



**HAL**  
open science

## Reconstruction de Phase Garantie par Descente Miroir

Jean-Jacques Godeme, Jalal Fadili, Xavier Buet, Myriam Zerrad, Michel  
Lequime, Claude Amra

► **To cite this version:**

Jean-Jacques Godeme, Jalal Fadili, Xavier Buet, Myriam Zerrad, Michel Lequime, et al.. Reconstruction de Phase Garantie par Descente Miroir. 2022. hal-03622580v3

**HAL Id: hal-03622580**

**<https://hal.science/hal-03622580v3>**

Preprint submitted on 17 Oct 2022 (v3), last revised 20 Jun 2024 (v6)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Provable Phase Retrieval with Mirror Descent

Jean-Jacques Godeme\*    Jalal Fadili\*    Xavier Buet<sup>†</sup>    Myriam Zerrad<sup>†</sup>  
Michel Lequime<sup>†</sup>    Claude Amra<sup>†</sup>

**Abstract.** In this paper, we consider the problem of phase retrieval, which consists of recovering an  $n$ -dimensional real vector from the magnitude of its  $m$  linear measurements. We propose a mirror descent (or Bregman gradient descent) algorithm based on a wisely chosen Bregman divergence, hence allowing to remove the classical global Lipschitz continuity requirement on the gradient of the non-convex phase retrieval objective to be minimized. We apply the mirror descent for two random measurements: the i.i.d standard Gaussian and those obtained by multiple structured illuminations through Coded Diffraction Patterns (CDP). For the Gaussian case, we show that when the number of measurements  $m$  is large enough, then with high probability, for almost all initializers, the algorithm recovers the original vector up to a global sign change. For both measurements, the mirror descent exhibits a local linear convergence behaviour with a dimension-independent convergence rate. Our theoretical results are finally illustrated with various numerical experiments, including an application to the reconstruction of images in precision optics.

**Key words.** Phase retrieval, Inverse problems, Mirror descent, Random measurements.

## 1 Introduction

### 1.1 Problem statement and motivations

In this work, we study phase retrieval which is an ill-posed inverse problem which consists in recovering a general signal from the intensity of its  $m$  linear measurements, i.e., from phaseless observations. Historically, the first application of phase retrieval started with X-ray crystallography, and it now permeates many areas of imaging science with applications that include diffraction imaging, astronomical imaging, microscopy to name just a few; see [54, 37, 46] and references therein. One of the main applications motivating our work originates from precision in optics. Often components (e.g., interference filters) exhibit optical losses of order  $10^{-6}$  of the incident power. Super-polished surfaces are commonly used to circumvent this issue. Indeed, their roughness (responsible for losses by optical scattering) is very low compared to the illumination wavelength. Therefore, it is crucial to know how to characterize the roughness of polished surfaces. To do so, light scattering is ideal among the existing techniques because it is fast and non-invasive. The surface is illuminated with a laser source, and the diffusion is measured by moving a detector. Then the power spectral density of the surface topography can be directly measured thanks to the electromagnetic theory of light scattering; see [3, 17] and references therein for a detailed description.

Our focus in this paper will be on the case of real signals. Formally, suppose  $\bar{x} \in \mathbb{R}^n$  is a signal and that we are given information about the squared modulus of the inner product between  $\bar{x}$  and  $m$

---

\*Normandie Univ, ENSICAEN, CNRS, GREYC, France. e-mail: jean-jacques.godeme@unicaen.fr, Jalal.Fadili@ensicaen.fr.

<sup>†</sup>Aix-Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, Marseille, France. `firstname.lastname@fresnel.fr`.

sensing/measurement vectors  $(a_r)_{r \in [m]}$ . The phase retrieval problem can be cast as:

$$\begin{cases} \text{Recover } \bar{x} \in \mathbb{R}^n \text{ from the measurements } y \in \mathbb{R}^m \\ y[r] = |a_r^* \bar{x}|^2, \quad r \in [m] \end{cases} \quad (\text{GeneralPR})$$

where  $y[r]$  is the  $r$ -th entry of the vector  $y$ .

Since  $\bar{x}$  is real-valued, the best one can hope is to ensure that  $\bar{x}$  is uniquely determined by  $y$  up to a global sign. Phase retrieval is in fact an ill-posed inverse problem in general and is known to be NP-hard [53]. Thus, one of the major challenges is to design efficient recovery algorithms and find conditions on  $m$  and  $(a_r)_{r \in [m]}$  which guarantee exact (up to a global sign change) and robust recovery.

## 1.2 Prior work

Our review here is by no means exhaustive and the interested reader should refer to the following references for comprehensive reviews [54, 37, 31, 58].

**Feasibility formulation of constrained phase retrieval** In the one-dimensional case with Fourier measurements, it was shown by [1, 2, 62] that the phase retrieval problem without any a priori constraints lacks uniqueness (up to trivial ambiguities). This fundamental barrier does not apply in higher dimensions as pointed out in [16], and uniqueness was shown to hold "generically" in [7].

To circumvent this barrier, workarounds have been proposed that involve adding a constraint either implicitly or explicitly. Phase retrieval is then formulated as a feasibility problem, that is, as finding some point in the intersection of the set of points satisfying the constraints implied by the data measurements in (GeneralPR), and the set of points satisfying constraints expressing some prior knowledge on the object to recover, such as support, band-limitedness, non-negativity, sparsity, etc. The Gerchberg and Saxton algorithm [33], proposed in the early 70's in the optics literature, is an alternating projection algorithm to solve such a feasibility problem. Improved variants include Fienup's basic input-output and the hybrid input-output (HIO) [27, 32, 28]. For the case of a support constraint alone, it has been identified by [10] that HIO corresponds to the now well-known Douglas-Rachford algorithm. Other fixed-point iterations based on projections that apply to constrained phase retrieval have also been proposed, such as the HPR scheme [11], or RAAR [44] which is a relaxation of Douglas-Rachford. Thanks to a wealth of results in the variational analysis community, some convergence properties of these algorithms for the phase retrieval problem are now known. We know for instance that alternating projections applied to consistent phase retrieval problems is locally linearly convergent at points of intersection provided that the constraints do not intersect tangentially [30, 12, 50, 45, 39, 40]. Similar results are also known for Douglas-Rachford [36, 51]. Establishing global convergence guarantees is however difficult, and translating the non-tangential intersection into conditions on  $m$  and  $(a_r)_{r \in [m]}$  remains open.

**Unconstrained phase retrieval** In the unconstrained setting of (GeneralPR), the dominant approach in computational phase retrieval is to take more measurements to ensure well-posedness and improve the performance of phase retrieval algorithms. This idea of oversampling has been known for a while, and for instance in non-crystallographic modalities [47]. From a theoretical point of view, for the case where  $(a_r)_{r \in [m]}$  is a frame (redundant complete system), the authors in [6, 5] derived various necessary and sufficient conditions for the uniqueness of the solution, as well as algebraic polynomial-time numerical algorithms valid for very specific choices of  $(a_r)_{r \in [m]}$ . This approach is however of theoretical interest only and has drawbacks for instance that it requires specific types of measurements that cannot be realized in most applications of interest.

A very different route consists in considering that the measurement vectors  $(a_r)_{r \in [m]}$  are sampled from an appropriate distribution, and then showing that when  $m$  is on the order of  $n$  (up to polylogarithmic

factors), then with high probability the original vector can be recovered exactly up to sign or phase change in the complex case, from the magnitude measurements in (GeneralPR). This can be done either through semidefinite convex relaxation or by directly attacking the non-convex formulation of the phase retrieval problem.

**Convex relaxation** The key ingredient is to use a well-known trick turning a quadratic function on  $\mathbb{R}^n$ , such as in the data measurements in (GeneralPR), into a linear function on the space of  $n \times n$  matrices [13, 34]. Thus the recovery of a vector from quadratic measurements is lifted into that of recovering a rank-one Hermitian semidefinite positive (SDP) matrix from affine constraints, and the rank-one constraint is then relaxed into a convenient convex one. The two most popular methods in this line are PhaseLift [23] and PhaseCut [61]. Both approaches are inspired by the matrix completion problem [21] and they differ in the way factorization takes place. Exact and robust recovery with random Gaussian or CDP (Coded Diffraction Patterns) measurements using PhaseLift was established in [22, 20, 18]. For Gaussian measurements, [22] showed that exact recovery by PhaseLift holds for a sampling complexity bound  $m \gtrsim n \log(n)$ . This has been improved to a universal result with  $m \gtrsim n$ . Exact recovery by PhaseLift for CDP measurements was established in [18] for  $m \gtrsim n \log^4(n)$ , and has been improved to  $m \gtrsim n \log^2(n)$  in [35]. While SDP based relaxations lead to solving tractable convex problems, the prospect of squaring the number of unknowns make them computationally prohibitive and impractical as  $n$  increases. Since then, more direct non-convex methods are again being proposed.

**Nonconvex formulations** The general strategy here is to use an initialization techniques that land one in a neighborhood of the optimal solution (up to global sign or phase change) where a usual iterative procedure from nonlinear programming with carefully chosen parameters can perform reliably.

In [19], the authors use a spectral initialization and propose a gradient-descent type algorithm (Wirtinger flow) for solving the general complex phase retrieval problem by casting it as

$$\min_{z \in \mathbb{C}^n} f(z) \stackrel{\text{def}}{=} \frac{1}{4m} \sum_{r=1}^m (y[r] - |a_r^* z|^2)^2. \quad (1.1)$$

For an appropriate (Wirtinger) gradient-descent step-size, they showed that with high probability, the scheme converges linearly to the true vector (up to a global phase change) for both Gaussian and CDP measurements provided that  $m$  is on the order of  $n$  up to polylogarithmic terms. A truncated version of the Wirtinger flow was proposed in [25] which uses careful selection rules providing a tighter initial guess, better descent directions and step-sizes, and thus enhanced performance. For Gaussian measurements, truncated Wirtinger flow was also shown to converge linearly to the correct solution and is robust to noise provided that  $m \gtrsim n$ . Other variants of Wirtinger flow possibly and/or other initializations were proposed in [64] and [63], and were shown to enjoy similar guarantees in the noiseless case for Gaussian measurements. The Polyak subgradient method to minimize  $\frac{1}{m} \sum_{r=1}^m |y[r] - \langle a_r, z \rangle|^2$  was proposed and analyzed in [29] for noiseless real phase retrieval with isotropic sub-gaussian measurements. When properly initialized, its linear convergence was also shown for  $m \gtrsim n$ . Observe that this algorithm requires the knowledge of the minimal value of the phase retrieval objective, which is obviously 0 for the noiseless case.

An alternating minimization strategy, alternating between phase update and vector update, with a resampling-based initialization has been proposed in [49] and was shown to enjoy noiseless exact recovery for  $m \gtrsim n \log(n)^3$ . A truncated version of the spectral initialization followed by alternating projection was also proposed in [60] with exact recovery guarantees for Gaussian measurements under the sample complexity bound  $m \gtrsim n$ .

The authors in [55] studied the landscape geometry of the nonconvex objective in (1.1) for Gaussian measurements. They showed that for large enough number of measurements, i.e.,  $m \gtrsim n \log(n)^3$ , there

are no spurious local minimizers, all global minimizers are equal to the correct signal  $\bar{x}$ , up to a global sign or phase, and the objective function has a negative directional curvature around each saddle point (that we coin strict saddles in our paper). This allowed them to describe and analyze a second-order trust-region algorithm to find a global minimizer without special initialization. The work of [26] provides an analysis of global convergence properties of gradient descent for (real) Gaussian measurements and heavily relying on Gaussianity of the initialization. They required a sampling complexity bound  $m \gtrsim npoly \log(m)$  without making explicit the linear local convergence rate.

### 1.3 Contributions and relation to prior work

In this paper, we consider the real<sup>1</sup> phase retrieval problem that we formulate (**GeneralPR**) as the minimization problem (2.8). Inspired by [15], we propose a mirror descent (or Bregman gradient descent) algorithm with backtracking associated to a wisely chosen Bregman divergence, hence removing the classical global Lipschitz continuity requirement on the gradient of the nonconvex objective in (2.8).

In the deterministic case, we show that for almost all initializers, bounded iterates of our algorithm converge to a critical point where the objective has no direction of negative curvature, i.e., a critical point which is not a strict saddle point. In addition, under a local relative strong convexity assumption, we also show that our mirror descent scheme exhibits a local linear convergence behaviour.

In the case of i.i.d standard Gaussian measurements, provided that the the number  $m$  of sensing vectors is large enough, it turns out that the iterates of our algorithm are bounded and that the set of critical points is the union of  $\{\pm\bar{x}\}$  and the set of strict saddle points. This together with the above deterministic guarantees ensures that with high probability, for almost all initializers, our mirror descent recovers the original vector  $\bar{x}$  up to a global sign change, and exhibits a local linear convergence behaviour with a dimension-independent convergence rate. This is in contrast with the Wirtinger flow [19] which requires spectral initialization and whose local convergence rate degrades with the dimension, though the latter aspect has been improved in the truncated Wirtinger flow [25] and the Polyak subgradient method [29] (both also require spectral initialization). Our results are also far more general than those of [26] requiring for instance a smaller sampling complexity bound and assuming any random initialization provided that it is drawn from a distribution that has a density *w.r.t* the Lebesgue measure, *i.e.* the Gaussian nature of initialization in [26] is irrelevant in our context.

For both CDP and Gaussian measurements, we show that one can afford a smaller sampling complexity bound but at the price of using an appropriate spectral initialization procedure to find an initial guess near the solution before applying our scheme. Starting from this initial guess, mirror descent then converges linearly to the true vector up to a global sign change.

Though we focus on Gaussian measurements when establishing the global recovery properties of our mirror descent algorithm, our theory extends to the situation where the  $a_r$ 's are i.i.d sub-Gaussian random vectors. The case where  $a_r$ 's are drawn from the CDP model is, however, far more challenging. One of the main difficulties is that several of our arguments rely on uniform bounds, for instance on the Hessian, that need to hold simultaneously for all vectors  $x \in \mathbb{R}^n$  with high probability. But the CDP model bears much less randomness to exploit for establishing such bounds with reasonable sampling complexity bounds. Whether this is possible or not is an open problem that we leave to future research.

### 1.4 Paper organization

The rest of the paper is organized as follows. In Section 2, we describe the mirror descent algorithm with backtracking and establish its global and local convergence guarantees in the deterministic case. We then turn to the case of random measurements in Section 3 where we provide sample complexity bounds for the

---

<sup>1</sup>This is motivated by main application in light scattering where the roughness of a surface to be recovered is real.

deterministic guarantees to hold with high probability. Section 4 is devoted to the numerical experiments. The proofs of technical results are collected in the appendix.

## 2 Deterministic Phase Retrieval

**Notations** We denote  $\langle \cdot, \cdot \rangle$  the scalar product and  $\|\cdot\|$  the corresponding norm.  $B(x, r)$  is the corresponding ball of radius centered at  $x$  and  $\mathbb{S}^{n-1}$  is the corresponding unit sphere. We use the shorthand notation  $[m] = \{1, \dots, m\}$ . The  $i$ -th entry of a vector  $x$  is denoted  $x[i]$ . Given a matrix  $M$ ,  $M^\top$  is its transpose and  $M^*$  is its adjoint (transpose conjugate). Let  $\lambda_{\min}(M)$  and  $\lambda_{\max}(M)$  be respectively the smallest and the largest eigenvalues of  $M$ . For two real symmetric matrices  $M$  and  $N$ ,  $M \succeq N$  if  $M - N$  is positive semidefinite.  $\text{int}$  is the interior of a set. We denote by  $\Gamma_0(\mathbb{R}^n)$  the class of proper lower semicontinuous convex function.  $\text{dom}(f)$  is the domain of the function  $f$ . Recall that the set of critical points of  $f \in C^1(\mathbb{R}^n)$  is  $\text{crit}(f) = \{x \in \mathbb{R}^n : \nabla f(x) = 0\}$ .

Let us denote the set of true vectors by  $\bar{\mathcal{X}} = \{\pm \bar{x}\}$ . For any vector  $x \in \mathbb{R}^n$ , the distance to the set of true vectors is

$$\text{dist}(x, \bar{\mathcal{X}}) \stackrel{\text{def}}{=} \min \{\|x - \bar{x}\|; \|x + \bar{x}\|\}. \quad (2.1)$$

We will also use the shorthand notation:  $A$  is the  $m \times n$  matrix with  $a_r^*$ 's as its rows.

### 2.1 Bregman toolbox

For any  $\phi : \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  such that  $\phi \in \Gamma_0(\mathbb{R}^n) \cap C^1(\text{int}(\text{dom}(\phi)))$ , we define a proximity measure associated with  $\phi$ .

**Definition 2.1. (Bregman divergence)** The Bregman divergence associated with  $\phi$  is defined as  $D_\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow ]-\infty, +\infty]$ :

$$D_\phi(x, u) \stackrel{\text{def}}{=} \begin{cases} \phi(x) - \phi(u) + \langle \nabla \phi(u), x - u \rangle & \text{if } (x, u) \in \text{dom}(\phi) \times \text{int}(\text{dom}(\phi)), \\ +\infty & \text{otherwise.} \end{cases} \quad (2.2)$$

The classical euclidean distance is generated by the energy entropy  $\phi(x) = \frac{1}{2} \|x\|^2$ . More examples of entropies and associated Bregman divergences can be found in [57]. Clearly,  $D_\phi$  is not a distance (it is not symmetric in general for example).

We now collect some of the properties of the Bregman divergence that are will be useful in our context. See [56, 24] and [8, Proposition 2.10] for the last claim.

**Proposition 2.2. (Properties of the Bregman divergence)**

- (i)  $D_\phi$  is nonnegative if and only if  $\phi$  is convex. If in addition  $\phi$  is strictly convex,  $D_\phi$  vanishes if and only if its arguments are equal.
- (ii) *Linear additivity:* for any  $\alpha, \beta \in \mathbb{R}$  and any functions  $\phi_1$  and  $\phi_2$  we have

$$D_{\alpha\phi_1 + \beta\phi_2}(x, u) = \alpha D_{\phi_1}(x, u) + \beta D_{\phi_2}(x, u), \quad (2.3)$$

for all  $(x, u) \in (\text{dom}\phi_1 \cap \text{dom}\phi_2)^2$  such that both  $\phi_1$  and  $\phi_2$  are differentiable at  $u$ .

- (iii) *The three-point identity:* for any  $x \in \text{dom}(\phi)$  and  $u, z \in \text{int}(\text{dom}(\phi))$ , we have

$$D_\phi(x, z) - D_\phi(x, u) - D_\phi(u, z) = \langle \nabla \phi(u) - \nabla \phi(z), x - u \rangle, \quad (2.4)$$

- (iv) Suppose that  $\phi$  is also  $C^2(\text{int}(\text{dom}(\phi)))$  and  $\nabla^2 \phi(x)$  is positive definite for any  $x \in \text{int}(\text{dom}(\phi))$ . Then for every convex compact subset  $\Omega \subset \text{int}(\text{dom}(\phi))$ , there exists  $0 < \theta_\Omega \leq \Theta_\Omega < +\infty$  such that for all  $x, u \in \Omega$ ,

$$\frac{\theta_\Omega}{2} \|x - u\|^2 \leq D_\phi(x, u) \leq \frac{\Theta_\Omega}{2} \|x - u\|^2. \quad (2.5)$$

We are now ready to extend the gradient Lipschitz continuity property to the Bregman setting, that we coin relative smoothness. The notion of relative smoothness is key to the analysis of differentiable but not Lipschitz-smooth optimization problems. The earliest reference to this notion can be found in an economics paper [14] where it is used to address a problem in game theory involving fisher markets. Later on it was developed in [9, 15] and then in [43], although first coined relative smoothness in [43].

**Definition 2.3. ( $L$ -relative smoothness)**

Let  $\phi \in \Gamma_0(\mathbb{R}^n) \cap C^1(\text{int}(\text{dom}(\phi)))$ , and  $g$  be a proper and lower semicontinuous function such that  $\text{dom}(\phi) \subset \text{dom}(g)$  and  $g \in C^1(\text{int}(\text{dom}(\phi)))$ .  $g$  is called  $L$ -relatively smooth w.r.t  $\phi$  on  $\text{int}(\text{dom}(\phi))$  if there exists  $L > 0$  such that  $L\phi - g$  is convex on  $\text{int}(\text{dom}(\phi))$  i.e.,

$$D_g(x, u) \leq LD_\phi(x, u) \quad \forall x, u \in \text{int}(\text{dom}(\phi)). \quad (2.6)$$

When  $\phi$  is the energy entropy, i.e.,  $\phi = \frac{1}{2} \|\cdot\|^2$ , one recovers the standard descent lemma implied by Lipschitz continuity of the gradient of  $g$ .

In an analogous way, we also generalize the standard local strong convexity property to a relative version w.r.t to an entropy or kernel  $\phi$ .

**Definition 2.4. (Local relative strong convexity)** Let  $\phi \in \Gamma_0(\mathbb{R}^n) \cap C^1(\text{int}(\text{dom}(\phi)))$ , and  $g$  be a proper and lower semicontinuous function such that  $\text{dom}(\phi) \subset \text{dom}(g)$  and  $g \in C^1(\text{int}(\text{dom}(\phi)))$ . For  $\sigma > 0$ , we say that a function  $g$  is locally  $\sigma$ -strongly convex relatively to  $\phi$  at  $x$  if there exists  $\delta > 0$  such that

$$D_g(x, u) \geq \sigma D_\phi(x, u) \quad \forall u \in B(x, \delta) \cap \text{int}(\text{dom}(\phi)). \quad (2.7)$$

## 2.2 Phase retrieval minimization problem

In this work, we cast (**GeneralPR**) as solving the following optimization problem

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) \stackrel{\text{def}}{=} \frac{1}{4m} \sum_{r=1}^m (y[r] - |(Ax)[r]|^2)^2 \right\}. \quad (2.8)$$

Observe that  $f \in C^2(\mathbb{R}^n)$  but is obviously non-convex. It is also clear that  $\nabla f$  is not Lipschitz continuous. This is the main motivation behind considering the framework of Bregman gradient descent. As we will see shortly,  $f$  has a relative smoothness property (see Definition 2.3 above) with respect to a well-chosen entropy function. In turn, relative smoothness will prove crucial for establishing descent properties of Bregman gradient descent, also known as, mirror descent.

Following [15], let us consider the following kernel or entropy function

$$\psi(x) = \frac{1}{4} \|x\|^4 + \frac{1}{2} \|x\|^2. \quad (2.9)$$

It is easy to show that  $\psi$  enjoys the following properties.

**Property 2.5. (Properties of  $\psi$ )**

- $\psi \in C^2(\mathbb{R}^n)$  and is 1-strongly convex,
- $\nabla\psi$  is Lipschitz over bounded subsets of  $\mathbb{R}^n$ .

It turns out that the objective  $f$  in (2.8) is smooth relatively to the entropy  $\psi$  defined in (2.9) on the whole space  $\mathbb{R}^n$ . This is stated in the following result whose proof is provided in Appendix A.1.

**Lemma 2.6.** *Let  $f$  and  $\psi$  as defined in (2.8) and (2.9) respectively.  $f$  is  $L$ -relatively smooth w.r.t to  $\psi$  on  $\mathbb{R}^n$  for any  $L \geq \frac{1}{m} \sum_{r=1}^m 3 \|a_r\|^4$ .*

This estimate of of the modulus of relative smoothness  $L$  in Lemma 2.6 is rather crude but has the advantage to not depend on the measurements  $y$ . A far sharper estimate will be provided in the case where the sensing vectors are random; see Section 3.

### 2.3 Mirror descent with backtracking

We recall the following mapping closely related to the Bregman gradient descent. For all  $x \in \mathbb{R}^n$  and any step-size  $\gamma > 0$ ,

$$T_\gamma(x) \stackrel{\text{def}}{=} \operatorname{argmin}_{u \in \mathbb{R}^n} \left\{ \langle \nabla f(x), u - x \rangle + \frac{1}{\gamma} D_\psi(u, x) \right\}. \quad (2.10)$$

The pair  $(f, \psi)$  defined in (2.8)-(2.9) satisfies [15, Assumptions  $A, B, C, D$ ] (in fact  $\psi$  is even strongly convex in our case). Therefore, it is straightforward to see that  $T_\gamma$  is a well-defined and single-valued on  $\mathbb{R}^n$ ; see [15, Lemma 3.1]. Moreover,  $\psi$  is a Legendre function and thus  $\nabla\psi$  is a bijection from  $\mathbb{R}^n$  to  $\mathbb{R}^n$ , and its inverse is the gradient of the Legendre-Fenchel conjugate  $\psi^*$ ; see [52, Theorem 26.5]. In turn, letting  $x^+ = T_\gamma(x)$ , the first order optimality condition for (2.10) yields

$$x^+ = F(x) \stackrel{\text{def}}{=} \nabla\psi^{-1}(\nabla\psi(x) - \gamma\nabla f(x)) = \nabla\psi^*(\nabla\psi(x) - \gamma\nabla f(x)). \quad (2.11)$$

Our mirror descent (or Bregman gradient descent) scheme with backtracking is summarized in Algorithm 1.

---

#### Algorithm 1: Mirror Descent for Phase Retrieval

---

**Parameters:**  $L_0 = L$  (see Lemma 2.6),  $\kappa \in ]0, 1[$ ,  $\xi \geq 1$ .

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

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

**repeat**

$L_k \leftarrow L_k / \xi$ ,  $\gamma_k = \frac{1-\kappa}{L_k}$   
         $x_{k+1} = F(x_k) = \nabla\psi^*(\nabla\psi(x_k) - \gamma_k \nabla f(x_k))$

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

$L_k \leftarrow \xi L_k$ ,  $\gamma_k = \frac{1-\kappa}{L_k}$   
         $x_{k+1} = F(x_k)$ .

---

Observe that Algorithm 1 cannot be trapped in the second loop thanks to Lemma 2.6. The version without backtracking is recovered by setting  $\xi = 1$  and using constant step-size verifying  $0 < \gamma < \frac{1}{L}$  where  $L$  is the global relative smoothness coefficient. Backtracking for an inertial version of the Bregman proximal gradient algorithm was used in [48].

It remains now to compute the mirror step. This amounts to finding a root of a third-order polynomial.

**Proposition 2.7. (Mirror step computation)**[15, Proposition 5.1] *Let  $x \in \mathbb{R}^n$  and  $p_\gamma(x) = \nabla\psi(x) - \gamma\nabla f(x)$ . Then computing (2.11) amounts to,*

$$x^+ = t^* p_\gamma(x), \quad (2.12)$$

where  $t^*$  is the unique real positive root of  $t^3 \|p_\gamma(x)\|^2 + t - 1 = 0$ .

### 2.4 Deterministic recovery guarantees by mirror descent

We start with the following key bound.

**Lemma 2.8.** *Let the sequence  $(x_k)_k$  generated by Algorithm 1. Then  $\forall u \in \mathbb{R}^n$*

$$D_\psi(u, x_{k+1}) + \gamma_k (f(x_{k+1}) - f(u)) \leq D_\psi(u, x_k) - \kappa D_\psi(x_{k+1}, x_k) - \gamma_k D_f(u, x_k). \quad (2.13)$$

See Section A.2 for the proof.

We pause to recall two notions that will be important in our convergence result.



**Definition 2.9. ( $f$ -attentive neighborhood)** A point  $u \in \mathbb{R}^n$  belongs to an  $f$ -attentive neighborhood of  $x \in \mathbb{R}^n$ , if there exist  $\delta > 0$  and  $\mu > 0$  such that  $u \in B(x, \delta)$  and  $f(x) < f(u) < f(x) + \mu$ .

**Definition 2.10. (Strict saddle points)** A point  $x_* \in \text{crit}(f)$  is a strict saddle point of  $f$  if  $\lambda_{\min}(\nabla^2 f(x_*)) < 0$ . The set of strict saddle points of  $f$  is denoted  $\text{strisad}(f)$ .

We are now ready to state our main convergence result.

**Theorem 2.11.** Let  $(x_k)_{k \in \mathbb{N}}$  be a bounded sequence generated by Algorithm 1 for the phase retrieval problem (GeneralPR). Then,

- (i) the sequence  $(f(x_k))_{k \in \mathbb{N}}$  is non-increasing,
- (ii) the sequence  $(x_k)_{k \in \mathbb{N}}$  has a finite length and converges to a point in  $\text{crit}(f)$ .
- (iii) Let  $r > 0$ . Assume that the initial point  $x_0$  is in the  $f$ -attentive neighborhood of  $x^* \in \text{Argmin}(f) \neq \emptyset$ , i.e.  $\exists \delta \in ]0, r[$  and  $\mu > 0$  such that  $x_0 \in B(x^*, \delta)$  and  $f(x_0) \in ]0, \mu[$ , then
  - (a)  $\forall k \in \mathbb{N}, x_k \in B(x^*, r)$ , and  $x_k$  converges to a global minimizer of  $f$ .
  - (b) Besides, if  $\exists \rho > 0$  such that  $f$  is  $\sigma$ -strongly convex relatively to  $\psi$  at  $x^*$  on  $B(x^*, \rho)$ , with  $r \leq \frac{\rho}{\max(\sqrt{\Theta(\rho)}, 1)}$ , where we recall  $\Theta(\rho)$  from Proposition 2.2(iv), then  $\forall k \in \mathbb{N}$

$$\|x_k - x^*\|^2 \leq \prod_{i=0}^{k-1} (1 - \sigma \gamma_i) \rho^2 \rightarrow 0. \quad (2.14)$$

- (iv) If  $L_k = L$ , then for Lebesgue almost all initializers  $x_0$ , the sequence  $(x_k)_{k \in \mathbb{N}}$  converges to an element in  $\text{crit}(f) \setminus \text{strisad}(f)$ .

See Section A.3 for the proof.

**Remark 2.12.**

- A standard assumption that automatically guarantees the boundedness of the sequence  $(x_k)_{k \in \mathbb{N}}$ , hence its convergence, is coercivity of  $f$ . Since the latter is a composition of a coercive function (a positive quartic function) and the linear operator  $A$  (recall that its rows are the  $a_r^*$ 's), coercivity of  $f$  amounts to injectivity of  $A$ . This is exactly what we will show in the random case when  $m$  is large enough.
- It is clear that  $\text{Argmin}(f) \neq \emptyset$  since  $\overline{\mathcal{X}} \subset \text{Argmin}(f)$  and the claim (iii) applies at  $\pm \bar{x}$  in which case one has exact recovery up to a global sign.
- A close inspection at the proof of Proposition 2.2(iv) shows that  $\Theta(\rho) = \sup_{x \in B(\bar{x}, \rho)} \|\nabla^2 \psi(x)\|$  does the job. In view of (A.3), it is easy to see that  $\Theta(\rho) \leq 6 \|\bar{x}\|^2 + 6\rho^2 + 1$ .
- Claim (iii) shows local linear convergence of  $x_k$  to  $x^*$ . Indeed,  $\sigma \leq L_k$  for any  $k$ , and thus  $1 - \sigma \gamma_k \in ]\kappa, 1[$ .
- Clearly, claim (iv) states that when the initial point is selected according to a distribution which has a density w.r.t the Lebesgue measure, then the sequence  $(x_k)_{k \in \mathbb{N}}$  converges to a point that avoids strict saddle points of  $f$ . This is a consequence of the centre stable manifold theorem applied to our mirror descent algorithm.
- When it will come to the phase retrieval problem from random measurements (see forthcoming section), in order to prove local linear convergence, the key argument will be to show that for a sufficient number of measurements, then w.h.p  $f$  is locally relatively strongly convex w.r.t  $\psi$ .

### 3 Random Phase Retrieval via Mirror Descent

#### 3.1 Framework

Throughout the paper, we will work under two random measurement models:

- (1) The sensing vectors are drawn i.i.d following a (real) standard Gaussian distribution. We can then rewrite the observation data as

$$y[r] = |a_r^\top \bar{x}|^2, \quad r \in [m], \quad (3.1)$$

where  $(a_r)_{r \in [m]}$  are i.i.d  $\mathcal{N}(0, 1)$ .

- (2) The Coded Diffraction Patterns (CDP) model, as considered for instance in [18]. The idea is to modulate the signal before diffraction in the case of the Fourier transform measurements. The observation model is then

$$y = (|\mathcal{F}(D_p \bar{x})[j]|^2)_{j,p} = \left( \left| \sum_{\ell=0}^{n-1} \bar{x}_\ell d_p[\ell] e^{-i \frac{2\pi j \ell}{n}} \right|^2 \right)_{j,p}. \quad (3.2)$$

where  $j \in \{0, \dots, n-1\}$  and  $p \in \{0, \dots, P-1\}$ ,  $D_p$  is a real diagonal matrix with the modulation pattern  $d_p$  on its diagonal, and  $\mathcal{F}$  is the discrete Fourier transform.  $P$  is the number of coded patterns/masks and the total number of measurements is then  $m = nP$ . The modulation patterns  $(d_p)_{p \in [P]}$  are i.i.d copies of the same random vector  $d$  satisfying the following assumption:

**Assumption A.**

(A.1)  $d$  is symmetric and  $\exists M > 0$  such that  $|d| \leq M$ .

(A.2) Moments conditions:  $\mathbb{E}(d) = 0$  and  $\mathbb{E}(d^4) = 2\mathbb{E}(d^2)^2$ . Without loss of generality, we assume  $\mathbb{E}(d^2) = 1$ .

For example, we can take ternary random variables with values in  $\{-1, 0, 1\}$  with probabilities  $\{\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\}$ . We refer to [18] for other modulation patterns.

When the number of measurements is large enough for both measurements models, we will be able to establish local convergence properties of Algorithm 1 provided it is initialized with a good guess. For this, we use a spectral initialization method; see for instance [19, 25, 49, 64, 63, 60]. The procedure consists of taking  $x_0$  as the leading eigenvector of a specific matrix as described in Algorithm 2.

---

**Algorithm 2:** Spectral Initialization.

---

**Input:**  $y[r], r = 1, \dots, m$ .

**Output:**  $x_0$

Set  $\lambda^2 = n \frac{\sum_r y[r]}{\sum_r \|a_r\|^2}$ ;

Take  $x_0$  the top eigenvector of  $Y = \frac{1}{m} \sum_{r=1}^m y[r] a_r a_r^*$  normalized to  $\|x_0\| = \lambda$ .

---

We are now ready to state our main results for each measurement model.

### 3.2 Gaussian measurements

Before stating our result, we consider the following events which will be helpful in our proofs. For this, we fix  $\varrho \in ]0, 1[$  and  $\lambda \in ]0, 1[$ .

- The event

$$\mathcal{E}_{strictsad} = \{\text{crit}(f) = \bar{\mathcal{X}} \cup \text{strisad}(f)\} \quad (3.3)$$

means that the set of critical points of the function  $f$  is reduced to  $\{\pm \bar{x}\}$  and the set of strict saddle points.

- The event

$$\mathcal{E}_{conH} = \left\{ \forall x \in \mathbb{R}^n, \quad \|\nabla^2 f(x) - \mathbb{E}(\nabla^2 f(x))\| \leq \varrho \left( \|x\|^2 + \|\bar{x}\|^2 / 3 \right) \right\} \quad (3.4)$$

captures the deviation of the Hessian of  $f$  around its expectation.

- The event

$$\mathcal{E}_{inj} = \left\{ \forall x \in \mathbb{R}^n, \quad (1 - \varrho) \|x\|^2 \leq \frac{1}{m} \|Ax\|^2 \right\} \quad (3.5)$$

represents injectivity of the measurement matrix  $A$ .

- $\mathcal{E}_{smad}$  is the event on which the function  $f$  is  $L$ -relatively smooth w.r.t  $\psi$  in the sense of Definition 2.3, with  $L = 3 + \varrho \max(\|\bar{x}\|^2/3, 1)$ .
- $\mathcal{E}_{scvx}$  is the event on which  $f$  is  $\sigma$ -strongly convex relatively to  $\psi$  locally on  $B(\bar{\mathcal{X}}, \rho)$  in the sense of Definition 2.4, with  $\sigma = (\lambda \min(\|\bar{x}\|^2, 1) - \varrho \max(\|\bar{x}\|^2/3, 1))$  and  $\rho = \frac{1-\lambda}{\sqrt{3}} \|\bar{x}\|$ .
- We end up by denoting

$$\mathcal{E}_{conv} = \mathcal{E}_{strictsad} \cap \mathcal{E}_{conH} \cap \mathcal{E}_{inj} \cap \mathcal{E}_{smad} \cap \mathcal{E}_{scvx}. \quad (3.6)$$

Our main result for Gaussian measurements is the following.

**Theorem 3.1.** Fix  $\lambda \in ]0, 1[$  and  $\varrho \in ]0, \lambda \min(\|\bar{x}\|^2, 1)/(2 \max(\|\bar{x}\|^2/3, 1))]$ . Let  $(x_k)_{k \in \mathbb{N}}$  be the sequence generated by Algorithm 1.

- (i) If the number of measurements  $m$  is large enough i.e.  $m \geq C(\varrho)n \log^3(n)$ , then for almost all initializers  $x_0$  of Algorithm 1 used with constant step-size  $\gamma_k \equiv \gamma = \frac{1-\kappa}{3+\varrho \max(\|\bar{x}\|^2/3, 1)}$ , for any  $\kappa \in ]0, 1[$ , we have

$$\text{dist}(x_k, \bar{\mathcal{X}}) \rightarrow 0,$$

and  $\exists K \geq 0$ , large enough such that  $\forall k \geq K$ ,

$$\text{dist}^2(x_k, \bar{\mathcal{X}}) \leq (1 - \nu)^{k-K} \rho^2, \quad (3.7)$$

where

$$\nu = \frac{(1 - \kappa) \left( \lambda \min(\|\bar{x}\|^2, 1) - \varrho \max(\|\bar{x}\|^2/3, 1) \right)}{3 + \varrho \max(\|\bar{x}\|^2/3, 1)}. \quad (3.8)$$

This holds with a probability at least  $1 - 2e^{-\frac{m(\sqrt{1+\varrho}-1)^2}{8}} - 5e^{-\zeta n} - 4/n^2 - c/m$ , where  $C(\varrho), c$  and  $\zeta$  are numerical positive constants.

- (ii) Suppose moreover that  $\varrho$  obeys (B.12). When  $m \geq C(\varrho)n \log(n)$ , if Algorithm 1 is initialized with the spectral method in Algorithm 2, then with probability at least  $1 - 2e^{-\frac{m(\sqrt{1+\varrho}-1)^2}{8}} - 5e^{-\zeta n} - 4/n^2$  ( $\zeta$  is a fixed numerical constant), (3.7) holds for all  $k \geq K = 0$ .

Before proving our result, the following remarks are in order.

**Remark 3.2.**

- In the regime of claim (i), when  $x_0$  is chosen uniformly at random, Algorithm 1 provably converges to a global minimum up to a sign change. In this case any initialization strategy becomes superfluous, though the number of measurements required then is slightly (polylogarithmically) higher than with spectral initialization.
- In the regime of claim (ii), one has to use a spectral initialization to find a good initial guess, from which mirror descent converges locally linearly to a point around  $\bar{x}$  up to global sign change.
- When the true vector norm is one, as assumed in many works, the convergence rate takes the simple form  $\left(1 - \frac{(1-\kappa)(\lambda-\varrho)}{3+\varrho}\right) \leq \frac{2}{3} + O((1-\lambda) + \kappa + \varrho)$ .
- It is also important to point out that our convergence rate is independent from the dimension  $n$  of the signal without any specific strategy. This is in stark contrast with the Wirtinger flow [19, 18], whose convergence rate is  $\left(1 - \frac{cst}{n}\right)$  and thus dimension-dependent. Such dependence was removed for the truncated Wirtinger flow with Gaussian measurements [25].

To close these remarks, we strongly believe that handling the geometry of the problem through the framework of mirror/Bregman gradient descent with a wisely chosen entropy/kernel  $\psi$  is a key for this better behaviour in our case.

**Proof.**

(i) Assume for this claim that  $\mathcal{E}_{conv}$  holds true; we will show later that this is indeed the case *w.h.p* when the number of measurements is as large as prescribed. The proof then consists in combining Theorem 2.11 and the characterization of the structure of  $\text{crit}(f)$ .

- Global convergence of the iterates: under event  $\mathcal{E}_{inj}$  (see (3.5)), the sequence  $(x_k)_{k \in \mathbb{N}}$  generated by Algorithm 1 is bounded; see the discussion in Remark 2.12. Since  $\mathcal{E}_{smad}$  holds, Theorem 2.11(i)-(ii) ensure that the sequence  $(x_k)_{k \in \mathbb{N}}$  converges to  $x^* \in \text{crit}(f)$  and the induced sequence  $(f(x_k))_{k \in \mathbb{N}}$  converges to  $f(x^*)$ .
- Since  $\mathcal{E}_{strictsad}$  holds also, we have by Theorem 2.11(iv) that for almost all initial points  $x_0$ , the sequence  $(x_k)_{k \in \mathbb{N}}$  converges to an element of  $\text{crit}(f) \setminus \text{strisad}(f) = \bar{\mathcal{X}}$ . We assume *w.l.o.g* that  $x_k \rightarrow \bar{x}$  whence  $\|x_k - \bar{x}\| \rightarrow 0$ , and  $f(x_k) \rightarrow \min(f) = 0$ . Therefore, for  $\eta \leq \frac{\rho}{\sqrt{\max(\Theta(\rho), 1)}}$ , there exists  $\exists K = K(\eta)$  such that,

$$\forall k \geq K, \|x_k - \bar{x}\| < \eta \text{ and } f(x_k) \in ]0, \eta[, \quad (3.9)$$

*i.e.*, for  $k \geq K$ ,  $x_k$  is in an  $f$ -attentive neighborhood of  $\bar{x}$ .

- Thanks to  $\mathcal{E}_{scvx}$ ,  $f$  is  $\sigma$ -strongly convex relatively to  $\psi$  locally on  $B(\bar{x}, \rho)$  with  $\sigma$  and  $\rho$  as given in that event. It then follows from Theorem 2.11(iii) that,  $\forall k > K$  and  $\gamma_k \equiv \frac{(1-\kappa)}{3+\varrho \max(\|\bar{x}\|^2/3, 1)}$ , we have

$$\begin{aligned} D_\psi(\bar{x}, x_{k+1}) &\leq (1-\nu) D_\psi(\bar{x}, x_k) \\ &\leq (1-\nu)^{k-K} D_\psi(\bar{x}, x_K). \end{aligned}$$

Moreover by (2.5) and 1-strong convexity of  $\psi$ , for all  $k \geq K$

$$\begin{aligned} \text{dist}^2(x_k, \bar{\mathcal{X}}) &\leq \|x_k - \bar{x}\|^2 \leq 2D_\psi(\bar{x}, x_k) \leq (1-\nu)^{k-K} \Theta(\rho) \eta^2, \\ &\leq (1-\nu)^{k-K} \rho^2. \end{aligned}$$

To conclude this part of the proof we need to compute the probability that the event  $\mathcal{E}_{conv}$  occurs. We have,

$$\begin{aligned} \mathcal{E}_{conv} &= \mathcal{E}_{strictsad} \cap \mathcal{E}_{conH} \cap \mathcal{E}_{inj} \cap \mathcal{E}_{smad} \cap \mathcal{E}_{scvx}, \\ &= \mathcal{E}_{strictsad} \cap \mathcal{E}_{conH} \cap \mathcal{E}_{inj}, \end{aligned}$$

since  $\mathcal{E}_{smad} \subset \mathcal{E}_{conH}$  and  $\mathcal{E}_{scvx} \subset \mathcal{E}_{conH}$  thanks to Lemma B.5 and Lemma B.6 respectively. Owing to Lemma B.3, the event  $\mathcal{E}_{conH}$  holds true with a probability at least  $1 - 5e^{-\zeta n} - \frac{4}{n^2}$ , where  $\zeta$  is a fixed numerical constant, with the proviso that  $m \geq C(\varrho)n \log(n)$ .

On the other hand, Lemma B.4 tells us that, when  $m \geq \frac{16}{\varrho^2}n$ , the event  $\mathcal{E}_{inj}$  is true with a probability at least  $1 - 2e^{-\frac{m(\sqrt{1+\varrho}-1)^2}{8}}$ . The study of the critical points of the objective  $f$ , see [55, Theorem 2.2], shows that when  $m \geq C(\varrho)n \log^3(n)$ , the event  $\mathcal{E}_{strictsad}$  holds true with a probability  $1 - \frac{c}{m}$  (where  $c$  a fixed numerical constant). Using a union bound,  $\mathcal{E}_{conv}$  occurs with the stated high probability provided that  $m \geq C(\varrho)n \log(n)$  for a large enough numerical constant  $C(\varrho)$ .

- (ii) The proof of this claim is similar to the last part of claim (i) except that now, we invoke Lemma B.7(iii) to see that with probability at least  $1 - 2e^{-\frac{m(\sqrt{1+\varrho}-1)^2}{8}} - 5e^{-\zeta n} - \frac{4}{n^2}$ , the initial guess  $x_0$  obtained by spectral initialization belongs to  $B\left(\bar{\mathcal{X}}, \frac{\rho}{\sqrt{\max(\Theta(\rho), 1)}}\right)$ . We can now follow the reasoning in the last item of the proof of statement (i) to conclude. □

### 3.3 CDP measurements

Our main result for the CDP measurements model is the following.

**Theorem 3.3.** *Let  $\varrho \in ]0, 1[$  and  $(x_k)_{k \in \mathbb{N}}$  be the sequence generated by Algorithm 1.*

- (i) *If the number of patterns  $P$  satisfies  $P \geq C(\varrho) \log(n)$ , then with a probability at least  $1 - 1/n^2$ , for almost all initializers  $x_0$  of Algorithm 1 used with constant step-size  $\gamma_k \equiv \gamma = \frac{1-\kappa}{L}$ , for any  $\kappa \in ]0, 1[$  and  $L$  given by Lemma 2.6,  $(x_k)_{k \in \mathbb{N}}$  converges to an element in  $\text{crit}(f) \setminus \text{strisad}(f)$ .*
- (ii) *Let  $\delta \in ]0, \min(\|\bar{x}\|^2, 1)/2[$ . There exists  $\rho_\delta > 0$  such that if  $\varrho$  is small enough (i.e., it satisfies (B.24)) and  $P \geq C(\varrho)n \log^3(n)$ , and if Algorithm 1 is initialized with the spectral method in Algorithm 2, then with probability at least  $1 - \frac{4P+1}{n^3} - \frac{1}{n^2}$*

$$\text{dist}^2(x_k, \bar{x}) \leq \prod_{i=0}^{k-1} (1 - \nu_i) \rho_\delta^2, \quad \forall k \geq 0, \quad (3.10)$$

where

$$\nu_i = \frac{(1 - \kappa) \left( \min(\|\bar{x}\|^2, 1) - 2\delta \right)}{(1 + \delta)L_i}. \quad (3.11)$$

Let us first discuss this result and compare it to the one for Gaussian measurements.

#### Remark 3.4.

- As far as global recovery guarantees are concerned, Theorem 3.3(i) does not ensure exact recovery of  $\pm \bar{x}$ . This is in contrast with the Gaussian model where this was established in Theorem 3.1(i). As we pointed out earlier in the introduction section, one of the main difficulties is that several of our arguments in the Gaussian case rely on uniform bounds, for instance on the Hessian and gradient, that need to hold simultaneously for all vectors  $x \in \mathbb{R}^n$  w.h.p. Unfortunately, the CDP model enjoys much much less randomness to exploit in the mathematical analysis making this very challenging. Nevertheless, numerical evidence in the next section suggests that global exact recovery (without spectral initialization) holds for the CDP model as well.
- Theorem 3.3(ii) ensures local linear convergence to the true vectors  $\pm \bar{x}$  when our algorithm is initialized with the spectral method. The convergence rate is expressed in terms of the step-sizes  $\gamma_i = \frac{1-\kappa}{L_i}$ , where the  $L_i$ 's are expected to be much smaller than  $L$  in Lemma 2.6. It is tempting to use  $2(1 + \delta)^2$ , the local relative smoothness constant in (B.20), as an upper-bound estimate of the  $L_i$ 's. But one has to keep in mind that this is valid only locally on  $B(\pm \bar{x}, \rho_\delta)$ , and thus one cannot use it when iterating from  $x_k$  to  $x_{k+1}$ . In our numerical experiments, we nevertheless observe that the linear convergence rate in (3.11) is well estimated by  $\left( 1 - \frac{(1-\kappa)(\min(\|\bar{x}\|^2, 1) - 2\delta)}{2(1+\delta)^3} \right)$ . When  $\|\bar{x}\| \leq 1$ , this rate reads  $\left( 1 - \frac{(1-\kappa)(1-2\delta)}{2(1+\delta)^3} \right) \leq \frac{1}{2} + O(\kappa + \delta)$ .

#### Proof.

- (i) Under the bound on  $P$ , we know from Lemma B.10 that the measurement operator  $A$  is injective with probability at least  $1 - 1/n^2$ . On this event, the objective  $f$  is coercive, and thus the sequence  $(x_k)_{k \in \mathbb{N}}$  generated by Algorithm 1 is bounded. Since  $f$  is  $L$ -relatively smooth according to Lemma 2.6, Theorem 2.11(i)-(ii) ensure that the sequence  $(x_k)_{k \in \mathbb{N}}$  converges to  $x^* \in \text{crit}(f)$  and the induced sequence  $(f(x_k))_{k \in \mathbb{N}}$  converges to  $f(x^*)$ . Then using Theorem 2.11(iv) we get the statement.
- (ii) By Lemma B.12(iii), we have that the spectral initialization guess  $x_0$  belongs  $B(\bar{\mathcal{X}}, \frac{\rho_\delta}{\sqrt{\max(\Theta(\rho_\delta), 1)}}$ ) with probability larger than  $1 - \frac{4P+1}{n^3} - \frac{1}{n^2}$ . Moreover, we know from

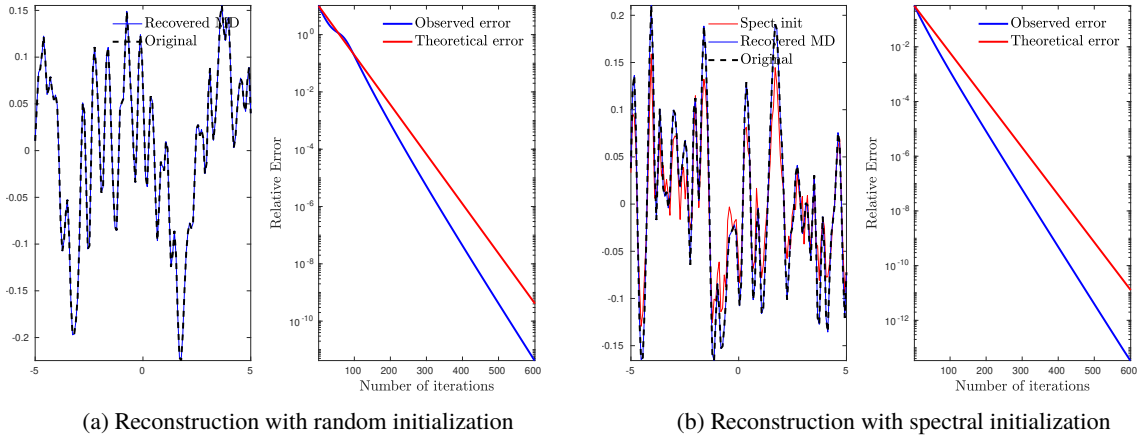


Figure 1: Reconstruction of signal from Gaussian measurements.

Lemma B.11 that with probability at least  $1 - \frac{4P+1}{2n^3}$ ,  $f$  is  $\sigma$ -strongly convex relatively to  $\psi$  locally on  $B(\bar{\mathcal{X}}, \rho_\delta)$  with  $\sigma = \frac{(\min(\|\bar{x}\|^2, 1) - 2\delta)}{1+\delta}$ . The rest of the proof follows the same reasoning as in the last item of the proof of statement Theorem 3.1(i). We omit the details.  $\square$

## 4 Numerical experiments

In this section, we discuss some numerical experiments to illustrate the efficiency of our phase recovery algorithm. We use the standard normal Gaussian and we consider the CDP Model with the random variable  $d$  taking values in  $\{-1, 0, 1\}$  with probability  $\{1/4, 1/2, 1/4\}$ . In each instance, we measured the relative error of our algorithm to the original signal using the following formula,

$$\frac{\text{dist}(x, \bar{\mathcal{X}})}{\|\bar{x}\|}. \quad (4.1)$$

In the experiments, we set  $\|\bar{x}\| = 1$ .

### 4.1 Reconstruction of 1D signals

#### 4.1.1 Gaussian measurements

The goal is to recover a one-dimensional signal with  $n = 128$  from Gaussians measurements. Figure 1(a) shows the reconstruction result from  $m = 2 \times 128 \times \log^3(128)$  without spectral initialization. Algorithm 1 was initialized with a vector drawn from the uniform distribution, and used with 600 iterations and a constant step-size  $\gamma = \frac{0.99}{3}$ . As predicted by Theorem 3.1(i), one can observe from Figure 1(a) that we have exact recovery, and after  $\sim 90$  iterations, the iterates enter a linear convergence regime.

Figure 1(b) displays the results for the case where  $m = 2 \times 128 \times \log(128)$ , and Algorithm 1 was applied with the same parameters as above except that the spectral initialization method was used to get the initial guess. As anticipated by Theorem 3.1(ii), we again have exact recovery with a linear convergence behaviour starting from the initial guess.

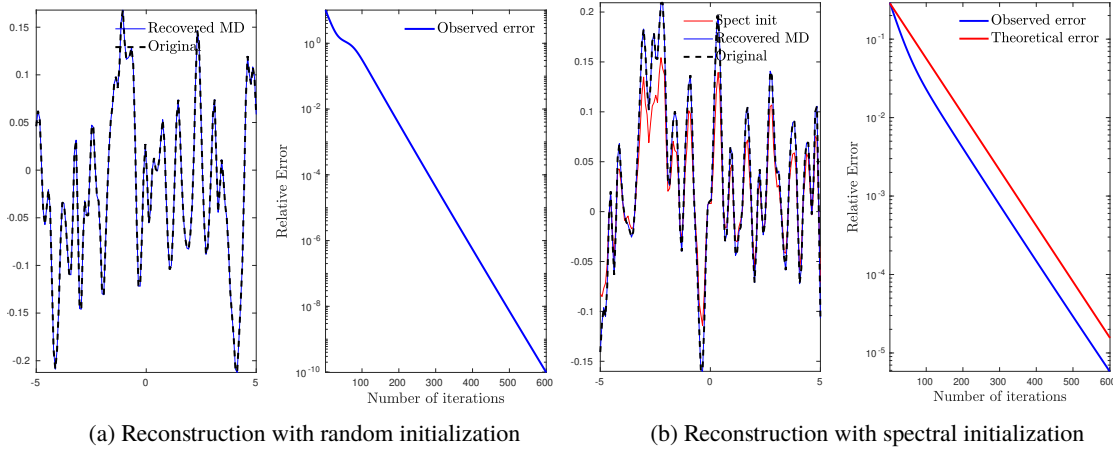


Figure 2: Reconstruction of a signal from from the CDP measurements model.

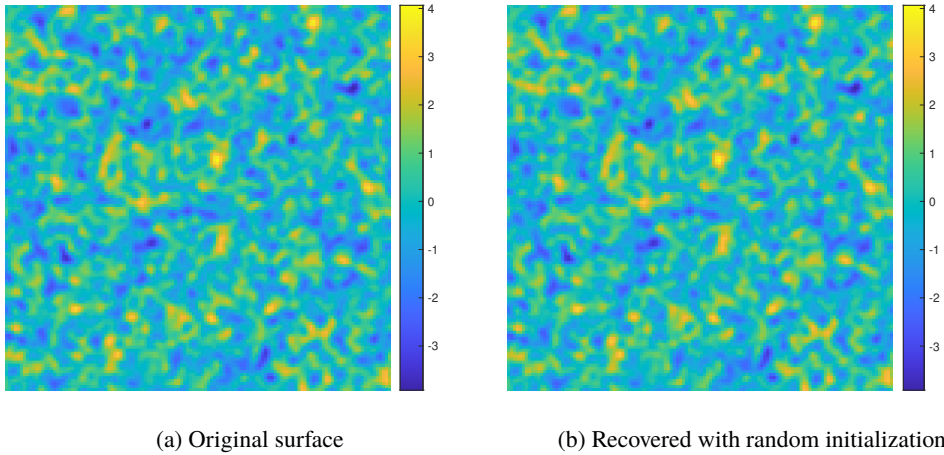


Figure 3: Roughness surface profile reconstruction by solving the phase retrieval problem from the CDP measurement model using mirror descent with random initialization.

#### 4.1.2 CDP measurements

We carried out the same experiment with the CDP measurements where we took  $P = 7 \times \log^3(128)$  masks and  $\gamma = \frac{0.99}{2}$ . The results are shown in Figure 2. The same conclusions drawn in the Gaussian case remain true for the CDP model. The results with spectral initialization depicted in Figure 2(b) are in agreement with those of Theorem 3.3(i). As for random uniform initialization, the results of Figure 2(a) provide numerical evidence that our algorithm enjoys global exact recovery properties, though this is so far not justified by our theoretical analysis.

### 4.2 Recovery of the roughness of a 2D surface (light scattering)

In this experiment, we simulated a rough surface as a  $256 \times 256$  Gaussian random field. The goal is to recover this surface profile from the magnitude of the measurements according to the CDP model with  $P = 100$  masks. The initial guess was drawn from the uniform distribution. The recovery results are displayed in Figure 3 hence confirming effectiveness of our algorithm.

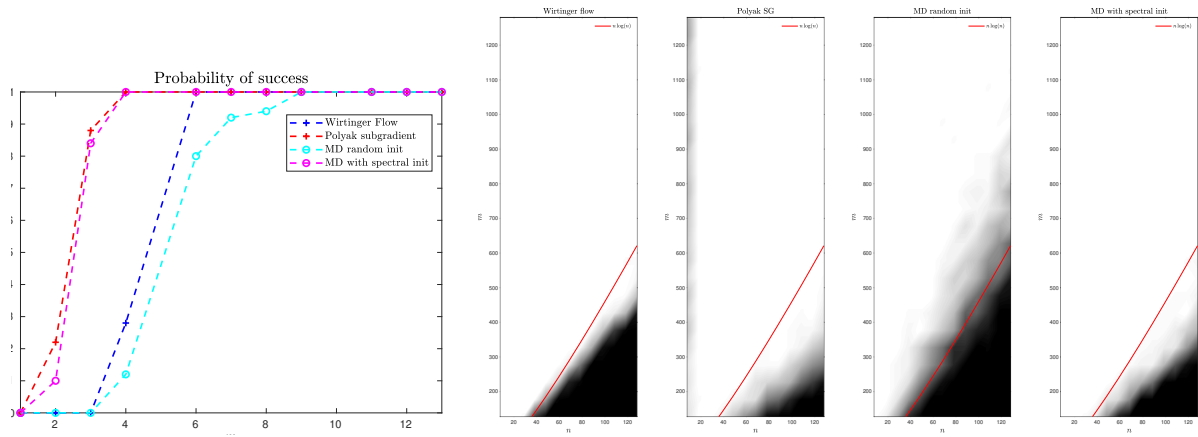


Figure 4: Phase diagrams for Gaussian measurements.

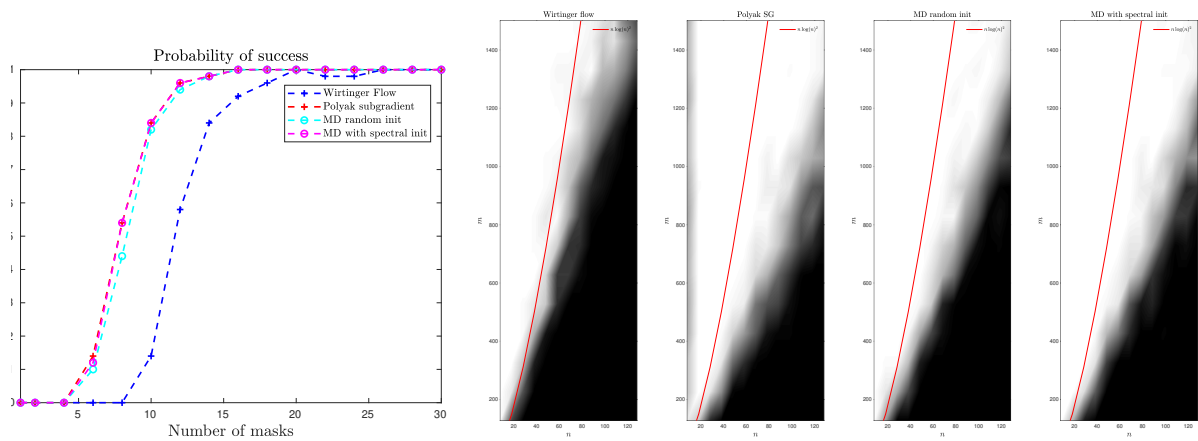


Figure 5: Phase diagrams for measurements from the CDP model.



### 4.3 Phase diagrams and comparison with other algorithms

We now compare mirror descent with the Wirtinger flow [19] and the Polyak subgradient method used in [29] (the latter can only be applied in the noiseless case) for phase recovery. These algorithms were initialized with spectral initialization. For mirror descent, we applied it both with spectral initialization and with random uniform initialization. For this, we report the results of an experiment designed to estimate the phase retrieval probability for each of the four compared algorithms as we vary  $n$  and  $m$ . The results are depicted in Figure 4 (resp. Figure 5) for the Gaussian case (resp. the CDP case). For each pair  $(n, m)$ , we generated 100 instances and solved them with each algorithm. Each diagram shows the empirical probability (among the 100 instances) that the corresponding algorithm successfully recovers the original vector up to a global sign change. We declared that a signal is recovered if the relative error (4.1) is less than  $10^{-5}$ . The grayscale of each point in the diagrams reflects the observed probability of success, from 0% (black) to 100% (white). The solid curve marks the prediction of the phase transition edge. On the left panel of Figure 4 and Figure 5, we also plot a profile of the phase diagram extracted at  $n = 128$ .

One clearly sees a phase transition phenomenon which is in agreement with the predicted sample complexity bound shown as a solid line. For Gaussian measurements, one clearly sees that mirror descent performs as well as Polyak subgradient method, when both are initialized by the spectral method. Both algorithms are better than the Wirtinger flow. Mirror descent with uniform random initialization has a recovery performance rather close to that of the Wirtinger flow (with spectral initialization), with a transition to success occurring at a higher threshold compared to the version of mirror descent with spectral initialization. This is in agreement with our theoretical findings. On the other hand, for CDP measurements, mirror descent with uniform random initialization shows comparable performance to the version with spectral initialization as well as to Polyak subgradient method (with spectral initialization), confirming numerically that spectral initialization does not seem to be mandatory for mirror descent with CDP measurements.

## Appendices

### A Proofs for the Deterministic Case

Let us start this section by recalling our objective function *i.e.*,

$$\forall x \in \mathbb{R}^n, \quad f(x) = \frac{1}{4m} \sum_{r=1}^m (|a_r^* x|^2 - y[r])^2 = \frac{1}{4m} \sum_{r=1}^m (|a_r^* x|^2 - |a_r^* \bar{x}|^2)^2, \quad (\text{A.1})$$

The following expressions give the gradients and Hessians of  $f$  and  $\psi$  that will be used throughout. For all  $\forall x \in \mathbb{R}^n$ , we have

$$\nabla f(x) = \frac{1}{m} \sum_{r=1}^m (|a_r^* x|^2 - |a_r^* \bar{x}|^2) a_r a_r^*, \quad \nabla^2 f(x) = \frac{1}{m} \sum_{r=1}^m (3|a_r^* x|^2 - |a_r^* \bar{x}|^2) a_r a_r^*, \quad (\text{A.2})$$

$$\nabla \psi(x) = \left( \|x\|^2 + 1 \right) x, \quad \nabla^2 \psi(x) = \left( \|x\|^2 + 1 \right) \text{Id} + 2xx^\top. \quad (\text{A.3})$$

Let start with the following useful lemma to compare the Bregman divergences of smooth functions.

**Lemma A.1.** *Let  $g, \phi \in C^2(\mathbb{R}^n)$ . If  $\forall u \in \mathbb{R}^n, \nabla^2 g(u) \preceq \nabla^2 \phi(u)$  for all  $u$  in the segment  $[x, z]$ , then,*

$$D_g(x, z) \leq D_\phi(x, z). \quad (\text{A.4})$$

**Proof.** The result comes from the Taylor-MacLaurin expansion. Indeed we have  $\forall x, z \in \mathbb{R}^n$

$$\begin{aligned} D_g(x, z) &= g(x) - g(z) - \langle \nabla g(z), x - z \rangle \\ &= \int_0^1 (1 - \tau) \langle x - z, \nabla^2 g(z + \tau(x - z))(x - z) \rangle d\tau, \end{aligned}$$

and thus

$$\begin{aligned} D_\phi(x, z) - D_g(x, z) &= \\ &= \int_0^1 (1 - \tau) \langle x - z, (\nabla^2 \phi(z + \tau(x - z)) - \nabla^2 g(z + \tau(x - z)))(x - z) \rangle d\tau. \end{aligned}$$

The positive semidefiniteness assumption implies the claim.  $\square$

### A.1 Proof of Lemma 2.6

**Proof.** Our proof is different from that of [15, Lemma 5.1] and gives a better estimate of  $L$ . Since  $y$  has positive entries, we have for all  $x, u \in \mathbb{R}^n$ ,

$$\begin{aligned} \langle u, \nabla^2 f(x)u \rangle &= \frac{1}{m} \sum_{r=1}^m (3|a_r^* x|^2 - y[r]) |a_r^* u|^2 \\ &\leq \frac{1}{m} \sum_{r=1}^m 3|a_r^* x|^2 |a_r^* u|^2 \\ &\leq \|x\|^2 \|u\|^2 \frac{1}{m} \sum_{r=1}^m 3 \|a_r\|^4. \end{aligned}$$

On the other hand,

$$\begin{aligned} \langle u, \nabla^2 \psi(x)u \rangle &= (\|x\|^2 + 1) \|u\|^2 + 2|\langle x, u \rangle|^2 \\ &\geq \|x\|^2 \|u\|^2 \end{aligned}$$

Thus for any  $L \geq \frac{1}{m} \sum_{r=1}^m 3 \|a_r\|^4$ , we have for all  $x \in \mathbb{R}^n$

$$\nabla^2 f(x) \preceq L \nabla^2 \psi(x). \quad (\text{A.5})$$

We conclude by invoking Lemma A.1 with  $g = f$  and  $\phi = L\psi$ , and Proposition 2.2(ii).  $\square$

### A.2 Proof of Lemma 2.8

**Proof.** From the update of  $x_{k+1}$ , we have  $\nabla \psi(x_k) - \nabla \psi(x_{k+1}) = \gamma_k \nabla f(x_k)$ , and multiplying both sides by  $x_{k+1} - u$ , we get

$$\langle \nabla \psi(x_k) - \nabla \psi(x_{k+1}), x_{k+1} - u \rangle = \gamma_k \langle \nabla f(x_k), x_{k+1} - u \rangle. \quad (\text{A.6})$$

Using the three-point identity (2.4), we have

$$D_\psi(u, x_k) - D_\psi(u, x_{k+1}) - D_\psi(x_{k+1}, x_k) = \gamma_k \langle \nabla f(x_k), x_{k+1} - u \rangle, \quad (\text{A.7})$$

By the backtracking test, we have that  $f$  verifies the  $L_k$ -relative smoothness inequality *w.r.t*  $\psi$  at  $(x_{k+1}, x_k)$ , where  $L_k \leq L$ . It then follows that

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq \langle \nabla f(x_k), x_{k+1} - x_k \rangle + L_k D_\psi(x_{k+1}, x_k) \\ &= \langle \nabla f(x_k), x_{k+1} - u \rangle + \langle \nabla f(x_k), u - x_k \rangle + L_k D_\psi(x_{k+1}, x_k), \end{aligned} \quad (\text{A.8})$$

Plugging (A.7) into (A.8), we arrive at

$$\begin{aligned}
& \gamma_k (f(x_{k+1}) - f(x_k)) \\
& \leq D_\psi(u, x_k) - D_\psi(u, x_{k+1}) - D_\psi(x_{k+1}, x_k) + \gamma_k \langle \nabla f(x_k), u - x_k \rangle + \gamma_k L_k D_\psi(x_{k+1}, x_k) \\
& \leq D_\psi(u, x_k) - D_\psi(u, x_{k+1}) - (1 - \gamma_k L_k) D_\psi(x_{k+1}, x_k) + \gamma_k \langle \nabla f(x_k), u - x_k \rangle \\
& \leq D_\psi(u, x_k) - D_\psi(u, x_{k+1}) - \kappa D_\psi(x_{k+1}, x_k) + \gamma_k \langle \nabla f(x_k), u - x_k \rangle.
\end{aligned}$$

Therefore

$$\begin{aligned}
\gamma_k (f(x_{k+1}) - f(u)) & \leq D_\psi(u, x_k) - D_\psi(u, x_{k+1}) - \kappa D_\psi(x_{k+1}, x_k) \\
& \quad + \gamma_k (f(x_k) - f(u) + \langle \nabla f(x_k), u - x_k \rangle) \\
& = D_\psi(u, x_k) - D_\psi(u, x_{k+1}) - \kappa D_\psi(x_{k+1}, x_k) - \gamma_k D_f(u, x_k).
\end{aligned}$$

□

### A.3 Proof of Theorem 2.11

**Proof.**

**(i)-(ii)** The objective function  $f$  in (2.8) is a real polynomial, hence obviously semi-algebraic. It then follows that  $f$  satisfies the Kurdyka-Łojasiewicz (KL) property [41, 42]. Combining this with Lemma 2.8, which ensures that the sequence  $(x_k)_{k \in \mathbb{N}}$  is a gradient-like descent sequence, and 1-strong convexity of the entropy  $\psi$ , the proof of (i)-(ii) are similar to those of [15, Proposition 4.1, Theorem 4.1] with slight modifications to handle backtracking.

**(iii)-(a)** The proof of this claim follows the same steps as the proof of [4, Theorem 2.12] using again that  $f$  is a continuous function which satisfies the KL property, that  $\min f = 0$  and that  $(x_k)_{k \in \mathbb{N}}$  is a gradient-like descent sequence thanks to Lemma 2.8.

**(iii)-(b)** We verify by induction that  $x_k \in B(x^*, \rho), \forall k \in \mathbb{N}$ . Observe first that  $x_0 \in B(x^*, r) \subset B(x^*, \rho)$  since  $r \leq \frac{\rho}{\max(\sqrt{\Theta(\rho)}, 1)} \leq \rho$ . Suppose now that for  $k \geq 0$ ,  $x_i \in B(x^*, \rho)$  for all  $i \leq k$ . From Lemma 2.8 and the optimality of  $x^*$  we have

$$\begin{aligned}
D_\psi(x^*, x_{k+1}) & \leq D_\psi(x^*, x_{k+1}) - (1 - \gamma_k L) D_\psi(x_{k+1}, x_k) - \gamma_k D_f(x^*, x_k) \\
& \leq D_\psi(x^*, x_k) - \gamma_k D_f(x^*, x_k) \\
& \leq (1 - \gamma_k \sigma) D_\psi(x^*, x_k) \\
& \leq \prod_{i=0}^k (1 - \gamma_i \sigma) D_\psi(x^*, x_0) \\
& \leq D_\psi(x^*, x_0),
\end{aligned} \tag{A.9}$$

where we used the positivity of  $D_\psi$  and the relative strong convexity at  $x_i \in B(x^*, \rho)$ . Now invoking Proposition 2.2(iv), we have

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 & \leq 2D_\psi(x^*, x_k) \leq 2 \prod_{i=0}^k (1 - \gamma_i \sigma) D_\psi(x^*, x_0) \\
& \leq \Theta(\rho) \|x_0 - x^*\|^2 \leq \frac{\Theta(\rho)}{\max(\Theta(\rho), 1)} \rho^2 \leq \rho^2,
\end{aligned}$$

which entails that  $x_i \in B(x^*, \rho)$  for all  $i \leq k + 1$  as desired. (2.14) follows by combining (A.9), strong convexity of  $\psi$  and that  $2D_\psi(x^*, x_0) \leq \rho^2$ .

(iv) We need the following lemma which is an extension of [38, Proposition 10] to the more general  $L$ -relatively smooth case.

**Lemma A.2.** *Let  $F$  be defined as in (2.11) then,*

- (a)  $\forall x \in \mathbb{R}^n, \det DF(x) \neq 0,$
- (b)  $\text{strisad}(f) \subset U_F \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n : F(x) = x, \max_i |\lambda_i(DF(x))| > 1\}.$

**Proof of Lemma A.2.** Recall that  $F(x) = (\nabla\psi)^{-1}(\nabla\psi(x) - \gamma\nabla f(x))$ . Denote  $G(x) \stackrel{\text{def}}{=} \nabla\psi(x) - \gamma\nabla f(x)$  so that  $F(x) = (\nabla\psi)^{-1} \circ G(x)$ .

- (a) Since  $\psi$  is  $C^2$  function, and thus  $\nabla\psi$  is  $C^1$ , and as  $\psi$  is strongly convex, the inverse function theorem ensures that  $(\nabla\psi)^{-1}$  is a local diffeomorphism<sup>2</sup>. Therefore to have  $\det DF(x) \neq 0$ , it suffices to show that  $G$  is a local diffeomorphism *i.e.*  $\forall x \in \mathbb{R}^n, DG(x)$  is an invertible linear transformation. We have  $DG(x) = \nabla^2\psi(x) - \gamma\nabla^2 f(x)$ , and the  $L$ -relative smoothness property of  $f$  w.r.t  $\psi$  (see (A.5) in the proof of Lemma 2.6) implies that

$$DG(x) = \nabla^2\psi(x) - \gamma\nabla^2 f(x) \succeq (1 - \gamma L)\nabla^2\psi(x) = \kappa\nabla^2\psi(x) \succeq \kappa\text{Id} \succ 0.$$

where we used 1-strong convexity of  $\psi$  and that  $\gamma L = 1 - \kappa \in ]0, 1[$ .

- (b) For  $x_\star \in \text{strisad}(f)$ , we have  $F(x_\star) = x_\star$  since  $\text{strisad}(f) \subset \text{crit}(f)$ . It remains to show that  $\det DF(x_\star)$  has an eigenvalue of magnitude greater than one. We have,

$$\begin{aligned} DF(x_\star) &\stackrel{(\text{Chain rule})}{=} \nabla^2\psi^{-1}(G(x_\star))DG(x_\star), \\ &= \nabla^2\psi^{-1}(x_\star) (\nabla^2\psi(x_\star) - \gamma\nabla^2 f(x_\star)), \\ &= \text{Id} - \gamma\nabla^2\psi(x_\star)^{-1}\nabla^2 f(x_\star). \end{aligned}$$

Denote for short  $H_\psi = \nabla^2\psi(x_\star)$ . We then have

$$H_\psi^{1/2}DF(x_\star)H_\psi^{-1/2} = \text{Id} - \gamma H_\psi^{-1/2}\nabla^2 f(x_\star)H_\psi^{-1/2}.$$

$H_\psi^{1/2}DF(x_\star)H_\psi^{-1/2}$  is symmetric. Let  $v' = H_\psi^{1/2}v$  with  $v$  a unit-norm eigenvector associated to a strictly negative eigenvalue of  $\nabla^2 f(x_\star)$ . By the Courant-Fisher min-max theorem, we have

$$\begin{aligned} \lambda_{\min}(H_\psi^{-1/2}\nabla^2 f(x_\star)H_\psi^{-1/2}) &\leq \langle v', H_\psi^{-1/2}\nabla^2 f(x_\star)H_\psi^{-1/2}v' \rangle \\ &= \langle v, \nabla^2 f(x_\star)v \rangle < 0. \end{aligned}$$

In turn,  $1 - \gamma\lambda_{\min}(H_\psi^{-1/2}\nabla^2 f(x_\star)H_\psi^{-1/2}) > 1$  is an eigenvalue of  $H_\psi^{1/2}DF(x_\star)H_\psi^{-1/2}$ . Since,  $H_\psi^{1/2}DF(x_\star)H_\psi^{-1/2}$  is similar to  $DF(x_\star)$ , we conclude. □

To show (iv), we combine claim (ii), Lemma A.2 and the centre stable manifold theorem (see [38, Corollary 1]) which allows to conclude that  $\{x_0 \in \mathbb{R}^n : \lim_{k \rightarrow \infty} F^k(x_0) \in \text{strisad}(f)\}$  has measure zero. □

## B Proofs for Random Measurements

### B.1 Gaussian measurements

In this section, we assume that the sensing vectors  $(a_r)_{r \in [m]}$  follow the i.i.d standard Gaussian model.

<sup>2</sup>Recall that we have already argued that  $\psi$  is a Legendre function and thus  $\nabla\psi$  is a bijection from  $\mathbb{R}^n$  to  $\mathbb{R}^n$  with inverse  $(\nabla\psi)^{-1} = \nabla\psi^*$ ; see [52, Theorem 26.5]

### B.1.1 Expectation and deviation of the Hessian

The next lemma gives the expression of the expectation of  $\nabla^2 f(x)$ .

**Lemma B.1.** (*Expectation of the Hessian*) Under the Gaussian model, we have

$$\mathbb{E}(\nabla^2 f(x)) = 3 \left( 2xx^\top + \|x\|^2 \text{Id} \right) - 2\bar{x}\bar{x}^\top - \|\bar{x}\|^2 \text{Id}. \quad (\text{B.1})$$

**Proof.** In view of (A.2), it is sufficient to compute

$$\mathbb{E} \left( \frac{1}{m} \sum_{r=1}^m |a_r^\top x|^2 a_r a_r^\top \right).$$

Computing this expectation is standard using independence and a simple moment calculation, which gives

$$\mathbb{E} \left( \frac{1}{m} \sum_{r=1}^m |a_r^\top x|^2 a_r a_r^\top \right) = 2xx^\top + \|x\|^2 \text{Id}. \quad (\text{B.2})$$

□

We now turn our attention to the concentration of the Hessian of  $f$  around its mean. We start with following key lemma.

**Lemma B.2.** Fix  $\varrho \in ]0, 1[$ . If the number of samples obeys  $m \geq C(\varrho)n \log n$ , for some sufficiently large  $C(\varrho) > 0$ , then

$$\left\| \frac{1}{m} \sum_{r=1}^m |a_r^\top x|^2 a_r a_r^\top - \left( 2xx^\top + \|x\|^2 \text{Id} \right) \right\| \leq \frac{\varrho}{3} \|x\|^2.$$

holds simultaneously for all  $x \in \mathbb{R}^n$  with a probability at least  $1 - 5e^{-\zeta n} - \frac{4}{n^2}$ , where  $\zeta$  is a fixed numerical constant.

**Proof.** We follow a similar strategy to that of [19, Section A.4]. By a homogeneity argument and isotropy of the Gaussian distribution, it is sufficient to establish the claim for  $x = e_1$ , i.e. that

$$\left\| \frac{1}{m} \sum_{r=1}^m |a_r[1]|^2 a_r a_r^\top - \left( 2e_1 e_1^\top + \text{Id} \right) \right\| \leq \frac{\varrho}{3}. \quad (\text{B.3})$$

Since the matrix in (B.3) is symmetric, its spectral norm can be computed via the associated quadratic form, and (B.3) amounts to showing that

$$V(v) \stackrel{\text{def}}{=} \left| \frac{1}{m} \sum_{r=1}^m |a_r[1]|^2 |a_r^\top v|^2 - (1 + 2v[1]^2) \right| \leq \frac{\varrho}{3}$$

for all  $v \in \mathbb{S}^{n-1}$ . The rest of the proof shows this claim.

Let  $\tilde{a}_r = (a_r[2], \dots, a_r[n])$  and  $\tilde{v} = (v[2], \dots, v[n])$ . We rewrite

$$|a_r^\top v|^2 = \left( a_r[1]v[1] + \tilde{a}_r^\top \tilde{v} \right)^2 = (a_r[1]v[1])^2 + \left( \tilde{a}_r^\top \tilde{v} \right)^2 + 2a_r[1]v[1]\tilde{a}_r^\top \tilde{v}.$$

We plug this decomposition into  $V(v)$  to get

$$\begin{aligned} V(v) &= \left| \frac{1}{m} \sum_{r=1}^m a_r[1]^4 v[1]^2 + \frac{1}{m} \sum_{r=1}^m a_r[1]^2 (\tilde{a}_r^\top \tilde{v})^2 + 2 \frac{1}{m} \sum_{r=1}^m |a_r[1]|^3 v[1] \tilde{a}_r^\top \tilde{v} - \left( \|\tilde{v}\|^2 + 3v[1]^2 \right) \right|, \\ &\leq \left| \frac{1}{m} \sum_{r=1}^m a_r[1]^4 - 3 \right| v[1]^2 + \left| \frac{1}{m} \sum_{r=1}^m a_r[1]^2 - 1 \right| \|\tilde{v}\|^2 + 2 \left| \frac{1}{m} \sum_{r=1}^m |a_r[1]|^3 v[1] \tilde{a}_r^\top \tilde{v} \right| \\ &\quad + \left| \frac{1}{m} \sum_{r=1}^m a_r[1]^2 \left( \tilde{a}_r^\top \tilde{v} - \|\tilde{v}\|^2 \right) \right|. \end{aligned}$$

If  $X \sim \mathcal{N}(0, 1)$  we have  $\mathbb{E}(X^{2p}) = \frac{(2p)!}{2^p p!}$ , and in particular  $\mathbb{E}(X^2) = 1$  and  $\mathbb{E}(X^4) = 3$ . By the Tchebyshev's inequality and a union bound argument,  $\forall \varepsilon > 0$ , and a constant  $C(\varepsilon) \approx \max(26, \frac{96}{\varepsilon^2})$  such that when  $m \geq C(\varepsilon)n$  we have,

$$\begin{aligned} \frac{1}{m} \sum_{r=1}^m (a_r[1]^4 - 3) < \varepsilon, \quad \frac{1}{m} \sum_{r=1}^m (a_r[1]^2 - 1) < \varepsilon, \quad \frac{1}{m} \sum_{r=1}^m a_r[1]^6 \leq 20 \\ \text{and } \max_{1 \leq r \leq m} |a_r[1]| \leq \sqrt{10 \log m}. \end{aligned}$$

Each of these event happens with probability at least  $1 - \frac{1}{n^2}$ , and thus their intersection occurs with a probability at least  $1 - \frac{4}{n^2}$ . On this intersection event, we have

$$V(v) \leq \varepsilon(v[1]^2 + \|\tilde{v}\|^2) + 2 \left| \frac{1}{m} \sum_{r=1}^m a_r[1]^3 v[1] \tilde{a}_r^\top \tilde{v} \right| + \left| \frac{1}{m} \sum_{r=1}^m a_r[1]^2 \left( \tilde{a}_r^\top \tilde{v} - \|\tilde{v}\|^2 \right) \right|.$$

On the one hand, by a Hoeffding-type inequality ([59, Proposition 5.10]), we have

$$\forall \varrho' > 0, \quad \left| \frac{1}{m} \sum_{r=1}^m a_r[1]^3 v[1] \tilde{a}_r^\top \tilde{v} \right| < \varrho' |v[1]| \|\tilde{v}\|^2,$$

with a probability  $1 - ee^{-\zeta'n} \geq 1 - 3e^{-\zeta'n}$ , when  $m \geq C(\varrho') \sqrt{n \sum_{r=1}^m a_r[1]^6}$  with  $C(\varrho') \approx \frac{1}{\varrho'^2}$  and  $\zeta' > 2$  an absolute constant.

On the other hand, by Bernstein-type inequality ([59, Proposition 5.16]), we have

$$\forall \varrho' > 0, \quad \left| \frac{1}{m} \sum_{r=1}^m a_r[1]^2 \left( \tilde{a}_r^\top \tilde{v} - \|\tilde{v}\|^2 \right) \right| \leq \varrho' \|\tilde{v}\|^2,$$

with a probability  $1 - 2e^{-\zeta'n}$ , when  $m \geq C(\varrho') \left( \sqrt{n \sum_{r=1}^m a_r[1]^4} + n \max_{1 \leq r \leq m} a_r[1]^2 \right)$  with  $C(\varrho') \approx \frac{1}{\varrho'^2}$ .

Overall, for any  $v \in \mathbb{S}^{n-1}$ , we have with probability at least  $1 - 5e^{-\zeta'n}$

$$V(v) \leq \varepsilon + 3\varrho'.$$

At this stage, we use a covering argument ([59, Lemma 5.4]) with an  $\frac{1}{2}$ -net whose cardinality is smaller than  $5^n$ . Therefore, choosing  $\varepsilon = \varrho'$  and  $\varrho = 12\varrho'$  we get the claim where  $\zeta = \zeta' - \log(5) > 0$  since  $\zeta' > 2$  in the Hoeffding and Bernstein inequalities used above.  $\square$

**Lemma B.3. (Concentration of the Hessian)** Fix  $\varrho \in ]0, 1[$ . If the number of samples obeys  $m \geq C(\varrho)n \log n$ , for some sufficiently large constant  $C(\varrho) > 0$ , then

$$\|\nabla^2 f(x) - \mathbb{E}(\nabla^2 f(x))\| \leq \varrho \left( \|x\|^2 + \frac{\|\bar{x}\|^2}{3} \right) \quad (\text{B.4})$$

holds simultaneously for all  $x \in \mathbb{R}^n$  with a probability at least  $1 - 5e^{-\zeta n} - \frac{4}{n^2}$ , where  $\zeta$  is a fixed numerical constant.

**Proof.** Recall  $\nabla^2 f(x)$  from (A.2). By the triangle inequality and Lemma B.1, we have

$$\begin{aligned} \|\nabla^2 f(x) - \mathbb{E}(\nabla^2 f(x))\| &\leq 3 \left\| \frac{1}{m} \sum_{r=1}^m |a_r^\top x|^2 a_r a_r^\top - (2xx^\top + \|x\|^2 \text{Id}) \right\| \\ &\quad + \left\| \frac{1}{m} \sum_{r=1}^m |a_r^\top \bar{x}|^2 a_r a_r^\top - (2\bar{x}\bar{x}^\top + \|\bar{x}\|^2 \text{Id}) \right\|. \end{aligned}$$

The claim is then a consequence of Lemma B.2.  $\square$

### B.1.2 Injectivity of the measurement operator

The next result shows that when the number of measurements is large enough, the measurement matrix  $A$  (whose rows are the  $a_r^\top$ 's) is injective *w.h.p.*

**Lemma B.4.** Fix  $\varrho \in ]0, 1[$ . Assume that  $m \geq \frac{16}{\varrho^2}n$ . Then

$$(1 - \varrho) \|x\|^2 \leq \frac{1}{m} \|Ax\|^2 \leq (1 + \varrho) \|x\|^2, \quad \forall x \in \mathbb{R}^n. \quad (\text{B.5})$$

This happens with a probability at least  $1 - 2e^{-mt^2/2}$  with  $\frac{\varrho}{4} = t^2 + t$ .

**Proof.** This is a consequence of very standard deviation inequalities on the singular values of Gaussian random matrices; see [22, Lemma 3.1] for a similar statement.  $\square$

### B.1.3 Relative smoothness

For the Gaussian phase retrieval, we have the following refined dimension-independent estimate of the relative smoothness modulus, which is much better than the bound of Proposition 2.6.

**Lemma B.5.** Fix  $\varrho \in ]0, 1[$ . If the event  $\mathcal{E}_{\text{conH}}$  defined by (3.4) holds true then,

$$D_f(x, z) \leq \left( 3 + \varrho \max(\|\bar{x}\|^2/3, 1) \right) D_\psi(x, z), \quad \forall x, z \in \mathbb{R}^n. \quad (\text{B.6})$$

**Proof.** Using (3.4), Lemma B.1 and (A.3), we have

$$\begin{aligned} \forall x \in \mathbb{R}^n, \quad \nabla^2 f(x) &\leq \mathbb{E}(\nabla^2 f(x)) + \varrho \left( \|x\|^2 + \frac{\|\bar{x}\|^2}{3} \right) \text{Id}, \\ &\leq 3 \left( 2xx^\top + \|x\|^2 \text{Id} \right) - 2\bar{x}\bar{x}^\top - \|\bar{x}\|^2 \text{Id} \\ &\quad + \varrho \max(\|\bar{x}\|^2/3, 1) \left( \|x\|^2 + 1 \right) \text{Id}, \\ &\leq 3 \left( 2xx^\top + (\|x\|^2 + 1) \text{Id} \right) + \varrho \max(\|\bar{x}\|^2/3, 1) \nabla^2 \psi(x), \\ &= 3\nabla^2 \psi(x) + \varrho \max(\|\bar{x}\|^2/3, 1) \nabla^2 \psi(x). \end{aligned} \quad (\text{B.7})$$

We conclude by applying Lemma A.1.  $\square$

### B.1.4 Local relative strong convexity

The next proposition establishes strong convexity relatively to the entropy  $\psi$  locally around  $\bar{\mathcal{X}}$ . In view of strong 1-convexity of  $\psi$ , our result also implies local strong convexity with respect to the euclidian norm as shown in [19, 55].

**Lemma B.6.** *Fix  $\lambda \in ]0, 1[$  and  $\varrho \in ]0, \lambda \min(\|\bar{x}\|^2, 1)/(2 \max(\|\bar{x}\|^2/3, 1))$ . If the event  $\mathcal{E}_{\text{conH}}$  defined by (3.4) holds true then for all  $x, z \in B\left(\bar{x}, \frac{1-\lambda}{\sqrt{3}} \|\bar{x}\|\right)$  and  $x, z \in B\left(-\bar{x}, \frac{1-\lambda}{\sqrt{3}} \|\bar{x}\|\right)$ ,*

$$D_f(x, z) \geq \left( \lambda \min(\|\bar{x}\|^2, 1) - \varrho \max(\|\bar{x}\|^2/3, 1) \right) D_\psi(x, z). \quad (\text{B.8})$$

Observe that if  $\|\bar{x}\| = 1$  the above result has a simpler statement. In particular,  $\varrho$  must lie in  $]0, \lambda[$ , and the local relative strong convexity modulus is  $\lambda - \varrho$  on a ball of radius  $\frac{1-\lambda}{\sqrt{3}} \|\bar{x}\|$  around  $\bar{\mathcal{X}}$ .

**Proof.** We embark from (3.4) and Lemma B.1 to infer that  $\forall x \in \mathbb{R}^n$

$$\nabla^2 f(x) \succeq -\varrho \left( \|x\|^2 + \frac{\|\bar{x}\|^2}{3} \right) \text{Id} + 3 \left( 2xx^\top + \|x\|^2 \text{Id} \right) - \left( 2\bar{x}\bar{x}^\top + \|\bar{x}\|^2 \text{Id} \right) \quad (\text{B.9})$$

$$\succeq -\varrho \max(\|\bar{x}\|^2/3, 1) \nabla^2 \psi(x) + 3 \left( 2xx^\top + \|x\|^2 \text{Id} \right) - \left( 2\bar{x}\bar{x}^\top + \|\bar{x}\|^2 \text{Id} \right). \quad (\text{B.10})$$

We then obtain, for any  $v \in \mathbb{S}^{n-1}$

$$v^\top \nabla^2 f(x) v + \varrho \max(\|\bar{x}\|^2/3, 1) v^\top \nabla^2 \psi(x) v \geq 3 \left( 2 \left( v^\top x \right)^2 + \|x\|^2 \right) - \left( 2 \left( v^\top \bar{x} \right)^2 + \|\bar{x}\|^2 \right).$$

Let  $\rho > 0$  small enough, to be made precise later. Thus for any  $x = \pm \bar{x} + \rho v$  we get

$$\begin{aligned} & v^\top \nabla^2 f(x) v + \varrho \max(\|\bar{x}\|^2/3, 1) v^\top \nabla^2 \psi(x) v \\ & \geq 6 \left( v^\top \bar{x} \right)^2 + 6\rho^2 \pm 12\rho v^\top \bar{x} + 3 \|\bar{x}\|^2 \pm 6\rho v^\top \bar{x} + 3\rho^2 - 2 \left( v^\top \bar{x} \right)^2 - \|\bar{x}\|^2 \\ & = 4 \left( v^\top \bar{x} \right)^2 + 9\rho^2 \pm 18\rho v^\top \bar{x} + 2 \|\bar{x}\|^2. \end{aligned}$$

From (A.3), we also have

$$v^\top \nabla^2 \psi(x) v = \|x\|^2 + 1 + 2 \left( v^\top x \right)^2 = 2 \left( v^\top \bar{x} \right)^2 + 3\rho^2 + \pm 6\rho v^\top \bar{x} + \|\bar{x}\|^2 + 1.$$

Consider first the case where  $\|\bar{x}\| \geq 1$ . We then get

$$\begin{aligned} & v^\top \left( \nabla^2 f(x) - \left( \lambda - \varrho \max(\|\bar{x}\|^2/3, 1) \right) \nabla^2 \psi(x) \right) v \\ & \geq 2(2 - \lambda) \left( v^\top \bar{x} \right)^2 + 3(3 - \lambda)\rho^2 \pm 6(3 - \lambda)\rho v^\top \bar{x} + (2 - \lambda) \|\bar{x}\|^2 - \lambda \\ & = 2(2 - \lambda) \left( v^\top \bar{x} \right)^2 + 3(3 - \lambda)\rho^2 \pm 6(3 - \lambda)\rho v^\top \bar{x} + 2(1 - \lambda) \|\bar{x}\|^2 + \lambda(\|\bar{x}\|^2 - 1) \\ & \geq 2(2 - \lambda) \left( v^\top \bar{x} \right)^2 + 3(3 - \lambda)\rho^2 \pm 6(3 - \lambda)\rho v^\top \bar{x} + 2(1 - \lambda) \|\bar{x}\|^2. \end{aligned}$$

We claim that

$$\inf_{v \in \mathbb{S}^{n-1}} 2(2 - \lambda) \left( v^\top \bar{x} \right)^2 + 3(3 - \lambda)\rho^2 \pm 6(3 - \lambda)\rho v^\top \bar{x} + 2(1 - \lambda) \|\bar{x}\|^2 \geq 0$$



for  $\rho$  small enough. Let  $v^\top \bar{x} = \alpha \|\bar{x}\|$ , where  $\alpha \in [-1, 1]$  and  $\rho = \beta \|\bar{x}\|$ . Thus

$$2(2 - \lambda) \left( v^\top \bar{x} \right)^2 + 3(3 - \lambda) \rho^2 \pm 6(3 - \lambda) \rho v^\top \bar{x} + 2(1 - \lambda) \|\bar{x}\|^2 = \\ (2(2 - \lambda) \alpha^2 + 3(3 - \lambda) \beta^2 \pm 6(3 - \lambda) \alpha \beta + 2(1 - \lambda)) \|\bar{x}\|^2.$$

Minimizing the last term for  $\alpha$  and substituting back, we have after simple algebra that

$$2(2 - \lambda) \alpha^2 + 3(3 - \lambda) \beta^2 \pm 6(3 - \lambda) \alpha \beta + 2(1 - \lambda) \geq 2(1 - \lambda) - \phi(\lambda) \beta^2.$$

where we set the function  $\phi : t \in ]0, 1[ \mapsto \frac{36(3-t)^2}{8(2-t)} - 3(3-t) \in \mathbb{R}_+$ . It can be easily shown that  $\sup_{]0, 1[} \phi(t) = \phi(1) = 12$ . In turn, we have

$$2(2 - \lambda) \alpha^2 + 3(3 - \lambda) \beta^2 \pm 6(3 - \lambda) \alpha \beta + 2(1 - \lambda) \geq 0$$

since we assumed that  $\rho \leq \frac{1-\lambda}{\sqrt{3}} \|\bar{x}\| \leq \frac{2(1-\lambda)}{\sqrt{\phi(\lambda)}} \|\bar{x}\|$ .

Let us now turn to the case where  $\|\bar{x}\| \leq 1$ . We then have

$$v^\top \left( \nabla^2 f(x) - \lambda \left( \|\bar{x}\|^2 - \varrho \max(\|\bar{x}\|^2 / 3, 1) \right) \nabla^2 \psi(x) \right) v \\ \geq 2(2 - \lambda \|\bar{x}\|^2) \left( v^\top \bar{x} \right)^2 + 3(3 - \lambda \|\bar{x}\|^2) \rho^2 \pm 6(3 - \lambda \|\bar{x}\|^2) \rho v^\top \bar{x} + 2 \|\bar{x}\|^2 - \lambda \|\bar{x}\|^4 - \lambda \|\bar{x}\|^2 \\ \geq 2(2 - \lambda \|\bar{x}\|^2) \left( v^\top \bar{x} \right)^2 + 3(3 - \lambda \|\bar{x}\|^2) \rho^2 \pm 6(3 - \lambda \|\bar{x}\|^2) \rho v^\top \bar{x} + 2(1 - \lambda) \|\bar{x}\|^2 + \lambda \left( \|\bar{x}\|^2 - \|\bar{x}\|^4 \right) \\ \geq 2(2 - \lambda \|\bar{x}\|^2) \left( v^\top \bar{x} \right)^2 + 3(3 - \lambda \|\bar{x}\|^2) \rho^2 \pm 6(3 - \lambda \|\bar{x}\|^2) \rho v^\top \bar{x} + 2(1 - \lambda) \|\bar{x}\|^2 \\ = \left( 2(2 - \lambda \|\bar{x}\|^2) \alpha^2 + 3(3 - \lambda \|\bar{x}\|^2) \beta^2 \pm 6(3 - \lambda \|\bar{x}\|^2) \alpha \beta + 2(1 - \lambda) \right) \|\bar{x}\|^2.$$

Arguing as in the first case, we have

$$2(2 - \lambda \|\bar{x}\|^2) \alpha^2 + 3(3 - \lambda \|\bar{x}\|^2) \beta^2 \pm 6(3 - \lambda \|\bar{x}\|^2) \alpha \beta + 2(1 - \lambda) \geq 2(1 - \lambda) - \phi(\lambda \|\bar{x}\|^2) \beta^2.$$

Thus, the right hand side is non-negative since

$$\beta \leq \frac{1 - \lambda}{\sqrt{3}} \leq \frac{2(1 - \lambda)}{\sqrt{\phi(\lambda \|\bar{x}\|^2)}},$$

where we used that  $\|\bar{x}\|^2 \leq 1$  in the argument of  $\phi$ .

Overall, we have shown that

$$v^\top \left( \nabla^2 f(x) - \left( \lambda \min(\|\bar{x}\|^2, 1) / 2 - \varrho \max(\|\bar{x}\|^2 / 3, 1) \right) \nabla^2 \psi(x) \right) v \geq 0$$

for all  $v \in \mathbb{S}^{n-1}$  and  $\rho \leq \frac{1-\lambda}{\sqrt{3}} \|\bar{x}\|$ . We complete the proof by invoking Lemma A.1. □

### B.1.5 Spectral initialization

We now show that the initial guess  $x_0$  generated by spectral initialization (Algorithm 2) belongs to a small  $f$ -attentive neighbourhood of  $\bar{\mathcal{X}}$ .

**Lemma B.7.** Fix  $\varrho \in ]0, 1[$ . If the number of samples obeys  $m \geq C(\varrho)n \log n$ , for some sufficiently large constant  $C(\varrho) > 0$ , then with probability at least  $1 - 2e^{-\frac{m(\sqrt{1+\varrho}-1)^2}{8}} - 5e^{-\zeta n} - \frac{4}{n^2}$ , where  $\zeta$  is a fixed numerical constant,  $x_0$  satisfies:

(i)  $\text{dist}(x_0, \bar{\mathcal{X}}) \leq \eta_1(\varrho) \|\bar{x}\|$ , where

$$\eta_1 : ]0, 1[ \rightarrow ]0, 1[$$

$$\varrho \mapsto \left( \sqrt{2 - 2\sqrt{1 - \varrho}} + \varrho/2 \right), \quad (\text{B.11})$$

which is an increasing function.

(ii)  $f(x_0) \leq \left( 3 + \varrho \max(\|\bar{x}\|^2/3, 1) \right) \frac{\Theta(\eta_1(\varrho)\|\bar{x}\|)}{2} \eta_1(\varrho)^2 \|\bar{x}\|^2$ .

(iii) Besides, for  $\lambda \in ]0, 1[$ , if

$$\varrho \leq \eta_1^{-1} \left( \frac{1 - \lambda}{\sqrt{3(6(1 + (1 - \lambda)^2/3) + 1)}} \frac{1}{\max(\|\bar{x}\|, 1)} \right), \quad (\text{B.12})$$

then with the same probability as above  $x_0 \in B \left( \bar{\mathcal{X}}, \frac{\rho}{\max(\sqrt{\Theta(\rho)}, 1)} \right)$  where  $\rho = \frac{1 - \lambda}{\sqrt{3}} \|\bar{x}\|$ .

**Proof.** (i) Denote the matrix

$$Y = \frac{1}{m} \sum_{r=1}^m y[r] a_r a_r^\top = \frac{1}{m} \sum_{r=1}^m |a_r^\top \bar{x}|^2 a_r a_r^\top.$$

By Lemma B.2, we have *w.h.p*

$$\|Y - \mathbb{E}(Y)\| \leq \varrho \|\bar{x}\|^2.$$

Let  $\tilde{x}$  be the eigenvector associated with the largest eigenvalue  $\tilde{\lambda}$  of  $Y$  such that  $\|\tilde{x}\| = \|\bar{x}\|$  (obviously  $\tilde{\lambda}$  is nonnegative since  $Y$  is semidefinite positive). Then,

$$\begin{aligned} \varrho \|\bar{x}\|^2 &\geq \|Y - \mathbb{E}(Y)\| \geq \|\bar{x}\|^{-2} \left| \tilde{x}^\top \left( Y - 2\bar{x}\bar{x}^\top - \|\bar{x}\|^2 \text{Id} \right) \tilde{x} \right| \\ &= \|\bar{x}\|^{-2} \left| \tilde{\lambda} \|\bar{x}\|^2 - 2(\tilde{x}^\top \bar{x})^2 - \|\bar{x}\|^4 \right|. \end{aligned}$$

Hence

$$2(\tilde{x}^\top \bar{x})^2 \geq \tilde{\lambda} \|\bar{x}\|^2 - (1 + \varrho) \|\bar{x}\|^4.$$

Moreover, using Lemma B.2 again entails that *w.h.p*

$$\tilde{\lambda} \|\bar{x}\|^2 \geq \bar{x}^\top Y \bar{x} \geq \bar{x}^\top \left( 2\bar{x}\bar{x}^\top + \|\bar{x}\|^2 \text{Id} \right) \bar{x} - \varrho \|\bar{x}\|^4 = (3 - \varrho) \|\bar{x}\|^4.$$

Combining the last two inequalities, we get

$$(\tilde{x}^\top \bar{x})^2 \geq (1 - \rho) \|\bar{x}\|^4.$$

It then follows that

$$\text{dist}(\tilde{x}, \bar{\mathcal{X}}) \leq \sqrt{2 - 2\sqrt{1 - \rho}} \|\bar{x}\|.$$

By definition of  $x_0$  in Algorithm 2,  $x_0 = \sqrt{m^{-1} \sum_r y[r]} \frac{\tilde{x}}{\|\tilde{x}\|}$ , and thus *w.h.p*

$$\|x_0 - \tilde{x}\| = \left| \sqrt{\frac{m^{-1} \sum_r y[r]}{\|\tilde{x}\|^2}} - 1 \right| \|\tilde{x}\| = \left| \sqrt{\frac{m^{-1} \|A\tilde{x}\|^2}{\|\tilde{x}\|^2}} - 1 \right| \|\tilde{x}\| \leq \varrho/2 \|\tilde{x}\|,$$

where we used Lemma B.4. In turn,

$$\text{dist}(x_0, \bar{\mathcal{X}}) \leq \text{dist}(\tilde{x}, \bar{\mathcal{X}}) + \|x_0 - \tilde{x}\| \leq \left( \sqrt{2 - 2\sqrt{1 - \rho}} + \varrho/2 \right) \|\bar{x}\|.$$

(ii) Under our sampling complexity bound, event  $\mathcal{E}_{conH}$  defined by (3.4) holds true *w.h.p.* . It then follows from Lemma B.5 applied at  $\bar{x}$  and  $x_0$ , that

$$Df(x_0, \bar{x}) \leq \left(3 + \varrho \max(\|\bar{x}\|^2/3, 1)\right) D_\psi(x_0, \bar{x}). \quad (\text{B.13})$$

Since  $f(\bar{x}) = 0$  and  $\nabla f(\bar{x}) = 0$ , we obtain from Proposition 2.2(iv) that

$$\begin{aligned} f(x_0) &\leq \left(3 + \varrho \max(\|\bar{x}\|^2/3, 1)\right) D_\psi(x_0, \bar{x}) \\ &\leq \left(3 + \varrho \max(\|\bar{x}\|^2/3, 1)\right) \frac{\Theta(\eta_1(\varrho) \|\bar{x}\|)}{2} \eta_1(\varrho)^2 \|\bar{x}\|^2. \end{aligned} \quad (\text{B.14})$$

(iii) In view of (i), it is sufficient to show that  $\eta_1(\varrho) \|\bar{x}\| \leq \frac{\rho}{\max(\sqrt{\Theta(\rho)}, 1)}$ . Since from Proposition 2.2(iv) (see also Remark 2.12) we have

$$\Theta(\rho) \leq 6(\|\bar{x}\|^2 + \rho^2) + 1 \leq (6(1 + (1 - \lambda)^2/3) + 1) \max(\|\bar{x}\|^2, 1),$$

and  $\eta_1$  is an increasing function, we conclude.  $\square$

## B.2 Proofs for the CDP model

In this section, we assume that the sensing vectors  $(a_r)_{r \in [m]}$  follow the CDP model introduced in Section 3.1.

### B.2.1 Expectation and deviation of the Hessian

**Lemma B.8. (Expectation of the Hessian)** *Under the CDP measurement model, the following holds*

$$\mathbb{E}(\nabla^2 f(x)) = 3 \left( xx^\top + \|x\|^2 \text{Id} \right) - \bar{x} \bar{x}^\top - \|\bar{x}\|^2 \text{Id}. \quad (\text{B.15})$$

**Proof.** From [18, Lemma 3.1], we have

$$\forall x \in \mathbb{R}^n, \quad \mathbb{E} \left( \frac{1}{nP} \sum_{j,p=1}^{n,P} |f_j^* D_p x|^2 D_p f_j f_j^* D_p \right) = xx^\top + \|x\|^2 \text{Id}. \quad (\text{B.16})$$

Combining this with (A.2) yields the claim.  $\square$

Unlike the Gaussian model, it turns out that it is very challenging to concentrate the Hessian of  $f$  around its mean simultaneously for all vectors  $x \in \mathbb{R}^n$  with non-trivial sampling complexity bounds. The main reason is that the CDP model does not have enough randomness to be used in the mathematical analysis. However, one can still do that for a fixed vector  $x$ . The next lemma gives the Hessian deviation at  $\pm \bar{x}$ .

**Lemma B.9. (Concentration of the Hessian)** *Fix  $\delta \in ]0, 1[$ . If the number of patterns obeys  $P \geq C(\delta) \log^3(n)$ , then with a probability at least  $1 - \frac{4P+1}{2n^3}$*

$$\|\nabla^2 f(\bar{x}) - \mathbb{E}(\nabla^2 f(\bar{x}))\| \leq \delta \|\bar{x}\|^2. \quad (\text{B.17})$$

**Proof.** Let  $f_j^*$  be the rows of the discrete Fourier transform, *i.e.*  $f_j[\ell] = e^{i \frac{2\pi j \ell}{n}}$ . With a slight adaptation to the real case of the argument in [19, Section A.4.1], we deduce that

$$\left\| \frac{1}{nP} \sum_{j,p}^{n,P} |f_j^* D_p \bar{x}|^2 D_p f_j f_j^* D_p - \left( \bar{x} \bar{x}^\top + \|\bar{x}\|^2 \text{Id} \right) \right\| \leq \frac{\delta}{2} \|\bar{x}\|^2, \quad (\text{B.18})$$

provided that  $P \geq C(\delta) \log^3(n)$  with a probability at least  $1 - \frac{4P+1}{2n^3}$ . Combining this with Lemma B.8, we conclude.  $\square$

### B.2.2 Injectivity of the measurement operator

We now establish that for  $m$  large enough, the measurement matrix  $A$  is injective *w.h.p.* Recall that the rows of  $A$  are the  $a_r^*$ 's.

**Lemma B.10.** *Fix  $\varrho \in ]0, 1[$ . Assume that  $P \geq C(\varrho) \log(n)$ . Then with a probability at least  $1 - 1/n^2$*

$$(1 - \varrho) \|x\|^2 \leq \frac{1}{m} \|Ax\|^2 \leq (1 + \varrho) \|x\|^2, \quad \forall x \in \mathbb{R}^n. \quad (\text{B.19})$$

**Proof.** This is a consequence of the fact that

$$\left\| \frac{1}{m} A^* A - \text{Id} \right\| \leq \varrho$$

with the claimed probability. Indeed, as for [18, Lemma 3.3], the covariance matrix  $\frac{1}{m} A^* A$  is diagonal with i.i.d diagonal entries whose expectation is  $\mathbb{E}(d^2) = 1$ , and the statement follows from Hoeffding's inequality and a union bound.  $\square$

### B.2.3 Local relative smoothness and relative strong convexity

We now turn to proving local relative smoothness and relative strong convexity near the true vectors. Unlike the Gaussian case, we only have a local version for relative smoothness. The reason behind this, as discussed above, is that it seems very hard to have a uniform concentration bound for the Hessian of  $f$  around its mean for the CDP model. To circumvent this, we use a continuity argument.

**Lemma B.11.** *Fix  $\delta \in ]0, \min(\|\bar{x}\|^2, 1)/2[$ . Suppose that (B.17) holds. Then there exists  $\rho_\delta > 0$  such that for all  $x, z \in B(\bar{x}, \rho_\delta)$  and  $x, z \in B(-\bar{x}, \rho_\delta)$*

$$\frac{(\min(\|\bar{x}\|^2, 1) - 2\delta)}{1 + \delta} D_\psi(x, z) \leq D_f(x, z) \leq 2(1 + \delta)^2 D_\psi(x, z). \quad (\text{B.20})$$

Observe that while in the Gaussian case, the ball radius on which relative strong convexity holds is fixed and explicit, for the CDP model, we only know it exists and it depends on  $\delta$ .

**Proof.** We prove the claim for  $\bar{x}$  and the same holds obviously around  $-\bar{x}$ . Using (B.17) and (A.3) gives

$$\begin{aligned} \nabla^2 f(\bar{x}) &\leq \mathbb{E}(\nabla^2 f(\bar{x})) + \delta \|\bar{x}\|^2 \text{Id} \\ &= 2\bar{x}\bar{x}^\top + (2 + \delta) \|\bar{x}\|^2 \text{Id} \\ &\leq (2 + \delta) \left( 2\bar{x}\bar{x}^\top + (\|\bar{x}\|^2 + 1) \text{Id} \right), \\ &= (2 + \delta) \nabla^2 \psi(\bar{x}). \end{aligned}$$

Again, from (B.17) and (A.3), we get

$$\nabla^2 f(\bar{x}) \succeq \mathbb{E}(\nabla^2 f(\bar{x})) - \delta \|\bar{x}\|^2 \text{Id} \succeq 2(\bar{x}\bar{x}^\top + \|\bar{x}\|^2 \text{Id}) - \delta \nabla^2 \psi(\bar{x}) \text{Id}.$$

If  $\|\bar{x}\| \geq 1$ , we arrive at

$$\nabla^2 f(\bar{x}) \succeq 2\bar{x}\bar{x}^\top + (\|\bar{x}\|^2 + 1) \text{Id} - \delta \nabla^2 \psi(\bar{x}) \text{Id} = (1 - \delta) \nabla^2 \psi(\bar{x}).$$

If  $\|\bar{x}\| \leq 1$ , we have

$$\begin{aligned} \nabla^2 f(\bar{x}) - \left( \|\bar{x}\|^2 - \delta \right) \nabla^2 \psi(\bar{x}) &\succeq 2\bar{x}\bar{x}^\top + 2\|\bar{x}\|^2 \text{Id} - 2\|\bar{x}\|^2 \bar{x}\bar{x}^\top - \|\bar{x}\|^4 \text{Id} - \|\bar{x}\|^2 \text{Id} \\ &= 2(1 - \|\bar{x}\|^2) \bar{x}\bar{x}^\top + (\|\bar{x}\|^2 - \|\bar{x}\|^4) \text{Id} \succeq 0. \end{aligned}$$

Therefore

$$\left(\min(\|\bar{x}\|^2, 1) - \delta\right) \nabla^2 \psi(\bar{x}) \preceq \nabla^2 f(\bar{x}) \preceq (2 + \delta) \nabla^2 \psi(\bar{x}). \quad (\text{B.21})$$

Combining (B.21) with continuity of  $\nabla^2 f$  and 1-strong convexity of  $\psi$ ,  $\exists \rho_\delta > 0$  such that  $\forall x \in B(\bar{x}, \rho_\delta)$  we have

$$\left(\min(\|\bar{x}\|^2, 1) - 2\delta\right) \nabla^2 \psi(\bar{x}) \preceq \nabla^2 f(\bar{x}) - \delta \text{Id} \preceq \nabla^2 f(x) \preceq \nabla^2 f(\bar{x}) + \delta \text{Id} \preceq 2(1 + \delta) \nabla^2 \psi(\bar{x}). \quad (\text{B.22})$$

Continuity of  $\nabla^2 \psi$  and 1-strong convexity of  $\psi$  also yield that  $\forall x \in B(\bar{x}, \rho_\delta)$

$$\nabla^2 \psi(\bar{x}) \preceq \nabla^2 \psi(x) + \delta \text{Id} \preceq (1 + \delta) \nabla^2 \psi(x) \quad \text{and} \quad \nabla^2 \psi(x) \preceq \nabla^2 \psi(\bar{x}) + \delta \text{Id} \preceq (1 + \delta) \nabla^2 \psi(\bar{x}). \quad (\text{B.23})$$

Combining (B.22) and (B.23), we obtain that  $\forall x \in B(\bar{x}, \rho_\delta)$ ,

$$\frac{\left(\min(\|\bar{x}\|^2, 1) - 2\delta\right)}{1 + \delta} \nabla^2 \psi(x) \preceq \nabla^2 f(x) \preceq 2(1 + \delta)^2 \nabla^2 \psi(x).$$

Invoking Lemma A.1, we get the statement.  $\square$

## B.2.4 Spectral initialization

We now show the analogue of Lemma B.7 for the CDP measurement model.

**Lemma B.12.** *Fix  $\varrho \in ]0, 1[$ . If the number of patterns obeys  $P \geq C(\varrho)n \log^3(n)$ , for some sufficiently large constant  $C(\varrho) > 0$ , then with probability at least  $1 - \frac{4P+1}{n^3} - \frac{1}{n^2}$ ,  $x_0$  satisfies:*

(i)  $\text{dist}(x_0, \bar{\mathcal{X}}) \leq \eta_1(\varrho) \|\bar{x}\|$ , where  $\eta_1$  is the function defined in (B.11).

*Let  $\delta \in ]0, \min(\|\bar{x}\|^2, 1)/2[$  and  $\rho_\delta$  is the neighbourhood radius in Lemma B.11. Suppose that  $\varrho$  is sufficiently small, i.e.*

$$\varrho \leq \min \left( \delta, \eta_1^{-1} \left( \frac{\rho_\delta / \|\bar{x}\|}{\sqrt{6(\|\bar{x}\|^2 + \rho_\delta^2) + 1}} \right) \right). \quad (\text{B.24})$$

*Then, with the same probability as above,*

(ii)  $f(x_0) \leq 2(1 + \delta^2) \frac{\Theta(\eta_1(\varrho)\|\bar{x}\|)}{2} \eta_1(\varrho)^2 \|\bar{x}\|^2$ ;

(iii)  $x_0 \in B \left( \bar{\mathcal{X}}, \frac{\rho_\delta}{\max(\sqrt{\Theta(\rho_\delta)}, 1)} \right)$ .

**Proof.** The proof of this claim is similar to that of Lemma B.7 for the Gaussian case, where we now invoke Lemma B.9 and Lemma B.10 for statement (i). For the last two claims, we also use Lemma B.11 and that  $\varrho$  is small enough as prescribed.  $\square$

**Acknowledgments.** The authors thank the French National Research Agency (ANR) for funding the project FIRST (ANR-19-CE42-0009).

## References

- [1] J. Akutowicz, E. On the Determination of the Phase of a Fourier Integral, I. *Transactions of the American Mathematical Society*, 83(1):179, September 1956.
- [2] J. Akutowicz, E. On the Determination of the Phase of a Fourier Integral, II. *Proceedings of the American Mathematical Society*, 8(2):234, April 1957.
- [3] C. Amra, M. Zerrad, S. Liukaityte, and M. Lequime. Instantaneous one-angle white-light scatterometer. *Opt. Express, OE*, 26(1):204–219, January 2018.
- [4] H. Attouch, J. Bolte, and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Math. Program.*, 137(1-2):91–129, February 2013.
- [5] R. Balan. Reconstruction of Signals from Magnitudes of Redundant Representations: The Complex Case. *Found Comput Math*, 16(3):677–721, June 2016.
- [6] R. Balan, P. Casazza, and D. Edidin. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3):345–356, May 2006.
- [7] R. Barakat and G. Newsam. Algorithms for reconstruction of partially known, band-limited Fourier-transform pairs from noisy data. *J. Opt. Soc. Am. A, JOSAA*, 2(11):2027–2039, November 1985.
- [8] H. Bauschke, H. and Adrian S. Lewis. Dykstras algorithm with bregman projections: A convergence proof. *Optimization*, 48(4):409–427, January 2000.
- [9] H. H. Bauschke, J. Bolte, and M. Teboulle. A Descent Lemma Beyond Lipschitz Gradient Continuity: First-Order Methods Revisited and Applications. *Mathematics of Operations Research*, page 20, 2016.
- [10] 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.
- [11] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Finding best approximation pairs relative to two closed convex sets in hilbert spaces. *J. Approx. Theory*, 127:178–192, 2004.
- [12] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang. Restricted normal cones and the method of alternating projections: applications. *Set-Valued and Variational Analysis*, 21:475–501, 2013.
- [13] A. Ben-Tal and A. S. Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. MPS-SIAM series on optimization. Society for Industrial and Applied Mathematics, 2001.
- [14] Benjamin Birnbaum, Nikhil R Devanur, and Lin Xiao. Distributed algorithms via gradient descent for fisher markets. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 127–136, 2011.
- [15] 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.
- [16] Yu.M. Bruck and L.G. Sodin. On the ambiguity of the image reconstruction problem. *Optics Communications*, 30(3):304–308, 1979.
- [17] Xavier Buet, Myriam Zerrad, Michel Lequime, Gabriel Soriano, Jean-Jacques Godeme, Jalal Fadili, and Claude Amra. Immediate and one-point roughness measurements using spectrally shaped light. *Opt. Express*, 30(10):16078–16093, May 2022.
- [18] E. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277–299, September 2015.
- [19] E. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Trans. Inform. Theory*, 61(4):1985–2007, 2015.

- [20] E. Candès and Xiaodong Li. Solving Quadratic Equations via PhaseLift When There Are About as Many Equations as Unknowns. *Found. Comput. Math.*, 2014.
- [21] E. Candès and Y. Plan. Matrix Completion With Noise. *Proceedings of the IEEE*, 98:925–936, 2010.
- [22] E. Candès, T. Strohmer, and V. Voroninski. PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.
- [23] Emmanuel J. Candès, Yonina C. Eldar, Thomas Strohmer, and Vladislav Voroninski. Phase retrieval via matrix completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225, 2013.
- [24] G. Chen and M. Teboulle. Convergence Analysis of a Proximal-Like Minimization Algorithm Using Bregman Functions. *SIAM J. Optim.*, 3(3):538–543, August 1993.
- [25] Y. Chen and E. Candès. Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems. *Comm. Pure Appl. Math.*, 70(5):822–883, May 2017.
- [26] Y. Chen, Y. Chi, J. Fan, and C. Ma. Gradient Descent with Random Initialization: Fast Global Convergence for Nonconvex Phase Retrieval. *Math. Program.*, 176(1-2):5–37, July 2019.
- [27] T. R. Crimmins and J. R. Fienup. Ambiguity of phase retrieval for functions with disconnected support. *J. Opt. Soc. Am.*, 71(8):1026, August 1981.
- [28] T. R. Crimmins and J. R. Fienup. Uniqueness of phase retrieval for functions with sufficiently disconnected support. *J. Opt. Soc. Am.*, 73(2):218, February 1983.
- [29] D. Davis, D. Drusvyatskiy, and C. Paquette. The nonsmooth landscape of phase retrieval. *IMA Journal of Numerical Analysis*, 40(4):2652–2695, October 2020.
- [30] D. Drusvyatskiy, A. D. Ioffe, and A. S. Lewis. Transversality and alternating projections for nonconvex sets. *Found. Comput. Math.*, 15(6):1637–1651, 2015.
- [31] A. Fannjiang and T. Strohmer. The numerics of phase retrieval. *Acta Numerica*, 29:125–228, May 2020.
- [32] J. R. Fienup. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21(15):2758, August 1982.
- [33] R. Gerchberg and W. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35(2):237, 1972.
- [34] X. Goemans, M. and P. Williamson, D. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.
- [35] D. Gross, F. Kraher, and R. Kueng. Improved Recovery Guarantees for Phase Retrieval from Coded Diffraction Patterns. *Applied and Computational Harmonic Analysis*, 42(1):37–64, January 2017.
- [36] R. Hesse and D. R. Luke. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *SIAM J. Optim.*, 23(4):2397–2419, 2013.
- [37] K. Jaganathan, Y. C. Eldar, and B. Hassibi. Phase retrieval: An overview of recent developments. In A. Stern, editor, *Optical Compressive Imaging*. CRC Press, 2016.
- [38] J. Lee, D., I. Panageas, G. Piliouras, M. Simchowitz, M. Jordan, I., and B. Recht. First-order methods almost always avoid strict saddle points. *Mathematical Programming*, 176(1):311–337, 2019.
- [39] A. S. Lewis, D. R. Luke, , and J. Malick. Local linear convergence of alternating and averaged projections. *Found. Comput. Math.*, 9(4):485–513, 2009.
- [40] A. S. Lewis and J. Malick. Alternating projections on manifolds. *Math. Oper. Res.*, 33:216–234, 2008.
- [41] S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. In *Les Équations aux Dérivées Partielles*, pages 87–89. Editions du Centre National de la Recherche Scientifique, 1963.

- [42] S. Łojasiewicz. Ensembles semi-analytiques. *Lectures Notes IHES (Bures-sur-Yvette)*, 1965.
- [43] Haihao Lu, Robert M. Freund, and Yurii Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIAM Journal on Optimization*, 28(1):333–354, 2018.
- [44] D. R. Luke. Finding best approximation pairs relative to a convex and a prox-regular set in hilbert space. *SIAM J. Optim.*, 19(2):714–739, 2008.
- [45] D. R. Luke. Local linear convergence of approximate projections onto regularized sets. *Nonlinear Anal.*, 75:1531–1546, 2012.
- [46] D. R. Luke. Phase Retrieval, What’s New? *SIAG/OPT Views and News*, 25(1):1–6, 2017.
- [47] J. Miao, P. Charalambous ad J. Kirz, and D. Sayre. Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400:342–344, 1999.
- [48] C. Mukkamala, M., P. Ochs, T. Pock, and S. Sabach. Convex-Concave Backtracking for Inertial Bregman Proximal Gradient Algorithms in Nonconvex Optimization. *SIAM Journal on Mathematics of Data Science*, 2(3):658–682, January 2020.
- [49] P. Netrapalli, P. Jain, and S. Sanghavi. Phase retrieval using alternating minimization. *IEEE Transactions on Signal Processing*, 63(18):4814–4826, 2015.
- [50] D. Noll and A. Rondepierre. On local convergence of the method of alternating projections. *Found. Comput. Math.*, 16(2):425–455, 2016.
- [51] H. Phan. Linear convergence of the douglas-rachford method for two closed sets. *Optimization*, 65:369–385, 2016.
- [52] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [53] H. Sahinoglou and S. Cabrera. On phase retrieval of finite-length sequences using the initial time sample. *IEEE Transactions on Circuits and Systems*, 38(5):954–958, 1991.
- [54] Y Shechtman, C. Eldar, Y., O. Cohen, N. Chapman, H., 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.
- [55] J. Sun, Q. Qu, and J. Wright. A Geometric Analysis of Phase Retrieval. *Found Comput Math*, 18(5):1131–1198, October 2018.
- [56] M. Teboulle. Entropic Proximal Mappings with Applications to Nonlinear Programming. *Mathematics of OR*, 17(3):670–690, August 1992.
- [57] M. Teboulle. A simplified view of first order methods for optimization. *Math. Program.*, 170(1):67–96, July 2018.
- [58] N. Vaswani. Non-Convex Structured Phase Retrieval. *arXiv:2006.13298 [cs, eess, math, stat]*, June 2020.
- [59] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv:1011.3027 [cs, math]*, November 2011.
- [60] I. Waldspurger. Phase Retrieval With Random Gaussian Sensing Vectors by Alternating Projections. *IEEE Transactions on Information Theory*, 64(5):3301–3312, May 2018.
- [61] I. Waldspurger, A. d’Aspremont, and S. Mallat. Phase recovery, MaxCut and complex semidefinite programming. *Math. Program.*, 149(1):47–81, February 2015.
- [62] A. Walther. The Question of Phase Retrieval in Optics. *Optica Acta: International Journal of Optics*, 10(1):41–49, January 1963.



- [63] G. Wang, B. Giannakis, G., and C. Eldar, Y. Solving Systems of Random Quadratic Equations via Truncated Amplitude Flow. *arXiv:1605.08285 [cs, math, stat]*, August 2017.
- [64] H. Zhang, Y. Liang, and Y. Chi. A Nonconvex Approach for Phase Retrieval: Reshaped Wirtinger Flow and Incremental Algorithms. *Journal of Machine Learning Research*, 18(141):1–35, 2017.