un dipôle)**.** *Soit $\pi = a\delta_t - b\delta_s$ un dipôle h un noyau. Nous avons alors que,*

$$\|\pi\|_h \leq |a-b| + |a|\|\delta_s - \delta_t\|_h. \tag{B.3}$$

Preuve. Remarquons que $a\delta_t - b\delta_s = a(\delta_t - \delta_s) + (a-b)\delta_s$. En utilisant l'inégalité triangulaire,

$$\|\pi\|_h \leq \|(a-b)\delta_s\|_h + \|a(\delta_t - \delta_s)\|_h = |a-b| + |a|\|\delta_t - \delta_s\|_h. \tag{B.4}$$

□

Lemme B.2 (Norme d'un dipôle pour un noyau Gaussien)**.** *Soit $\pi = a\delta_t - b\delta_s$ un ξ_0-dipôle tel que $0 \leq \xi_0 \leq \xi$ et soit h un noyau Gaussien de variance $2\sigma^2$. Nous avons alors les propriétés suivantes :*

$$(i) \qquad \|\pi\|_h^2 = (a-b)^2 + 2ab(1 - \mathrm{e}^{-\frac{\xi_0^2}{2\sigma^2}}), \tag{B.5}$$

$$(ii) \qquad \|\pi\|_h^2 \leq (a-b)^2 + 2|ab|(1 - \mathrm{e}^{-\frac{\xi^2}{2\sigma^2}}). \tag{B.6}$$

Preuve. En partant de $\|\pi\|_h^2$, nous développons cette norme pour obtenir (B.5),

$$\|\pi\|_h^2 = \|a\delta_t - b\delta_s\|_h^2 \tag{B.7}$$

$$= a^2\|\delta_t\|_h^2 + b^2\|\delta_s\|_h^2 - 2ab\langle\delta_t, \delta_s\rangle_h \tag{B.8}$$

$$= a^2 + b^2 - 2ab\mathrm{e}^{-\frac{\|t-s\|_2^2}{2\sigma^2}} \tag{B.9}$$

$$= a^2 + b^2 - 2ab + 2ab - 2ab\mathrm{e}^{-\frac{\|t-s\|_2^2}{2\sigma^2}} \tag{B.10}$$

$$\|\pi\|_h^2 = (a-b)^2 + 2ab(1 - \mathrm{e}^{-\frac{\xi_0^2}{2\sigma^2}}). \tag{B.11}$$

De plus, si $\xi_0 \leq \xi$, alors $\mathrm{e}^{-\xi_0^2/(2\sigma^2)} \geq \mathrm{e}^{-\xi^2/(2\sigma^2)}$ et

$$2ab(1 - \mathrm{e}^{-\frac{\xi_0^2}{2\sigma^2}}) \leq 2|a||b|(1 - \mathrm{e}^{-\frac{\xi^2}{2\sigma^2}}). \tag{B.12}$$

Nous injectons (B.11) dans (B.12) pour obtenir (B.6). □

Lemme B.3 (Borne du produit scalaire entre deux réponses impulsionnelles)**.** *Soient δ_s, δ_t deux mesures de Dirac avecs, $t \in \mathbb{R}^d$. Soit A un opérateur linéaire avec une γ-RIP sur $\mathcal{S}(\Sigma_{K,\epsilon})$ (voir Définition 4). Alors,*

$$(1+\gamma)\langle\delta_s, \delta_t\rangle_h - 2\gamma \leq \langle A\delta_s, A\delta_t\rangle \leq (1-\gamma)\langle\delta_s, \delta_t\rangle_h + 2\gamma, \tag{B.13}$$

et

$$|\langle A\delta_s, A\delta_t\rangle| \leq (1-\gamma)\langle\delta_s, \delta_t\rangle_h + 2\gamma. \tag{B.14}$$

Preuve. Commençons par développer $\langle A\delta_s, A\delta_t\rangle$,

$$\langle A\delta_s, A\delta_t\rangle = \frac{1}{2}\left(\|A\delta_s\|_2^2 + \|A\delta_t\|_2^2 - \|A(\delta_s - \delta_t)\|_2^2\right) \tag{B.15}$$

la γ-RIP et le Lemme 3.1 impliquent alors

$$\langle A\delta_s, A\delta_t\rangle \geq \frac{1}{2}\left((1-\gamma)\|\delta_s\|_h^2 + (1-\gamma)\|\delta_t\|_h^2 - (1+\gamma)\|(\delta_s - \delta_t)\|_h^2\right) \tag{B.16}$$

$$= (1-\gamma) - \frac{1}{2}(1+\gamma)\|(\delta_s - \delta_t)\|_h^2 \tag{B.17}$$

$$= 1 - \gamma - \frac{1}{2}(1+\gamma)\left(\|\delta_s\|_h^2 + \|\delta_t\|_h^2 - 2\langle\delta_s, \delta_t\rangle_h\right) \tag{B.18}$$

$$= 1 - \gamma - (1+\gamma)(1 - \langle\delta_s, \delta_t\rangle_h) \tag{B.19}$$

$$= (1+\gamma)\langle\delta_s, \delta_t\rangle_h - 2\gamma. \tag{B.20}$$

De la même manière,

$$\langle A\delta_s, A\delta_t\rangle \leq \frac{1}{2}\left((1+\gamma)\|\delta_s\|_h^2 + (1+\gamma)\|\delta_t\|_h^2 - (1-\gamma)\|(\delta_s-\delta_t)\|_h^2\right) \tag{B.21}$$

$$= (1+\gamma) - \frac{1}{2}(1-\gamma)\|(\delta_s-\delta_t)\|_h^2 \tag{B.22}$$

$$= 1+\gamma - \frac{1}{2}(1-\gamma)\left(\|\delta_s\|_h^2 + \|\delta_t\|_h^2 - 2\langle\delta_s,\delta_t\rangle_h\right) \tag{B.23}$$

$$= 1+\gamma - (1-\gamma)(1-\langle\delta_s,\delta_t\rangle_h) \tag{B.24}$$

$$= (1-\gamma)\langle\delta_s,\delta_t\rangle_h + 2\gamma. \tag{B.25}$$

En combinant ces deux bornes sur $\langle A\delta_s, A\delta_t\rangle$, nous venons de montrer le premier résultat de ce Lemme. Maintenant, pour borner $|\langle A\delta_s, A\delta_t\rangle|$, nous comparons les bornes inférieures et supérieures. Si $(1+\gamma)\langle\delta_s,\delta_t\rangle_h - 2\gamma \geq 0$, alors

$$|(1-\gamma)\langle\delta_s,\delta_t\rangle_h + 2\gamma| \geq |(1+\gamma)\langle\delta_s,\delta_t\rangle_h - 2\gamma|. \tag{B.26}$$

Dans le cas contraire, considérons le cas où $(1+\gamma)\langle\delta_s,\delta_t\rangle_h - 2\gamma \leq 0$. Comme $\gamma < 1$, nous avons $(1-\gamma)\langle\delta_s,\delta_t\rangle_h + 2\gamma > 0$. En prenant la différence entre les deux valeurs absolues, nous obtenons

$$|(1-\gamma)\langle\delta_s,\delta_t\rangle_h + 2\gamma| - |(1+\gamma)\langle\delta_s,\delta_t\rangle_h - 2\gamma|$$
$$= (1-\gamma)\langle\delta_s,\delta_t\rangle_h + 2\gamma - (-1)\times\left((1+\gamma)\langle\delta_s,\delta_t\rangle_h - 2\gamma\right) \tag{B.27}$$

En simplifiant certains termes qui s'annulent, nous obtenons

$$|(1-\gamma)\langle\delta_s,\delta_t\rangle_h + 2\gamma| - |(1+\gamma)\langle\delta_s,\delta_t\rangle_h - 2\gamma| = 2\langle\delta_s,\delta_t\rangle_h \geq 0. \tag{B.28}$$

Finalement, comme dans les deux cas (respectivement $(1+\gamma)\langle\delta_s,\delta_t\rangle_h - 2\gamma \geq 0$ et $(1+\gamma)\langle\delta_s,\delta_t\rangle_h - 2\gamma \leq 0$), nous avons que $|(1-\gamma)\langle\delta_s,\delta_t\rangle_h + 2\gamma| \geq |(1+\gamma)\langle\delta_s,\delta_t\rangle_h - 2\gamma|$ (respectivement (B.26) et (B.28)), nous concluons alors que

$$|\langle A\delta_s, A\delta_t\rangle| \leq \max\left(|(1-\gamma)\langle\delta_s,\delta_t\rangle_h + 2\gamma|, |(1+\gamma)\langle\delta_s,\delta_t\rangle_h - 2\gamma|\right) \tag{B.29}$$

$$= (1-\gamma)\langle\delta_s,\delta_t\rangle_h + 2\gamma. \tag{B.30}$$

□

Lemme B.4 (Borne du produit scalaire entre deux dipôles)**.** *Soient π_1, π_2 deux dipôles ϵ-séparés. Soit A un opérateur linéaire avec une γ-RIP sur $\mathcal{S}(\Sigma_{K,\epsilon})$ (voir Définition 4). Alors nous avons que,*

$$(1+\gamma)\langle\pi_1,\pi_2\rangle_h - \gamma(\|\pi_1\|_h^2 + \|\pi_2\|_h^2)$$
$$\leq \langle A\pi_1, A\pi_2\rangle \leq$$
$$(1-\gamma)\langle\pi_1,\pi_2\rangle_h + \gamma(\|\pi_1\|_h^2 + \|\pi_2\|_h^2) \tag{B.31}$$

et

$$|\langle A\pi_1, A\pi_2\rangle| \leq |\langle\pi_1,\pi_2\rangle_h| + \gamma\left|\|\pi_1\|_h^2 + \|\pi_2\|_h^2 - \langle\pi_1,\pi_2\rangle_h\right|. \tag{B.32}$$

Preuve. Avec la γ-RIP, nous avons,

$$\langle A\pi_1, A\pi_2\rangle = \frac{1}{2}\left(\|A\pi_1\|_2^2 + \|A\pi_2\|_2^2 - \|A(\pi_1-\pi_2)\|_2^2\right) \tag{B.33}$$

$$\geq \frac{1}{2}\Big((1-\gamma)(\|\pi_1\|_h^2 + \|\pi_2\|_h^2) - (1+\gamma)\|\pi_1-\pi_2\|_h^2\Big). \tag{B.34}$$

Nous développons la norme de la différence entre deux dipôles,

$$\|\pi_1 - \pi_2\|_h^2 = \|\pi_1\|_h^2 + \|\pi_2\|_h^2 - 2\langle \pi_1, \pi_2 \rangle_h. \tag{B.35}$$

En injectant ce résultat dans notre inégalité précédente (B.34), nous obtenons,

$$\begin{aligned}\langle A\pi_1, A\pi_2 \rangle &\geq \frac{1}{2}\Big((1-\gamma)\left(\|\pi_1\|_h^2 + \|\pi_2\|_h^2\right) - (1+\gamma)(\|\pi_1\|_h^2 + \|\pi_2\|_h^2 - 2\langle \pi_1, \pi_2 \rangle_h)\Big) &\text{(B.36)}\\ &= (1+\gamma)\langle \pi_1, \pi_2 \rangle_h - \gamma(\|\pi_1\|_h^2 + \|\pi_2\|_h^2). &\text{(B.37)}\end{aligned}$$

De la même manière,

$$\langle A\pi_1, A\pi_2 \rangle \leq \frac{1}{2}\left((1+\gamma)(\|\pi_1\|_h^2 + \|\pi_2\|_h^2) - (1-\gamma)\|\pi_1 - \pi_2\|_h^2\right). \tag{B.38}$$

Nous injectons dans cette inégalité le résultat (B.35) et nous obtenons

$$\begin{aligned}\langle A\pi_1, A\pi_2 \rangle &\leq \frac{1}{2}\Big((1+\gamma)(\|\pi_1\|_h^2 + \|\pi_2\|_h^2) - (1-\gamma)(\|\pi_1\|_h^2 + \|\pi_2\|_h^2 - 2\langle \pi_1, \pi_2 \rangle_h)\Big) &\text{(B.39)}\\ &= (1-\gamma)\langle \pi_1, \pi_2 \rangle_h + \gamma(\|\pi_1\|_h^2 + \|\pi_2\|_h^2). &\text{(B.40)}\end{aligned}$$

Les bornes obtenues précédemment (B.37) et (B.40), combinées et réécrites, nous donnent alors que

$$\begin{aligned}\gamma\left(\langle \pi_1, \pi_2 \rangle_h - \|\pi_1\|_h^2 - \|\pi_2\|_h^2\right) &\\ \leq \langle A\pi_1, A\pi_2 \rangle - \langle \pi_1, \pi_2 \rangle_h &\leq \\ \gamma\left(\|\pi_1\|_h^2 + \|\pi_2\|_h^2 - \langle \pi_1, \pi_2 \rangle_h\right). &\end{aligned} \tag{B.41}$$

Comme les bornes inférieures et supérieures sont symétriques, nous simplifions l'écriture avec la valeur absolue des bornes,

$$|\langle A\pi_1, A\pi_2 \rangle - \langle \pi_1, \pi_2 \rangle_h| \leq \gamma\left|\|\pi_1\|_h^2 + \|\pi_2\|_h^2 - \langle \pi_1, \pi_2 \rangle_h\right|. \tag{B.42}$$

En utilisant l'inégalité triangulaire, nous obtenons

$$|\langle A\pi_1, A\pi_2 \rangle| - |\langle \pi_1, \pi_2 \rangle_h| \leq |\langle A\pi_1, A\pi_2 \rangle - \langle \pi_1, \pi_2 \rangle_h|. \tag{B.43}$$

Ce qui implique que

$$|\langle A\pi_1, A\pi_2 \rangle| \leq |\langle \pi_1, \pi_2 \rangle_h| + \gamma\left|\|\pi_1\|_h^2 + \|\pi_2\|_h^2 - \langle \pi_1, \pi_2 \rangle_h\right|. \tag{B.44}$$

□

B.2 Preuve des résultats intermédiaires

Définissons d'abord l'ensemble $Z_{l,\xi,\frac{\epsilon}{3}}$ composé de dipôles ξ-concentrés, et pair à pair $\frac{\epsilon}{3}$-séparés,

$$\begin{aligned}Z_{l,\xi,\frac{\epsilon}{3}} := \Bigg\{ \sum_{i=1}^{l} \pi_i : \pi_i = a_i\delta_{t_i} - b_i\delta_{s_i};\ a,b \in \mathbb{R}^l;\ t_i \in \mathcal{D}; s_i \in \mathbb{R}^d;\ \forall i \in \{1,\dots,l\},&\\ \|t_i - s_i\|_2 \leq \xi;\ \forall i,j \in \{1,\dots,l\}, i \neq j, \pi_i, \pi_j \text{ sont } \frac{\epsilon}{3}\text{-séparés.} \Bigg\}.&\end{aligned} \tag{B.45}$$

Chaque élément de cet ensemble peut être vu comme la différence de deux signaux $x_1, x_2 \in \Sigma_{K,\epsilon}$ de telle manière à ce que chaque impulsion δ_{t_i} du signal x_1 soit appairée à une impulsion δ_s de x_2 avec $\|t - s\|_2 < \frac{\epsilon}{3}$. Cet ensemble nous sera utile par la suite car nous pourrons ainsi étudier ces différences de signaux comme des sommes de dipôles.

Lemme B.5. *Soit δ_s une mesure de Dirac et $z_l = \sum_{i=1}^{l} \pi_i \in Z_{l,\xi,\frac{\epsilon}{3}}$ tels que $\forall i = 1, \dots, l$, δ_s et π_i sont $\frac{\epsilon}{3}$-séparés et $\xi < \epsilon/3$. Supposons que l'opérateur linéaire A a une γ-RIP sur $\mathcal{S}(\Sigma_{K,\frac{\epsilon}{3}})$ (voir Définition 4). Posons $D = \max_{i\in\{1,\dots,l\}} \|\pi_i\|_h / \|a\|_\infty$ où $a \in \mathbb{R}^K$. Alors,*

$$\frac{\|Az_l\|_2^2}{\|a\|_\infty^2} \leq l(1+\gamma)(1+(l-1)\mu)D^2. \tag{B.46}$$

En outre, nous obtenons

$$\frac{|\langle A\delta_s, Az_l\rangle|}{\|a\|_\infty} \leq \kappa_l := (1+\gamma)\mu l D + \gamma l\left(1+D^2\right). \tag{B.47}$$

Preuve. À partir de la γ-RIP,

$$\|Az_l\|_2^2 = \|A\sum_{i=1}^{l}\pi_i\|_2^2 \leq (1+\gamma)\|\sum_{i=1}^{l}\pi_i\|_h^2. \tag{B.48}$$

Grâce au Lemme 3.2, nous obtenons

$$\frac{\|Az_l\|_2^2}{\|a\|_\infty^2} \leq (1+\gamma)(1+(l-1)\mu)\sum_{i=1}^{l}\frac{\|\pi_i\|_h^2}{\|a\|_\infty^2} \leq l(1+\gamma)(1+(l-1)\mu)D^2. \tag{B.49}$$

Cela conclu la première partie de la preuve.

Nous cherchons maintenant à borner $|\langle A\delta_s, Az_l\rangle|$. À partir de l'inégalité triangulaire,

$$\frac{|\langle A\delta_s, Az_l\rangle|}{\|a\|_\infty} \leq \sum_{i=1}^{l}|\langle A\delta_s, A\frac{\pi_i}{\|a\|_\infty}\rangle|. \tag{B.50}$$

Puis, en utilisant le résultat (B.32) du Lemme B.4,

$$\frac{|\langle A\delta_s, Az_l\rangle|}{\|a\|_\infty} \leq \sum_{i=1}^{l}\left(\left|\langle\delta_s, \frac{\pi_i}{\|a\|_\infty}\rangle_h\right| + \gamma\left|\|\delta_s\|_h^2 + \frac{\|\pi_i\|_h^2}{\|a\|_\infty^2} - \langle\delta_s, \frac{\pi_i}{\|a\|_\infty}\rangle_h\right|\right). \tag{B.51}$$

Puisque δ_s et les $\pi_i, i = 1, \dots, l$ sont $\frac{\epsilon}{3}$-séparés, nous avons $\langle\delta_s, \pi_i\rangle_h \leq \mu\|\delta_s\|_h\|\pi_i\|_h = \mu\|\pi_i\|_h$. Nous injectons ceci dans notre inégalité (B.51) pour obtenir

$$\frac{|\langle A\delta_s, Az_l\rangle|}{\|a\|_\infty} \leq \sum_{i=1}^{l}\left(\mu\frac{\|\pi_i\|_h}{\|a\|_\infty} + \gamma\left(1 + \frac{\|\pi_i\|_h^2}{\|a\|_\infty^2} + \mu\frac{\|\pi_i\|_h}{\|a\|_\infty}\right)\right). \tag{B.52}$$

En réarrangeant les termes dans (B.52), nous obtenons

$$\frac{|\langle A\delta_s, Az_l\rangle|}{\|a\|_\infty} \leq \sum_{i=1}^{l}\left((1+\gamma)\mu\frac{\|\pi_i\|_h}{\|a\|_\infty} + \gamma\left(1 + \frac{\|\pi_i\|_h^2}{\|a\|_\infty^2}\right)\right) \tag{B.53}$$

$$\leq (1+\gamma)\mu l D + \gamma l\left(1+D^2\right). \tag{B.54}$$

□

Le Lemme suivant nous donne une condition sur ξ pour borner le produit scalaire d'un dipôle ξ-séparé.

Lemme B.6 (Condition sur ξ). *Soient $k+l=K$ avec $k,l,K\in\mathbb{N}$ et $K>0$. Soit $\alpha\geq 1$ et*

$$\xi \geq \sqrt{2\sigma^2 \ln\left(\frac{1}{1-(4K\alpha-1)\,(\mu+\gamma)}\right)}, \tag{B.55}$$

avec $\mu,\gamma\in\mathbb{R}_*^+$ tels que $\mu+\gamma<\frac{1}{4K\alpha-1}$. Considérons κ_l et D définis dans le Lemme B.5 et supposons que $D<1$. Alors

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < \frac{1-((2k-1)\alpha-1)(1-\gamma)\mu-(4k\alpha-1)\gamma-2\kappa_l\alpha}{1-\gamma}. \tag{B.56}$$

Preuve. En utilisant l'hypothèse que $D<1$, nous obtenons alors que $\kappa_l<(1+\gamma)\mu l+2\gamma l$. Ainsi (B.56) est vérifiée si

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < \frac{1-((2k-1)\alpha-1)(1-\gamma)\mu-(4k\alpha-1)\gamma-2\alpha(1+\gamma)\mu l-4\alpha\gamma l}{1-\gamma} \tag{B.57}$$

$$= \frac{1-[((2k-1)\alpha-1)(1-\gamma)+2\alpha(1+\gamma)l]\mu-[(4k+4l)\alpha-1]\gamma}{1-\gamma}. \tag{B.58}$$

Sachant que nous supposons que $0<\gamma<1$, cette inégalité est vérifiée si

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < 1-[((2k-1)\alpha-1)(1-\gamma)+2\alpha(1+\gamma)l]\mu-[(4k+4l)\alpha-1]\gamma. \tag{B.59}$$

En utilisant notre hypothèse sur γ, nous pouvons alors borner certains termes de notre inégalité précédente,

$$1-\gamma<1 \qquad \text{et} \qquad 2\alpha(1+\gamma)<4\alpha. \tag{B.60}$$

Ainsi, (B.59) est vérifiée si l'inégalité suivante est vraie,

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < 1-[(2k-1)\alpha-1+4\alpha l]\mu-[(4k+4l)\alpha-1]\gamma. \tag{B.61}$$

Finalement, l'égalité précédente est vérifiée si les inégalités suivantes sont vraies,

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < 1-[4k\alpha+4l\alpha-1]\mu-[4(k+l)\alpha-1]\gamma \tag{B.62}$$

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < 1-[4(k+l)\alpha-1](\mu+\gamma). \tag{B.63}$$

En utilisant notre hypothèse que $k+l=K$, nous obtenons alors la condition que

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} \leq 1-(4K\alpha-1)(\mu+\gamma); \tag{B.64}$$

ce qui est exactement (B.55) (en prenant le logarithme népérien). □

Le Lemme suivant nous donne une borne sur la cohérence mutuelle μ en réécrivant certaines conditions sur $\mu+\gamma$ et κ_l.

Lemme B.7 (Condition sur γ et μ). *Soient $k+l=K$ avec $k,l,K\in\mathbb{N}$, $K>0$. Soient $a\in\mathbb{R}^k$ et $\alpha\geq 1$. Soit $\kappa_l:=(1+\gamma)\mu l D+\gamma l(1+D^2)$ et supposons que $D<1$. Supposons également que*

$$\gamma+\mu<\frac{1}{4K\alpha-1} \qquad \text{avec} \qquad \gamma,\mu\in\mathbb{R}^+. \tag{B.65}$$

Alors

$$\mu<\frac{1-(4k\alpha-1)\gamma-2\kappa_l\alpha}{(2k\alpha-1)(1-\gamma)}. \tag{B.66}$$

Preuve. Puisque $D < 1$, nous avons alors que

$$\kappa_l = (1+\gamma)\mu l D + \gamma l(1+D^2) < (1+\gamma)\mu l + 2\gamma l \tag{B.67}$$

Nous pouvons alors en déduire que notre conclusion (B.66) est vérifiée si

$$\mu < \frac{1-(4k\alpha-1)\gamma - 2\alpha((1+\gamma)\mu l + 2\gamma l)}{(2k\alpha-1)(1-\gamma)}. \tag{B.68}$$

Parce que $0 < \gamma < 1$, nous avons alors que (B.68) est garantie si

$$(2k\alpha-1)\mu < 1-(4k\alpha-1)\gamma - 2\alpha((1+\gamma)\mu l + 2\gamma l); \tag{B.69}$$

qui, puisque $1+\gamma < 2$, est à son tour garantie si

$$(2k\alpha-1)\mu < 1-(4k\alpha-1)\gamma - 4\alpha l(\mu+\gamma). \tag{B.70}$$

En réarrangeant les termes de cette inégalité, nous obtenons alors que,

$$((2k+4l)\alpha-1)\mu + (4(k+l)\alpha-1)\gamma < 1. \tag{B.71}$$

Puisque $0 \leq k, l \leq K$ et $k+l = K$, l'inégalité précédente est garantie si

$$(4K\alpha-1)\mu + (4K\alpha-1)\gamma < 1, \tag{B.72}$$

ce qui est équivalent à notre hypothèse $\mu+\gamma < 1/(4K\alpha-1)$. □

Le Lemme suivant nous donne une borne numérique pour la norme de la différence entre les amplitudes du signal à reconstruire et de leurs estimations par la méthode des moindres carrés.

Lemme B.8 (Borne sur les amplitudes estimées)**.** *Soient* $y = \sum_{i=1}^{k} a_i \delta_{t_i} \in \Sigma_{K,\epsilon}$, *et* $s_1, \ldots, s_l \in \mathbb{R}^d$, *avec* $l, k \in \mathbb{N}$ *tels que* $k+l = K \geq 2$. *Supposons que* $\|s_i - t_i\|_2 \leq \xi < \epsilon/3$ *et* $\max_{i \neq j} |\langle \delta_{s_i}, \delta_{s_j} \rangle| \leq \mu$ *pour tout* $1 \leq i, j \leq l$. *Supposons également que* $\mu+\gamma \leq \frac{1}{5K(4K\alpha-1)}$ *avec* $\alpha \geq 1$. *De plus, supposons que l'opérateur linéaire* A *a une* γ*-RIP sur* $\mathcal{S}(\Sigma_{K,\frac{\epsilon}{3}})$ *(voir Définition 4). Soient*

$$\xi := \sqrt{2\sigma^2 \ln\left(\frac{1}{(1-(4K\alpha-1)(\mu+\gamma))}\right)}. \tag{B.73}$$

et

$$b \in \arg\min_{\tilde{b} \in \mathbb{R}^l} \left\| A(\sum_{i=1}^{l} \tilde{b}_i \delta_{s_i} - a_i \delta_{t_i}) - A \sum_{i=l+1}^{K} a_i \delta_{t_i} \right\|_2^2. \tag{B.74}$$

Alors

$$\frac{\|b - a_{1:l}\|_2}{\|a\|_\infty} \leq 0.77. \tag{B.75}$$

Preuve. Pour cette preuve, nous notons la borne supérieure de $\mu+\gamma \leq \frac{1}{\beta K(4K-1)}$ avec $\beta = 5$ pour simplifier les notations.

Soit $\pi = b_i \delta_{s_i} - a_i \delta_{t_i}$. Avec le Lemme B.1, nous avons que $\|\pi_i\|_h = \|b_i\delta_{s_i} - a_i\delta_{t_i}\|_h = \|(b_i - a_i)\delta_{s_i} - a_i(\delta_{t_i} - \delta_{s_i})\|_h \leq \|(b_i-a_i)\delta_{s_i}\|_h + |a_i|\|(\delta_{t_i} - \delta_{s_i})\|_h = |b_i - a_i| + |a_i|\|\delta_{t_i} - \delta_{s_i}\|_h$.

Considérons maintenant la matrice $B = (A\delta_{s_1}, \dots, A\delta_{s_l})$ avec B^H sa transposée conjuguée, nous avons que $b = (B^H B)^{-1} B^H y$ est la solution de l'estimation par moindres carrés (B.74) avec $y = A\sum_{i=1}^{K} a_i \delta_{t_i}$.

La norme $\|(B^H B)^{-1}\|_2^2$ est liée à la plus petite valeur propore de $B^H B$ que nous notons $\lambda_{min}(B^H B)$. En effet, nous avons que

$$\|(B^H B)^{-1}\|_{op}^2 = \frac{1}{|\lambda_{min}(B^H B)|^2}. \tag{B.76}$$

Afin de pouvoir trouver une borne sur $\lambda_{min}(B^H B)$, nous utilisons le théorème des cercles de Gershgorin [9]. Il existe alors $1 \leq i \leq l$, tels que,

$$\left|\lambda_{min}(B^H B) - \|A\delta_{s_i}\|_2^2\right| \leq \sum_{j \neq i} \left|\langle A\delta_{s_i}, A\delta_{s_j}\rangle\right|. \tag{B.77}$$

En utilisant l'inégalité triangulaire, nous avons

$$\left|\lambda_{min}(B^H B) - \|A\delta_{s_i}\|_2^2\right| \geq \|A\delta_{s_i}\|_2^2 - |\lambda_{min}(B^H B)|. \tag{B.78}$$

En multipliant par -1 et en injectant (B.77), nous obtenons

$$\left|\lambda_{min}(B^H B)\right| - \|A\delta_{s_i}\|_2^2 \geq -\sum_{j \neq i} \left|\langle A\delta_{s_i}, A\delta_{s_j}\rangle\right| \tag{B.79}$$

$$\left|\lambda_{min}(B^H B)\right| \geq \|A\delta_{s_i}\|_2^2 - \sum_{j \neq i} \left|\langle A\delta_{s_i}, A\delta_{s_j}\rangle\right| \tag{B.80}$$

$$\left|\lambda_{min}(B^H B)\right| \geq \|A\delta_{s_i}\|_2^2 - (l-1) \max_{i \neq j} \left|\langle A\delta_{s_i}, A\delta_{s_j}\rangle\right|. \tag{B.81}$$

En utilisant le Lemme B.3, nous avons que $\left|\langle A\delta_{s_i}, A\delta_{s_j}\rangle\right| \leq (1-\gamma)\langle \delta_{s_i}, \delta_{s_j}\rangle + 2\gamma$. Ainsi, avec la RIP, nous obtenons

$$|\lambda_{min}(B^H B)| \geq (1-\gamma)\|\delta_{s_i}\|_h^2 - 2(l-1)\gamma - (l-1)(1-\gamma) \max_{i \neq j} |\langle \delta_{s_i}, \delta_{s_j}\rangle| \tag{B.82}$$

$$= 1 - (2l-1)\gamma - (l-1)(1-\gamma) \max_{i \neq j} |\langle \delta_{s_i}, \delta_{s_j}\rangle|. \tag{B.83}$$

Avec la cohérence mutuelle (voir (3.4)) et grâce à l'hypothèse sur $\gamma + \mu$, nous avons

$$|\lambda_{min}(B^H B)| \geq 1 - (2l-1)\gamma - (l-1)(1-\gamma)\mu \tag{B.84}$$

$$\geq 1 - (2l-1)\gamma - (l-1)\mu \tag{B.85}$$

$$\geq 1 - 2K\gamma - K\mu \geq 1 - 2K(\gamma + \mu) \tag{B.86}$$

$$\geq 1 - \frac{2K}{\beta K(4K-1)} \geq 1 - \frac{2}{\beta(4K-1)}. \tag{B.87}$$

Comme $K \geq 2$, nous pouvons conclure que

$$|\lambda_{min}(B^H B)| \geq 1 - \frac{2}{7\beta}. \tag{B.88}$$

Maintenant que nous avons borné la plus petite valeur singulière de $B^H B$, nous pouvons alors calculer et borner $\|b - a_{1:l}\|_2^2$. Nous avons que

$$b - a_{1:l} = (B^H B)^{-1} B^H y - a_{1:l} = (B^H B)^{-1}(B^H y - B^H B a_{1:l}). \tag{B.89}$$

D'une part, nous avons

$$B^H y = \begin{bmatrix} \overline{A\delta_{s_1}} \\ \hline \vdots \\ \hline \overline{A\delta_{s_l}} \end{bmatrix} A \sum_{j=1}^{K} a_j \delta_{t_j} \quad \text{c.-à-d.} \quad \left[B^H y\right]_i = \sum_{j=1}^{K} a_j \langle A\delta_{s_i}, A\delta_{t_j}\rangle. \tag{B.90}$$

D'une autre part,

$$B^H B a_{1:l} = \begin{bmatrix} \langle A\delta_{s_1}, A\delta_{s_1}\rangle & \cdots & \langle A\delta_{s_1}, A\delta_{s_l}\rangle \\ \vdots & \ddots & \vdots \\ \langle A\delta_{s_l}, A\delta_{s_1}\rangle & \cdots & \langle A\delta_{s_l}, A\delta_{s_l}\rangle \end{bmatrix} \begin{bmatrix} a_1 \\ \hline \vdots \\ \hline a_l \end{bmatrix}. \tag{B.91}$$

Par conséquent,

$$\left[B^H B a_{1:l}\right]_i = \sum_{j=1}^{l} a_j \langle A\delta_{s_i}, A\delta_{s_j}\rangle. \tag{B.92}$$

En utilisant (B.89) et en combinant (B.90) avec (B.92), nous obtenons alors,

$$\|b - a_{1:l}\|_2^2 = \|(B^H B)^{-1}(B^H y - B^H B a_{1:l})\|_2^2 \tag{B.93}$$

$$\leq \sum_{i=1}^{l} \left| \sum_{j=1}^{K} a_j \langle A\delta_{s_i}, A\delta_{t_j}\rangle - \sum_{j=1}^{l} a_j \langle A\delta_{s_i}, A\delta_{s_j}\rangle \right|^2 \frac{1}{|\lambda_{min}(B^H B)|^2} \tag{B.94}$$

$$= \sum_{i=1}^{l} \left| \sum_{j=1}^{l} a_j \langle A\delta_{s_i}, A(\delta_{t_j} - \delta_{s_j})\rangle + \sum_{j=l+1}^{K} a_j \langle A\delta_{s_i}, A\delta_{t_j}\rangle \right|^2 \frac{1}{|\lambda_{min}(B^H B)|^2}. \tag{B.95}$$

Par inégalité triangulaire,

$$\|b - a_{1:l}\|_2^2 \leq \sum_{i=1}^{l} \left(\sum_{j=1}^{l} |a_j| |\langle A\delta_{s_i}, A(\delta_{t_j} - \delta_{s_j})\rangle| + \sum_{j=l+1}^{K} |a_j| |\langle A\delta_{s_i}, A\delta_{t_j}\rangle| \right)^2 \frac{1}{|\lambda_{min}(B^H B)|^2}. \tag{B.96}$$

Pour trouver la borne supérieure à cette équation, nous bornons chaque terme séparément. Premièrement, pour tout $1 \leq i \leq l,\ l+1 \leq j \leq K$, grâce au Lemme B.3, nous avons,

$$|\langle A\delta_{s_i}, A\delta_{t_j}\rangle| \leq (1-\gamma)\langle \delta_{s_i}, \delta_{t_j}\rangle_h + 2\gamma \leq (1-\gamma)\mu + 2\gamma. \tag{B.97}$$

Ensuite, pour tout $1 \leq i \leq l, 1 \leq j \leq l$, nous distinguons les cas où $i = j$ et $i \neq j$. Nous utilisons l'inégalité de Cauchy-Schwarz et le Lemme B.4,

$$|\langle A\delta_{s_i}, A(\delta_{t_j} - \delta_{s_j})\rangle| \leq \begin{cases} \|A\delta_{s_i}\|_2 \|A(\delta_{t_j} - \delta_{s_j})\|_2, & i = j \\ \left|\langle \delta_{s_i}, \delta_{t_j} - \delta_{s_j}\rangle_h\right| & \\ \quad + \gamma \left| \|\delta_{s_i}\|_h^2 + \|\delta_{t_j} - \delta_{s_j}\|_h^2 - \langle \delta_{s_i}, \delta_{t_j} - \delta_{s_j}\rangle_h \right|, & i \neq j \end{cases} \tag{B.98}$$

Avec la propriété que $|\langle \pi_1, \pi_2\rangle| \leq \mu \|\pi_1\|_h \|\pi_2\|_h$ pour π_1, π_2, deux dipôles $\frac{\epsilon}{3}$-séparés de notre Hypothèse 1, l'inégalité de Cauchy-Schwarz, et $\|A\pi\|_2 \leq \sqrt{1+\gamma}\|\pi\|_h$ avec la RIP,

nous obtenons alors la borne suivante,

$$|\langle A\delta_{s_i}, A(\delta_{t_j}-\delta_{s_j})\rangle| \leq \begin{cases} (1+\gamma)\|\delta_{t_j}-\delta_{s_j}\|_h, & i=j \\ \mu\|\delta_{t_j}-\delta_{s_j}\|_h & \\ \quad +\gamma\left(1+\|\delta_{t_j}-\delta_{s_j}\|_h^2+\left|\langle\delta_{s_i},\delta_{t_j}-\delta_{s_j}\rangle_h\right|\right), & i\neq j \end{cases} \tag{B.99}$$

$$\leq \begin{cases} (1+\gamma)\|\delta_{t_j}-\delta_{s_j}\|_h, & i=j \\ \mu\|\delta_{t_j}-\delta_{s_j}\|_h & \\ \quad +\gamma\left(1+\|\delta_{t_j}-\delta_{s_j}\|_h^2+\mu\|\delta_{t_j}-\delta_{s_j}\|_h\right), & i\neq j \end{cases} \tag{B.100}$$

$$= \begin{cases} (1+\gamma)\|\delta_{t_j}-\delta_{s_j}\|_h, & i=j \\ \mu(1+\gamma)\|\delta_{t_j}-\delta_{s_j}\|_h+\gamma\left(1+\|\delta_{t_j}-\delta_{s_j}\|_h^2\right), & i\neq j. \end{cases} \tag{B.101}$$

Nous utilisons le fait que $\|s_i-t_i\|_2 \leq \xi = \sqrt{2\sigma^2\ln\left(\frac{1}{(1-(4K\alpha-1)(\mu+\gamma))}\right)}$ de notre hypothèse (B.73),

$$|\langle A\delta_{s_i}, A(\delta_{t_j}-\delta_{s_j})\rangle| \leq \begin{cases} (1+\gamma)\sqrt{2(1-\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}})}, & i=j \\ \mu(1+\gamma)\sqrt{2(1-\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}})}+\gamma+2\gamma(1-\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}}), & i\neq j \end{cases} \tag{B.102}$$

$$\leq \begin{cases} (1+\gamma)\sqrt{2(4K\alpha-1)(\mu+\gamma)}, & i=j \\ \mu(1+\gamma)\sqrt{2(4K\alpha-1)(\mu+\gamma)} & \\ \quad +\gamma+2\gamma(4K\alpha-1)(\mu+\gamma), & i\neq j. \end{cases} \tag{B.103}$$

Avec l'hypothèse sur $\mu+\gamma$, nous avons,

$$|\langle A\delta_{s_i}, A(\delta_{t_j}-\delta_{s_j})\rangle| \leq \begin{cases} (1+\gamma)\sqrt{\frac{2}{\beta K}}, & i=j \\ \mu(1+\gamma)\sqrt{\frac{2}{\beta K}}+(1+\frac{2}{\beta K})\gamma, & i\neq j. \end{cases} \tag{B.104}$$

En injectant les bornes obtenues (B.97) et (B.104) dans l'inégalité (B.96), nous obtenons

$$\|b-a_{1:l}\|_2^2 \leq \frac{1}{|\lambda_{min}(B^HB)|^2}\sum_{i=1}^{l}\Bigg[(1+\gamma)\sqrt{\frac{2}{\beta K}}|a_i| + \left(\mu(1+\gamma)\sqrt{\frac{2}{\beta K}}+(1+\frac{2}{\beta K})\gamma\right)\sum_{\substack{j=1\\ i\neq j}}^{l}|a_j| + \Big((1-\gamma)\mu+2\gamma\Big)\sum_{j=l+1}^{K}|a_j|\Bigg]^2. \tag{B.105}$$

En divisant les deux membres de l'inégalité par $\|a\|_\infty^2$, nous obtenons la borne supérieure suivante,

$$\frac{\|b-a_{1:l}\|_2^2}{\|a\|_\infty^2} \leq \frac{l}{|\lambda_{min}(B^HB)|^2}\Bigg[(1+\gamma)\sqrt{\frac{2}{\beta K}} + \left(\mu(1+\gamma)\sqrt{\frac{2}{\beta K}}+(1+\frac{2}{\beta K})\gamma\right)(l-1) + \Big((1-\gamma)\mu+2\gamma\Big)(K-l)\Bigg]^2. \tag{B.106}$$

En prenant la racine carré des deux membres, nous avons

$$\frac{\|b-a_{1:l}\|_2}{\|a\|_\infty} \leq \frac{\sqrt{l}}{|\lambda_{min}(B^HB)|}\Bigg[(1+\gamma)\sqrt{\frac{2}{\beta K}} + \left(\mu(1+\gamma)\sqrt{\frac{2}{\beta K}} + (1+\frac{2}{\beta K})\gamma\right)(l-1) + \Big((1-\gamma)\mu+2\gamma\Big)(K-l)\Bigg]. \quad \text{(B.107)}$$

Nous prenons le maximum entre $\mu(1+\gamma)\sqrt{\frac{2}{\beta K}} + (1+\frac{2}{\beta K})\gamma$ et $(1-\gamma)\mu+2\gamma$ pour obtenir

$$\frac{\|b-a_{1:l}\|_2}{\|a\|_\infty} \leq \frac{\sqrt{l}}{|\lambda_{min}(B^HB)|}\Bigg[(1+\gamma)\sqrt{\frac{2}{\beta K}} + \max\left(\mu(1+\gamma)\sqrt{\frac{2}{\beta K}} + (1+\frac{2}{\beta K})\gamma, (1-\gamma)\mu+2\gamma\right)(K-1)\Bigg]. \quad \text{(B.108)}$$

En remarquant que

$$c_1 := \mu(1+\gamma)\sqrt{\frac{2}{\beta K}} + \left(1+\frac{2}{\beta K}\right)\gamma \leq 2\mu\sqrt{\frac{2}{\beta K}} + \left(1+\frac{2}{\beta K}\right)\gamma \leq 2(\mu+\gamma) \quad \text{(B.109)}$$

et

$$c_2 := (1-\gamma)\mu + 2\gamma \leq 2(\mu+\gamma), \quad \text{(B.110)}$$

nous avons alors $\max(c_1, c_2) \leq 2(\mu+\gamma)$ et

$$\frac{\|b-a_{1:l}\|_2}{\|a\|_\infty} \leq \frac{1}{|\lambda_{min}(B^HB)|}\sqrt{l}\left((1+\gamma)\sqrt{\frac{2}{\beta K}} + 2(\mu+\gamma)K\right) \quad \text{(B.111)}$$

$$\leq \frac{1}{|\lambda_{min}(B^HB)|}\sqrt{K}\left((1+\gamma)\sqrt{\frac{2}{\beta K}} + \frac{2}{\beta(4K\alpha-1)}\right) \quad \text{(B.112)}$$

$$\leq \frac{1}{|\lambda_{min}(B^HB)|}\left((1+\gamma)\sqrt{\frac{2}{\beta}} + \frac{2\sqrt{K}}{\beta(4K-1)}\right). \quad \text{(B.113)}$$

Pour $K \geq 2$, nous pouvons alors borner γ par

$$\gamma \leq \mu+\gamma \leq \frac{1}{\beta K(4K\alpha-1)} \leq \frac{1}{14\beta}. \quad \text{(B.114)}$$

Nous obtenons, pour $K \geq 2$,

$$\frac{\|b-a_{1:l}\|_2}{\|a\|_\infty} \leq \frac{\left(1+\frac{1}{14\beta}\right)\sqrt{\frac{2}{\beta}} + \frac{2\sqrt{K}}{\beta(4K-1)}}{|\lambda_{min}(B^HB)|} \leq \frac{\left(1+\frac{1}{14\beta}\right)\sqrt{\frac{2}{\beta}} + \frac{2\sqrt{2}}{7\beta}}{|\lambda_{min}(B^HB)|}. \quad \text{(B.115)}$$

En utilisant la borne (B.88) sur $\lambda_{min}(B^HB)$, nous en déduisons que

$$\frac{\|b-a_{1:l}\|_2}{\|a\|_\infty} \leq \frac{1}{1-\frac{2}{7\beta}}\left(\left(1+\frac{1}{14\beta}\right)\sqrt{\frac{2}{\beta}} + \frac{2\sqrt{2}}{7\beta}\right). \quad \text{(B.116)}$$

En choisissant $\beta = 5$, nous trouvons une valeur numérique pour la borne,

$$\frac{\|b-a_{1:l}\|_2}{\|a\|_\infty} \leq 0.77\,. \quad \text{(B.117)}$$

□

B.3 Preuve de l'hérédité avec amplitudes non fixées (Théorème 3.3)

Le proposition suivante indique que pour n'importe quel signal ϵ-séparé avec un résidu contrôlé, le résultat donné par le maximum de corrélation entre ce signal et une impulsion se trouve dans un périmètre de rayon ξ d'une impulsion de ce signal. Il s'agit de l'argument principal dans la démonstration du Théorème 3.3 (dont la preuve se trouve dans la Section B.4). Pour la preuve de l'hérédité avec amplitudes fixées pour le Théorème 3.4, voir la Section B.5.

Proposition B.1. *Soient $k, l, K \in \mathbb{N}$ tels que $k+l = K \geq 2$. Soit $x_{k,\epsilon} = \sum_{i=1}^{k} a_i \delta_{t_i} \in \Sigma_{K,\epsilon}$ où $\|a\|_\infty = |a_1| \geq |a_2| \geq \ldots \geq |a_K|$. Notons $\alpha = |a_1|/|a_K|$. Supposons que l'opérateur linéaire A a une γ-RIP sur $\mathcal{S}(\Sigma_{K,\frac{\epsilon}{3}})$ avec un noyau Gaussien h de variance σ^2.*
Soit

$$\xi := \sqrt{2\sigma^2 \ln\left(\frac{1}{(1-(4K\alpha-1)(\mu+\gamma))}\right)}. \tag{B.118}$$

Soit $X = x_{k,\epsilon} + z_l$ où $z_l = \sum_{i=k+1}^{K} a_i \delta_{t_i} - b_i \delta_{s_i} \in Z_{l,\xi,\frac{\epsilon}{3}}$ et $s^* \in \arg\max_{\tilde{s} \in \mathbb{R}^d} |\langle A\delta_{\tilde{s}}, AX\rangle|$.
Supposons

$$\mu + \gamma < \frac{1}{5K\,(4K\alpha - 1)} \tag{B.119}$$

et

$$\epsilon^2 > 18\sigma^2 \ln\left(\frac{10}{9}\right). \tag{B.120}$$

Alors, il existe un $i_0 \in \{1, \ldots, k\}$ tel que $\|s^* - t_{i_0}\|_2 < \xi < \frac{\epsilon}{3}$.

Preuve. **Bornes préliminaires.**

— Avec nos hypothèses (B.118) et (B.119), nous avons, pour $K \geq 2$,

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} = (1-(4K\alpha-1)(\mu+\gamma)) > 1 - \frac{1}{5K} \geq \frac{9}{10}. \tag{B.121}$$

Nous obtenons

$$2\left(1 - \mathrm{e}^{-\frac{\xi^2}{2\sigma^2}}\right) \leq 2\left(1 - \frac{9}{10}\right) = \frac{1}{5}. \tag{B.122}$$

En utilisant le Lemme B.2, nous avons

$$\frac{\|a_i\delta_{t_i} - b_i\delta_{s_i}\|_h}{\|a\|_\infty} \leq \frac{|a_i|\|\delta_{t_i} - \delta_{s_i}\|_h + |b_i - a_i|}{\|a\|_\infty}. \tag{B.123}$$

D'une part, $|a_i|/\|a\|_\infty \leq 1$. D'une autre part, utiliser le Lemme B.8 nous donne la garantie que $\|b - a_{1:l}\|_2/\|a\|_\infty \leq 0.77$. Nous pouvons alors en déduire que $|b_i - a_i|/\|a\|_\infty \leq 0.77$. Ceci combiné à (B.123), nous obtenons

$$\frac{\|a_i\delta_{t_i} - b_i\delta_{s_i}\|_h}{\|a\|_\infty} \leq \|\delta_{t_i} - \delta_{s_i}\|_h + 0.77 \leq \frac{1}{5} + 0.77 < 1. \tag{B.124}$$

Par conséquent,

$$D := \max_{i \in \{k+1,\ldots,K\}} \frac{\|a_i\delta_{t_i} - b_i\delta_{s_i}\|_h}{\|a\|_\infty} < 1. \tag{B.125}$$

L'inégalité (B.124) nous sera utile pour invoquer le Lemme B.6 ainsi que le Lemme B.7 dans la suite de cette preuve.

— À partir de (B.119), nous déduisons de cette hypothèse que

$$\mu + \gamma \leq \frac{1}{4K\alpha - 1}. \tag{B.126}$$

du Lemme B.7 est vérifié.

— Nous avons que (B.121) implique :

$$\frac{1}{1 - (4K\alpha - 1)(\mu + \gamma)} < \frac{10}{9} \tag{B.127}$$

et

$$\xi = \sqrt{2\sigma^2 \ln\left(\frac{1}{1 - 4(4K\alpha - 1)(\mu + \gamma)}\right)} < \sqrt{2\sigma^2 \ln\left(\frac{10}{9}\right)} = \frac{1}{3}\sqrt{18\sigma^2 \ln\left(\frac{10}{9}\right)}. \tag{B.128}$$

En utilisant l'hypothèse (B.120), nous obtenons

$$\xi < \frac{\epsilon}{3}. \tag{B.129}$$

Preuve principale de la Proposition B.1. Nous divisons l'ensemble $\mathbb{R}^d$ entre trois sets :

— $E_1 = \{s \in \mathbb{R}^d : \exists i_0 \in \{1, \ldots, k\} \ / \ \|s - t_{i_0}\|_2 < \xi\}$;

— $E_2 = \{s \in \mathbb{R}^d : \exists i_0 \in \{1, \ldots, k\} \ / \ \xi \leq \|s - t_{i_0}\|_2 < \frac{\epsilon}{3}\}$;

— $E_3 = \{s \in \mathbb{R}^d : \forall t_i, \|s - t_i\|_2 \geq \frac{\epsilon}{3}\}$.

Nous notons que E_1, E_2 et E_3 sont pair à pair disjoints et que $E_1 \cup E_2 \cup E_3 = \mathbb{R}^n$. De plus, vu que $\xi < \frac{\epsilon}{3}$, l'ensemble E_2 est non vide.

a) Construction de la preuve Notre idée dans cette preuve est de montrer que si $s^* \in \underset{\tilde{s} \in E_1 \cup E_2 \cup E_3}{\arg\max} |\langle A\delta_{\tilde{s}}, AX\rangle|$ alors $s^* \in E_1$. Pour procéder, nous séparons la preuve en deux parties.

Tout d'abord, nous choisissons un élément $s_1 \in E_1$ et n'importe quel élément $s_2 \in E_2$. Ainsi, en montrant que $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_2}, AX\rangle|$, nous rejetons la possibilité que $s^* \in E_2$.

Ensuite, nous choisissons un élément $s_1 \in E_1$ et n'importe quel élément $s_3 \in E_3$. En montrant que $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_3}, AX\rangle|$, nous rejetons également la possibilité que $s^* \in E_3$.

Ainsi, puisque $s^* \notin E_2 \cup E_3$ et que $E_1 \cup E_2 \cup E_3 = \mathbb{R}^n$, alors$s^* \in E_1$, c'est à dire, $\exists i_0 \ / \ \|s^* - t_{i_0}\|_2 < \xi$.

b) Comparaison entre $s_1 \in E_1$ et $s_2 \in E_2$ Soient $s_1 \in E_1$ et i_0 tels que $\|s_1 - t_{i_0}\|_2 \leq \xi$. Comme s^* ne change pas si l'on multiplie X par -1 (grâce à la valeur absolue), nous supposons que $a_0 \geq 0$ sans perte de généralité. Nous avons alors

$$|\langle A\delta_{s_1}, AX\rangle| \geq \langle A\delta_{s_1}, AX\rangle \tag{B.130}$$

$$= \langle A\delta_{s_1}, Aa_{i_0}\delta_{t_{i_0}} + A\sum_{i=1; i\neq i_0}^{k} a_i\delta_{t_i} + Az_l\rangle \tag{B.131}$$

$$= \langle A\delta_{s_1}, Aa_{i_0}\delta_{t_{i_0}}\rangle + \langle A\delta_{s_1}, A\sum_{i=1; i\neq i_0}^{k} a_i\delta_{t_i}\rangle + \langle A\delta_{s_1}, Az_l\rangle. \tag{B.132}$$

Nous nous concentrons d'abord sur le terme $\langle A\delta_{s_1}, A\sum_{i=1;i\neq i_0}^{k} a_i\delta_{t_i}\rangle$. Nous découpons cette somme en deux termes, avec $I = \{i \neq i_0 : a_i \geq 0\}$ et $J = \{i \neq i_0 : a_i < 0\}$ pour obtenir,

$$\langle A\delta_{s_1}, A\sum_{i=1;i\neq i_0}^{k} a_i\delta_{t_i}\rangle = \sum_{i=1;i\neq i_0}^{k} a_i\langle A\delta_{s_1}, A\delta_{t_i}\rangle \tag{B.133}$$

$$= \sum_{i\in I} |a_i|\langle A\delta_{s_1}, A\delta_{t_i}\rangle - \sum_{i\in J} |a_i|\langle A\delta_{s_1}, A\delta_{t_i}\rangle. \tag{B.134}$$

En utilisant le Lemme B.3, nous avons

$$\begin{aligned}\langle A\delta_{s_1}, A\sum_{i=1;i\neq i_0}^{k} a_i\delta_{t_i}\rangle &\geq (1+\gamma)\sum_{i\in I} |a_i|\langle\delta_{s_1}, \delta_{t_i}\rangle_h - 2\gamma\sum_{i\in I}|a_i| \\ &\quad - (1-\gamma)\sum_{i\in J} |a_i|\langle\delta_{s_1}, \delta_{t_i}\rangle_h - 2\gamma\sum_{i\in J}|a_i| \end{aligned} \tag{B.135}$$

$$\begin{aligned}&= (1+\gamma)\sum_{i\in I} |a_i|\langle\delta_{s_1}, \delta_{t_i}\rangle_h - 2\gamma\sum_{i\in I}|a_i| \\ &\quad - (1-\gamma)\sum_{i\in J} |a_i|\mathrm{e}^{-\frac{\|s_1-t_i\|_2^2}{2\sigma^2}} - 2\gamma\sum_{i\in J}|a_i|. \end{aligned} \tag{B.136}$$

Puisque nous avons que $\|t_{i_0} - s_1\|_2 \leq \xi < \frac{\epsilon}{3}$ et que les t_i sont tous deux à deux ϵ-séparés, nous utilisons alors l'inégalité triangulaire pour obtenir avec $i \neq i_0$,

$$\|t_{i_0}-s_1\|_2+\|s_1-t_i\|_2 \geq \|t_{i_0}-t_i\|_2 \quad \text{c.-à-d.} \quad \frac{\epsilon}{3}+\|s_1-t_i\|_2 > \epsilon \quad \text{so that} \quad \|s_1-t_i\|_2 > \frac{2\epsilon}{3}. \tag{B.137}$$

Cela implique de pour $i \neq i_0$, avec le point (3.) du Lemme 3.1, nous avons $\mathrm{e}^{-\|s_1-t_i\|_2^2/(2\sigma^2)} \leq \mu$. En injectant ce résultat dans (B.136), nous obtenons

$$\begin{aligned}\langle A\delta_{s_1}, A\sum_{i=1;i\neq i_0}^{k} a_i\delta_{t_i}\rangle &\geq (1+\gamma)\sum_{i\in I} |a_i|\langle\delta_{s_1}, \delta_{t_i}\rangle_h \\ &\quad - 2\gamma\sum_{i\in I}|a_i| - (1-\gamma)\mu\sum_{i\in J}|a_i| - 2\gamma\sum_{i\in J}|a_i|. \end{aligned} \tag{B.138}$$

Puisque $\forall i \in I, \langle\delta_{s_1}, \delta_{t_i}\rangle_h \geq 0$ et $|I| + |J| = k - 1$, nous avons que

$$\langle A\delta_{s_1}, A\sum_{i=1;i\neq i_0}^{k} a_i\delta_{t_i}\rangle \geq -(1-\gamma)\mu\sum_{i\in J}|a_i| - 2\gamma\sum_{i=1;i\neq i_0}^{k}|a_i|. \tag{B.139}$$

Dans le pire des cas, toutes les amplitudes a_i sont négatives (c.-à-d. $|J| = k - 1$), cela implique que

$$\langle A\delta_{s_1}, A\sum_{i=1;i\neq i_0}^{k} a_i\delta_{t_i}\rangle \geq -\left((1-\gamma)\mu + 2\gamma\right)\sum_{i=1;i\neq i_0}^{k}|a_i|. \tag{B.140}$$

En revenant à $|\langle A\delta_{s_1}, AX\rangle|$, nous utilisons le Lemme B.5 avec la borne supérieure $\kappa_l \geq |\langle A\delta_{s_1}, Az_l\rangle|/\|a\|_\infty$ définie dans (B.47). En utilisant le Lemme B.3, avec (B.132) et (B.140),

nous obtenons alors que $a_{i_0} \geq 0$,

$$|\langle A\delta_{s_1}, AX\rangle| \geq (1+\gamma)|a_{i_0}|\langle \delta_{s_1}, \delta_{t_{i_0}}\rangle_h - 2|a_{i_0}|\gamma - ((1-\gamma)\mu + 2\gamma) \sum_{i=1;i\neq i_0}^{k} |a_i| - \|a\|_\infty \kappa_l \tag{B.141}$$

$$= (1+\gamma)|a_{i_0}|\mathrm{e}^{-\frac{\|s_1 - t_{i_0}\|_2^2}{2\sigma^2}} - (1-\gamma)\mu \sum_{i=1;i\neq i_0}^{k} |a_i| - 2\gamma \sum_{i=1}^{k} |a_i| - \|a\|_\infty \kappa_l. \tag{B.142}$$

Maintenant que nous avons obtenu une borne inférieure pour $|\langle A\delta_{s_1}, AX\rangle|$, nous procédons maintenant au calcul d'une borne supérieure pour $|\langle A\delta_{s_2}, AX\rangle|$,

$$|\langle A\delta_{s_2}, AX\rangle| = \left| \sum_{i=1}^{k} a_i \langle A\delta_{s_2}, A\delta_{t_i}\rangle + \langle A\delta_{s_2}, Az\rangle \right| \leq \sum_{i=1}^{k} |a_i \langle A\delta_{s_2}, A\delta_{t_i}\rangle| + |\langle A\delta_{s_2}, Az\rangle|. \tag{B.143}$$

Il existe un j_0 tel que $\xi \leq \|t_{j_0} - s^*\|_2 < \epsilon/3$. Avec le Lemme B.5, nous obtenons une borne supérieure de $|\langle A\delta_{s_2}, Az\rangle| / \|a\|_\infty$,

$$|\langle A\delta_{s_2}, AX\rangle| \leq |a_{j_0}||\langle A\delta_{s_2}, A\delta_{t_{j_0}}\rangle| + \sum_{i=1;i\neq j_0}^{k} |a_i||\langle A\delta_{s_2}, A\delta_{t_i}\rangle| + \|a\|_\infty \kappa_l. \tag{B.144}$$

Nous utilisons la borne supérieure donnée par le Lemme B.3 pour obtenir,

$$\begin{aligned} |\langle A\delta_{s_2}, AX\rangle| \leq{}& (1-\gamma)|a_{j_0}|\langle \delta_{s_2}, \delta_{t_{j_0}}\rangle_h + 2\gamma|a_{j_0}| \\ &+ (1-\gamma) \sum_{i=1;i\neq j_0}^{k} |a_i|\langle \delta_{s_2}, \delta_{t_i}\rangle_h + 2\gamma \sum_{i=1;i\neq j_0}^{k} |a_i| + \|a\|_\infty \kappa_l \end{aligned} \tag{B.145}$$

$$\begin{aligned} |\langle A\delta_{s_2}, AX\rangle| \leq{}& (1-\gamma)|a_{j_0}|\mathrm{e}^{-\frac{\|s_2 - t_{j_0}\|_2^2}{2\sigma^2}} + 2\gamma|a_{j_0}| \\ &+ (1-\gamma) \sum_{i=1;i\neq j_0}^{k} |a_i|\mathrm{e}^{-\frac{\|s_2 - t_i\|_2^2}{2\sigma^2}} + 2\gamma \sum_{i=1;i\neq j_0}^{k} |a_i| + \|a\|_\infty \kappa_l. \end{aligned} \tag{B.146}$$

Puisque $\|t_{j_0} - s_2\|_2 \leq \frac{\epsilon}{3} < \frac{\epsilon}{2}$, grâce à l'inégalité triangulaire, nous obtenons que pour $i \neq j_0$,

$$\|t_{j_0} - s_2\|_2 + \|s_2 - t_i\|_2 \geq \|t_{j_0} - t_i\|_2 \quad \text{c.-à-d.} \quad \frac{\epsilon}{2} + \|s_2 - t_i\|_2 > \epsilon \quad \text{donc} \quad \|s_2 - t_i\|_2 > \frac{\epsilon}{2}. \tag{B.147}$$

En utilisant la propriété dans (B.146) et par le fait que, comme dans (3.4), pour tout $i \neq j_0$, $\mathrm{e}^{-\|s_2 - t_{j_0}\|_2^2/(2\sigma^2)} \leq \mu$, nous obtenons

$$|\langle A\delta_{s_2}, AX\rangle| \leq (1-\gamma)|a_{j_0}|\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} + (1-\gamma)\mu \sum_{i=1;i\neq j_0}^{k} |a_i| + 2\gamma \sum_{i=1}^{k} |a_i| + \|a\|_\infty \kappa_l. \tag{B.148}$$

Une condition suffisante pour prouver que $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_2}, AX\rangle|$ est de montrer que l'inégalité centrale suivante est vraie,

$$\begin{aligned} &|\langle A\delta_{s_1}, AX\rangle| \geq (1+\gamma)|a_{i_0}|\mathrm{e}^{-\frac{\|s_1 - t_{i_0}\|_2^2}{2\sigma^2}} - (1-\gamma)\mu \sum_{i=1;i\neq i_0}^{k} |a_i| - 2\gamma \sum_{i=1}^{k} |a_i| - \|a\|_\infty \kappa_l > \\ &(1-\gamma)|a_{j_0}|\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} + (1-\gamma)\mu \sum_{i=1;i\neq j_0}^{k} |a_i| + 2\gamma \sum_{i=1}^{k} |a_i| + \|a\|_\infty \kappa_l \geq |\langle A\delta_{s_2}, AX\rangle|. \end{aligned} \tag{B.149}$$

Considérons maintenant n'importe quel $s_1 \in E_1$, si pour tout $s_2 \in E_2$, l'inégalité précédente (B.149) est vérifié, alors $s^* \notin E_2$. Nous prenons alors $s_1 = t_1 = t_{i_0} \in E_1$ avec $i_0 = 1$ et, nous réécrivons l'inégalité centrale dans (B.149) comme

$$(1-\gamma)|a_{j_0}|\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < (1+\gamma)|a_1| - (1-\gamma)\mu\left(\sum_{i=2}^{k}|a_i| + \sum_{i=1;i\neq j_0}^{k}|a_i|\right) - 4\gamma\sum_{i=1}^{k}|a_i| - 2\|a\|_\infty\kappa_l. \quad \text{(B.150)}$$

En divisant les deux membres de l'inégalité par $(1-\gamma)|a_{j_0}|$, nous obtenons alors que l'inégalité précédente est vérifiée si la suivante est vérifiée,

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < \frac{1}{(1-\gamma)|a_{j_0}|}\left[|a_1| - (1-\gamma)\mu\left(\sum_{i=2}^{k}|a_i| + \sum_{i=1;i\neq j_0}^{k}|a_i|\right) - \gamma\left(4\left(\sum_{i=1}^{k}|a_i|\right) - |a_1|\right) - 2\|a\|_\infty\kappa_l\right]. \quad \text{(B.151)}$$

Si nous réécrivons $\sum_{i=1}^{k}|a_i|$ par $\|a\|_1$, nous obtenons alors

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} = \frac{1}{1-\gamma}\left[\frac{|a_1|}{|a_{j_0}|} - (1-\gamma)\mu\left(\frac{\|a\|_1 - |a_1|}{|a_{j_0}|} + \frac{\|a\|_1 - |a_{j_0}|}{|a_{j_0}|}\right) - \frac{4\|a\|_1 - |a_1|}{|a_{j_0}|}\gamma - 2\frac{\|a\|_\infty}{|a_{j_0}|}\kappa_l\right]. \quad \text{(B.152)}$$

Avec $\alpha = |a_1|/|a_k| = \|a\|_\infty / \min_i |a_i|$ alors $(\|a\|_1 - |a_1|)/|a_{j_0}| \leq (k-1)\alpha$, $(\|a\|_1 - |a_{j_0}|)/|a_{j_0}| \leq k\alpha - 1$ et $(4\|a\|_1 - |a_1|)/|a_{j_0}| \leq (4k-1)\alpha$. Nous en déduisons alors que l'inégalité précédente est vérifiée si

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < \frac{1 - ((2k-1)\alpha - 1)(1-\gamma)\mu - (4k\alpha - 1)\gamma - 2\kappa_l\alpha}{1-\gamma}. \quad \text{(B.153)}$$

Grâce au Lemme B.6, puisque (B.118) est vérifiée et que nous avons montré avec (B.125) que $D < 1$, nous avons alors que (B.153) est vérifiée et que à son tour $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_2}, AX\rangle|$ l'est également. Nous en déduisons alors que $|\langle A\delta_{s^*}, AX\rangle| \geq |\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_2}, AX\rangle|$.

c) Comparaison entre $s_1 \in E_1$ et $s_3 \in E_3$ Dans ce cas là, nous utilisons également la même borne inférieure (B.142) pour $|\langle A\delta_{s_1}, AX\rangle|$. Soit $s_3 \in E_3$,

$$|\langle A\delta_{s_3}, AX\rangle| = \left|\sum_{i=1}^{k} a_i\langle A\delta_{s_3}, A\delta_{t_i}\rangle + \langle A\delta_{s_3}, Az_l\rangle\right| \leq \sum_{i=1}^{k}|a_i\langle A\delta_{s_3}, A\delta_{t_i}\rangle| + |\langle A\delta_{s_3}, Az_l\rangle|. \quad \text{(B.154)}$$

Nous utilisons le Lemme B.5 et la borne supérieure κ_l définie dans (B.47) pour obtenir

$$|\langle A\delta_{s_3}, AX\rangle| \leq \sum_{i=1}^{k}|a_i||\langle A\delta_{s_3}, A\delta_{t_i}\rangle| + \|a\|_\infty\kappa_l. \quad \text{(B.155)}$$

En utilisant le Lemme B.3,

$$|\langle A\delta_{s_3}, AX\rangle| \leq (1-\gamma)\sum_{i=1}^{k}|a_i|\langle \delta_{s_3}, \delta_{t_i}\rangle_h + 2\gamma\sum_{i=1}^{k}|a_i| + \|a\|_\infty \kappa_l. \tag{B.156}$$

Puisque $s_3 \in E_3$, nous avons alors que $\forall i \in \{1,\dots,K\}, \|s_3 - t_i\|_2 \geq \frac{\epsilon}{3}$ et grâce au Lemme 3.1 (iii), nous obtenons

$$|\langle A\delta_{s_3}, AX\rangle| \leq (1-\gamma)\mu\sum_{i=1}^{k}|a_i| + 2\gamma\sum_{i=1}^{k}|a_i| + \|a\|_\infty \kappa_l. \tag{B.157}$$

Une condition suffisante pour prouver que $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_3}, AX\rangle|$ est de montrer que l'inégalité suivante est vraie,

$$(1+\gamma)|a_{i_0}|\mathrm{e}^{-\frac{\|s_1-t_{i_0}\|_2^2}{2\sigma^2}} - (1-\gamma)\mu\sum_{i=i;i\neq i_0}^{k}|a_i| - 2\gamma\sum_{i=1}^{k}|a_i| - \|a\|_\infty\kappa_l > (1-\gamma)\mu\sum_{i=1}^{k}|a_i| + 2\gamma\sum_{i=1}^{k}|a_i| + \|a\|_\infty\kappa_l. \tag{B.158}$$

Considérons maintenant n'importe quel $s_1 \in E_1$. Si pour tout $s_3 \in E_3$, l'inégalité précédente (B.158) est vérifiée, alors $s^* \notin E_3$. Prenons alors $s_1 = t_1 = t_{i_0} \in E_1$ avec $i_0 = 1$, nous réécrivons (B.158) comme

$$(1+\gamma)|a_1| - 2\kappa_l\|a\|_\infty > (1-\gamma)\mu\left(2\sum_{i=1}^{k}|a_i| - |a_1|\right) + 4\gamma\sum_{i=1}^{k}|a_i|. \tag{B.159}$$

En isolant μ, nous obtenons

$$\mu < \frac{|a_1| - (4\sum_{i=1}^{k}|a_i| - |a_1|)\gamma - 2\kappa_l\|a\|_\infty}{(2\sum_{i=1}^{k}|a_i| - |a_1|)(1-\gamma)} = \frac{1 - (4\|a\|_1/|a_1| - 1)\gamma - 2\kappa_l\|a\|_\infty/|a_1|}{(2\|a\|_1/|a_1| - 1)(1-\gamma)}. \tag{B.160}$$

Cette inégalité est vérifiée si

$$\mu < \frac{1 - (4k\alpha - 1)\gamma - 2\kappa_l\alpha}{(2k\alpha - 1)(1-\gamma)}. \tag{B.161}$$

Comme montré précédemment dans le Lemme B.7, grâce à notre borne supérieure sur $\mu + \gamma$ (dans (B.126)) et au fait que $D < 1$ (dans (B.125)), nous avons que (B.161) est vraie. Nous en déduisons alors que $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_3}, AX\rangle|$ et $|\langle A\delta_{s^*}, AX\rangle| \geq |\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_3}, AX\rangle|$.

d) Conclusion : Nous avons finalement montré que pour tout élément $s \in E_2 \cup E_3$, il existe un élément $s_1 \in E_1$ tel que $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_s, AX\rangle|$. Cela implique que s^* appartient nécessairement à l'ensemble E_1 comme $E_1 \cap (E_2 \cup E_3) = \emptyset$, c'est à dire qu'il existe un i_0 tel que $\|s^* - t_{i_0}\|_2 < \xi$. □

B.4 Preuve principale des Théorèmes 3.3 et 3.4

Nous pouvons maintenant prouver le Théorème 3.3 avec un raisonnement par récurrence.

Preuve du Théorème (3.3). Définissons l'ensemble suivant

$$\mathcal{X}_{K,l}^{\epsilon,\xi} := \Bigg\{ x_{k,\epsilon} + z_{l,\xi,\frac{\epsilon}{3}} : k = K-l;\ x_{k,\epsilon} = \sum_{i=1}^{k} a_i \delta_{t_i} \in \Sigma_{K,\epsilon};$$
$$z_{l,\xi,\frac{\epsilon}{3}} = \sum_{i=k+1}^{k+l} a_i \delta_{t_i} - b_i \delta_{s_i} \in Z_{l,\xi,\frac{\epsilon}{3}};\ \{t_i, i = 1,\dots,K\} \text{ deux à deux } \epsilon\text{-séparés} \Bigg\}. \quad \text{(B.162)}$$

Nous définissons également $X_{K,0} := x_{K,\epsilon}$ et, pour $l \in \{1,\dots,K-1\}$, $X_{K,l} = x_{K,\epsilon} - \sum_{i=1}^{l} b_i \delta_{s_i^*}$, où b_i correspond à l'amplitude associée à la sortie d'OP-COMP à l'itération i. Notre but est donc de montrer par récurrence que pour tout $l = 0,\dots,K$, nous avons que $X_{K,l} \in \mathcal{X}_{K,l}^{\epsilon,\xi}$.

a) Définition de la propriété de récurrence : Pour tout $l \in \{0,\dots,K\}, X_{K,l} \in \mathcal{X}_{K,l}^{\epsilon,\xi}$. Soit $\mathcal{P}(l) : X_{K,l} \in \mathcal{X}_{K,l}^{\epsilon,\xi}$. Nous cherchons à appliquer un raisonnement par récurrence sur l.

b) Initialisation de la récurrence $\mathcal{P}(0)$: Nous avons $X_{K,0} = x_{K,\epsilon} \in \Sigma_{K,\epsilon}$. En utilisant la définition de $\mathcal{X}_{K,0}^{\epsilon,\xi}$, nous avons que $X_{K,0} = x_{K,\epsilon} + z_{0,\xi,\frac{\epsilon}{3}} \in \mathcal{X}_{K,0}^{\epsilon,\xi}$ avec $z_{0,\xi,\frac{\epsilon}{3}} = 0$, $\mathcal{P}(0)$ est vraie.

c) Hérédité de la récurrence : Soit un $l \in \{0,\dots,K-1\}$ et $k = K-l$. Supposons que $\mathcal{P}(l)$ soit vraie. Nous cherchons donc à montrer que $\mathcal{P}(l+1)$ est également vraie.
Afin d'alléger les notations, nous considérons la permutation $\tilde{t} \in \mathbb{R}^{K\times d}$ des positions t (et leur permutations associées $\tilde{a} \in \mathbb{R}^K$ des amplitudes a) de telle manière à ce que pour les l premières sorties de COMP sans descente $s_1^*,\dots,s_l^*$, nous ayons $\forall i = 1,\dots,l+1,\ \|\tilde{t}_{K-l+i} - s_i^*\|_2 < \xi$. Les hypothèses de la Proposition B.1 sont vérifiées pour $X_{K,l}$, donc il existe alors un $i_0 \in \{1,\dots,k-l\}$ tel que $\|s_{l+1}^* - \tilde{t}_{i_0}\|_2 < \xi < \frac{\epsilon}{3}$ et

$$X_{K,l+1} = x_{K,\epsilon} - \sum_{j=1}^{l+1} b_j \delta_{s_j^*} = \sum_{i=1}^{K} \tilde{a}\delta_{\tilde{t}_i} - \sum_{j=1}^{l+1} b_j \delta_{s_j^*} \quad \text{(B.163)}$$
$$= \underbrace{\sum_{i=1;\, i\neq i_0}^{k} \tilde{a}_i \delta_{\tilde{t}_i}}_{=x_{k-1,\epsilon}\in\Sigma_{k-1,\epsilon}} + \underbrace{\sum_{i=K-l+1}^{K} \tilde{a}_i \delta_{\tilde{t}_i} - \sum_{j=1}^{l} b_j \delta_{s_j^*} + \tilde{a}_{i_0}\delta_{\tilde{t}_{i_0}} - b_{l+1}\delta_{s_{l+1}^*}}_{=z_{l,\xi,\frac{\epsilon}{3}}\in Z_{l,\xi,\frac{\epsilon}{3}}}. \quad \text{(B.164)}$$

Pour montrer que $z_{l,\xi,\frac{\epsilon}{3}} + \tilde{a}_{i_0}\delta_{\tilde{t}_{i_0}} - b_{l+1}\delta_{s_{l+1}^*} \in Z_{l+1,\xi,\frac{\epsilon}{3}}$, comme nous nous sommes assurés que $\pi_{l+1} := \tilde{a}_{i_0}\delta_{\tilde{t}_{i_0}} - b_{l+1}\delta_{s_{l+1}^*}$ est un dipôle ξ-concentré, nous avons juste besoin de vérifier que π_{l+1} est $\frac{\epsilon}{3}$-séparé deux à deux avec tout les autres dipôles $\pi_i := \tilde{a}_{K-l+i}\,\delta_{\tilde{t}_{K-l+i}} - b_i\delta_{s_i^*}$ pour $i = 1,\dots,l$.

d) Dipôles $\frac{\epsilon}{3}$-séparés En utilisant l'inégalité triangulaire inversée, nous avons pour tout $i \in \{1, \dots, l\}$, $\|\tilde{t}_i - s^*_{l+1}\|_2 \geq \|\tilde{t}_i - \tilde{t}_{i_0}\|_2 - \|\tilde{t}_{i_0} - s^*_{l+1}\|_2 \geq \epsilon - \xi > \epsilon - \frac{\epsilon}{3} > \frac{\epsilon}{3}$.
Pour contrôler les distances $\|s^*_{l+1} - s^*_i\|_2$ pour $i = 1, \dots, l$, nous utilisons le fait que $\|\tilde{t}_{K-l+i} - s^*_i\|_2 \leq \xi < \frac{\epsilon}{3}$ (de $\mathcal{P}(l)$). En commençant par $\|\tilde{t}_{i_0} - \tilde{t}_{K-l+i}\|_2 > \epsilon$, nous avons que

$$\|\tilde{t}_{i_0} - \tilde{t}_{K-l+i}\|_2 > \epsilon \tag{B.165}$$

c.-à-d.
$$\|\tilde{t}_{i_0} - s^*_{l+1} + s^*_{l+1} - s^*_i + s^*_i - \tilde{t}_{K-l+i}\|_2 > \epsilon \tag{B.166}$$

alors
$$\|\tilde{t}_{i_0} - s^*_{l+1}\|_2 + \|s^*_{l+1} - s^*_i\|_2 + \|s^*_i - \tilde{t}_{K-l+i}\|_2 > \epsilon \tag{B.167}$$

ainsi
$$\frac{\epsilon}{3} + \|s^*_{l+1} - s^*_i\|_2 + \frac{\epsilon}{3} > \epsilon \tag{B.168}$$

donc
$$\|s^*_{l+1} - s^*_i\|_2 > \frac{\epsilon}{3}. \tag{B.169}$$

Nous avons montré que tout élément de l'ensemble $\{s^*_i, i = 1, \dots, l\}$ est $\frac{\epsilon}{3}$-séparé avec s^*_{l+1}. Puisque les supports des dipôles $\{\pi_i, i = 1, \dots, l\}$ sont $\frac{\epsilon}{3}$-séparés avec le support du dipôle π_{l+1}, nous avons que $z_{l,\xi,\frac{\epsilon}{3}} + \tilde{a}_{i_0}\delta_{\tilde{t}_{i_0}} - b_{l+1}\delta_{s^*_{l+1}} \in Z_{l+1,\xi,\frac{\epsilon}{3}}$.
Finalement, en posant $z_{l+1,\xi,\frac{\epsilon}{3}} = z_{l,\xi,\frac{\epsilon}{3}} + \tilde{a}_{i_0}\delta_{\tilde{t}_{i_0}} - b_{l+1}\delta_{s^*_{l+1}}$, nous obtenons que

$$X_{K,l+1} = x_{k-1,\epsilon} + z_{l+1,\xi,\frac{\epsilon}{3}} \in \mathcal{X}^{\epsilon,\xi}_{K,l+1}. \tag{B.170}$$

Cela équivaut à dire que la propriété $\mathcal{P}(l+1)$ est vraie, terminant ainsi l'hérédité de la récurrence.

e) Conclusion Comme la base (b) et l'hérédité (c) sont tout deux vrais, par récurrence, la propriété $\mathcal{P}(l)$ est donc vraie pour tout $l \in \{0, \dots, K\}$, ce qui implique que pour tout $l \in \{0, \dots, K\}$, il existe un $i \in \{0, \dots, K\}$, $\|s^*_l - t_i\|_2 \leq \xi$. □

B.5 Preuve de l'hérédité avec amplitudes fixées (Théorème 3.4)

Cette preuve est en essence similaire (dans sa structure) à la preuve du Théorème 3.3. Pour ce qui est du raisonnement par récurrence, il est identique et n'a donc pas besoin d'être modifié. Nous avons seulement besoin de modifier l'hérédité avec la Proposition suivante.

Proposition B.2. *Soient $k, l, K \in \mathbb{N}$ tels que $k + l = K \geq 2$. Soit $x_{k,\epsilon} = \sum_{i=1}^{k} \delta_{t_i} \in \Sigma_{K,\epsilon}$. Supposons que l'opérateur linéaire A a une γ-RIP sur $\mathcal{S}(\Sigma_{K,\frac{\epsilon}{3}})$ avec un noyau Gaussien h de variance σ^2. Soit*

$$\xi := \sqrt{2\sigma^2 \ln\left(\frac{1}{(1 - (4K-1)(\mu+\gamma))}\right)}. \tag{B.171}$$

Soient $X = x_{k,\epsilon} + z_l$ où $z_l = \sum_{i=k+1}^{K} \delta_{t_i} - \delta_{s_i} \in Z_{l,\xi,\frac{\epsilon}{3}}$ et $s^* \in \arg\max_{\tilde{s} \in \mathbb{R}^d} |\langle A\delta_{\tilde{s}}, AX\rangle|$.

Supposons

$$\mu + \gamma < \frac{1}{10\,(4K-1)} \tag{B.172}$$

et

$$\epsilon^2 > 18\sigma^2 \ln\left(\frac{10}{9}\right). \tag{B.173}$$

Alors, il existe un $i_0 \in \{1, \dots, k\}$ tel que $\|s^* - t_{i_0}\|_2 < \xi < \frac{\epsilon}{3}$.

Cette preuve est similaire à celle de la Proposition B.1. Les seuls éléments qui diffèrent sont les hypothèses qui sont moins strictes.

Preuve. **Bornes préliminaires.**

— Avec nos hypothèses (B.171) et (B.172), nous avons que, pour $K \geq 2$,

$$e^{-\frac{\xi^2}{2\sigma^2}} = 1 - (4K-1)(\mu+\gamma) > 1 - \frac{1}{10} = \frac{9}{10}. \tag{B.174}$$

Nous obtenons

$$2\left(1 - e^{-\frac{\xi^2}{2\sigma^2}}\right) \leq 2\left(1 - \frac{9}{10}\right) = \frac{1}{5}. \tag{B.175}$$

Nous pouvons alors borner $\|\delta_{t_i} - \delta_{s_i}\|_h$ pour tout $i \in \{k+1, \dots, K\}$,

$$\|\delta_{t_i} - \delta_{s_i}\|_h = \|\delta_{t_i}\|_h + \|\delta_{s_i}\|_h - 2\langle \delta_{t_i}, \delta_{s_i} \rangle_h = 2(1 - e^{\frac{\|t_i - s_i\|_2^2}{2\sigma^2}}) \tag{B.176}$$

$$\text{et donc} \quad \|\delta_{t_i} - \delta_{s_i}\|_h < 2\left(1 - e^{-\frac{\xi^2}{2\sigma^2}}\right) = \frac{1}{5}. \tag{B.177}$$

Ainsi,

$$D := \max_{i \in \{k+1,\dots,K\}} \|\delta_{t_i} - \delta_{s_i}\|_h < 1. \tag{B.178}$$

Cette inégalité (B.178) nous sera utile pour utiliser le Lemme B.6 ainsi que le Lemme B.7 plus tard dans cette preuve.

— À partir de (B.172), nous pouvons déduire que l'hypothèse

$$\mu + \gamma \leq \frac{1}{4K\alpha - 1}. \tag{B.179}$$

du Lemme B.7 est vérifiée.

— Nous avons que (B.174) implique $\left(1 - (4K-1)(\mu+\gamma)\right)^{-1} < \frac{10}{9}$, et donc

$$\xi = \sqrt{2\sigma^2 \ln\left(\frac{1}{1-(4K-1)(\mu+\gamma)}\right)} < \sqrt{2\sigma^2 \ln\left(\frac{10}{9}\right)} = \frac{1}{3}\sqrt{18\sigma^2 \ln\left(\frac{10}{9}\right)}. \tag{B.180}$$

En utilisant l'hypothèse (B.173), nous obtenons $\xi < \frac{\epsilon}{3}$.

Preuve principale de la Proposition B.2. Nous divisons l'espace $\mathbb{R}^d$ en trois ensembles :

— $E_1 = \{s \in \mathbb{R}^d : \exists i_0 \in \{1, \dots, k\} \,/\, \|s - t_{i_0}\|_2 < \xi\}$;

— $E_2 = \{s \in \mathbb{R}^d : \exists i_0 \in \{1, \dots, k\} \,/\, \xi \leq \|s - t_{i_0}\|_2 < \frac{\epsilon}{3}\}$;

— $E_3 = \{s \in \mathbb{R}^d : \forall t_i, \|s - t_i\|_2 \geq \frac{\epsilon}{3}\}$.

Notons que E_1, E_2 et E_3 sont pair à pair disjoints et que $E_1 \cup E_2 \cup E_3 = \mathbb{R}^n$. De plus, puisque $\xi < \frac{\epsilon}{3}$, l'ensemble E_2 est non vide.

b) Comparaison entre $s_1 \in E_1$ et $s_2 \in E_2$ Soient $s_1 \in E_1$ et i_0 tels que $\|s_1 - t_{i_0}\|_2 \leq \xi$. Nous avons alors

$$|\langle A\delta_{s_1}, AX\rangle| \geq \langle A\delta_{s_1}, AX\rangle = \langle A\delta_{s_1}, A\delta_{t_{i_0}} + A\sum_{i=1;i\neq i_0}^{k} \delta_{t_i} + Az_l\rangle \tag{B.181}$$

$$= \langle A\delta_{s_1}, A\delta_{t_{i_0}}\rangle + \langle A\delta_{s_1}, A\sum_{i=1;i\neq i_0}^{k} \delta_{t_i}\rangle + \langle A\delta_{s_1}, Az_l\rangle. \tag{B.182}$$

En utilisant le Lemme B.3, nous obtenons que

$$\langle A\delta_{s_1}, A\sum_{i=1;i\neq i_0}^{k} \delta_{t_i}\rangle \geq (1+\gamma)\sum_{i=1;i\neq i_0}^{k} \langle \delta_{s_1}, \delta_{t_i}\rangle_h - \sum_{i=1;i\neq i_0}^{k} 2\gamma \geq -2\gamma(k-1). \tag{B.183}$$

En revenant à $|\langle A\delta_{s_1}, AX\rangle|$, nous utilisons le Lemme B.5 avec la borne supérieure $\kappa_l \geq |\langle A\delta_{s_1}, Az_l\rangle|$ définie dans (B.47). En utilisant le Lemme B.3, avec les inégalités (B.182) et (B.183), nous obtenons,

$$|\langle A\delta_{s_1}, AX\rangle| \geq (1+\gamma)\langle \delta_{s_1}, \delta_{t_{i_0}}\rangle_h - 2\gamma - 2\gamma(k-1) - \kappa_l \tag{B.184}$$

$$= (1+\gamma)\mathrm{e}^{-\frac{\|s_1 - t_{i_0}\|_2^2}{2\sigma^2}} - 2\gamma k - \kappa_l. \tag{B.185}$$

Maintenant que nous avons la borne inférieure pour $|\langle A\delta_{s_1}, AX\rangle|$, cherchons une borne supérieure pour $|\langle A\delta_{s_2}, AX\rangle|$,

$$|\langle A\delta_{s_2}, AX\rangle| = \left|\sum_{i=1}^{k}\langle A\delta_{s_2}, A\delta_{t_i}\rangle + \langle A\delta_{s_2}, Az\rangle\right| \leq \sum_{i=1}^{k}|\langle A\delta_{s_2}, A\delta_{t_i}\rangle| + |\langle A\delta_{s_2}, Az\rangle|\,. \tag{B.186}$$

Il existe un j_0 tel que $\xi \leq \|t_{j_0} - s^*\|_2 < \frac{\epsilon}{3}$. Avec le Lemme B.5, nous obtenons une borne supérieure pour $|\langle A\delta_{s_2}, Az\rangle|$,

$$|\langle A\delta_{s_2}, AX\rangle| \leq |\langle A\delta_{s_2}, A\delta_{t_{j_0}}\rangle| + \sum_{i=1;i\neq j_0}^{k} \langle A\delta_{s_2}, A\delta_{t_i}\rangle| + \kappa_l. \tag{B.187}$$

We utilisons la borne supérieure donnée par le Lemme B.3 pour obtenir,

$$\begin{aligned}|\langle A\delta_{s_2}, AX\rangle| \leq &(1-\gamma)\langle \delta_{s_2}, \delta_{t_{j_0}}\rangle_h + 2\gamma + (1-\gamma)\sum_{i=1;i\neq j_0}^{k} \langle \delta_{s_2}, \delta_{t_i}\rangle_h \\ &+ 2\gamma\sum_{i=1;i\neq j_0}^{k} 1 + \kappa_l\end{aligned} \tag{B.188}$$

$$\begin{aligned}\leq &(1-\gamma)\mathrm{e}^{-\frac{\|s_2 - t_{j_0}\|_2^2}{2\sigma^2}} + 2\gamma + (1-\gamma)\sum_{i=1;i\neq j_0}^{k} \mathrm{e}^{-\frac{\|s_2 - t_i\|_2^2}{2\sigma^2}} \\ &+ 2\gamma(k-1) + \kappa_l.\end{aligned} \tag{B.189}$$

Puisque $\|t_{j_0} - s_2\|_2 \leq \frac{\epsilon}{3}$, grâce à l'inégalité triangulaire, nous avons, pour $i \neq j_0$,

$$\|t_{j_0} - s_2\|_2 + \|s_2 - t_i\|_2 \geq \|t_{j_0} - t_i\|_2 \quad \text{alors} \quad \frac{\epsilon}{3} + \|s_2 - t_i\|_2 > \epsilon \quad \text{donc} \quad \|s_2 - t_i\|_2 > \frac{2\epsilon}{3} > \frac{\epsilon}{3}. \tag{B.190}$$

En utilisant ce résultat dans l'inégalité (B.189) et grâce au fait que, comme dans (3.4), $\forall i \neq j_0,\ \mathrm{e}^{-\|s_2-t_i\|_2^2/(2\sigma^2)} \leq \mu$, nous obtenons

$$|\langle A\delta_{s_2}, AX\rangle| \leq (1-\gamma)\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} + (1-\gamma)\mu k + 2\gamma k + \kappa_l. \tag{B.191}$$

Une condition suffisante pour prouver $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_2}, AX\rangle|$ est

$$|\langle A\delta_{s_1}, AX\rangle| \geq (1+\gamma)\mathrm{e}^{-\frac{\|s_1-t_{i_0}\|_2^2}{2\sigma^2}} - 2\gamma k - \kappa_l > (1-\gamma)\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} + (1-\gamma)\mu k + 2\gamma k + \kappa_l \geq |\langle A\delta_{s_2}, AX\rangle|\,. \tag{B.192}$$

Considérons n'importe quel élément $s_1 \in E_1$, si pour tout élément $s_2 \in E_2$, l'inégalité précédente (B.192) est vérifiée, alors $s^* \notin E_2$. Maintenant, prenons $s_1 = t_1 = t_{i_0} \in E_1$ avec $i_0 = 1$, nous réécrivons l'inégalité centrale dans (B.192) comme

$$(1-\gamma)\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < (1+\gamma) - (1-\gamma)\mu k - 4\gamma k - 2\kappa_l. \tag{B.193}$$

En divisant les deux membres par $1-\gamma$ et en réarrangeant quelques termes, nous obtenons

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < \frac{1}{1-\gamma}\Big(1 - (1-\gamma)\mu k - (4k-1)\gamma - 2\kappa_l\Big). \tag{B.194}$$

Avec $\alpha \geq 1$, nous déduisons que l'inégalité précédente est vérifiée si

$$\mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} < \frac{1 - ((2k-1)\alpha - 1)(1-\gamma)\mu - (4k\alpha - 1)\gamma - 2\kappa_l\alpha}{1-\gamma}. \tag{B.195}$$

Grâce au Lemme B.6, puisque (B.171) est vérifiée et que nous avons montré que $D < 1$ avec (B.178), nous avons que (B.195) est vérifiée et à son tour $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_2}, AX\rangle|$ l'est également. Nous déduisons alors que $|\langle A\delta_{s^*}, AX\rangle| \geq |\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_2}, AX\rangle|$.

c) Comparaison entre $s_1 \in E_1$ et $s_3 \in E_3$ Dans ce cas là, nous utilisons la même borne inférieure (B.185) pour $|\langle A\delta_{s_1}, AX\rangle|$. Let $s_3 \in E_3$,

$$|\langle A\delta_{s_3}, AX\rangle| = \left|\sum_{i=1}^{k}\langle A\delta_{s_3}, A\delta_{t_i}\rangle + \langle A\delta_{s_3}, Az_l\rangle\right| \leq \sum_{i=1}^{k}|\langle A\delta_{s_3}, A\delta_{t_i}\rangle| + |\langle A\delta_{s_3}, Az_l\rangle|\,. \tag{B.196}$$

Nous utilisons le Lemme B.5 et la borne supérieure κ_l définie dans (B.47) pour obtenir

$$|\langle A\delta_{s_3}, AX\rangle| \leq \sum_{i=1}^{k}\langle A\delta_{s_3}, A\delta_{t_i}\rangle| + \kappa_l. \tag{B.197}$$

En utilisant le Lemme B.3,

$$|\langle A\delta_{s_3}, AX\rangle| \leq (1-\gamma)\sum_{i=1}^{k}\langle \delta_{s_3}, \delta_{t_i}\rangle_h + \sum_{i=1}^{k} 2\gamma + \kappa_l = (1-\gamma)\sum_{i=1}^{k}\mathrm{e}^{-\frac{\|s_3-t_i\|_2^2}{2\sigma^2}} + 2\gamma k + \kappa_l. \tag{B.198}$$

Puisque $s_3 \in E_3$, nous avons que $\forall i \in \{1,\ldots,K\}, \|s_3 - t_i\|_2 \geq \frac{\epsilon}{3}$ et grâce à la cohérence mutuelle de notre Hypothèse 1, nous obtenons

$$|\langle A\delta_{s_3}, AX\rangle| \leq (1-\gamma)\mu k + 2\gamma k + \kappa_l. \tag{B.199}$$

Une condition suffisante pour prouver que $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_3}, AX\rangle|$ est

$$(1+\gamma)\mathrm{e}^{-\frac{\|s_1-t_{i_0}\|_2^2}{2\sigma^2}} - 2\gamma k - \kappa_l > (1-\gamma)\mu k + 2\gamma k + \kappa_l. \tag{B.200}$$

Considérons n'importe quel élément $s_1 \in E_1$. Si pour tout élément $s_3 \in E_3$, l'inégalité précédente (B.200) est vérifiée, alors $s^* \notin E_3$. Maintenant, prenons $i_0 = 1$ et $s_1 = t_{i_0} \in E_1$, nous réécrivons (B.200) comme

$$(1+\gamma) > (1-\gamma)\mu k + 4\gamma k + 2\kappa_l. \tag{B.201}$$

Sachant que $\alpha \geq 1$, ceci est vérifié si

$$\mu < \frac{1-(4k\alpha-1)\gamma - 2\kappa_l\alpha}{(2k\alpha-1)(1-\gamma)}. \tag{B.202}$$

Comme montré dans le Lemme B.7, grâce à (B.179) et au fait que $D < 1$ (de (B.178)), nous obtenons que (B.202) est vraie. Nous en déduisons que $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_3}, AX\rangle|$ et donc $|\langle A\delta_{s^*}, AX\rangle| \geq |\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_{s_3}, AX\rangle|$.

d) Conclusion Nous avons montré que pour tout élément $s \in E_2 \cup E_3$, il existe un élément $s_1 \in E_1$ tel que $|\langle A\delta_{s_1}, AX\rangle| > |\langle A\delta_s, AX\rangle|$. Cela implique que s^* appartient nécessairement à l'ensemble E_1 comme $E_1 \cap (E_2 \cup E_3) = \emptyset$, c'est à dire qu'il existe un i_0 tel que $\|s^* - t_{i_0}\|_2 < \xi$. □

Bibliographie

[1] T. Alamo, D. Limon, and P. Krupa. Restart fista with global linear convergence. In *2019 18th European Control Conference (ECC)*, pages 1969–1974, 2019.

[2] Z. Allen-Zhu, Z. Qu, P. Richtárik, and Y. Yuan. Even faster accelerated coordinate descent using non-uniform sampling. In *ICML*, volume 48, pages 1110–1119. PMLR, 2016.

[3] J.-F. Aujol, C. Dossal, H. Labarrière, and A. Rondepierre. FISTA restart using an automatic estimation of the growth parameter. working paper or preprint, May 2022.

[4] A. Auslender. Asymptotic properties of the fenchel dual functional and applications to decomposition problems. *Journal of Optimization Theory and Applications*, 73 :427–449, 1992.

[5] D. Axelrod. Cell-substrate contacts illuminated by total internal reflection fluorescence. *The Journal of cell biology*, 89(1) :141–145, 1981.

[6] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11) :2419–2434, 2009.

[7] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1) :183–202, 2009.

[8] A. Beck and L. Tetruashvili. On the convergence of block coordinate descent type methods. *SIAM Journal on Optimization*, 23(4) :2037–2060, 2013.

[9] H. E. Bell. Gershgorin's theorem and the zeros of polynomials. *The American Mathematical Monthly*, 72(3) :292–295, 1965.

[10] P.-J. Bénard, Y. Traonmilin, and J.-F. Aujol. Fast off-the-grid sparse recovery with over-parametrized projected gradient descent. In *2022 30th EUSIPCO*, pages 2206–2210. IEEE, 2022.

[11] P.-J. Bénard, Y. Traonmilin, and J.-F. Aujol. Code for the experiments. https://github.com/pjbenard/opCOMP_PGD_microscopy, 2023. [Online].

[12] P.-J. Bénard, Y. Traonmilin, and J.-F. Aujol. Code of the experiments. https://github.com/pjbenard/opCOMP_PGD_BCD, 2024. [Online].

[13] P.-J. Bénard, Y. Traonmilin, and J.-F. Aujol. Code of the experiments. https://github.com/pjbenard/opCOMP_PGD_sketching, 2024. [Online].

[14] P.-J. Bénard, Y. Traonmilin, and J. F. Aujol. Projected Block Coordinate Descent for sparse spike estimation. In *2024 32nd EUSIPCO*. IEEE, Feb. 2024. https://hal.science/hal-04462779.

[15] P.-J. Bénard, Y. Traonmilin, and J.-F. Aujol. Sketched over-parametrized projected gradient descent for sparse spike estimation. preprint, https://hal.science/hal-04584951/, Apr. 2024.

[16] P.-J. Bénard, Y. Traonmilin, J.-F. Aujol, and E. Soubies. Estimation of off-the grid sparse spikes with over-parametrized projected gradient descent : theory and application. *Inverse Problems*, 40(5) :055010, 2024.

[17] D. P. Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3) :334–334, 1997.

[18] T. Blumensath. Sampling and reconstructing signals from a union of linear subspaces. *IEEE Transactions on Information Theory*, 57(7) :4660–4671, 2011.

[19] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier analysis and Applications*, 14 :629–654, 2008.

[20] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Applied and computational harmonic analysis*, 27(3) :265–274, 2009.

[21] S. Bourguignon, H. Carfantan, and J. Idier. A sparsity-based method for the estimation of spectral lines from irregularly sampled data. *IEEE Journal of Selected Topics in Signal Processing*, 1(4) :575–585, 2007.

[22] J.-F. Cai, T. Wang, and K. Wei. Spectral compressed sensing via projected gradient descent. *SIAM Journal on Optimization*, 28(3) :2625–2653, 2018.

[23] E. Candes and J. Romberg. l1-magic : Recovery of sparse signals via convex programming. *URL : www. acm. caltech. edu/l1magic/downloads/l1magic. pdf*, 4(14) :16, 2005.

[24] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6) :906–956, 2014.

[25] E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12) :4203–4215, 2005.

[26] A. Chambolle and C. H. Dossal. On the convergence of the iterates of" fista". *Journal of Optimization Theory and Applications*, 166(3) :25, 2015.

[27] B. Chen, S. He, Z. Li, and S. Zhang. Maximum block improvement and polynomial optim. *SIAM Journal on Optimization*, 22(1) :87–107, 2012.

[28] L. Chen and Y. Gu. The convergence guarantees of a non-convex approach for sparse recovery. *IEEE Transactions on Signal Processing*, 62(15) :3754–3767, 2014.

[29] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1) :129–159, 2001.

[30] Y. Chen and M. J. Wainwright. Fast low-rank estimation by projected gradient descent : General statistical and algorithmic guarantees. *arXiv preprint arXiv :1509.03025*, 2015.

[31] L. Chizat. Sparse optimization on measures with over-parameterized gradient descent. *Mathematical Programming*, 194(1-2) :487–532, 2022.

[32] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation : A SIAM Interdisciplinary Journal*, 4(4) :1168–1200, 2005.

[33] M. F. Da Costa and Y. Chi. Local geometry of nonconvex spike deconvolution from low-pass measurements. *IEEE Journal on Selected Areas in Information Theory*, 4 :1–15, 2023.

[34] Y. De Castro and F. Gamboa. Exact reconstruction using beurling minimal extrapolation. *Journal of Mathematical Analysis and applications*, 395(1) :336–354, 2012.

[35] Y. De Castro, F. Gamboa, D. Henrion, and J.-B. Lasserre. Exact solutions to super resolution on semi-algebraic domains in higher dimensions. *IEEE Transactions on Information Theory*, 63(1) :621–630, 2016.

[36] V. Debarnot and P. Weiss. Blind inverse problems with isolated spikes. *Information and Inference : A Journal of the IMA*, 12(1) :26–71, 2023.

[37] Q. Denoyelle, V. Duval, G. Peyré, and E. Soubies. The Sliding Frank-Wolfe Algorithm and its Application to Super-Resolution Microscopy. *Inverse Problems*, 36(1) :014001, 2019.

[38] D. L. Donoho. Superresolution via sparsity constraints. *SIAM journal on Mathematical Analysis*, 23(5) :1309–1331, 1992.

[39] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell 1$ minimization. *Proceedings of the National Academy of Sciences*, 100(5) :2197–2202, 2003.

[40] C. Dossal. A necessary and sufficient condition for exact sparse recovery by $\ell 1$ minimization. *Comptes Rendus Mathematique*, 350(1-2) :117–120, 2012.

[41] C. Dossal and S. Mallat. Sparse spike deconvolution with minimum scale. In *Signal processing with adaptive sparse structured representations (SPARS workshop)*, pages 1–4, 2005.

[42] F. Dublanchet, J. Idier, and P. Duwaut. Direction-of-arrival and frequency estimation using poisson-gaussian modeling. In *ICASSP*, volume 5, pages 3501–3504. IEEE, 1997.

[43] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32(1) :407–499, 2004.

[44] C. Elvira, R. Gribonval, C. Soussen, and C. Herzet. OMP and continuous dictionaries : Is k-step recovery possible ? In *ICASSP*, pages 5546–5550. IEEE, 2019.

[45] C. Elvira, R. Gribonval, C. Soussen, and C. Herzet. When does OMP achieve exact recovery with continuous dictionaries ? *Applied and Computational Harmonic Analysis*, 51 :39, 2021.

[46] A. Flinth, F. de Gournay, and P. Weiss. Grid is good : Adaptive refinement algorithms for off-the-grid total variation minimization, 2023. preprint.

[47] M. Fontaine, C. Vanwynsberghe, A. Liutkus, and R. Badeau. Scalable source localization with multichannel α-stable distributions. In *2017 25th EUSIPCO*, pages 11–15, 2017.

[48] M. Fontaine, C. Vanwynsberghe, A. Liutkus, and R. Badeau. Sketching for nearfield acoustic imaging of heavy-tailed sources. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 80–88. Springer, 2017.

[49] S. Foucart and H. Rauhut. *An invitation to compressive sensing*. Springer, 2013.

[50] M. Frank, P. Wolfe, et al. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2) :95–110, 1956.

[51] G. Gasso, A. Rakotomamonjy, and S. Canu. Recovering sparse signals with a certain family of nonconvex penalties and dc programming. *IEEE Transactions on Signal Processing*, 57(12) :4686–4698, 2009.

[52] M. Golbabaee and M. E. Davies. Inexact gradient projection and fast data driven compressed sensing. *IEEE Transactions on Information Theory*, 64(10) :6707–6721, 2018.

[53] R. Gribonval, G. Blanchard, N. Keriven, and Y. Traonmilin. Compressive Statistical Learning with Random Feature Moments. *Mathematical Statistics and Learning*, 3(2) :113–164, 2021.

[54] R. Gribonval, G. Blanchard, N. Keriven, and Y. Traonmilin. Statistical learning guarantees for compressive clustering and compressive mixture modeling. *Mathematical Statistics and Learning*, 3(2) :165–257, 2021.

[55] R. Gribonval, A. Chatalic, N. Keriven, V. Schellekens, L. Jacques, and P. Schniter. Sketching Data Sets for Large-Scale Learning : Keeping only what you need. *IEEE Signal Processing Magazine*, 38(5) :12–36, Sept. 2021. For an extended version of this article that contains additional references and more in-depth discussions on a variety of topics, see the long version published on HAL (https ://hal.inria.fr/hal-02909766) and arXiv (2008.01839).

[56] B. Huang, W. Wang, M. Bates, and X. Zhuang. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*, 319(5864) :810–813, 2008.

[57] N. Keriven and R. Gribonval. Compressive gaussian mixture estimation by orthogonal matching pursuit with replacement. In *SPARS*, 2015.

[58] N. Keriven, R. Gribonval, G. Blanchard, et al. Spikes super-resolution with random fourier sampling. In *SPARS workshop*, 2017.

[59] N. Keriven, N. Tremblay, Y. Traonmilin, and R. Gribonval. Compressive k-means. In *ICASSP*, pages 6369–6373. IEEE, 2017.

[60] C. Lage, N. Pustelnik, J. M. Arbona, and B. Audit. Space-scale hybrid continuous-discrete sliding frank-wolfe method. *IEEE Signal Processing Letters*, 31 :830–834, 2024.

[61] Y. Li and S. Osher. Coordinate descent optimization for l1 minimization with application to compressed sensing ; a greedy algorithm. *Inverse Problems and Imaging*, 3(3) :487–503, 2009.

[62] L. Meier, S. Van De Geer, and P. Bühlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society Series B : Statistical Methodology*, 70(1) :53–71, 2008.

[63] NASA. James webb gallery. https://webbtelescope.org/, 2024. [Online].

[64] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4) :1574–1609, 2009.

[65] Y. Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2) :341–362, 2012.

[66] J. Nutini, I. Laradji, and M. Schmidt. Let's make block coordinate descent converge faster : Faster greedy rules, message-passing, active-set complexity, and superlinear convergence. *Journal of Machine Learning Research*, 23(131) :1–74, 2022.

[67] J. Nutini, M. Schmidt, I. Laradji, M. Friedlander, and H. Koepke. Coordinate descent converges faster with the gauss-southwell rule than random selection. In *ICML*, pages 1632–1641. PMLR, 2015.

[68] B. O'donoghue and E. Candes. Adaptive restart for accelerated gradient schemes. *Foundations of computational mathematics*, 15 :715–732, 2015.

[69] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit : Recursive function approximation with applications to wavelet decomposition. In *Proceedings of 27th Asilomar conference on signals, systems and computers*, pages 40–44. IEEE, 1993.

[70] M. Q. Pham, L. Duval, C. Chaux, and J.-C. Pesquet. A primal-dual proximal algorithm for sparse template-based adaptive filtering : Application to seismic multiple removal. *IEEE Transactions on Signal Processing*, 62(16) :4256–4269, 2014.

[71] Z. Qin, K. Scheinberg, and D. Goldfarb. Efficient block-coordinate descent algorithms for the group lasso. *Mathematical Programming Computation*, 5(2) :143–169, 2013.

[72] S. Rama Prasanna Pavani, M. A Thompson, J. S Biteen, S. Lord, N. Liu, R. Twieg, R. Piestun, and W. Moerner. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proceedings of the National Academy of Sciences of the United States of America*, 106 :2995–9, 03 2009.

[73] P. Richtárik and M. Takáč. Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Mathematical Programming*, 144(1-2) :1–38, 2014.

[74] D. Sage, T.-A. Pham, H. Babcock, T. Lukes, T. Pengo, J. Chao, R. Velmurugan, A. Herbert, A. Agrawal, S. Colabrese, et al. Super-resolution fight club : assessment of 2d and 3d single-molecule localization microscopy software. *Nature methods*, 16(5) :387–395, 2019.

[75] S. Sardy, A. G. Bruce, and P. Tseng. Block coordinate relaxation methods for nonparametric wavelet denoising. *Journal of computational and graphical statistics*, 9(2) :361–379, 2000.

[76] V. Schellekens and L. Jacques. Compressive classification (machine learning without learning). In *ITWIST'18 workshop*, 2018.

[77] V. Schellekens and L. Jacques. Compressive learning of generative networks. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, volume 28, 2020.

[78] H. Shi, Y. Traonmilin, and J.-F. Aujol. Compressive learning for patch-based image denoising. *SIAM Journal on Imaging Sciences*, 15(3) :1184–1212, 2022.

[79] H. Shi, Y. Traonmilin, and J.-F. Aujol. Compressive learning of deep regularization for denoising. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 162–174. Springer, 2023.

[80] H.-J. M. Shi, S. Tu, Y. Xu, and W. Yin. A primer on coordinate descent algorithms. Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University, 2016.

[81] E. Soubies, L. Blanc-Féraud, and G. Aubert. A continuous exact ℓ_0 penalty (CEL0) for least squares regularized problem. *SIAM Journal on Imaging Sciences*, 8(3) :1607–1639, 2015.

[82] E. Soubies, L. Blanc-Féraud, and G. Aubert. A unified view of exact continuous penalties for ℓ_2-ℓ_0 minimization. *SIAM Journal on Optimization*, 27(3) :2034–2060, 2017.

[83] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B : Statistical Methodology*, 58(1) :267–288, 1996.

[84] Y. Traonmilin and J.-F. Aujol. The basins of attraction of the global minimizers of the non-convex sparse spike estimation problem. *Inverse Problems*, 36(4) :045003, feb 2020.

[85] Y. Traonmilin, J.-F. Aujol, P.-J. Bénard, and A. Leclaire. On strong basins of attractions for non-convex sparse spike estimation : upper and lower bounds. *Journal of Mathematical Imaging and Vision*, 66(1) :57–74, 2024.

[86] Y. Traonmilin, J.-F. Aujol, and A. Leclaire. Projected gradient descent for non-convex sparse spike estimation. *IEEE Signal Processing Letters*, 27 :1110–1114, 2020.

[87] Y. Traonmilin, J.-F. Aujol, and A. Leclaire. The basins of attraction of the global minimizers of non-convex inverse problems with low-dimensional models in infinite dimension. *Information and Inference : A Journal of the IMA*, 12(1) :113–156, 2023.

[88] J. Tropp. Greed is good : algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10) :2231–2242, 2004.

[89] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*, 53(12) :4655–4666, 2007.

[90] P. Tseng and S. Yun. A coordinate gradient descent method for nonsmooth separable minimization. *Mathematical Programming*, 117 :387–423, 2009.

[91] S. Vaiter, G. Peyré, C. Dossal, and J. Fadili. Robust sparse analysis regularization. *IEEE Transactions on information theory*, 59(4) :2001–2016, 2012.

[92] S. J. Wright. Coordinate descent algorithms. *Mathematical programming*, 151(1) :3–34, 2015.

[93] T. T. Wu and K. Lange. Coordinate descent algorithms for lasso penalized regression. *The Annals of Applied Statistics*, 2(1) :224, 2008.

[94] Y. Xu and W. Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 6(3) :1758–1789, 2013.