Relaxed-Reflect-Reflect (RRR) for Phase Retrieval

Eitan Levin (netID: eitanl)
Advised by Tamir Bendory, PACM

November 1, 2019

1 Problem Statement

We consider the problem of phase retrieval: Given a sensing matrix $A \in \mathbb{C}^{m\times n}$ and a ground truth vector $x^* \in \mathbb{C}^n$, we measure $|Ax^*|$ and seek a point z in the intersection $z \in \mathcal{M} \cap \mathrm{col}(A)$ where $\mathcal{M} = \{z \in \mathbb{C}^m : \ |z| = |Ax^*|\}$. The sensing matrix is usually taken to be either i.i.d. Gaussian or an oversampled DFT matrix [2], the latter being especially relevant for applications in optics and crystallography [6, 9]. The intersection $\mathcal{M} \cap \mathrm{col}(A)$ is never singleton, since if $z^* \in \mathrm{col}(A) \cap \mathcal{M}$ then $e^{i\theta} z^* \in \mathrm{col}(A) \cap \mathcal{M}$ for any global phase θ. It has been shown however that $\mathrm{col}(A) \cap \mathcal{M}$ is singleton modulo global phase in various settings [1, 2, 3].

The naive approach to this feasibility problem, called Grechberg-Saxton (GS) in the phase retrieval literature, is alternating projections [6]. Define the projection $P_A(z) = AA^\dagger z$ onto $\mathrm{col}(A)$, where $A^\dagger$ is the pseudoinverse of A, and the projection $P_{\mathcal{M}}(z)_i = |(Ax^*)_i| \frac{z_i}{|z_i|}$ onto M. The GS algorithm then iterates $z^{t+1} = P_A P_{\mathcal{M}}(z^t)$. Unfortunately, this is known to quickly converge to suboptimal local minima, and in practice is only used to refine a solution [6, 10].

Instead, several algorithms are used in practice that are all closely related [10, 6]. The two most ones are Fineup's Hybrid Input-Output (HIO) [7] which iterates

$$z^{t+1} = P_A P_{\mathcal{M}}(z^t) + (I - P_A)(I - \eta_1 P_{\mathcal{M}})(z^t),$$

with constant $\eta_1 \in [0.5, 1]$, and Relaxed-Reflect-Reflect (RRR) [4, 5] which iterates

$$z^{t+1} = (1 - \eta_2)z + \eta_2\Big(P_A P_{\mathcal{M}}(z) + (I - P_A)(I - P_{\mathcal{M}})(z)\Big),$$

with $\eta_2 \in (0, 2)$. Notably, when $\eta_1 = \eta_2 = 1$ both algorithms coincide with Douglas-Rachford splitting applied to $F(z) = \mathbb{1}_{\mathrm{col}(A)} + \mathbb{1}_{\mathcal{M}}$, where $\mathbb{1}_C$ is an indicator function for the set C. These and other variants of Douglas-Rachford have been observed in practice to consistently converge to a solution in the intersection $\mathrm{col}(A) \cap \mathcal{M}$ (at least in the absence of noise) [6]. We therefore attempt to gain some insight into the performance of these algorithms and how they could be improved.

2 Basic Analysis

When $\eta_1 = \eta_2 = 1$, our iterations become $z^{t+1} = P_A P_{\mathcal{M}}(z^t) + P_A^c P_{\mathcal{M}}^c(z)$ where $P_A^c = I - P_A$ is projection onto $\mathrm{col}(A)^\perp$ and in analogy $P_{\mathcal{M}}^c = I - P_{\mathcal{M}}$. This is a sum of two orthogonal vectors - one in $\mathrm{col}(A)$ coinciding with the GS iterations, and one in $\mathrm{col}(A)^\perp$ which evidently makes the GS iterations convergent. This additional component is described in connection to HIO as "negative feedback" in $\mathrm{col}(A)^\perp$, or alternatively in terms of a saddle point problem [10].

To analyze the above algorithms, we introduce the cost function

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

which, when A, z are real and $z_i \neq 0$ for any i, $f(z)$ is differentiable with $\nabla f(z) = P_A(z) + P_{\mathcal{M}}(z) - 2P_A P_{\mathcal{M}}(z)$ so the RRR iterations become $z^{t+1} = z^t - \eta \nabla f(z^t)$, i.e. (sub)gradient descent. Unfortunately, in the complex case it can be shown that the RRR (and HIO) iterations are not gradients (see Appendix A.1), but nonetheless we observe that our cost $f(z)$ is still useful - see Sect. 3.

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

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 A.2 □

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 A.3. □

Theorem 1. ([8, 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 A.4. □

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 [10]. 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 A.5. □

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 A.6. □

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 $s(z) = sign(\langle z, Ax^*\rangle)$. 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 $z^* \in$ col(A) $\cap\, \alpha\mathcal{M}$ with $\alpha > 0$, and letting the solution be z^*/α, we have $E > 0$.*

Proof. See Appendix A.7. □

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 $P_{\mathcal{M}}(z) \in col(A) \cap \mathcal{M}$ 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 A.8. □

3 Improvements

We choose the RRR parameter $\eta \in (0, 2)$ using backtracking line search on our cost function $f(z)$. Since $f(z)$ can be negative whereas $f(z^*) = 0$ for any z^* corresponding to a solution, we choose η to minimize $|f(z)|$ instead. Since our function is nonconvex, we observe that at some points we cannot get reasonable decrease in our function, if at all, so we detect these points by setting a lower bound on the step size we estimate, and if this lower bound is met we set the step size to $\eta = 1$, which is large enough to escape the bad point and optimal when near a solution by Lemma 3. Our algorithm is given in Algorithm 1. This method can be applied even in the complex case, even though the expression we use for $\nabla f(z)$ is no longer the gradient of $f(z)$.

As is evident from Lemmas 1,6, RRR does not terminate when $P_{\mathcal{M}}(z) \in \text{col}(A) \cap \mathcal{M}$ or when $P_A(z) \in \text{col}(A) \cap \mathcal{M}$, and from numerical experiments it can continue for arbitrarily many iterations when initialized from such points (when these points are sufficiently far from solutions as per Lemma 3). In practice for the real case, we observe that $P_{\mathcal{M}}(z)$ becomes a solution before convergence, while $P_A(z)$ does not. We hypothesize that this is cause by the expression of $||\nabla f(z)||_2^2$ - see Appendix A.8 - which depends on $||P_{\mathcal{M}}(z) - P_A P_{\mathcal{M}}(z)||_2^2$ measuring how close $P_{\mathcal{M}}(z)$ is to a solution, but not on $||P_A(z) - P_{\mathcal{M}} P_A(z)||_2^2$ which measures a similar quantity for $P_A(z)$. We therefore check the former norm to see if $P_{\mathcal{M}}(z)$ is close to being a solution, in which case we replace $z \mapsto P_{\mathcal{M}}(z)$. This check does not increase our cost-per-iteration as we compute the relevant quantity for $\nabla f(z)$ anyway.

We show the results of our algorithm on both real and complex problems with random Gaussian sensing matrices in Figs. **??**, **??**, where we get an improvement of at least a factor of 2 in iteration count with a similar cost-per-iteration.

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] 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.

[3] 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.

[4] V. Elser. Matrix product constraints by projection methods. *Journal of Global Optimization*, 68(2):329–355, 2017.

[5] V. Elser. The complexity of bit retrieval. *IEEE Transactions on Information Theory*, 64(1):412–428, 2018.

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

[7] J. R. Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, 1982.

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

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

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

A Appendix - Proofs

A.1 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 [10]. However, $P_A P_{\mathcal{M}}(z)$ is not a gradient, as can be seen by comparing mixed Wirtinger derivatives:

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

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.]

A.2 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.

A.3 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.

A.4 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.

A.5 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:

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

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 $\lim_{\eta\to\infty} f(z - \eta d) = \infty$.

A.6 Corollary 1

For RRR, we have $d = \nabla f(z) = P_A(z) + P_{\mathcal{M}}(z) - 2P_A P_{\mathcal{M}}(z)$, then $P_A(d) = P_A(z) - P_A P_{\mathcal{M}}(z)$, so $P_A(d) = 0$ 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 $P_{\mathcal{M}}(z) \in \mathrm{col}(A) \cap \mathcal{M}$.

A.7 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,

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

and

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

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,

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

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,

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

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

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

which satisfies similar properties.

A.8 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 $P_{\mathcal{M}}(z) \in \mathrm{col}(A) \cap \mathcal{M}$. Taking $\epsilon \leq \epsilon_1$, we then have

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

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

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

so $||x - P_{\mathcal{M}}(x+y)||_2 \geq |(Ax^*)_j|$. Hence, if we choose $\epsilon < \min(\epsilon_1, |(Ax^*)_1|, \ldots, |(Ax^*)_m|)$ then we must have $\mathrm{sign}(z_i) = \mathrm{sign}(P_A(z)_i + P_A^c(z)_i) = \mathrm{sign}(P_A(z)_i)$ 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 $u \in \mathrm{col}(A)^{\perp}$ 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, $\mathrm{sign}(P_{\mathcal{M}}(z) + u + P_A^c(z)) = \mathrm{sign}(z - w + u) = \mathrm{sign}(z)$. 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

$$I = \{i \in \{1, \ldots, m\} : \mathrm{sign}(P_A^c(z)_i) \neq \mathrm{sign}(z_i),\ \mathrm{sign}(w_i) = \mathrm{sign}(z_i),\ |P_A^c(z)_i| \geq |(Ax^*)_i|\},$$

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

- $\mathrm{sign}(P_A^c(z)_i) = \mathrm{sign}(z_i)$: in which case $\mathrm{sign}(z^*) = \mathrm{sign}(P_{\mathcal{M}}(z)_i + (1 - \alpha) P_A^c(z)_i) = \mathrm{sign}\left[(|(Ax^*)_i| + (1-\alpha)|P_A^c(z)_i|)\mathrm{sign}(z_i)\right] = \mathrm{sign}(z_i)$,

- $\mathrm{sign}(P_A^c(z)_i) = -\mathrm{sign}(z_i)$ and $\mathrm{sign}(w_i) = -\mathrm{sign}(z_i)$: in which case $\mathrm{sign}(z^*) = \mathrm{sign}(z_i - w_i - \alpha P_A^c(z)) = \mathrm{sign}\left[(|z_i| + |w_i| - \alpha|P_A^c(z)|)\mathrm{sign}(z_i)\right] = \mathrm{sign}(z_i)$ as $|z_i| \geq |P_A^c(z)_i| > \alpha|P_A^c(z)_i|$,

- $|P_A^c(z)_i| < |(Ax^*)_i|$: in which case $\mathrm{sign}(P_{\mathcal{M}}(z)_i + (1 - \alpha) P_A^c(z)_i) = \mathrm{sign}[(|(Ax^*)_i| \pm (1 - \alpha)|P_A^c(z)|)\mathrm{sign}(z_i)] = \mathrm{sign}(z_i)$.

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

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

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

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

as $|(Ax^*)_i| - |P_A^c(z)_i| + \epsilon(|P_A^c(z)_i|/d) > \epsilon\,(|P_A^c(z)_i/d - 1) \geq 0$. 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

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

If $\min_i |z_i| \geq \epsilon$, we must have $I = \emptyset$, as if $i \in I$ then $|z_i| = |P_{\mathcal{M}}(z)_i + w_i + P_A^c(z)_i| = |(Ax^*)_i| + |w_i| - |P_A^c(z)_i| \leq |w_i| < \epsilon$, 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 $||z - z^*||_2 = ||w||_2 < \epsilon$.

Algorithm 1 Line Search for RRR

Input: α, β, δ

Output: η

$\eta \leftarrow 2$ ▷ Initialization

while $|f(z - \eta\nabla f(z))| > |f(z)| - \alpha\eta||\nabla f(z)||_2^2$ and $\eta > \delta$ **do**

  $\eta \leftarrow \beta\eta$

end while

if $\eta \leq \delta$ **then**

  $\eta \leftarrow 1$

end if

In our numerical experiments, we set $\delta = 0.2$, $\beta = 0.75$ and $\alpha = 1/10$ in the real case and $\alpha = 1/100$ in the complex case.