Provable Phase retrieval via Mirror descent

J-J. Godeme[1]

Joint work: J. Fadili[1], X. Buet[2], M. Zerrad[2], M. Lequime[2] and C. Amra[2]

[1]Normandie Univ., ENSICAEN, UNICAEN, CNRS, GREYC, France

[2]Aix-Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, Marseille, France

GdR-Mathematics of Optimization and Applications (Nice, 11-14th/10/2022)

October 14, 2022

Groupe de Recherche En Informatique, Image, automatique et instrumentation de Caen (UMR6072). CNRS-ENSICAEN-Université de Caen Basse Normandie

Outline

Introduction

Problem Statement

Goal: To Recover $\bar{x} \in \mathbb{R}^n$ from the measurements

$$y_r = |\langle a_r, \bar{x} \rangle|^2 = |a_r^* \bar{x}|^2, \quad r \in [m], \tag{PR}$$

where $(a_r)_{r\in[m]}$ are the sensing vectors.

Introduction

Problem Statement

Goal: To Recover $\bar{x} \in \mathbb{R}^n$ from the measurements

$$y_r = |\langle a_r, \bar{x}\rangle|^2 = |a_r^* \bar{x}|^2, \quad r \in [m], \tag{PR}$$

where $(a_r)_{r\in[m]}$ are the sensing vectors.

We cast it as solving the following least squares problem:

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{4m}\sum_{r=1}^{m}\left(y_r - |a_r^*\bar{x}|^2\right)^2. \tag{$\mathcal{P}$}$$

Introduction

Problem Statement

Goal: To Recover $\bar{x} \in \mathbb{R}^n$ from the measurements

$$y_r = |\langle a_r, \bar{x} \rangle|^2 = |a_r^* \bar{x}|^2, \quad r \in [m], \tag{PR}$$

where $(a_r)_{r \in [m]}$ are the sensing vectors.

We cast it as solving the following least squares problem:

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{4m} \sum_{r=1}^{m} \left(y_r - |a_r^* \bar{x}|^2 \right)^2. \tag{$\mathcal{P}$}$$

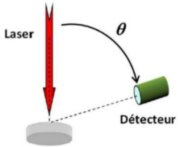
Keys Observations

- One can only hope to recover $\bar{x}$ up to global sign change.
- f is C^2, but ∇f is not Lipschitz.
- f is non-convex.

Application to Light scattering

Light Scattering with CONCEPT team (Fresnel Institute)

- Perfomance in industry depend on the structure of the materials; small defects can yields to big problems.
- Light Scattering is a technique to determine non destructively the roughness of a given polished surface.



$$\nu_d = \frac{n_j \sin \theta_j^d}{\lambda} \quad j = 0,1$$

$s_e(x,y)$ *Profile of the illumination beam*

$h(x,y)$ *Topography of the surface*

$$d\Phi_0^d \propto \frac{1}{S} \left| \left[\hat{h} \star \hat{s}_e \right]_{\vec{\nu}_d} \right|^2 = \gamma_e(\vec{\nu}_d)$$

C. Amra, M. Lequime and M. Zerrad, « electromagnetic Optics of Thin-Film Coatings », Cambridge University Press, Cambridge (2021)

Prior work

How do we design a scalable and efficient numerical scheme to solve the problem of phase retrieval?

Wirtinger Flow (Gradient descent) (Candès et al. 2015)

Find an initial guess near the solution and apply Gradient descent.

Wirtinger Flow (Gradient descent) (Candès et al. 2015)

Find an initial guess near the solution and apply Gradient descent.

Algorithm

Algorithm 2 Wirtinger Flow Procedure

Input: $y_r, r = 1, \ldots, m,\ \lambda^2 = n\frac{\sum_r y_r}{\sum_r \|a_r\|^2},\ \gamma > 0$

- x_0 as top eigenvector of $Y = \frac{1}{m}\sum_{r=1}^{m} y_r a_r a_r^*$ normalized to $\|x_0\| = \lambda$.
- Compute $x_{k+1} = x_k - \gamma\nabla f(x_k)$

Let us define:

$$\forall x \in \mathbb{R}^n, \quad \mathrm{dist}(x, \bar{x}) = \min\{\|x - \bar{x}\|, \|x + \bar{x}\|\} \tag{1}$$

Wirtinger flow (Gradient descent)

Theorem (Candès et al. 2015)

When the number of measurements $m \geq cn\log(n)$ for the Gaussian case (resp. $m \geq cn\log(n)^3$ for the CDP model). Then *w.h.p* the spectral estimate x_0 satisfies the following

$$\mathrm{dist}(x_0, \bar{x}) \leq \frac{1}{8}\|\bar{x}\|, \tag{2}$$

Besides, if the stepsize $\gamma = \frac{c_1}{n}$ for some fixed numerical constant c_1, then *w.h.p* the iterates of the gradient descent satisfies

$$\mathrm{dist}(x_k, \bar{x}) \leq \frac{\|\bar{x}\|}{8}\left(1 - \frac{c_1}{4n}\right)^{k/2}. \tag{3}$$

Wirtinger flow (Gradient descent)

Theorem (Candès et al. 2015)

When the number of measurements $m \geq cn\log(n)$ for the Gaussian case (resp. $m \geq cn\log(n)^3$ for the CDP model). Then *w.h.p* the spectral estimate x_0 satisfies the following

$$\operatorname{dist}(x_0, \bar{x}) \leq \frac{1}{8} \|\bar{x}\|, \tag{2}$$

Besides, if the stepsize $\gamma = \frac{c_1}{n}$ for some fixed numerical constant c_1, then *w.h.p* the iterates of the gradient descent satisfies

$$\operatorname{dist}(x_k, \bar{x}) \leq \frac{\|\bar{x}\|}{8} \left(1 - \frac{c_1}{4n}\right)^{k/2}. \tag{3}$$

The iterates of Gradient descent are really slow as n grows!!

Main Challenges

1. f is C^2 but ∇f is not Lipschitz continuous

Main Challenges

1. f is C^2 but ∇f is not Lipschitz continuous $\Rightarrow$ precludes simple gradient descent.

Main Challenges

1. f is C^2 but ∇f is not Lipschitz continuous $\Rightarrow$ precludes simple gradient descent.
2. f nonconvex

Main Challenges

1. f is C^2 but ∇f is not Lipschitz continuous $\Rightarrow$ precludes simple gradient descent.
2. f nonconvex $\Rightarrow$ how to avoid special techniques to find a good initial guess?

Main Challenges

1. f is C^2 but ∇f is not Lipschitz continuous $\Rightarrow$ precludes simple gradient descent.
2. f nonconvex $\Rightarrow$ how to avoid special techniques to find a good initial guess?

Key Idea: Change of geometry

- Associate to f the "nice" entropy $\psi(x) = \frac{1}{4}\|x\|^4 + \frac{1}{2}\|x\|^2$.
- ψ is smooth and strongly convex on $\mathbb{R}^n$.
- f has the *relative* gradient Lipchitz continuity property with respect to ψ. (To be explained shortly.)

Bregman Toolbox

To any function $g : \mathbb{R}^n \to (-\infty, +\infty]$ such that $g \in C^1(\mathbb{R}^n)$, we define:

Definition

The Bregman proximity distance generated by a function g is given by:

$$D_g(x, y) = g(x) - g(y) - \langle \nabla g(y); x - y \rangle \tag{4}$$

Bregman Toolbox

To any function $g : \mathbb{R}^n \to (-\infty, +\infty]$ such that $g \in C^1(\mathbb{R}^n)$, we define:

Definition

The Bregman proximity distance generated by a function g is given by:

$$D_g(x, y) = g(x) - g(y) - \langle \nabla g(y); x - y \rangle \tag{4}$$

Properties of the Bregman distance

- This proximity measure is not symmetric in general.
- g is convex if and only if $D_g(x, y) \geq 0, \forall x, y \in \mathbb{R}^n$.

Generalization of gradient Lipschitz continuity

Definition (Relative smoothness)

A pair of function (ϕ, g) satisfy the $L-$smooth adaptable (L-smad) condition (or *relative* smoothness) if there exists $L > 0$ such that $L\phi - g$ and $L\phi + g$ are convex *i.e.*,

$$|D_g(x, y)| \leq LD_\phi(x, y) \quad \forall x, y \in \mathbb{R}^n. \tag{5}$$

Generalization of gradient Lipschitz continuity

Definition (Relative smoothness)

A pair of function (ϕ, g) satisfy the $L-$smooth adaptable (L-smad) condition (or *relative* smoothness) if there exists $L > 0$ such that $L\phi - g$ and $L\phi + g$ are convex *i.e.*,

$$|D_g(x, y)| \leq LD_\phi(x, y) \quad \forall x, y \in \mathbb{R}^n. \tag{5}$$

When $\phi(x) = \frac{1}{2}\|x\|^2$, we recover the classical definition. Since (5) is true $\forall x, y$ we deduce,

$$\langle x - y; \nabla g(x) - \nabla g(y) \rangle \leq L \|x - y\|^2,$$

this fact implies that,

$$\|\nabla g(x) - \nabla g(y)\| \leq L \|x - y\|, \quad \forall x, y \in \mathbb{R}^n, \tag{6}$$

Generalization of strong convexity

Definition (Relative Strong Convexity)

A function g is said to be relatively strongly convex with respect to another function ϕ if there exists $\sigma > 0$ such that $g - \sigma\phi$ is convex *i.e.*,

$$\sigma D_\phi(x, y) \leq D_g(x, y) \quad \forall x, y \in \mathbb{R}^n. \tag{7}$$

Generalization of strong convexity

Definition (Relative Strong Convexity)

A function g is said to be relatively strongly convex with respect to another function ϕ if there exists $\sigma > 0$ such that $g - \sigma\phi$ is convex *i.e.*,

$$\sigma D_\phi(x, y) \leq D_g(x, y) \quad \forall x, y \in \mathbb{R}^n. \tag{7}$$

When $\phi(x) = \frac{1}{2}\|x\|^2$, we recover the classical definition A function g is $\sigma-$strongly convex if the function $g - \frac{\sigma}{2}\|.\|^2$ is convex.

Phase retrieval via Mirror descent

To,

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{4m}\sum_{r=1}^{m}\left(y_r - |a_r^* x|^2\right)^2. \qquad (\mathcal{P})$$

We associate

$$\psi(x) = \frac{1}{4}\|x\|^4 + \frac{1}{2}\|x\|^2, \qquad (8)$$

Lemma (Bolte et al. 2018)

The function f is *relatively* smooth with respect to the entropy ψ with $L = \frac{3}{m}\sum_{r=1}^{m}\|a_r\|^4$.

Phase retrieval via Mirror descent

Algorithm

Algorithm 3 Mirror Descent with backtracking for Phase retrieval

Parameters: $L_0 > 0$, $\kappa > 0$ (small), $\xi \geq 1$,
Initialization: $x_0 \in \mathbb{R}^n$,
for: $k = 0, 1, \ldots$ **do**
Repeat until: $D_f(x_{k+1}, x_k) > L_k D_\psi(x_{k+1}, x_k)$

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

end

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

end.

Where we have: $\nabla\psi^* = \nabla\psi^{-1}$.

Deterministic Main Result

Theorem

Let $x^\star \in \mathrm{Argmin}(f) \neq \emptyset, r > 0$ and $(x_k)_k$ be a bounded sequence generated by the Algorithm 3, then

1. $(f(x_k))_k$ is nonincreasing, $(x_k)_k$ has a finite length and converges to a point in $\mathrm{crit}(f)$.

Deterministic Main Result

Theorem

Let $x^\star \in \mathrm{Argmin}(f) \neq \emptyset, r > 0$ and $(x_k)_k$ be a bounded sequence generated by the Algorithm 3, then

1. $(f(x_k))_k$ is nonincreasing, $(x_k)_k$ has a finite length and converges to a point in $\mathrm{crit}(f)$.
2. Assume that x_0 is the $f-$attentive neighborhood of $x^\star$ *i.e.*,$\exists\delta \in]0, r[$ and $\mu > 0$ such that $x_0 \in B(x^\star, \delta)$ and $f(x_0) \in]0, \mu[$ then,

Deterministic Main Result

Theorem

Let $x^\star \in \mathrm{Argmin}(f) \neq \emptyset, r > 0$ and $(x_k)_k$ be a bounded sequence generated by the Algorithm 3, then

1. $(f(x_k))_k$ is nonincreasing, $(x_k)_k$ has a finite length and converges to a point in $\mathrm{crit}(f)$.
2. Assume that x_0 is the $f-$attentive neighborhood of $x^\star$ *i.e.*,$\exists\delta \in]0, r[$ and $\mu > 0$ such that $x_0 \in B(x^\star, \delta)$ and $f(x_0) \in]0, \mu[$ then,
   - For all $k \in \mathbb{N}$, $x_k \in B(x^\star, r)$ and $\mathrm{dist}(x_k, x^\star) \to 0$.

Deterministic Main Result

Theorem

Let $x^\star \in \text{Argmin}(f) \neq \emptyset, r > 0$ and $(x_k)_k$ be a bounded sequence generated by the Algorithm 3, then

1. $(f(x_k))_k$ is nonincreasing, $(x_k)_k$ has a finite length and converges to a point in $\text{crit}(f)$.
2. Assume that x_0 is the $f-$attentive neighborhood of $x^\star$ *i.e.*,$\exists\delta \in]0, r[$ and $\mu > 0$ such that $x_0 \in B(x^\star, \delta)$ and $f(x_0) \in]0, \mu[$ then,
   - For all $k \in \mathbb{N}$, $x_k \in B(x^\star, r)$ and $\text{dist}(x_k, x^\star) \to 0$.
   - Besides, if $\exists\rho > 0$ such that f is locally $\sigma-$ strong convex relatively to ψ in $B(x^\star, \rho)$ with $r \leq \frac{\rho}{\max(\sqrt{\Theta(\rho)},1)}$ then $\forall k = 1, 2, \cdots$

$$\|x_k - x^\star\|^2 \leq \prod_{i=1}^{k-1} (1 - \sigma\gamma_i)\, \rho^2 \to 0. \tag{9}$$

Deterministic Main Result

Theorem

Let $x^\star \in \mathrm{Argmin}(f) \neq \emptyset, r > 0$ and $(x_k)_k$ be a bounded sequence generated by the Algorithm 3, then

1. $(f(x_k))_k$ is nonincreasing, $(x_k)_k$ has a finite length and converges to a point in $\mathrm{crit}(f)$.
2. Assume that x_0 is the $f-$attentive neighborhood of $x^\star$ *i.e.*,$\exists \delta \in]0, r[$ and $\mu > 0$ such that $x_0 \in B(x^\star, \delta)$ and $f(x_0) \in]0, \mu[$ then,
   - For all $k \in \mathbb{N}$, $x_k \in B(x^\star, r)$ and $\mathrm{dist}(x_k, x^\star) \to 0$.
   - Besides, if $\exists \rho > 0$ such that f is locally $\sigma-$ strong convex relatively to ψ in $B(x^\star, \rho)$ with $r \leq \frac{\rho}{\max(\sqrt{\Theta(\rho)},1)}$ then $\forall k = 1, 2, \cdots$

$$\|x_k - x^\star\|^2 \leq \prod_{i=1}^{k-1} (1 - \sigma\gamma_i)\, \rho^2 \to 0. \tag{9}$$

3. If $L_k \equiv L$ then for **Lebesgue almost all initializers** x_0, $x_k \to \widetilde{x} \in \mathrm{crit}(f)$ **where** $f(\widetilde{x})$ **has no direction of strictly negative curvature.**

If $\mathrm{crit}(f)\backslash\mathrm{strisad}(f) = \mathrm{Argmin}\, f$ then $x_k \to \widetilde{x} \in \mathrm{Argmin}\, f$.

Random Phase retrieval

Framework: Types of sensing vectors

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

$$y[r] = |{a_r}^\top \bar{x}|^2, \quad r \in [m], \tag{10}$$

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

- The Coded Diffraction Patterns (CDP) model. The observation model is

$$y = \left(|\mathcal{F}(D_p\bar{x})[j]|^2\right)_{j,p} = \left(\left|\sum_{\ell=0}^{n-1} \bar{x}_\ell d_p[\ell] e^{-i\frac{2\pi j\ell}{n}}\right|^2\right)_{j,p}. \tag{11}$$

where $j \in \{1,\dots,n\}$, $p \in \{0,\dots,P-1\}$ and $(d_p)_p$ are the mask random variables drawn i.i.d from an appropriate distribution.

Random Phase retrieval

Assumption

- (Boundness) $|d| \leq M$ for some positive constant M *i.e.* Subgaussian,
- (Moment control) $\mathbb{E}(d) = 0, \mathbb{E}(d^4) = 2\mathbb{E}^2(|d|^2)$.

Example: $d = \{-1, 0, 1\}$ with probability $\{\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\}$.

Gaussian Phase Retrieval

Theorem (Godeme et al. 2022)

Fix $\lambda \in]0,1[$ and $\varrho \in]0, \Upsilon(\lambda, \|\bar{x}\|)[$. Let $(x_k)_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 3.

1. If the number of measurements m satisfies $m \geq C(\varrho)n\log(n)^3$, then *w.h.p*, **for almost all initializers** x_0 of Algorithm 3 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 $\mathrm{dist}(x_k, \bar{x}) \to 0$, and $\exists K \geq 0$, large enough such that $\forall k \geq K$,
$$\mathrm{dist}^2(x_k, \bar{x}) \leq (1 - \nu(\kappa, \varrho, \|\bar{x}\|))^{k-K} \rho^2. \tag{12}$$
2. If $m \geq C(\varrho)n\log(n)$ and Algorithm 3 is initialized with the spectral method, then *w.h.p*, (13) holds for all $k \geq K = 0$.

Gaussian Phase Retrieval

Theorem (Godeme et al. 2022)

Fix $\lambda \in]0,1[$ and $\varrho \in]0, \Upsilon(\lambda, \|\bar{x}\|)[$. Let $(x_k)_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 3.

1. If the number of measurements m satisfies $m \geq C(\varrho)n\log(n)^3$, then *w.h.p*, **for almost all initializers** x_0 of Algorithm 3 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 $\mathrm{dist}(x_k, \bar{x}) \to 0$, and $\exists K \geq 0$, large enough such that $\forall k \geq K$,
$$\mathrm{dist}^2(x_k, \bar{x}) \leq \left(1 - \nu(\kappa, \varrho, \|\bar{x}\|)\right)^{k-K} \rho^2. \tag{12}$$
2. If $m \geq C(\varrho)n\log(n)$ and Algorithm 3 is initialized with the spectral method, then *w.h.p*, (13) holds for all $k \geq K = 0$.

$\Upsilon(\lambda, \|\bar{x}\|) = \frac{\lambda\min(\|\bar{x}\|^2,1)}{(2\max(\|\bar{x}\|^2/3,1))}$ and

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

Gaussian Phase retrieval

Remark

- Clearly when $m \geq C(\varrho)n\log(n)^3$ for almost all initializers, MD recover $\pm\bar{x}$ and any initialization becomes superfluous.
- When $\|\bar{x}\| = 1$, the convergence rate takes the simple form

$$\left(1 - \frac{(1-\kappa)(\lambda-\varrho)}{3+\varrho}\right) \approx \frac{2}{3}.$$

- Besides, our convergence rate is dimension-independent which is in clear contrast with the Wirtinger flow.

Coded Diffraction Patterns

Theorem (Godeme et al. 2022)

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

1. If the number of patterns P satisfies $P \geq C(\varrho)n\log(n)$, then *w.h.p,* **for almost all initializers** x_0 of Algorithm 3 used with constant step-size $\gamma_k \equiv \gamma = \frac{1-\kappa}{L}$, for any $\kappa \in]0,1[$, the sequence $(x_k)_{k\in\mathbb{N}}$ converges to an element in $\mathrm{crit}(f)\backslash\mathrm{strisad}(f)$.
2. There exists $\rho_\delta > 0$ such that if ϱ is small enough and $P \geq C(\varrho)n\log^3(n)$ and if Algorithm 3 is initialized with the spectral method, then *w.h.p,* we have,

$$\mathrm{dist}^2(x_k, \bar{x}) \leq \prod_{i=0}^{k-1} \left(1 - \nu_i(\kappa, \varrho, \|\bar{x}\|)\right) \rho_\delta^2. \tag{13}$$

Coded Diffraction Patterns

Theorem (Godeme et al. 2022)

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

1. If the number of patterns P satisfies $P \geq C(\varrho) n \log(n)$, then *w.h.p*, **for almost all initializers** x_0 of Algorithm 3 used with constant step-size $\gamma_k \equiv \gamma = \frac{1-\kappa}{L}$, for any $\kappa \in]0,1[$, the sequence $(x_k)_{k\in\mathbb{N}}$ converges to an element in $\mathrm{crit}(f) \backslash \mathrm{strisad}(f)$.
2. There exists $\rho_\delta > 0$ such that if ϱ is small enough and $P \geq C(\varrho) n \log^3(n)$ and if Algorithm 3 is initialized with the spectral method, then *w.h.p*, we have,
$$\mathrm{dist}^2(x_k, \bar{x}) \leq \prod_{i=0}^{k-1} \left(1 - \nu_i(\kappa, \varrho, \|\bar{x}\|)\right) \rho_\delta^2. \tag{13}$$

$\nu_i(\kappa, \varrho, \|\bar{x}\|) = \frac{(1-\kappa)\left(\min(\|\bar{x}\|^2, 1) - 2\delta\right)}{2(1+\delta)^3}$.

Coded Diffraction Patterns

Remark

- Local linear convergence to the true vector $\bar{x}$ up to a sign change with spectral initialization,

Coded Diffraction Patterns

Remark

- Local linear convergence to the true vector $\bar{x}$ up to a sign change with spectral initialization,
- When $\|\bar{x}\| = 1$, the dimension independent convergence rate is

$$\left(1 - \frac{(1-\kappa)(1-2\delta)}{2(1+\delta)^2}\right) \approx \frac{1}{2}. \tag{14}$$

Coded Diffraction Patterns

Remark

- Local linear convergence to the true vector $\bar{x}$ up to a sign change with spectral initialization,
- When $\|\bar{x}\| = 1$, the dimension independent convergence rate is

$$\left(1 - \frac{(1-\kappa)(1-2\delta)}{2(1+\delta)^2}\right) \approx \frac{1}{2}. \tag{14}$$

- Difficult to show global convergence to the true vectors $\pm\bar{x}$; due to the less randomness of the model.

Coded Diffraction Patterns

Remark

- Local linear convergence to the true vector $\bar{x}$ up to a sign change with spectral initialization,
- When $\|\bar{x}\| = 1$, the dimension independent convergence rate is

$$\left(1 - \frac{(1-\kappa)(1-2\delta)}{2(1+\delta)^2}\right) \approx \frac{1}{2}. \tag{14}$$

- Difficult to show global convergence to the true vectors $\pm\bar{x}$; due to the less randomness of the model.
- Numerical experiments (forthcoming session) show that we recover the true vectors even with random initialization.

Simulations: Gaussian model

We reconstruct a signal $\bar{x} \in \mathbb{R}^n$ from the Gaussian model with $n = 128$.

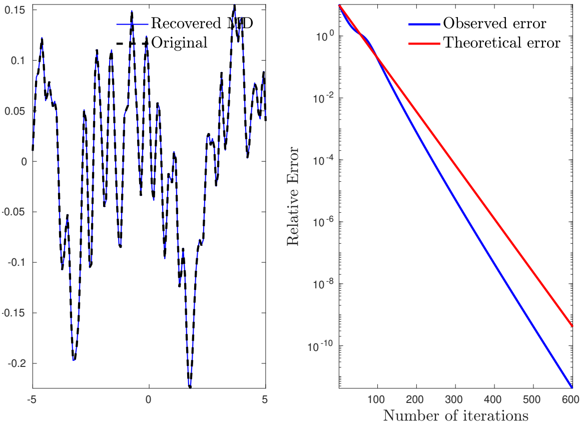


Figure: Reconstruction with random initialization from $m = 2 \times 128 \times \log(128)^3$.

Simulations: Gaussian model

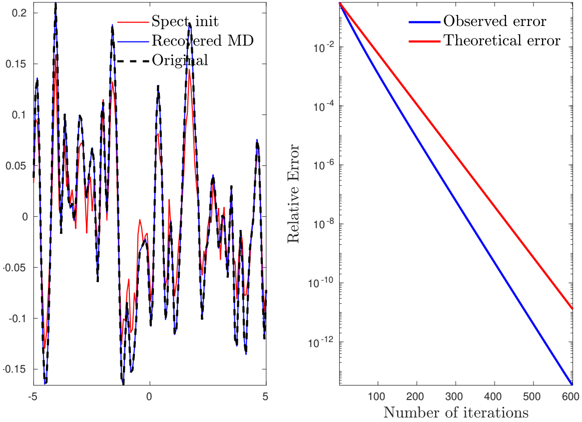


Figure: Reconstruction with spectral initialization from $m = 2 \times 128 \times \log(128)$.

Simulations: Gaussian model

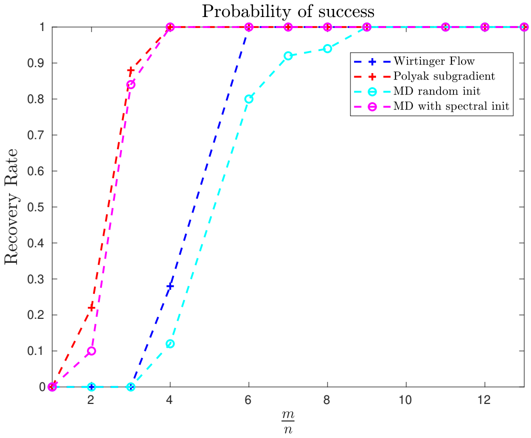


Figure: Phase transition for the Gaussian model.

Simulations: CDP model

We recover a random signal $\bar{x} \in \mathbb{R}^n$ from the Coded Diffraction Pattern Model with $n = 128$.

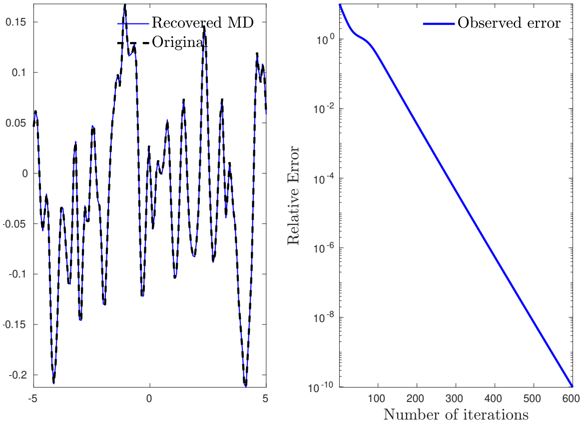


Figure: Reconstruction with random initialization from $P = 7 \times \log(128)^3$ patterns.

Simulations: CDP model

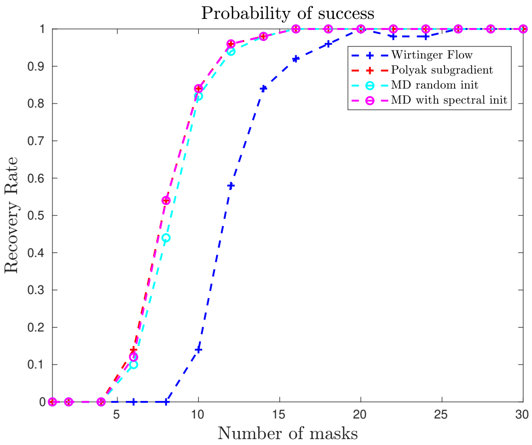


Figure: Phase transition of the CDP model.

Conclusion

Take away messages

- Solve the Phase retrieval using the Mirror descent algorithm with backtracking.

Conclusion

Take away messages

- Solve the Phase retrieval using the Mirror descent algorithm with backtracking.
- For almost all initializers, under a sufficient number of measurements Mirror descent converges to the true vector up to a signchange.

Conclusion

Take away messages

- Solve the Phase retrieval using the Mirror descent algorithm with backtracking.
- For almost all initializers, under a sufficient number of measurements Mirror descent converges to the true vector up to a signchange.
- Show local linear non-dependent dimension convergence rate.

Conclusion

Take away messages

- Solve the Phase retrieval using the Mirror descent algorithm with backtracking.
- For almost all initializers, under a sufficient number of measurements Mirror descent converges to the true vector up to a signchange.
- Show local linear non-dependent dimension convergence rate.
- Mirror descent with our well-chosen entropy ψ is the key to achieve this dimension-independent rate.

Conclusion

Perspectives

- Extend our global convergence result to the case of Coded Diffraction Pattern.

Conclusion

Perspectives

- Extend our global convergence result to the case of Coded Diffraction Pattern.
- Extend our results to the noisy measurements.

Conclusion

Perspectives

- Extend our global convergence result to the case of Coded Diffraction Pattern.
- Extend our results to the noisy measurements.
- Extend to the case of prior knowledge/regularization on the true signal .

Thanks!
Merci!
Akpe!
谢谢!
ありがとう！
¡Gracias!
Grazie!
Mulțmesc!

References I

Bolte, J. et al. (2018). "First Order Methods Beyond Convexity and Lipschitz Gradient Continuity with Applications to Quadratic Inverse Problems". In: *SIAM J. Optim.* 28.3, pp. 2131–2151. DOI: 10.1137/17M1138558. eprint: https://doi.org/10.1137/17M1138558. URL: https://doi.org/10.1137/17M1138558.

Candès, E. et al. (2015). "Phase Retrieval via Matrix Completion". In: *SIAM Rev.* DOI: 10.1137/151005099.

Godeme, J.-J. et al. (2022). "Provable Phase Retrieval via Mirror descent". In: *(to be submitted)*.