

Déconvolution Dynamique par Descente de Gradient Préconditionnée

Joseph GABET Charles SOUSSEN Maxime FERREIRA DA COSTA

Laboratoire des Signaux et Systèmes, CentraleSupélec, Université Paris–Saclay, CNRS, Gif-sur-Yvette, France

Résumé – Nous considérons le problème de la déconvolution de Diracs dans le domaine de Fourier lorsque leurs positions et amplitudes évoluent dynamiquement au cours du temps. Pour suivre efficacement cette évolution, l’originalité de notre approche réside dans l’introduction d’un algorithme de descente de gradient avec preconditionnement diagonal et momentum, une méthode de premier ordre exploitant la structure du problème pour accélérer la convergence. Notre contribution principale est de montrer que cette méthode peut être appliquée de manière efficace dans un cadre dynamique en réutilisant les estimations précédentes à chaque pas de temps. Nous illustrons empiriquement les bénéfices de notre méthode sur des signaux synthétiques.

Abstract – We consider the problem of deconvolving Diracs in the Fourier domain when their positions and amplitudes evolve dynamically over time. To efficiently track this evolution, the originality of our approach lies in the introduction of a gradient descent algorithm with diagonal preconditioning and momentum, a first-order method that leverages the problem’s structure to accelerate convergence. Our main contribution is to demonstrate that this method can be effectively applied in a dynamic setting by reusing previous estimates at each time step. We empirically illustrate the benefits of our method on synthetic signals.

1 Introduction

1.1 Contributions, état de l’art et organisation

La reconstruction de signaux composés d’impulsions discrètes à partir d’observations dans le domaine de Fourier constitue un problème fondamental en traitement du signal et en imagerie. Les approches classiques incluent notamment des méthodes convexes reposant sur une discrétisation spatiale fine (par exemple, minimisation de norme en variation totale sur grille), et des méthodes spectrales paramétriques comme MUSIC ou ESPRIT, qui ne reposent pas explicitement sur une grille mais sur la structure particulière du problème. Toutefois, la plupart de ces méthodes ont été développées pour des contextes statiques.

Dans de nombreux scénarios applicatifs pratiques tels que, les radars OFDM pour les communications sans fil [1] ou le suivi de cibles [2]; ou encore l’astronomie [3], les Diracs évoluent temporellement : leurs localisations et amplitudes ne sont alors plus fixes. La déconvolution dynamique vise alors à suivre ces évolutions tout en garantissant robustesse face au bruit et efficacité computationnelle, de manière à permettre un traitement en temps réel ou quasi temps réel. Quelques approches existent dans ce contexte dynamique, telles que des méthodes récursives basées sur des filtres de Kalman ou des extensions des méthodes spectrales classiques par fenêtre glissante, mais elles restent coûteuses ou sensibles lorsque les Diracs sont proches ou se confondent ponctuellement.

Notre principale contribution est d’introduire une méthode continue, paramétrique et sans grille, reposant sur une approche de premier ordre combinant preconditionnement diagonal adaptatif et momentum. Bien que non convexe, notre approche suit efficacement l’évolution temporelle des Diracs, tout en gardant un coût par itération très faible. Cette stratégie simple, rapide et robuste convient particulièrement au cadre dynamique étudié ici.

Ce travail a été partiellement financé par l’ANR (ANR-24-CE48-3094), l’AID (AID-2023639) et la chaire Orange “6G durable” de CentraleSupélec.

1.2 Notation et définitions

Les vecteurs sont notés en gras minuscules \mathbf{a} , et les matrices en gras majuscules \mathbf{A} . L’espace des matrices de taille $u \times v$ à coefficients dans un corps \mathbb{K} est noté $\mathcal{M}_{u,v}(\mathbb{K})$. La transposée et la transposée hermitienne de \mathbf{A} sont notées \mathbf{A}^\top et \mathbf{A}^H . On note $\mathbb{1}_r$ et $\mathbf{0}_r$ les vecteurs de \mathbb{C}^r tout en un ou tout en zéro. Le produit scalaire s’écrit $\langle \mathbf{a}, \mathbf{b} \rangle$, et $\mathbf{A} \otimes \mathbf{B}$ désigne le produit de Kronecker. Pour $a, b \in \mathbb{N}$, $\llbracket a; b \rrbracket$ est l’ensemble des entiers entre a et b , et $[a] = \llbracket 1; a \rrbracket$.

2 Formulation du problème

2.1 Formulation du problème statique

On considère un signal composé de r impulsions de Diracs, chacun d’eux étant paramétré par une amplitude complexe $a_\ell^* \in \mathbb{C}$ et un vecteur de localisation de dimension d , $\boldsymbol{\tau}_\ell^* \in \mathbb{R}^d$, pour $1 \leq \ell \leq r$. Soit $\mathbf{a}^* = [a_1^*, \dots, a_r^*]^\top \in \mathbb{C}^r$ le vecteur de toutes les amplitudes, et $\boldsymbol{\tau}^* = [\tau_{1;1}^*, \tau_{2;1}^*, \dots, \tau_{r;1}^*, \dots, \tau_{r;d}^*]^\top \in \mathbb{R}^{dr}$ le vecteur de localisation concaténé, où $\tau_{l;i}^*$ est la i -ème composante du vecteur de localisation de la l -ème impulsion.

Pour tout couple $(\mathbf{a}, \boldsymbol{\tau})$, nous définissons la mesure de Radon complexe parcimonieuse par

$$\mu(\mathbf{a}, \boldsymbol{\tau}) = \sum_{j=1}^r a_j \delta_{\boldsymbol{\tau}_j}, \quad (1)$$

où $\delta_{\boldsymbol{\tau}_j}$ désigne la distribution de Dirac centrée en $\boldsymbol{\tau}_j \in \mathbb{R}^d$.

Après convolution avec une fonction d’étalement du point (PSF) connue $g \in L_1(\mathbb{R}^d)$, le signal continu $s(\mathbf{x})$, défini sur \mathbb{R}^d , s’exprime par

$$s(\mathbf{x}) = \left(g \star \mu(\mathbf{a}^*, \boldsymbol{\tau}^*) \right)(\mathbf{x}), \quad (2)$$

où \star désigne l’opérateur de convolution sur \mathbb{R}^d .

Nous considérons par ailleurs des mesures \mathbf{y} acquises à partir du signal s observé dans le domaine de Fourier, un cadre très répandu en traitement du signal. La transformée de Fourier continue $\mathcal{F}(\cdot)$ d'une mesure de Radon est définie par

$$\mathcal{F}(\mu)(\mathbf{f}) = \int_{\mathbb{R}^d} e^{-2i\pi\langle \mathbf{f}, \mathbf{x} \rangle} d\mu(\mathbf{x}), \quad \forall \mathbf{f} \in \mathbb{R}^d. \quad (3)$$

On fixe $N = 2n + 1$, un nombre impair, et on suppose disposer de N mesures dans le domaine de Fourier, prises de manière uniforme sur la bande de fréquences $J_N = [-\frac{N}{2T}, \frac{N}{2T}]^d$, de sorte que l'ensemble d'échantillonnage $\Omega \subset \mathbb{R}^d$ est défini par $\Omega = \frac{1}{T} \llbracket -n, n \rrbracket^d$, c'est-à-dire le produit cartésien d'intervalles entiers en dimension d .

Soit $\mathbf{G} = \text{diag}([\mathcal{F}(g)(\mathbf{f})]_{\mathbf{f} \in \Omega})$ la matrice diagonale formée des valeurs de $\mathcal{F}(g)$ sur Ω . Le vecteur d'observation discret \mathbf{y} s'écrit alors

$$\mathbf{y} = \mathbf{G} \mathbf{V}_{\tau^*} \mathbf{a}^* + \mathbf{z}, \quad (4)$$

où $\mathbf{V}_{\tau} \in \mathbb{C}^{N^d \times r}$ est la matrice de Vandermonde dont les noeuds sont $e^{-2i\pi\langle \mathbf{f}, \tau_j \rangle}$ pour $\mathbf{f} \in \Omega$, et \mathbf{z} est un bruit additif, supposé arbitraire.

2.2 Passage au problème dynamique

Dans le problème dynamique, nous supposons que les paramètres $\boldsymbol{\theta}(t) = (\mathbf{a}(t), \boldsymbol{\tau}(t))$ évoluent en fonction d'une variable de temps t (distincte de la variable spatiale \mathbf{x}) et que la fonction $\boldsymbol{\theta} : t \mapsto (\mathbf{a}(t), \boldsymbol{\tau}(t)) \in \mathbb{C}^r \times \mathbb{R}^{r \times d}$ est Lipschitz.

Nous considérons alors le signal continu dynamique $s(\mathbf{x}, t)$, défini sur $\mathbb{R}^d \times \mathbb{R}^+$, s'exprime par

$$s(\mathbf{x}, t) = \left(g * \mu(\mathbf{a}^*(t), \boldsymbol{\tau}^*(t)) \right)(\mathbf{x}), \quad (5)$$

et nous observons à chaque pas de temps $t_i \in \mathbb{R}^+$, $i \in [T_{\max}]$ le vecteur discret dans le domaine de Fourier

$$\mathbf{y}(t_i) = \mathbf{G} \mathbf{V}_{\tau^*(t_i)} \mathbf{a}^*(t_i) + \mathbf{z}(t_i). \quad (6)$$

3 Descente de gradient préconditionnée avec momentum

3.1 Algorithme dans la cadre statique

Étant donné le modèle de mesure statique (4), nous introduisons la fonction de coût quadratique

$$\mathcal{L}_{\mathbf{y}}(\mathbf{a}, \boldsymbol{\tau}) = \frac{1}{2} \|\mathbf{G} \mathbf{V}_{\boldsymbol{\tau}} \mathbf{a} - \mathbf{y}\|_2^2. \quad (7)$$

Dans le cas d'un bruit gaussien, l'expression (7) correspond à la négative de la log-vraisemblance du modèle. Nous estimons les paramètres du signal $(\hat{\mathbf{a}}, \hat{\boldsymbol{\tau}})$ comme le minimiseur non contraint de (7). Puisque la fonction de coût (7) est non convexe, plusieurs minimiseurs locaux peuvent exister, et l'optimisation par méthode du premier ordre ne garantit pas la convergence globale. De plus, la fonction de coût souffre d'ambiguïtés de permutation sur les composantes de $\boldsymbol{\tau}$, ce qui implique l'existence de plusieurs minimiseurs globaux.

Pour les besoins de l'optimisation, nous vectorisons les paramètres d'entrée $(\mathbf{a}, \boldsymbol{\tau})$ dans la variable $\boldsymbol{\theta} = [\mathbf{a}^\top; \boldsymbol{\tau}^\top]^\top \in$

$\mathbb{C}^r \times \mathbb{R}^{rd}$. Nous présentons ci-après l'expression du gradient de la fonction de coût (7) par rapport aux amplitudes et aux positions. Pour ce faire, nous introduisons $\mathbf{M}_{\mathbf{a}} = \text{diag}([\mathbb{1}_r^\top; (\mathbb{1}_d \otimes \mathbf{a})^\top]^\top) \in \text{Diag}_{(d+1)r}(\mathbb{C})$, et nous définissons la matrice de dérivation suivant la j -ème dimension $\boldsymbol{\Lambda}_j = -2i\pi \text{diag}(\{\mathbf{f}_u\}_j)_{\mathbf{f}_u \in \Omega} \in \text{Diag}_{N^d}(\mathbb{C})$. De plus, nous posons $\mathbf{W}_{\boldsymbol{\tau}} = [\mathbf{V}_{\boldsymbol{\tau}} \mid \boldsymbol{\Lambda}_1 \mathbf{V}_{\boldsymbol{\tau}} \mid \dots \mid \boldsymbol{\Lambda}_d \mathbf{V}_{\boldsymbol{\tau}}] \in \mathcal{M}_{N^d, (d+1)r}(\mathbb{C})$. Ainsi, Le gradient de $\mathcal{L}_{\mathbf{y}}$, tel qu'exprimé dans l'équation (7), s'écrit alors

$$\nabla \mathcal{L}_{\mathbf{y}}(\boldsymbol{\theta}) = \mathbf{M}_{\mathbf{a}}^\text{H} \mathbf{W}_{\boldsymbol{\tau}}^\text{H} \left(\mathbf{W}_{\boldsymbol{\tau}} \mathbf{M}_{\mathbf{a}^*} \begin{bmatrix} \mathbf{a} \\ \mathbf{0}_{rd} \end{bmatrix} - \mathbf{W}_{\boldsymbol{\tau}^*} \mathbf{M}_{\mathbf{a}^*} \begin{bmatrix} \mathbf{a}^* \\ \mathbf{0}_{rd} \end{bmatrix} + \mathbf{z} \right). \quad (8)$$

Compte tenu de la nature hétérogène du vecteur $\boldsymbol{\theta}$, qui contient à la fois les positions et les amplitudes, la fonction de coût (7) ainsi que son gradient (8) sont, en général, mal conditionnés, notamment lorsque la plage dynamique est grande. Nous nous appuyons donc sur le *préconditionnement* — une méthode quasi-Newton — afin d'adapter la direction de descente au paysage local de la fonction de coût. Ceci est réalisé en multipliant, à l'étape k , le gradient par une matrice \mathbf{P}_k [4], qui dépend de l'estimation courante $\boldsymbol{\theta}_k$. On définit la fonction d'autocorrélation $K(\boldsymbol{\tau}) = \langle \mathbf{G} \mathbf{V}_{\mathbf{0}}; \mathbf{G} \mathbf{V}_{\boldsymbol{\tau}} \rangle$ et on choisit le preconditionneur

$$\mathbf{P}_k = \text{diag} \left(\begin{bmatrix} \mathbb{1}_r \\ [-\nabla^2 K(\mathbf{0})]_{1,1}^{-1} |\mathbf{a}_k|^{-2} \\ \vdots \\ [-\nabla^2 K(\mathbf{0})]_{d,d}^{-1} |\mathbf{a}_k|^{-2} \end{bmatrix} \right). \quad (9)$$

Comme la complexité de calcul de l'Algorithme 1 par itération est déterminée par celle de la multiplication matrice-vecteur dans la quatrième ligne, nous limitons notre analyse aux matrices de preconditionnement diagonales de sorte que le calcul de la direction de descente preconditionnée entraîne un coût de calcul supplémentaire marginal par rapport à la descente de gradient vanilla. Il a été montré [4] que lorsque l'on initialise la descente de gradient suffisamment proche du minimiseur global alors, sous des conditions raisonnable de séparation entre les impulsions de Dirac [5] et l'échelle dynamique du problème, la descente de gradient preconditionnée converge linéairement vers $\boldsymbol{\theta}^*$. Par ailleurs, il a été montré [6] qu'il était possible d'initialiser dans le bassin d'attraction de la fonction de perte (7) au temps t_0 rendant l'Algorithme 1 effectif en pratique.

Nous proposons d'incorporer une dynamique de momentum [7] — une méthode classique ayant récemment retrouvé un intérêt marqué dans le contexte de l'apprentissage automatique — dans l'Algorithme 1 (MPGD) afin d'accélérer plus encore la convergence. L'idée est d'ajouter un terme qui exploite la direction des mises à jour précédentes. Cela se traduit par la mise en œuvre de deux hyperparamètres : un taux d'apprentissage $\eta > 0$ et un coefficient de momentum $\nu \in [0; 1]$ ce qui permet d'obtenir en pratique le meilleur compromis entre stabilité et vitesse de convergence. Concrètement, la mise à jour complète des paramètres s'exprime par

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \eta \mathbf{P}_k \nabla \mathcal{L}_{\mathbf{y}}(\boldsymbol{\theta}_k) + \nu (\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}). \quad (10)$$

3.2 Algorithme dans le cadre dynamique

Étant donné que le modèle d'observation dynamique (6) correspond à une succession de problème statique (4), nous proposons de résoudre un problème statique à chaque pas de temps

Algorithm 1 Descente Préconditionnée avec Momentum

```

1: Entrées :  $\theta_0, y, K_{\max}, \eta, \nu$ .
2: Initialiser  $v \leftarrow 0$ .
3: Pour  $k \leftarrow 0$  jusqu'à  $K_{\max}$ 
4:   Calculer  $P_k$  comme en (9).
5:   Calculer  $\nabla \mathcal{L}_y(\theta_k)$ .
6:    $v \leftarrow \nu \cdot v - \eta P_k \nabla \mathcal{L}_y(\theta_k)$ .
7:    $\theta_{k+1} \leftarrow \theta_k + v$ .
8: FinPour
9: Sortie :  $\theta_{K_{\max}}$ .
  
```

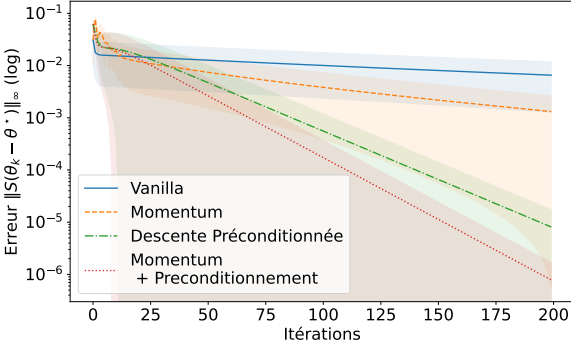


FIGURE 1 : Comparaison vitesse de convergence descentes de gradient (100 réalisation d'une expériences avec initialisation θ_0 aléatoires) en moyenne ± 1 écart-type. (sans bruit)

en ne considérant connue que l'estimation du pas de temps précédent. Cette procédure est décrit dans l'Algorithme 2.

Algorithm 2 Déconvolution Dynamique

```

1: Entrées :  $(A_0, \tau_0), K_{\max}, T_{\max}, \eta, \nu$ .
2: Initialiser  $\theta[0] \leftarrow (A_0, \tau_0)$ .
3: Pour  $i \leftarrow 0$  jusqu'à  $T_{\max}$ 
4:   On observe  $y[i]$ 
5:    $\theta[i+1] \leftarrow \text{MPGD}(\theta[i], y[i], K_{\max}, \eta, \nu)$ .
6: FinPour
7: Sortie :  $\{\hat{\theta}[i], i \in [T_{\max}]\}$ .
  
```

4 Résultats numériques

Le code est intégralement disponible sur Gitlab ¹.

Bien que la situation multidimensionnelle soit plus générale, le cas unidimensionnel demeure particulièrement intéressant, notamment parce que deux trajectoires partant de positions différentes et se dirigeant l'une vers l'autre ne peuvent pas s'éviter, ce qui complique grandement l'analyse du problème. Ainsi, l'ensemble des expérimentations sera mené dans le cadre unidimensionnel.

4.1 Comparaison des descentes de gradients

Dans un premier temps, nous allons chercher à démontrer expérimentalement l'avantage du préconditionnement et du momentum par rapport à une simple descente de gradient.

¹<https://gitlab-student.centralesupelec.fr/joseph.gabet/gretsi-2025/-/tree/main>

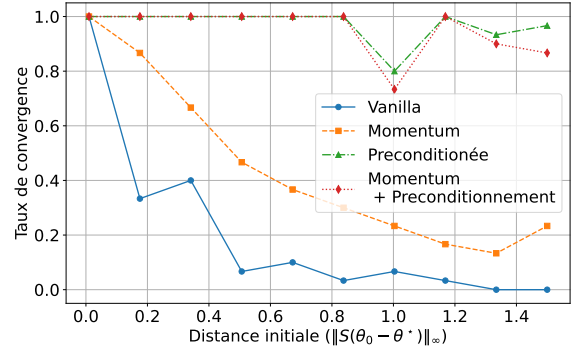


FIGURE 2 : Comparaison du bassin d'attraction des descentes de gradient. (sans bruit)

Pour ce faire, nous nous plaçons dans un cadre statique afin d'illustrer la supériorité de l'algorithme 1.

Nous mesurons le taux de contraction en termes de *pire distance relative* de tout paramètre par rapport à la vérité terrain. Cela permet un contrôle homogène de l'estimation des paramètres sur l'ensemble des impulsions et transcende certaines propriétés du problème considéré, telles que le nombre de sources ou la plage dynamique. Mathématiquement nous étudions la contraction de la suite d'itérés

$$\|S(\theta_k - \theta^*)\|_\infty = \max_j \left\{ \frac{|a_{k,j} - a_j^*|}{|a_j^*|}; \sqrt{-K''(0)} |\tau_{k,j} - \tau_j^*| \right\}.$$

La Figure 1 montre que la vitesse de convergence de la descente préconditionnée est linéaire comme exprimé par la théorie [4]. On observe également que le préconditionnement accélère considérablement la convergence. De plus, bien que l'ajout du momentum seul ne semble pas améliorer drastiquement la vitesse de convergence moyenne, apparaissant ainsi comme moins intéressant que le préconditionnement seul, l'ajout d'un faible momentum en complément du préconditionnement semble apporter un avantage substantiel. Pour chaque expérience les paramètres η et ν ont été fixé indépendamment, à 10^{-1} près de sorte à obtenir la meilleure vitesse de convergence moyenne sans diverger.

Outre la rapidité constatée en Figure 1, la Figure 2 montre que le bassin d'attraction induit par le préconditionnement est significativement supérieur à celui sans préconditionnement. L'ajout d'un momentum semble légèrement réduire la taille du bassin d'attraction, sans que cela paraisse significatif au regard de l'accélération apportée.

4.2 Déconvolution dynamique

Considérons alors $r = 3$ Diracs évoluant dynamiquement en amplitudes $a_i^*(t) \in \mathbb{R}^+$ et positions $\tau_i^*(t) \in [-10; 10]$, tel que les amplitudes sont normalisées et $\frac{\max_i a_i^*(t)}{\min_i a_i^*(t)} \geq \frac{3}{2}$. Les positions sont initialement éloignées, puis entrent dans un régime où elles se croisent à partir de l'itération 75 avant de se séparer de nouveau à partir de l'itération 120.

Les amplitudes optimales pour la fonction de perte (7) associées à une estimation $\hat{\tau}$ étant obtenues par moindres carrés, nous ne considérons que la distance entre $\hat{\tau}$ et la vérité-terrain τ^* . Pour lever l'ambiguïté due à la permutation des positions, nous utilisons la distance quadratique de pairage $md_2(\hat{\tau}, \tau^*)$, définie comme la distance euclidienne minimale entre $\hat{\tau}$ et toute permutation de τ^* .

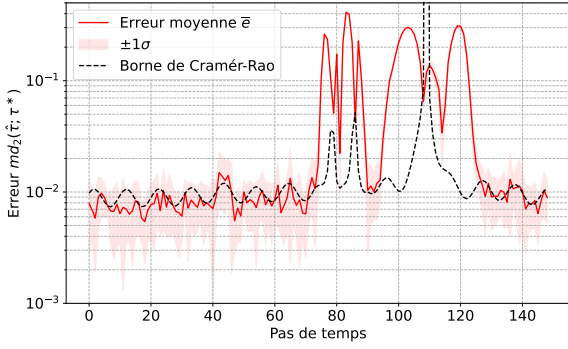


FIGURE 3 : Erreur moyenne (100 expériences) comparée à la borne de Cramér-Rao — SNR = 40 dB.

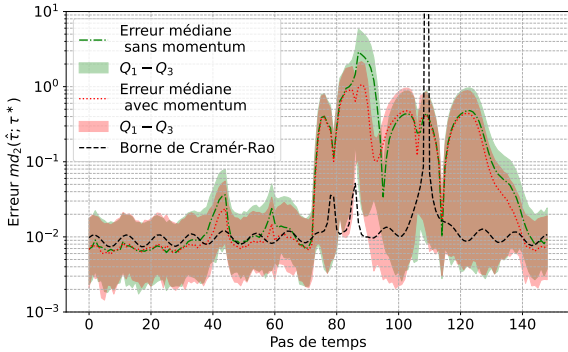


FIGURE 4 : Comparaison erreur descente préconditionnée avec et sans momentum, médiane et écart inter-quartiles (200 expériences) — SNR = 40 dB.

La Figure 4 compare les performances de l’algorithme 1 avec un grand nombre d’itérations (200) à la borne théorique de Cramér-Rao. On observe que l’algorithme suit précisément cette borne lorsque les sources sont suffisamment éloignées. Quand les sources se rapprochent, la qualité de la solution diminue tout en restant relativement proche de la borne. Dès que les sources s’éloignent à nouveau, l’algorithme retrouve des performances proche de la limite théorique. Cette expérience valide ainsi empiriquement la précision de l’algorithme dans un régime de convergence complète.

Dans un scénario embarqué temps réel, le temps de calcul entre chaque pas est souvent limité, imposant un faible nombre d’itérations de gradient. Afin d’illustrer l’importance de la vitesse de convergence, la Figure 4 compare les performances médianes de la descente préconditionnée avec et sans momentum, en fixant un budget de 40 pas de gradient. Ce nombre limité ne permet pas d’atteindre la borne de Cramér-Rao. Plusieurs expériences avec une même trajectoire mais avec un bruit aléatoire correspondant à un SNR de 40 dB montrent que l’ajout du momentum accélère la convergence, diminuant ainsi l’erreur médiane comparativement au préconditionnement seul.

De plus, une convergence trop lente augmente le risque de quitter le bassin d’attraction du minimiseur global entre deux pas de temps, pouvant mener à une divergence de l’algorithme. Pour quantifier ce risque, nous calculons le taux de décrochage en réalisant, au temps T_{\max} , une descente de gradient sur 200 pas et en vérifiant l’obtention d’une erreur autour de 10^{-1} (le bruit empêchant une convergence exacte vers θ^*). Le Tableau 1 montre que, dans la configuration de la Figure 4, le taux de

décrochage sans momentum est significativement supérieur à celui avec momentum.

Méthode	% décrochage
PGD sans momentum	8.0%
PGD avec momentum	1.5%

TABLE 1 : Impact du momentum sur le taux de décrochage

5 Conclusion

Dans cette étude, nous avons présenté une méthode de premier ordre pour résoudre un problème de déconvolution dynamique de Diracs dans le domaine de Fourier, en dimension finie quelconque. Nous avons montré expérimentalement la robustesse de cette approche en présence de bruit, tout en maintenant un budget de calcul réduit par pas de temps. De plus, l’algorithme proposé s’avère robuste même dans des situations difficiles où les Diracs se confondent à certains instants.

Cependant, notre étude présente certaines limites. Tout d’abord il serait intéressant d’étendre l’analyse du bassin d’attraction au cas où un terme de momentum est ajouté au préconditionnement, car seul le ce dernier a été étudié dans ce contexte. Ensuite, nous ne disposons pas d’analyse théorique de l’algorithme dans le cadre dynamique et plus particulièrement dans le cas où les trajectoires se croisent.

Références

- [1] Christian R. Berger, Zhaocheng Zhou, Jingwen Huang, Shengli Zhou, and Peter Willett. Sparse channel estimation for OFDM : Overcomplete dictionaries and superresolution methods. *IEEE Transactions on Signal Processing*, 61(23) :5956–5968, 2013.
- [2] Reinhard Heckel and Helmut Bolcskei. Compressive sensing radar imaging of moving targets. *2010 IEEE Radar Conference*, pages 852–857, 2010.
- [3] Florent Mertens, Audrey Repetti, Jasleen Birdi, Arwa Dabbech, and Yves Wiaux. Radio interferometric imaging with joint sparsity constraints. *Monthly Notices of the Royal Astronomical Society*, 500(1) :1037–1055, 2020.
- [4] Joseph Gabet and Maxime Ferreira Da Costa. Preconditioned gradient descent for sketched mixture learning. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2024.
- [5] Mathias Hockmann. Analysis of the sparse super resolution limit using the cramér-rao lower bound. *IEEE Transactions on Information Theory*, 71(1) :390–395, 2025.
- [6] Joseph Gabet, Meghna Kalra, Maxime Ferreira Da Costa, and Kiryung Lee. Global Convergence of ESPRIT with Preconditioned First-Order Methods for Spike Deconvolution. preprint, February 2025.
- [7] B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5) :1–17, 1964.