A note on Douglas-Rachford, gradient descent, and phase retrieval

Eitan Levin and Tamir Bendory

The Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, 08540 USA

***Abstract*—TB: [The computational aspects of the phase retrieval problem received a lot of attention in the last decade. In particular, there was a deluge of papers employing and analyzing non-convex optimization techniques on different phase retrieval setups. However, in practice, phase retrieval practitioners apply a variety of different techniques, which can be understood as special cases of the Douglas-Rachford framework. In this work, we relate the two by showing that in some cases, Douglas-Rachford is a sub-gradient algorithm, aiming to find a zero of an objective function. The objective function is (we want to stress it is a surprising one.) On the contrary, we show that in other cases, Douglas-Rachford is not a sub-gradient algorithm of any objective function. Finally, we provide some basic analysis of the Douglas-Rachford algorithms for phase retrieval.]**

***Index Terms*—phase retrieval, Douglas-Rachford, sub-gradient descent**

I. Introduction

TB: [Phase retrieval]

Phase retrieval is the problem of recovering a signal from its Fourier magnitudes. This problem is plays a key role in a variety of applications, in particular in optics [14], [12], [13], [7] and signal processing [3], [2], [8], [4].

While almost all multidimensional signals are uniquely determined from their oversampled Fourier transform [Beinert], existing algorithms for recovery are poorly understood. In practice however, many heuristic algorithms exist that are use in the crystallography community, including HIO, RRR, and RAAR. These enjoy good empirical performance, recovering the correct solution every time, but no theoretical guarantees are known for them. For certain choices of their parameters, all three of the above algorithms coincide with with the Douglas-Rachford algorithm, although the theoretical guarantees associated with Douglas-Rachford do not apply to this problem because of its non-convexity.

In this paper, we focus on a particular heuristic algorithm called Relaxed-Reflect-Reflect (RRR) [Elser]. We formulate the algorithm as subgradient descent in a specific setting, and use backtracking line search to adaptively choose the step size, which coincides with the free 'relaxation' parameter introduced in [Elser]. Numerical experiments show that our algorithm can significantly reduce the iteration complexity and time for recovery. Furthermore, while it was previously reported that RRR has an exponential distribution of running time, our algorithm does not.

Because of the difficulty of the Fourier phase retrieval problem, a simplified version has been extensively studied in the optimization literature, often referred to as generalized phase retrieval. In this setting, the Fourier matrix is replaced by a random Gaussian matrix. Several algorithms have been proposed for this problem that enjoy global convergence guarantees. Unfortunately, these algorithms are unable to retrieve Fourier phases. We apply our algorithm to the generalized phase retrieval problems as well and demonstrate a consistent advantage.

[Things I think it worth mentioning in the manuscript: - explain the DR framework - how it fits phase retrieval - relation to ADMM (it is more well known in the community) - relation with alternating projection - we can also include the different ways to derive the GS algorithm.]

II. Problem Statement and Previous Algorithms

We formulate phase retrieval as a feasibility problem: Given a sensing matrix $A \in \mathbb{C}^{m\times n}$ and magnitudes $b \in \mathbb{R}^m_{\geq 0}$, we require a point $x \in \mathcal{M} \cap \operatorname{col}(A)$ where $\mathcal{M} = \{x \in \mathbb{C}^n : |x| = b\}$ and $|x|$ denotes entry-wise absolute value. The sensing matrix is usually taken to be either i.i.d. Gaussian or an oversampled DFT matrix [3], the latter being especially relevant for applications in optics and crystallography [6], [10]. The intersection $\mathcal{M} \cap \operatorname{col}(A)$ is never singleton, since if $z^* \in \operatorname{col}(A)\cap\mathcal{M}$ then $e^{i\theta}z^* \in \operatorname{col}(A)\cap\mathcal{M}$ for any global phase θ. It has been shown however that $\operatorname{col}(A) \cap \mathcal{M}$ is singleton modulo global phase in various settings [1], [3], [5]. [Expand on this?]

The naive approach to this problem, called Grechberg-Saxton (GS) or Error Reduction (ER) in the literature, consists of alternate projection between the two constraint set until convergence. Specifically, we introduce the projectors

$$P_A(x) = AA^\dagger x, \quad P_{\mathcal{M}}(x) = b \odot \operatorname{phase}(x), \tag{1}$$

where P_A projects onto $\operatorname{col}(A)$, $P_{\mathcal{M}}$ projections onto $\mathcal{M}$ and $\operatorname{phase}(x)_i = \frac{x_i}{|x_i|}$. Then, GS performs the iteration

$$x \mapsto P_A P_{\mathcal{M}}(x), \tag{2}$$

which are unfortunately known to quickly converge to sub-optimal local minima, and in practice only used to refine a solution [6], [11].

Instead, many algorithms used in practice are based on relaxing the Douglas-Rachford algorithm applied to the sum of characteristic functions $F(x) = \mathcal{I}_A(x) + \mathcal{I}_\mathcal{M}(x)$. Specifically, Douglas-Rachford for $F(x)$ iterates

$$x \mapsto \frac{1}{2}(I+R_A R_\mathcal{M})(x) = x+2P_A P_\mathcal{M}(x)-P_A(x)-P_\mathcal{M}(x) = P_A P_\mathcal{M}(x)+P_A^c P_\mathcal{M}^c(x), \tag{3}$$

where $R_A = 2P_A - I$, $P_A^c = I - P_A$ and similarly for $R_\mathcal{M}$ and $P_\mathcal{M}^c$, and we used the fact that P_A is linear. Many algorithms proceed to relax this iteration by introducing different free parameters[Rewrite in terms of reflections?]:

- Fienup's Hybrid Input-Output (HIO) algorithm proceeds by iterating
$$x \mapsto P_A P_\mathcal{M}(x) + P_A^c(I - \beta P_\mathcal{M})(x), \tag{4}$$
where P_A^c is the projection onto the complement of the support and β is a parameter controlling the "negative feedback",
- The Relaxed-Reflect-Reflect (RRR) algorithm iterates
$$x \mapsto x + \beta\left[2P_A P_\mathcal{M}(x) - P_A(x) - P_\mathcal{M}(x)\right], \tag{5}$$
,
- The Relaxed Averaged Alternating Reflections (RAAR) algorithm iterates
$$x \mapsto \beta\left[z + 2P_A P_\mathcal{M}(z) - P_A(z) - P_\mathcal{M}(z)\right]+(1-\beta)P_\mathcal{M}(z). \tag{6}$$

These algorithms apply more generally to problems involving finding a point in the intersection of two sets, as demonstrated in [Veit, PNAS] for the RRR algorithm [Actually, "difference map"].

In addition to the above algorithms that are based on relaxing Douglas-Rachford, several algorithms have been developed for the simpler randomized phase retrieval problem. These include Wirtinger Flow and Truncated Wirtinger Flow which are based on gradient descent, and PhaseLift based on semidefinite relaxation. While effective for solving the randomized phase retrieval problem, these algorithms fail in the Fourier phase retrieval problem. Furthermore, even for randomized phase retrieval, they are outperformed by the RRR algorithm, as demonstrated in [benchmarks] and in Sect. TKTK.

III. Basic analysis

We show several basic results about $f(z)$ (because of space limitations, all proofs are given in the appendix).

We first characterize the fixed points of RRR and their relations to the solution of the original feasibility:

Definition 1. A point $z \in \mathbb{C}^m$ is said to *correspond to a solution* if $P_A(z) = P_\mathcal{M}(z)$.

Lemma 1. *z^* is a fixed point of RRR or HIO (and hence a critical point of $f(z)$ in the real case) if and only if z^* corresponds to a solution.*

Proof. See Appendix IV-B □

Lemma 2. *z^* corresponds to a solution if and only if $z^* = y^* + w$ where $y^* \in col(A) \cap \mathcal{M}$ and $w \in col(A)^\perp$ satisfies either $phase(w_i) = phase(y_i^*)$ or $phase(w_i) = -phase(y_i^*)$ and $|w_i| < |(Ax^*)_i|$ for all $1 \leq i \leq m$.*

Proof. See Appendix IV-C. □

We characterize the convergence of RRR to a fixed point from a point sufficiently close to it [in its basin of attraction?]:

Theorem 1. ([9, Thm. 3]) *Suppose $A \in \mathbb{C}^{m\times n}$ with $m/n \geq 2$ is isometric, and $\eta \in (0, 1]$. Suppose $z^* \in col(A) \cap \mathcal{M}$. Then if z is sufficiently close to z^* (see reference for details), RRR converges linearly to z^*.*

In the real case, we can prove an even stronger result:

Lemma 3. *Suppose $z^* \in \mathbb{R}^m$ corresponds to a solution and $d = \min_i |z_i^*| > 0$. Then $f(z)$ is convex in the ℓ_2 ball of radius d about z^*, and 1-strongly convex when restricted to $col(A)$. Furthermore, if $||z - z^*||_2 < d$, and $\eta \in (0, 2)$, then RRR converges to a fixed point linearly, and for $\eta = 1$ after one iteration.*

Proof. See Appendix IV-D. □

As the above Lemma shows, every fixed point is a local minimum of $f(z)$ around which $f(z)$ is convex, making our formulation more stable than the saddle-point formulation in [11]. Note however that these are not the global minima of $f(z)$, as $f(z) = 0$ at any critical point, while $f(z) < 0$ for any suboptimal fixed point of GS. Nevertheless, we can show that $f(z)$ cannot escape to $-\infty$ along many directions:

Lemma 4. *In the real case, there exists large enough step size $\eta > 0$ such that $f(z - \eta d) > 0$ for any $z \in \mathbb{R}^m$, and any direction $d \in \mathbb{R}^m$ such that $d_i \neq 0$ for all i and either $P_A(d) \neq 0$ or $\langle d, P_\mathcal{M}(z)\rangle > 0$.*

Proof. See Appendix IV-E. □

Corollary 1. *For any $z \in \mathbb{R}^m$ such that $\nabla f(z)_i \neq 0$ for any i there exists a sufficiently large step size $\eta > 0$ such that $f(z - \eta\nabla f(z)) > 0$. Similarly, if z^+ is the next HIO iteration and $z_i^+ \neq z_i$ for any i, then either $P_\mathcal{M}(z) \in col(A) \cap \mathcal{M}$ is a solution or there exists a large enough $\eta > 0$ such that $f(z^+) > 0$.*

Proof. See Appendix IV-F. □

We also show that on average, the negative gradient direction is positively correlated with the vector from the current iterate to the nearest solution in $\mathrm{col}(A) \cap \mathcal{M}$:

Lemma 5. *In the real case, for any $z \in \mathbb{R}^m$ let $s(z) = \text{sign}\left(\langle z, Ax^*\rangle\right)$. Then both $E = \mathbb{E}_{z\sim\mathcal{N}(0,I)}\left[\langle -\nabla f(z), s(z)Ax^* - z\rangle\right]$ depends only on $|Ax^*|$ so can be computed in practice, and after possibly renormalizing the problem, i.e. solving for $z^* \in \text{col}(A) \cap \alpha\mathcal{M}$ with $\alpha > 0$, and letting the solution be z^*/α, we have $E > 0$.*

Proof. See Appendix IV-G. □

Finally, we show some stability for the RRR iterations, in the sense that if the gradient is sufficiently small than there is a solution nearby:

Lemma 6. *In the real case, there exists a sufficiently small $\epsilon > 0$ such that if $||\nabla f(z)||_2 < \epsilon$ then $P_{\mathcal{M}}(z) \in \text{col}(A) \cap \mathcal{M}$ is a solution. Furthermore, if $d = \min_i |(Ax^*)_i| > 0$ then there exists a point $z^* \in \mathbb{R}^m$ that corresponds to a solution such that $||z - z^*||_2 < \epsilon\left(1 + \frac{||P_A^c(z)||_2}{d}\right)$. If in addition $\min_i |z_i| \geq \epsilon$ then $||z - z^*||_2 < \epsilon$.*

Proof. See Appendix IV-H. □

IV. Proofs

A. RRR and HIO are not gradients in the complex case

Suppose $z_i \neq 0$ for any i. Note that $P_A(z), P_{\mathcal{M}}(z)$ are gradients, as shown in [11]. However, $P_A P_{\mathcal{M}}(z)$ is not a gradient, as can be seen by comparing mixed Wirtinger derivatives:

$$\begin{aligned}\frac{\partial}{\partial \overline{z_k}} P_A P_{\mathcal{M}}(z)_i &= -\frac{1}{2}(AA^\dagger)_{i,k}|y_k|\frac{z_k}{|z_k|\overline{z_k}},\\ \frac{\partial}{\partial \overline{z_i}} P_A P_{\mathcal{M}}(z)_k &= -\frac{1}{2}(AA^\dagger)_{k,i}|y_i|\frac{z_i}{|z_i|\overline{z_i}}\\ &= -\frac{1}{2}\overline{(AA^\dagger)}_{i,k}|y_i|\frac{z_i}{|z_i|\overline{z_i}},\end{aligned}$$

so $\frac{\partial}{\partial \overline{z_k}} P_A P_{\mathcal{M}}(z)_i \neq \frac{\partial}{\partial \overline{z_i}} P_A P_{\mathcal{M}}(z)_k$.

[In the real case, the derivative of the sign function is zero.]

B. Proof of Lemma 1

The condition for a fixed point of RRR, or a critical point of $f(z)$ in the real case, is equivalent to

$$2P_A P_{\mathcal{M}}(z) - P_{\mathcal{M}}(z) - P_A(z) = 0,$$

which after applying P_A and $I - P_A$ to both sides yields $P_A(z) = P_A P_{\mathcal{M}}(z)$ and $P_{\mathcal{M}}(z) = P_A P_{\mathcal{M}}(z)$, respectively, so $P_A(z) = P_{\mathcal{M}}(z)$ and z corresponds to a solution. Conversely, if z^* corresponds to a solution then $2P_A P_{\mathcal{M}}(z) = P_{\mathcal{M}}(z) + P_A(z)$ trivially.

The proof for HIO is almost exactly the same.

C. Proof of Lemma 2

If $z^* = y^* + w$ with y^*, w as hypothesized then $P_A(z^*) = y^*$ and $P_{\mathcal{M}}(z^*) = P_{\mathcal{M}}(y^*) = y^*$ as either $\text{phase}(y_i^* + w_i) = \text{phase}((|(Ax^*)_i| + |w_i|)\text{phase}(y_i^*)) = \text{phase}(y_i^*)$ or $\text{phase}(y_i^* + w_i) = \text{phase}((|(Ax^*)_i| - |w_i|)\text{phase}(y_i^*)) = \text{phase}(y_i^*)$ so $\text{phase}(z^*) = \text{phase}(y^*)$, and hence $P_A(z^*) = P_{\mathcal{M}}(z^*)$.

Conversely, if z^* corresponds to a solution, write $z^* = P_A(z^*) + P_A^c(z^*)$ and note that $P_{\mathcal{M}}(z^*) = P_A(z^*) = P_{\mathcal{M}} P_A(z^*)$ and hence $\text{phase}(P_A(z^*)_i + P_A^c(z^*)_i) = \text{phase}(P_A(z^*)_i)$, so either $\text{phase}(P_A^c(z^*)_i) = \text{phase}(P_A(z^*)_i)$ or $\text{phase}(P_A^c(z^*)_i) = -\text{phase}(P_A(z^*)_i)$ and $|P_A^c(z^*)_i| < |P_A(z^*)_i|$, as desired.

D. Proof of Lemma 3

Note that if $\text{sign}(z_j) \neq \text{sign}(z_j^*)$ for any j, then

$$||z - z^*||_2^2 = \sum_i |z_i - z_i^*|^2 \geq (|z_j| + |z_j^*|)^2 \geq |z_j^*|^2,$$

so if $||z - z^*||_2 < d$ we must have $\text{sign}(z_i) = \text{sign}(z_i^*)$ for all i and hence $P_{\mathcal{M}}(z) = P_{\mathcal{M}}(z^*) = P_A(z^*)$. Hence in this ℓ_2 ball we have

$$f(z) = \frac{1}{2}\left(||z - P_A(z^*)||_2^2 - ||(I - P_A)(z)||_2^2\right),$$

so $f(z)$ is infinitely differentiable with $\nabla f(z) = P_A(z - z^*)$ and $\nabla^2 f(z) = AA^\dagger \succeq 0$, so $f(z)$ is convex. Furthermore, when restricted to $\text{col}(A)$ all the eigenvalues of $AA^\dagger$ are 1 as it is a projection matrix onto $\text{col}(A)$, so $f(z)|_{\text{col}(A)}$ is 1-strongly convex.

If $||z - z^*||_2 < d$ and $\eta \in (0, 2)$, then

$$z^+ = (1 - \eta)P_A(z) + \eta P_A(z^*) + P_A^c(z)$$

so

$$\begin{aligned}||z^+ - z^*||_2^2 &= (1 - \eta)^2||P_A(z - z^*)||_2^2 + ||P_A^c(z - z^*)||_2^2\\ &< ||P_A(z - z^*)||_2^2 + ||P_A^c(z - z^*)||_2^2\\ &= ||z - z^*||_2^2\\ &< d.\end{aligned}$$

This implies that if we initialize z^0 such that $||z^0 - z^*||_2 < d$, and use constant step size $\eta \in (0, 2)$, then $z^t = (1-\eta)^t P_A(z^0 - z^*) + P_A(z^*) + P_A^c(z^0)$ so $z_\infty = \lim_{t\to\infty} z^t = P_A(z^*) + P_A^c(z) = P_{\mathcal{M}}(z^*) + P_A^c(z)$. Note that z_∞ corresponds to a solution by Lemma 1 and the fact that $\nabla f(z_\infty) = 0$. Also note that if $\eta = 1$, RRR converges to z_∞ in one iteration.

E. Proof of Lemma 4

For $\eta > \max_i |z_i/d_i|$, we have $P_{\mathcal{M}}(z - \eta d) = P_{\mathcal{M}}(-\eta d) = -P_{\mathcal{M}}(d)$. Then,

$$\begin{aligned}||(z - \eta z) - P_A P_{\mathcal{M}}(z - \eta z)||_2^2 &= ||z - \eta d + P_A P_{\mathcal{M}}(d)||_2^2\\ &= \eta^2||d||_2^2 + ||z + P_A P_{\mathcal{M}}(d)||_2^2 - 2\eta\langle d, z + P_A P_{\mathcal{M}}(d)\rangle,\end{aligned}$$

where the second term is a constant with respect to η. Similarly, since P_A is linear,

$$\begin{aligned}&||(z-\eta d)-P_A(z-\eta d)||_2^2 = ||z-\eta d - P_A(z)+\eta P_A(d)||_2^2\\ &= \eta^2||(I-AA^\dagger)d||_2^2 + ||z-P_A(z)||_2^2 - 2\eta\langle d, z-P_A(z)\rangle,\end{aligned}$$

and

$$\begin{aligned}&||z-\eta d - P_{\mathcal{M}}(z-\eta d)||_2^2 = ||z-\eta d + P_{\mathcal{M}}(d)||_2^2\\ &= \eta^2||d||_2^2 + ||z+P_{\mathcal{M}}(d)||_2^2 - 2\eta\langle d, z+P_{\mathcal{M}}(d)\rangle,\end{aligned}$$

so putting everything together:

$$\begin{aligned}f(z-\eta d) = &\frac{1}{2}\eta^2\|P_A(d)\|_2^2 - \eta\langle d, P_A(z+2P_{\mathcal{M}}(d)) - P_{\mathcal{M}}(z)\rangle\\ &+ c\end{aligned}$$

where

$$c = ||z+P_A P_{\mathcal{M}}(d)||_2^2 - \frac{1}{2}\left(||z-P_A(z)||_2^2 + ||z+P_{\mathcal{M}}(d)||_2^2\right)$$

is independent of η.

If $P_A(d)\neq 0$ then $\lim_{\eta\to\infty} f(z-\eta d)=\infty$. If $P_A(d)=0$ then $f(z-\eta d) = \eta\langle d, P_{\mathcal{M}}(z)\rangle + c$, so if $\langle d, P_{\mathcal{M}}(z)\rangle > 0$ we again have $\lim_{\eta\to\infty} f(z-\eta d)=\infty$.

F. Proof of Corollary 1

For RRR, we have

$$d = \nabla f(z) = P_A(z) + P_{\mathcal{M}}(z) - 2P_A P_{\mathcal{M}}(z),$$

then

$$P_A(d) = P_A(z) - P_A P_{\mathcal{M}}(z).$$

Therefore, $P_A(d)=0$ implies $\nabla f(z) = P_A^c P_{\mathcal{M}}(z)$ and hence $\langle \nabla f(z), P_{\mathcal{M}}(z)\rangle = ||P_A^c P_{\mathcal{M}}(z)||_2^2 \geq 0$. If $\langle \nabla f(z), P_{\mathcal{M}}(z)\rangle = 0$ then $P_{\mathcal{M}}(z) = P_A P_{\mathcal{M}}(z) = P_A(z)$, so in fact $\nabla f(z) = 0$ and $z\in \mathrm{col}(A)\cap\mathcal{M}$ is a solution.

For HIO, let $z\mapsto z + P_A(P_{\mathcal{M}}(z)-z)$ and $d\mapsto P_A^c P_{\mathcal{M}}(z)$ in Lemma 4, and note that $P_A(d)=0$ and $\langle d, P_{\mathcal{M}}(z)\rangle = ||P_A^c P_{\mathcal{M}}(z)||_2^2 = 0$ if and only if $P_{\mathcal{M}}(z) = P_A P_{\mathcal{M}}(z)\in \mathrm{col}(A)$, so $P_{\mathcal{M}}(z)\in\mathrm{col}(A)\cap\mathcal{M}$.

G. Proof of Lemma 5

Since $\nabla f(z) = P_A(z)+P_{\mathcal{M}}(z) - 2P_A P_{\mathcal{M}}(z)$, we have

$$\begin{aligned}&\langle -\nabla f(z), s(z)Ax^* - z\rangle\\ &= \langle P_A(P_{\mathcal{M}}(z)-z) - (I-P_A)P_{\mathcal{M}}(z), s(z)Ax^*-z\rangle\\ &= \langle P_{\mathcal{M}}(z), s(z)Ax^*\rangle - |\langle z, Ax^*\rangle| - 2\langle P_{\mathcal{M}}(z), P_A(z)\rangle\\ &+\langle P_{\mathcal{M}}(z), z\rangle + \langle z, P_A(z)\rangle.\end{aligned}$$

We now proceed term by term. First,

$$\begin{aligned}&\mathbb{E}_{z\sim\mathcal{N}(0,\sigma^2 I)}\left[\langle P_{\mathcal{M}}(z), s(z)Ax^*\rangle\right]\\ &= \sum_{i=1}^m (Ax^*)_i|(Ax^*)_i|\mathbb{E}[\mathrm{sign}(z_i)\mathrm{sign}(\langle z, Ax^*\rangle)],\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}[\mathrm{sign}(z_i)\mathrm{sign}(\langle z, Ax^*\rangle)]\\ &= \mathbb{P}(\langle z, Ax^*\rangle > 0,\ z_i>0) + \mathbb{P}(\langle z, Ax^*\rangle<0,\ z_i<0)\\ &- \mathbb{P}(\langle z, Ax^*\rangle>0,\ z_i<0) - \mathbb{P}(\langle z, Ax^*\rangle<0,\ z_i>0).\end{aligned}$$

Writing $\langle z, Ax^*\rangle = \sum_j (Ax^*)_j z_j = y_i + (Ax^*)_i z_i$ where $y_i = \sum_{j\neq i}(Ax^*)_j z_j$ is independent of z_i, and noting that $y_i\sim\mathcal{N}(0,||y^{(i)}||_2^2)$ where $y^{(i)}\in\mathbb{R}^{m-1}$ is obtained from Ax^* by deleting the ith entry, we have

$$\begin{aligned}&\mathbb{P}((Ax^*)_i z_i + y_i < 0,\ z_i<0) = \mathbb{P}((Ax^*)_i z_i + y_i>0,\ z_i>0)\\ &= \frac{1}{2\pi\sigma||y^{(i)}||_2}\int_0^\infty\int_{-(Ax^*)_i z_i}^\infty e^{-z_i^2/2\sigma^2}e^{-y_i^2/2||y^{(i)}||_2^2}\,dy_i\,dz_i\\ &= \frac{1}{4}+\frac{1}{2\pi}\tan^{-1}(\sigma(Ax^*)_i/||y^{(i)}||_2)\\ &= \frac{1}{4}+\frac{\mathrm{sign}(Ax^*)_i}{2\pi}\tan^{-1}(\sigma|Ax^*|_i/||y^{(i)}||_2),\end{aligned}$$

$$\begin{aligned}&\mathbb{P}(z_i+y_i>0,\ z_i<0) = \mathbb{P}(z_i+y_i<0,\ z_i>0)\\ &= \frac{1}{2\pi\sigma||y^{(i)}||_2}\int_0^\infty\int_{-\infty}^{-(Ax^*)_i z_i} e^{-z_i^2/2\sigma^2}e^{-y_i^2/2||y^{(i)}||_2^2}\,dy_i\,dz_i\\ &= \frac{1}{4}-\frac{\mathrm{sign}(Ax^*)_i}{2\pi}\tan^{-1}(\sigma|Ax^*|_i/||y^{(i)}||_2),\end{aligned}$$

so

$$\mathbb{E}[\mathrm{sign}(z_i)\mathrm{sign}(\langle z, Ax^*\rangle)] = \frac{2}{\pi}\mathrm{sign}(Ax^*)_i\tan^{-1}(\sigma|Ax^*|_i/||y^{(i)}||_2),$$

and hence

$$\mathbb{E}\left[\langle P_{\mathcal{M}}(z), s(z)Ax^*\rangle\right] = \frac{2}{\pi}\sum_{i=1}^m |Ax^*|_i^2\tan^{-1}(\sigma|Ax^*|_i/||y^{(i)}||_2).$$

Next,

$$\begin{aligned}\mathbb{E}[\langle P_{\mathcal{M}}(z), P_A(z)\rangle] &= \sum_{i=1}^m |Ax^*|_i\sum_{j=1}^m (AA^\dagger)_{i,j}\mathbb{E}[\mathrm{sign}(z_i)z_j]\\ &= \sigma\sqrt{\frac{2}{\pi}}\sum_{i=1}^m |Ax^*|_i(AA^\dagger)_{i,i},\end{aligned}$$

and

$$\mathbb{E}\left[\langle P_{\mathcal{M}}(z), z\rangle\right] = \sum_{i=1}^m |Ax^*|_i\mathbb{E}[|z_i|] = \sigma\sqrt{\frac{2}{\pi}}\sum_{i=1}^m |Ax^*|_i$$

and since $\langle z, Ax^*\rangle\sim\mathcal{N}(0,\sigma^2||Ax^*||_2^2)$,

$$\mathbb{E}[|\langle z, Ax^*\rangle|] = \sigma||Ax^*||_2\sqrt{\frac{2}{\pi}}.$$

Finally,

$$\mathbb{E}[\langle z, P_A(z)\rangle] = \sigma^2\mathrm{Tr}(AA^\dagger) = \sigma^2 n,$$

as $AA^\dagger$ is a projection matrix onto an n-dimensional subspace $\mathrm{col}(A)$.

Putting everything together,

$$\begin{aligned}E &= \mathbb{E}_{z\sim\mathcal{N}(0,\sigma^2 I)}\left[\langle -\nabla f(z), s(z)Ax^*-z\rangle\right]\\ &= \frac{2}{\pi}\sum_{i=1}^m |Ax^*|_i^2\tan^{-1}(\sigma|Ax^*|_i/||y^{(i)}||_2) - \sigma\sqrt{\frac{2}{\pi}}||Ax^*||_2\\ &- \sigma\sqrt{\frac{2}{\pi}}\sum_{i=1}^m\Big(2(AA^\dagger)_{i,i}-1\Big)|Ax^*|_i + \sigma^2 n.\end{aligned}$$

Since E depends only on $|Ax^*|$, it is computable in practice. Since the dominant term in E as $|Ax^*|$ grows is a positive quadratic, we conclude that for large enough α, the renormalization $|Ax^*| \mapsto \alpha|Ax^*|$ makes $E > 0$.

It is perhaps more meaningful to consider the quantity

$$E_A = \mathbb{E}_{z\sim\mathcal{N}(0,\sigma^2 I)}\left[\langle -P_A(\nabla f(z)), s(z)Ax^* - P_A(z)\rangle\right],$$

which measures the inner product between the projections onto $\text{col}(A)$ of the different vectors. In $\text{col}(A)$ there are only two points corresponding to solutions, namely $\pm Ax^*$, whereas in $\mathbb{R}^m$ any point $\pm Ax^* + w$ with $w \in \text{col}(A)^\perp$ corresponds to a solution. In that case

$$\begin{aligned} E_A &= \mathbb{E}[\langle P_{\mathcal{M}}(z), s(z)Ax^*\rangle] - \mathbb{E}[|\langle z, Ax^*\rangle|] \\ &\quad - \mathbb{E}[\langle P_{\mathcal{M}}(z), P_A(z)\rangle] + \mathbb{E}[\langle z, P_A(z)\rangle] \\ &= \frac{2}{\pi}\sum_{i=1}^{m} |Ax^*|_i^2 \tan^{-1}(\sigma|Ax^*|_i/||y^{(i)}||_2) \\ &\quad - \sigma\sqrt{\frac{2}{\pi}}||Ax^*||_2 - \sigma\sqrt{\frac{2}{\pi}}\sum_{i=1}^{m} |Ax^*|_i (AA^\dagger)_{i,i} + \sigma^2 n, \end{aligned}$$

which satisfies similar properties.

H. Proof of Lemma 6

Note that

$$||\nabla f(z)||_2^2 = ||P_{\mathcal{M}}(z) - P_A P_{\mathcal{M}}(z)||_2^2 + ||P_A(z) - P_A P_{\mathcal{M}}(z)||_2^2,$$

so $||P_{\mathcal{M}}(z) - P_A P_{\mathcal{M}}(z)||_2 \le ||\nabla f(z)||_2$ and $||P_A(z) - P_A P_{\mathcal{M}}(z)||_2 \le ||\nabla f(z)||_2$. Then note that $||P_{\mathcal{M}}(z) - P_A P_{\mathcal{M}}(z)||_2$ depends only on the signs of z, and hence takes at most 2^m values, one of which is zero. Therefore, there exists ϵ_1 such that if $||P_{\mathcal{M}}(z) - P_A P_{\mathcal{M}}(z)||_2 < \epsilon_1$ then in fact $P_{\mathcal{M}}(z) = P_A P_{\mathcal{M}}(z)$ (namely, the second-to-smallest value in its value set) and so $P_{\mathcal{M}}(z) \in \text{col}(A) \cap \mathcal{M}$. Taking $\epsilon \le \epsilon_1$, we then have

$$\begin{aligned} ||P_A(z) - P_A P_{\mathcal{M}}(z)||_2^2 &= ||P_A(z) - P_{\mathcal{M}}(z)||_2^2 \\ &= ||P_A(z) - P_{\mathcal{M}}\Big(P_A(z) + P_A^c(z)\Big)||_2^2 \\ &< \epsilon. \end{aligned}$$

For general vectors $x, y \in \mathbb{R}^m$, note that if $\text{sign}(x_j + y_j) \neq \text{sign}(x_j)$ for any $1 \le j \le m$, then

$$\begin{aligned} ||x - P_{\mathcal{M}}(x+y)||_2^2 &= \sum_{i=1}^{m} \Big| |x_i|\text{sign}(x_i) - |(Ax^*)_i|\text{sign}(x_i + y_i)\Big|^2 \\ &\ge \Big| |x_j| + |(Ax^*)_j| \Big|^2 \ge |(Ax^*)_j|^2, \end{aligned}$$

so $||x - P_{\mathcal{M}}(x+y)||_2 \ge |(Ax^*)_j|$. Hence, if we choose $\epsilon < \min(\epsilon_1, |(Ax^*)_1|, \ldots, |(Ax^*)_m|)$ then we must have $\text{sign}(z_i) = \text{sign}(P_A(z)_i + P_A^c(z)_i) = \text{sign}(P_A(z)_i)$ for all i so $P_{\mathcal{M}}(z) = P_{\mathcal{M}} P_A(z) = P_A P_{\mathcal{M}}(z)$.

Let $w = P_A(z) - P_{\mathcal{M}}(z)$, and note that $||w||_2 < \epsilon$ so $|w_i| < \epsilon$ for all i. Let $z^* = P_{\mathcal{M}}(z) + u + P_A^c(z)$ where $u \in \text{col}(A)^\perp$ is a small perturbation. We will show that there exists u with small $||u||_2$ such that z^* corresponds to a solution and is close to z. Clearly, $P_A(z^*) = P_{\mathcal{M}}(z) \in \mathcal{M}$. We need to show that $P_{\mathcal{M}}(z^*) = P_{\mathcal{M}}(z)$, or equivalently, $\text{sign}(P_{\mathcal{M}}(z) + u + P_A^c(z)) = \text{sign}(z - w + u) = \text{sign}(z)$. Here we shall focus on u of the form $u = -\alpha P_A^c(z)$ for small $\alpha \in (0, 1)$, so $z^* = P_{\mathcal{M}}(z) + (1-\alpha)P_A^c(z)$. Let

$$\begin{aligned} I = \{i \in \{1, \ldots, m\} :\ &\text{sign}(P_A^c(z)_i) \neq \text{sign}(z_i), \\ &\text{sign}(w_i) = \text{sign}(z_i),\ |P_A^c(z)_i| \ge |(Ax^*)_i|\}, \end{aligned}$$

and note that if $i \notin I$ then either:

- $\text{sign}(P_A^c(z)_i) = \text{sign}(z_i)$: in which case
$$\begin{aligned} \text{sign}(z^*) &= \text{sign}(P_{\mathcal{M}}(z)_i + (1-\alpha)P_A^c(z)_i) \\ &= \text{sign}\left[(|(Ax^*)_i| + (1-\alpha)|P_A^c(z)_i|)\text{sign}(z_i)\right]; \\ &= \text{sign}(z_i) \end{aligned}$$
- $\text{sign}(P_A^c(z)_i) = -\text{sign}(z_i)$ and $\text{sign}(w_i) = -\text{sign}(z_i)$: in which case
$$\begin{aligned} \text{sign}(z^*) &= \text{sign}(z_i - w_i - \alpha P_A^c(z)) \\ &= \text{sign}\left[(|z_i| + |w_i| - \alpha|P_A^c(z)|)\text{sign}(z_i)\right] \\ &= \text{sign}(z_i) \end{aligned}$$
as $|z_i| \ge |P_A^c(z)_i| > \alpha|P_A^c(z)_i|$;
- $|P_A^c(z)_i| < |(Ax^*)_i|$: in which case
$$\begin{aligned} &\text{sign}(P_{\mathcal{M}}(z)_i + (1-\alpha)P_A^c(z)_i) \\ &= \text{sign}[(|(Ax^*)_i| \pm (1-\alpha)|P_A^c(z)|)\text{sign}(z_i)] \\ &= \text{sign}(z_i). \end{aligned}$$

Thus, in either case $\text{sign}(P_{\mathcal{M}}(z)_i + u_i + P_A^c(z)_i) = \text{sign}(z_i)$.

If $i \in I$, note first that $P_{\mathcal{M}} P_A(z) = P_{\mathcal{M}}(z)$ implies $\text{sign}(z_i) = \text{sign}(P_{\mathcal{M}}(z)_i + w_i + P_A^c(z)_i)$, and for $i \in I$ we get $\text{sign}(z_i) = \text{sign}\left[(|(Ax^*)_i| + |w_i| - |P_A^c(z)_i|)\text{sign}(z_i)\right]$ so $|P_A^c(z)_i| < |(Ax^*)_i| + |w_i| < |(Ax^*)_i| + \epsilon$, and hence

$$|(Ax^*)_i| \le |P_A^c(z)_i| < |(Ax^*)_i| + \epsilon, \quad \forall i \in I.$$

Letting $d = \min_i |(Ax^*)_i| > 0$ and $\alpha = \epsilon/d < 1$ so $z^* = P_{\mathcal{M}}(z) + (1 - \frac{\epsilon}{d})P_A^c(z)$, note that if $i \in I$ then

$$\begin{aligned} &\text{sign}\left[P_{\mathcal{M}}(z) + (1 - \frac{\epsilon}{d})P_A^c(z)\right] \\ &= \text{sign}\left[(|(Ax^*)_i| - (1 - \epsilon/d)|P_A^c(z)_i|)\text{sign}(z_i)\right] \\ &= \text{sign}(z_i) \end{aligned}$$

as

$$\begin{aligned} |(Ax^*)_i| - |P_A^c(z)_i| + \epsilon(|P_A^c(z)_i|/d) &> \epsilon\left(|P_A^c(z)_i/d - 1\right) \\ &\ge 0. \end{aligned}$$

If $i \notin I$, then a similar result was shown above. Hence $P_{\mathcal{M}}(z^*) = P_{\mathcal{M}}(z) = P_A(z^*)$ so z^* corresponds to a solution, and

$$||z - z^*||_2 \le ||w||_2 + \frac{\epsilon}{d}||P_A^c(z)||_2 < \epsilon + \frac{\epsilon}{d}||P_A^c(z)||_2.$$

If $\min_i |z_i| \ge \epsilon$, we must have $I = \emptyset$, as if $i \in I$ then $|z_i| = |P_{\mathcal{M}}(z)_i + w_i + P_A^c(z)_i| = |(Ax^*)_i| + |w_i| - |P_A^c(z)_i| \le |w_i| < \epsilon$, a contradiction. In that case we may set $\alpha = 0$ in the above and conclude that $z^* = P_{\mathcal{M}}(z) + P_A^c(z)$ corresponds to a solution and $||z - z^*||_2 = ||w||_2 < \epsilon$.

REFERENCES

[1] A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A. Nelson. Saving phase: Injectivity and stability for phase retrieval. *Applied and Computational Harmonic Analysis*, 37(1):106–125, 2014.

[2] B. Baykal. Blind channel estimation via combining autocorrelation and blind phase estimation. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 51(6):1125–1131, 2004.

[3] T. Bendory, R. Beinert, and Y. C. Eldar. Fourier phase retrieval: Uniqueness and algorithms. In *Compressed Sensing and its Applications*, pages 55–91. Springer, 2017.

[4] T. Bendory, N. Boumal, C. Ma, Z. Zhao, and A. Singer. Bispectrum inversion with application to multireference alignment. *IEEE Transactions on Signal Processing*, 66(4):1037–1050, 2017.

[5] A. Conca, D. Edidin, M. Hering, and C. Vinzant. An algebraic characterization of injectivity in phase retrieval. *Applied and Computational Harmonic Analysis*, 38(2):346–356, 2015.

[6] V. Elser, T.-Y. Lan, and T. Bendory. Benchmark problems for phase retrieval. *arXiv preprint arXiv:1706.00399*, 2017.

[7] C. Fienup and J. Dainty. Phase retrieval and image reconstruction for astronomy. *Image Recovery: Theory and Application*, 231:275, 1987.

[8] R. Lawrence. *Fundamentals of speech recognition*. Pearson Education India, 2008.

[9] J. Li and T. Zhou. On relaxed averaged alternating reflections (raar) algorithm for phase retrieval with structured illumination. *Inverse Problems*, 33(2):025012, 2017.

[10] D. R. Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse problems*, 21(1):37, 2004.

[11] S. Marchesini. Phase retrieval and saddle-point optimization. *JOSA A*, 24(10):3289–3296, 2007.

[12] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

[13] R. Trebino. *Frequency-resolved optical gating: the measurement of ultrashort laser pulses*. Springer Science & Business Media, 2012.

[14] A. Walther. The question of phase retrieval in optics. *Optica Acta: International Journal of Optics*, 10(1):41–49, 1963.