

# Reconstruction de Phase Garantie par Descente Miroir

Jean-Jacques GODEME<sup>1</sup>, Jalal FADILI<sup>1</sup>, Xavier BUET<sup>2</sup>, Myriam ZERRAD<sup>2</sup> Michel LEQUIME<sup>2</sup> Claude AMRA<sup>2</sup>

<sup>1</sup>Normandie Univ, ENSICAEN, UNICAEN, CNRS, GREYC, France  
6 Boulevard Maréchal Juin, 14000 Caen, France

<sup>2</sup>Aix-Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, Marseille, France  
52 Av. Escadrille Normandie Niemen, 13013 Marseille, France

jean-jacques.godeme@unicaen.fr, Jalal.Fadili@ensicaen.fr

**Résumé** – Dans ce travail, nous considérons le problème de reconstruction de phase, qui consiste à recouvrer un signal de dimension  $n$  à partir des modules de ses  $m$  mesures linéaires. Nous présentons une nouvelle approche permettant de se passer de la Lipschitz continuité globale du gradient de l’objectif non-convexe à minimiser. Pour ce faire, nous introduisons un algorithme de descente miroir non-euclidienne via la divergence de Bregman d’une entropie bien choisie. Nous montrons que lorsque le nombre de mesures  $m$  est suffisamment grand, alors avec grande probabilité, pour presque toute initialisation, l’algorithme de descente miroir, converge linéairement avec un taux indépendant de la dimension  $n$ , permettant ainsi de recouvrer le signal original à un changement de signe près. Nous énonçons nos résultats pour deux types de mesures: Gaussiennes iid et celles obtenues par des masques codés pour des transformées de Fourier aléatoires.

**Abstract** – We consider the problem of phase retrieval that consists in recovering an  $n$ -dimensional real vector from the magnitude of its  $m$ -linear measurements. This paper presents a new approach allowing to lift the classical global Lipschitz continuity requirement on the gradient of the non-convex objective to minimize. We propose a mirror descent algorithm based on a wisely chosen Bregman divergence. We show that when the number of measurements  $m$  is large enough, for almost all initializers, the mirror descent algorithm converges linearly with a dimension-independent convergence rate. Consequently, the original signal can be reconstructed exactly up to a global sign change. We state our results for two types of measurements: iid standard Gaussian and those obtained by Coded Diffraction Patterns (CDP) for Randomized Fourier Transform.

## 1 Reconstruction de phase déterministe

La reconstruction de phase est un problème inverse mal posé qui a de nombreuses applications, notamment la cristallographie, l’imagerie par diffraction, l’astronomie, la microscopie, et bien d’autres; voir [10] pour une revue. Il s’agit de reconstruire un signal de dimension  $n$  à partir des modules de ses  $m$  mesures linéaires. La principale motivation applicative de notre travail vient de l’optique et concerne la reconstruction du profil des irrégularités de surfaces presque lisses. La diffusion blanche [1] est une technique idéale pour les imager car rapide et non-invasive pour les surfaces optiques.

Plusieurs algorithmes de reconstruction de phase ont été proposés dans la littérature, notamment l’algorithme de réduction d’erreur de Gerchberg-Saxton et ses variantes comme Fienup. Ces algorithmes sont des versions en contexte non-convexe d’algorithmes d’optimisation tels que la projection alternée, Dykstra ou Douglas-Rashford; voir [2]. L’analyse des garanties de convergence et de reconstruction (exacte) de ces algorithmes reste toutefois difficile. Dans sa forme la plus générale, le problème de reconstruction de phase (voir  $(\mathcal{P})$ ) peut être vu comme un programme quadratique non-convexe réputé NP-difficile. Toutefois, toute une classe de relaxations convexes, comme PhaseLift ou Phasecut, a été développée pour les résoudre de

manière garantie; voir [7] pour une revue. Ces relaxations ne passent hélas pas à l’échelle même pour des problèmes de taille modérée. Dans [5], le flot de Wirtinger a été introduit pour résoudre le problème non-convexe de reconstruction de phase. Il s’agit d’un algorithme de descente de gradient en complexe initialisé suffisamment proche de l’ensemble des minimiseurs globaux, et pour lequel des garanties de reconstruction peuvent être établies pour des mesures aléatoires. L’algorithme que nous proposons s’inscrit dans cette lignée mais présentera des avantages importants, notamment l’indépendance du pas de descente et du taux de convergence de la dimension  $n$ .

### 1.1 Formulation du problème

Considérons le problème de reconstruction de phase :

$$\left\{ \begin{array}{l} \text{Reconstruire un signal } \bar{x} \in \mathbb{R}^n \text{ à partir de} \\ y_r = |\langle a_r, \bar{x} \rangle|^2, \quad r \in [m]. \end{array} \right. \quad (\mathcal{P})$$

Les vecteurs  $(a_r)_{r \in [m]}$ , dans  $\mathbb{R}^n$  ou  $\mathbb{C}^n$ , sont appelés les vecteurs de mesures. Le problème  $(\mathcal{P})$  peut être reformulé comme le problème d’optimisation suivant :

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) \stackrel{\text{def}}{=} \frac{1}{4m} \sum_{r=1}^m (y_r - |\langle a_r, x \rangle|^2)^2 \right\}. \quad (1)$$

La fonction  $f$  est  $C^1(\mathbb{R}^n)$  mais évidemment non-convexe. De plus  $f$  n'est pas à gradient globalement Lipschitz, ce qui proscriit l'utilisation d'une descente de gradient simple. C'est la raison principale pour laquelle nous proposons de modifier la géométrie du problème en allant vers une descente de gradient dite miroir. Pour cela, nous considérons la fonction  $\psi$ , appelée entropie, proposée dans [3]

$$\psi(x) \stackrel{\text{def}}{=} \frac{1}{4} \|x\|^4 + \frac{1}{2} \|x\|^2, \quad (2)$$

La fonction  $\psi$  est fortement convexe et  $C^1(\mathbb{R}^n)$ . Cette fonction jouera un rôle fondamental permettant de généraliser la notion de Lipschitz continuité du gradient (en fait le lemme dit de descente), en introduisant une relation d'ordre entre les divergences de Bregman de  $f$  et de  $\psi$  (voir Définition 1.1).

Rappelons que la divergence de Bregman associée à  $\phi \in C^1(\mathbb{R}^n)$  est définie par :

$$D_\phi(x, z) \stackrel{\text{def}}{=} \phi(x) - \phi(z) - \langle \nabla \phi(z), x - z \rangle, \quad \forall x, z \in \mathbb{R}^n. \quad (3)$$

La définition suivante, introduite dans [3], étend la propriété de Lipschitz continuité du gradient d'une fonction  $g$  relativement à une entropie  $\phi$  :

**Définition 1.1. ( $L$ -régularité relative)** Un couple de fonctions  $C^1(\mathbb{R}^n)$ ,  $(g, \phi)$ , satisfait la propriété de  $L$ -régularité relative s'il existe  $L > 0$  tel que  $L\phi - g$  soit convexe, *i.e.*

$$D_g(x, z) \leq LD_\phi(x, z), \quad \forall x, z \in \mathbb{R}^n. \quad (4)$$

Le lemme suivant suit d'un calcul direct.

**Lemme 1.2.** *Le couple  $(f, \psi)$  vérifie la  $L$ -régularité relative avec  $L \geq 3 \max_{1 \leq r \leq m} \|a_r\|^4$ .*

Ce lemme donne toutefois une estimation grossière. Dans le cas de mesures aléatoires, nous fournirons une borne bien plus précise.

Nous introduisons aussi une généralisation de la notion de forte convexité locale relativement à une entropie.

**Définition 1.3. (Forte convexité relative locale)** Une fonction  $g$  sera dite localement fortement convexe par rapport à  $\phi$  en un point  $x \in \mathbb{R}^n$  s'il existe un voisinage  $\mathcal{V}_x$  de  $x$  et  $\sigma > 0$  tels que  $g - \sigma\phi$  soit convexe sur  $\mathcal{V}_x$ , *i.e.*

$$D_g(x, z) \geq \sigma D_\phi(x, z), \quad \forall z \in \mathcal{V}_x. \quad (5)$$

Le schéma itératif que nous proposons pour résoudre (1) est résumé dans l'Algorithme 1 qui implémente une descente miroir avec backtracking.

## 1.2 Garanties déterministes

Supposons dans cette section que la suite  $(x_k)_{k \in \mathbb{N}}$  est générée par Algorithme 1. Pour énoncer les résultats de convergence, on rappelle la notion suivante.

---

**Algorithm 1:** Descente Miroir pour la reconstruction de phase

---

**Paramètres :**  $x_0 \in \mathbb{R}^n$ ,  $L_0 > 0$ ,  $\kappa > 0$  (petit),  $\eta \geq 1$  ;

**Initialisation :**  $x_0 \in \mathbb{R}^n$  ;

**for**  $k = 0, 1, \dots$  **do**

**repeat**

$L_k \leftarrow L_k / \eta$ ,  $\gamma_k = \frac{1-\kappa}{L_k}$  ;

$x_{k+1} = \nabla \psi^* (\nabla \psi(x_k) - \gamma_k \nabla f(x_k))$  ;

**until**  $D_f(x_{k+1}, x_k) > L_k D_\psi(x_{k+1}, x_k)$  ;

---

**Définition 1.4. (Topologie  $f$ -attentive)** Soit  $x^*$  un minimiseur de  $f$ ,  $x \in \mathbb{R}^n$  est dans le voisinage de  $x^*$  dans la topologie  $f$ -attentive, si pour tout  $r > 0$ , il existe  $u \in ]0, r[$  et  $\mu > 0$  tel que :  $\|x - x^*\| \leq u$  et  $f(x) \in ]f(x^*), f(x^*) + \mu[$ .

Nous avons les résultats de convergence suivants.

**Théorème 1.5.** *Soit  $(x_k)_{k \in \mathbb{N}}$  une suite générée par l'Algorithme 1. Les assertions suivantes sont vraies :*

(i) *la suite  $(f(x_k))_{k \in \mathbb{N}}$  converge.*

(ii) *la suite  $(x_k)_{k \in \mathbb{N}}$  est de longueur finie et converge vers un point critique de  $f$ .*

(iii) *Soit  $x^*$  un minimiseur global du problème (1), et supposons que le point initial  $x_0$  est dans le voisinage de  $x^*$  dans la topologie  $f$ -attentive, alors,*

(a)  *$\exists r > 0$ , tel que  $\forall k \in \mathbb{N}$ ,  $x_k \in B(x^*, r)$  et  $x_k \rightarrow x^*$ .*

(b) *De plus, si  $f$  est localement  $\sigma$ -fortement convexe par rapport à  $\psi$  alors,  $\exists q \in ]0, 1[$  et  $C > 0$  tel que  $\forall k \in \mathbb{N}$*

$$\|x_k - x^*\|^2 \leq C (1 - q)^k D_\psi(x^*, x_0). \quad (6)$$

Lorsque le pas descente  $\gamma_k = \frac{1-\kappa}{L}$ , où  $L$  est une estimation de la constante de  $L$ -régularité relative (voir Lemme 1.2 pour une estimation grossière), on a  $q = \sigma(1 - \kappa)/L \in ]0, 1[$ .

*Esquisse de preuve (voir [8] pour les détails).* Tout repose sur la relation de descente suivante, qui découle du Lemme 1.2 et de la mise à jour de  $x_k$  dans l'algorithme,

$$D_\psi(u, x_{k+1}) + \gamma_k (f(x_{k+1}) - f(u)) \leq D_\psi(u, x_k) - \gamma_k D_f(u, x_k) - (1 - \gamma_k L_k) D_\psi(x_{k+1}, x_k), \quad (7)$$

valable en un  $u \in \mathbb{R}^n$  quelconque. On applique ensuite (7) en  $u = x_k$  pour avoir (i). Pour (ii), on utilise aussi le fait que  $f$  est algébrique et que  $\psi$  est fortement convexe. Pour les assertions dans (iii), on on applique (7) en  $u = x^*$ , et on utilise (i) pour avoir (a). La forte convexité relative locale de  $f$  par rapport  $\psi$  et la forte convexité de cette dernière donne ensuite (b).  $\square$

## 2 Cas des mesures aléatoires

L'idée directrice maintenant est de vérifier que les conditions du Théorème 1.5, en particulier celles de (iii), sont vérifiées

avec grande probabilité lorsque les  $(a_r)_{r \in [m]}$  sont tirées aléatoirement à partir d'une distribution appropriée, avec  $m$  suffisamment grand. Nous étudions ici deux types de mesures aléatoires :

1. Mesures Gaussiennes : les  $(a_r)_{r \in [m]}$  sont i.i.d suivant une loi Gaussienne standard.
2. Mesures CDP [4] : ce modèle est inspiré par l'imagerie diffractive. Il consiste à moduler le signal avant d'en prendre la transformée de Fourier. Dans ce cas, pour  $j \in \{0, \dots, n-1\}$  et  $p \in \{0, \dots, P-1\}$ , les mesures sont données par

$$y_{j,p} = |\mathcal{F}(D_p x)[j]|^2 = \left| \sum_{\ell=0}^{n-1} x[\ell] d_p[\ell] e^{-2i\pi j \ell / n} \right|^2. \quad (8)$$

Le nombre de mesures est alors  $m = nP$ , et les  $D_p$  sont des matrices diagonales, de diagonale  $d_p \in \mathbb{R}^n$  chacune, et les  $d_p$  sont des copies i.i.d d'un vecteur aléatoire  $d$ . Ces derniers vérifient :

- Hyp. 2.1.** (a)  $\exists M > 0$  tel que  $|d| \leq M$  ;  
(b) Conditions de moments :  $\mathbb{E}(d) = 0$  et  $\mathbb{E}(|d|^4) = 2(\mathbb{E}(d^2))^2$ .

A titre d'illustration, on peut utiliser  $d = \{-1, 0, 1\}$  avec probabilité  $\{\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\}$ .

## 2.1 Garanties pour les mesures aléatoires

Le problème de reconstruction de phase sur  $\mathbb{R}^n$  souffre d'une ambiguïté puisqu'un signal ne peut être construit qu'à un changement de signe global près. Pour le signal original  $\bar{x} \in \mathbb{R}^n$  dans  $(\mathcal{P})$ , on pose alors

$$\text{dist}(x, \bar{x}) = \min \{ \|x + \bar{x}\|; \|x - \bar{x}\| \}. \quad (9)$$

Nous sommes en mesure d'énoncer le résultat principal de ce travail.

**Théorème 2.2.** Soit  $\delta > 0$  suffisamment petit. Supposons que le nombre de mesures  $m$  est suffisamment grand i.e.  $m \geq C(\delta)n \log(n)^3$ . Alors avec grande probabilité, pour presque toute initialisation  $x_0$  de l'Algorithme 1,  $\text{dist}(x_k, \bar{x}) \rightarrow 0$  et  $\exists K \geq 0$  assez grand tel que  $\forall k \geq K$

$$\text{dist}^2(x_k, \bar{x}) = \mathcal{O} \left( \left( 1 - \frac{1}{3(3+\delta)} \right)^{k-K} \right). \quad (10)$$

En particulier, si l'Algorithme 1 est initialisé par la méthode spectrale de l'Algorithme 2, alors avec grande probabilité (10) est vraie pour tout  $k \geq 0$ .

Clairement, dès lors que  $x_0$  est tiré uniformément dans  $\mathbb{R}^n$ , notre algorithme convergera vers  $\bar{x}$  à un signe global près, avec une vitesse local linéaire. Une propriété remarquable de notre approche est que le taux de convergence linéaire local est indépendant de la dimension  $n$  du signal, alors que flot de Wirtinger proposé dans [5] possède un taux de convergence  $1 - \frac{\mu}{n}$ , pour une constante  $\mu > 0$ .

---

### Algorithm 2: Initialisation spectrale.

---

**Input:**  $y_r, r = 1, \dots, m$

**Output:**  $x_0$

Soit  $\lambda^2 = n \frac{\sum_r y_r}{\sum_r \|a_r\|^2}$  ;

Prendre  $x_0$  comme un vecteur propre associé à la plus grande valeur propre de la matrice

$$Y = \frac{1}{m} \sum_{r=1}^m y_r a_r a_r^* \text{ renormalisé à } \|x_0\| = \lambda.$$


---

*Esquisse de preuve (voir [8] pour les détails).* La preuve repose sur l'utilisation du Théorème 1.5 et du théorème de la variété centrale. Pour cela, nous opérons en plusieurs étapes :

1. Nous savons que  $x_k$  converge vers un point critique par Théorème 1.5(ii). On peut montrer que pour  $m$  suffisamment grand, avec grande probabilité, l'objective  $f$  ne possède que deux minima globaux en  $\pm \bar{x}$  et des points selles avec une courbure strictement négative dans au moins une direction. Ainsi, comme par [9, Corollaire 7], on sait que l'ensemble des  $x_0$  pour lesquels  $x_k$  converge vers de tels points selles stricts est de mesure de Lebesgue nulle, on obtient la première assertion.  
Pour montrer la convergence linéaire locale (10) :
2.  $L$ -régularité relative : on montre tout d'abord que pour  $\delta > 0$  suffisamment petit, si le nombre de mesures  $m$  est suffisamment élevé, alors avec grande probabilité :

$$D_f(x, z) \leq (3 + \delta) D_\psi(x, z), \forall x, z \in \mathbb{R}^n. \quad (11)$$

La preuve passe par une inégalité de concentration sur la hessienne de  $f$  autour de son espérance, et ensuite en bornant l'écart entre cette dernière et l'espérance de la hessienne de  $\psi$ . Observons que cette estimation est bien meilleure que celle du Lemme 1.2 puisqu'elle est indépendante de la dimension  $n$  et des observations  $(y_r)_{r \in [m]}$ .

3.  $f$  est localement fortement convexe relativement à  $\psi$  avec grande probabilité. En effet, pour  $\delta > 0$  suffisamment petit, si le nombre de mesures  $m$  est assez grand, alors avec grande probabilité  $\exists \epsilon > 0$  petit tel que :

$$\forall x \in B(\bar{x}, \epsilon), \quad \frac{1}{3} D_\psi(\bar{x}, x) \leq D_f(\bar{x}, x). \quad (12)$$

4. Pour conclure, puisque  $\text{dist}(x_k, \bar{x}) \rightarrow 0$ , il existe  $K \geq 0$  assez grand, qui dépend de  $\epsilon$ , tel que pour tout  $k \geq K$ ,  $x_k$  sera au voisinage de  $\pm \bar{x}$  dans la topologie attentive. Il suffit enfin combiner Théorème 1.5(iii) avec une borne d'union assurant que les trois événements ci-avant sont simultanément vrais avec grande probabilité.
5. Pour la dernière assertion, l'initialisation spectrale de l'Algorithme 2 assure que  $x_0$  sera au voisinage de  $\pm \bar{x}$  dans la topologie attentive avec grande probabilité si le nombre de mesures est suffisamment grand. □

## 2.2 Stabilité au bruit

Les résultats plus haut peuvent être étendus pour montrer la stabilité à un bruit additif. L'objectif dans ce cas est de recons-

truire  $\bar{x} \in \mathbb{R}^n$  à partir de

$$y_r = | \langle a_r, \bar{x} \rangle |^2 + \varepsilon_r, \quad r \in [m], \quad (\text{Noisy-PR})$$

où  $\varepsilon \in \mathbb{R}^m$  est le bruit, qui satisfait l'hypothèse suivante sur le rapport signal sur bruit :

**Hyp. 2.3.** *Il existe  $c_s \in ]0; \frac{1}{3(1+c_1)}$  [ telle que  $\|\varepsilon\|_\infty \leq c_s \|\bar{x}\|^2$ , avec  $c_1 > 0$  assez petit.*

On propose de résoudre le problème (1) avec l'Algorithme 1. Nous avons donc le résultat suivant assurant la reconstruction stable au bruit, que nous énonçons sans fournir de preuve compte tenu de la limitation de place, mais la démonstration suit des arguments inspirés de ceux esquissés pour le cas sans bruit (voir [8]).

**Théorème 2.4.** *Soit  $\delta > 0$  suffisamment petit. Supposons que le nombre de mesures  $m$  est suffisamment grand i.e.  $m \geq C(\delta)n \log(n)$  dans le cas gaussien (resp.  $m \geq C(\delta)n \log(n)^3$  dans le cas CDP). Si l'Algorithme 1 est initialisé avec la méthode spectrale de l'Algorithme 2, alors avec une probabilité d'au moins  $1 - 12e^{-\gamma n} - \frac{8}{n^2}$  dans le cas gaussien (resp.  $1 - \frac{2}{n^3}$  dans le cas CDP), on a,*

$$\text{dist}(x_k, \bar{x}) \leq 2\sqrt{2}c_s \|\bar{x}\| + \mathcal{O}\left((1 - q(\delta))^{k/2}\right), \quad (13)$$

avec  $q(\delta) = \frac{1 - (1 + c_1)c_s}{3(3 + \delta(1 + \|\varepsilon\|_\infty))} < 1$ .

### 3 Expériences numériques

**Reconstruction de signaux et d'images** La Figure 1 montre les résultats de construction d'un signal de taille  $n = 128$  à partir de  $m = \lceil 128 \times \log(128) \rceil^3$  mesures suivant le modèle CDP, sans et avec du bruit respectivement. La Figure 2 montre le résultat de reconstruction (sans bruit) pour une image 2D  $256 \times 256$  représentant une surface de rugosité à partir de mesures CDP avec 100 masques.

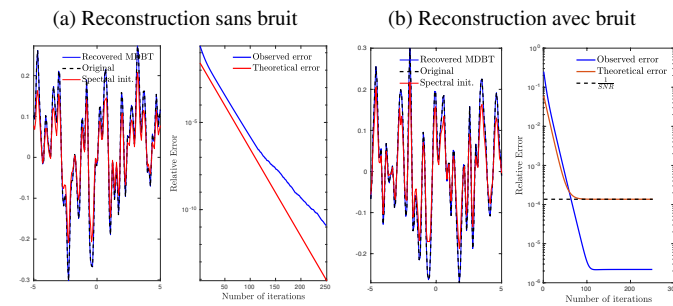


FIGURE 1 – Reconstruction à partir du modèle CDP.

Nous voyons clairement l'accord de ses simulations avec nos prédictions théoriques.

**Comparaison avec d'autres algorithmes** Nous avons aussi comparé notre algorithme avec le flot de Wirtinger de [5], et la méthode du sous-gradient de Polyak utilisée et analysée dans

(a) Surface originale

(b) Surface reconstruite

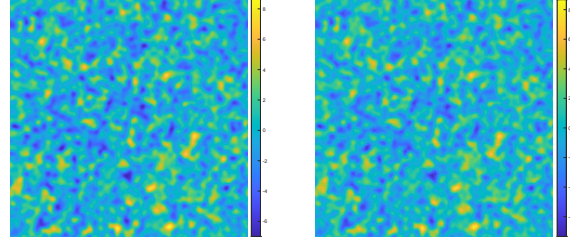


FIGURE 2 – Reconstruction de la surface de rugosité à partir de simulations.

[6]. Cette dernière ne peut s'appliquer que pour des mesures sans bruit. Nous présentons les diagrammes de transition de phase de chaque algorithme en Figure 3. On considère qu'un signal est parfaitement reconstruit si l'erreur relative est plus petite que  $10^{-5}$ . On trace le pourcentage de réussites parmi 100 instances du problème tirés aléatoirement pour chaque nombre de mesures/masques.

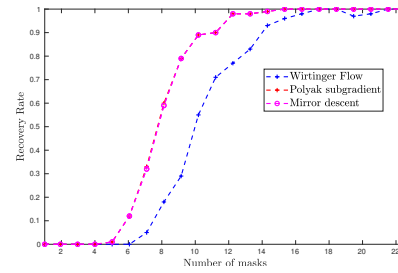


FIGURE 3 – Diagramme de transition de phase pour les mesures CDP.

### Références

- [1] C. Amra, M. Zerrad, S. Liukaityte, and M. Lequime. Instantaneous one-angle white-light scatterometer. *Opt. Express, OE*, 26(1) :204–219, January 2018.
- [2] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Phase retrieval, error reduction algorithm, and Fienup variants : a view from convex optimization. *J. Opt. Soc. Am. A*, 19(7) :1334, July 2002.
- [3] J. Bolte, S. Sabach, M. Teboulle, and Y. Vaisbourd. First order methods beyond convexity and lipschitz gradient continuity with applications to quadratic inverse problems. *SIAM J. Optim.*, 28(3) :2131–2151, 2018.
- [4] E. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2) :277–299, September 2015.
- [5] E. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via wirtinger flow : Theory and algorithms. *IEEE Trans. Inform. Theory*, 61(4) :1985–2007, 2015.
- [6] D. Davis, D. Drusvyatskiy, and C. Paquette. The nonsmooth landscape of phase retrieval. *IMA Journal of Numerical Analysis*, 40(4) :2652–2695, October 2020.
- [7] A. Fannjiang and T. Strohmer. The numerics of phase retrieval. *Acta Numerica*, 29 :125–228, May 2020.
- [8] J.-J. Godeme, J. Fadili, X. Buet, M. Zerrad, M. Lequime, and C. Amra. Provable phase retrieval via mirror descent. (*to be submitted*), 2022.
- [9] J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan, and B. Recht. First-order methods almost always avoid strict saddle points. *Mathematical Programming*, 176(1) :311–337, 2019.
- [10] Y. Shechtman, Y. C. Eldar, O. Cohen, N. Chapman, J. Miao, and M. Segev. Phase Retrieval with Application to Optical Imaging : A contemporary overview. *IEEE Signal Processing Magazine*, 32(3) :87–109, May 2015.