Local saddles of relaxed averaged alternating reflections algorithms on phase retrieval

Pengwen Chen
Applied mathematics, National Chung Hsing University, Taiwan
E-mail: pengwen@nchu.edu.tw

Abstract. Phase retrieval can be expressed as a non-convex constrained optimization problem to identify one phase minimizer on a torus. Many iterative transform techniques have been proposed to identify the minimizer, e.g., relaxed averaged alternating reflections (RAAR) algorithms. In this paper, we present one optimization viewpoint on the RAAR algorithm. RAAR algorithm is one alternating direction method of multipliers (ADMM) with one penalty parameter. Pairing with multipliers (dual vectors), phase vectors on the primal space are lifted to higher dimensional vectors, RAAR algorithm is one continuation algorithm, which searches for local saddles in the primal-dual space. The dual iteration approximates one gradient ascent flow, which drives the corresponding local minimizers in a positive-definite Hessian region. Altering penalty parameters, the RAAR avoids the stagnation of these corresponding local minimizers in the primal space and thus screens out many stationary points corresponding to non-local minimizers.

Keywords: Phase retrieval, relaxed averaged alternating reflections, alternating direction method of multipliers, Nash equilibrium, local saddles

1. Introduction

Phase retrieval has recently attracted attentions in the mathematics community (see one review [1] and references therein). The problem of phase retrieval is motivated by the inability of photo detectors to directly measure the phase of an electromagnetic wave at frequencies of THz (terahertz) and higher. The problem of phase retrieval aims to reconstruct an unknown object $x_0 \in \mathbb{C}^n$ from its magnitude measurement data $b = |A^*x_0|$, where $A \in \mathbb{C}^{n \times N}$ represents some isometric matrix and $A^*$ represents the Hermitian adjoint of $A$. Introduce one non-convex $N$-dimensional torus associated with its normalized torus $\mathcal{Z} := \{ z \in \mathbb{C}^N : |z| = b \}$, $\mathcal{U} := \{ u \in \mathbb{C}^N : |u| = 1 \}$. The whole problem is equivalent to reconstructing the missing phase information $u$ and the unknown object $x = x_0$ via solving the constrained least squares problem

$$
\min_{x \in \mathbb{C}^n, |u| = 1} \left\{ \| b \odot u - A^*x \|_2^2 : u \in \mathbb{C}^N \right\} = \min_{z \in \mathcal{Z}} \| A \perp z \|_2^2, \quad (1)
$$


Local saddles of RAAR algorithms

where \(A_\perp \in \mathbb{C}^{(N-n)\times N}\) is an isometric matrix with unitary matrix \([A^*, A_\perp]\),

\[
[A^*, A_\perp][A^*, A_\perp]^* = A_\perp A_\perp + A^* A = I.
\]

Here, \(b \odot u\) represents the component-wise multiplication between two vectors \(b, u\), respectively. The isometric condition is not very restrictive in applications, since Fourier transforms are commonly applied in phase retrieval. Even for non-Fourier transforms, we can still obtain equivalent problems via a QR-factorization, see \([2]\).

Let \(U^*\) denote the set in \(U\) consisting of all the local minimizers of \((1)\). A vector \(z^* \in Z\) minimizes \((1)\) is called a global solution. In the noiseless measurement case, \(A_\perp z^* = 0\) or \(z^* = A^* x^* = b \odot u^*\) for some \(u^* \in U\) and some \(x^* \in \mathbb{C}^n\). Numerically, it is a nontrivial task to obtain a global minimizer on the non-convex torus. The error reduction method is one traditional method \([3]\), which could produce a local solution of poor quality for \((1)\), if no proper initialization is taken. During last decades, researchers propose various spectral initialization algorithms to overcome this challenge\([4, 5, 6, 7, 2, 8, 9, 10, 11, 12]\). On the other hand, phase retrieval can be also tackled by another class of algorithms, including the well-known hybrid input-output algorithm(HIO)\([13, 14]\), the hybrid projection–reflection method\([15]\), Fourier Douglas-Rachford algorithm (FDR)\([16]\), alternating direction methods\([17]\) and relaxed averaged alternating reflections(RAAR) algorithms\([18]\). An important feature of these algorithms is the empirical ability to avoid local minima and converge to a global mini-mum for noise-free oversampled diffraction patterns. For instance, the empirical study of FDR indicates the disappearance of the stagnation at poor local solutions under sufficiently many random masks. A limit point of FDR is a global solution in \((1)\) and the limit point with appropriate spectral gap conditions reconstructs the phase retrieval solution \([16]\).

Traditional convergence study on Douglas-Rachford splitting algorithm \([19, 20]\) heavily relies on the convexity assumption. Noise-free measurement is a strict requirement for HIO and FDR, which motivates the proposal of relaxed averaged alternating reflections algorithm \([18, 21]\). Let \(A, B\) denote the sets \(\text{Range}(A^*)\) and \(Z\), respectively. Let \(P_A\) and \(P_B\) denote the projector on \(A\) and \(B\), respectively. Let \(R_A, R_B\) denote the reflectors corresponding to \(A, B\). With one parameter \(\beta \in (0, 1)\) relaxing the original feasibility problem (the intersection of \(A\) and \(B\)), the \(\beta\)-RAAR algorithm \([18]\) is defined as the iterations \(\{S^k(w) : k = 1, 2, \ldots\}\) for some initialization \(w \in \mathbb{C}^N\),

\[
S(w) = \beta \cdot \frac{1}{2}(R_A R_B + I)w + (1 - \beta) P_B w \quad (2)
\]

\[
= \frac{\beta}{2} \left\{ (2A^* A - I)(2b \odot \frac{w}{|w|} - w) + w \right\} + (1 - \beta)b \odot \frac{w}{|w|} \quad (3)
\]

\[
= \beta w + (1 - 2\beta)b \odot \frac{w}{|w|} + \beta A^* A(2b \odot \frac{w}{|w|} - w). \quad (4)
\]

Fourier Douglas-Rachford algorithm can be deemed as an extreme case of \(\beta\)-RAAR family with \(\beta = 1\). As RAAR converges to a fixed point \(w\), we could retrieve the phase information \(u = w/|w|\) for \((1)\). Any \(u\) in \(U\) yields a fixed point \(w\). Empirically, RAAR fixed points can produce local solutions of high quality, if a large value is properly chosen for \(\beta\), as reported in \([17, 21]\).
In this work, we disclose the relation between RAAR and the local minimizers \( z \) in Equation (1). As HIO can be reformulated as one alternating direction method of multipliers in [17], we identify RAAR as one ADMM with **penalty parameter** \( 1/\beta' = (1 - \beta)/\beta \) applied to the constrained optimization problem in (1), e.g., Theorem 2.5. This perspective links \( \beta \)-RAAR with a small parameter \( \beta \) to multiplier methods with large penalty \( \beta^{-1} \). It is known in optimization that convergence of a multiplier method relies on a sufficiently large penalty (e.g., see Prop. 2.7 in [22]). From this perspective, it is not surprising that the convergence of RAAR to its fixed point also requires a large penalty parameter. Actually, large penalty has been employed to ensure various ADMM iterations converging to stationary points [23, 24, 25]. For instance, ADMM [25] is applied to solve the minimization of nonconvex nonsmooth functions. Global convergence to a stationary point can be established, when sufficiently large penalty parameters are used.

Saddle plays a fundamental role in the theory and the application of convex optimization [26], in particular, the convergence of ADMM, e.g., [27]. For the application on phase retrieval, Sun et al. [28] conduct saddle analysis on a quadratic objective function of Gaussian measurements. The geometric analysis shows that with high probability the global solution is the one local minimizer, when \( N/n \) is sufficiently large. Most of critical points are saddles at actually. We believe that saddle analysis is also one key ingredient in explaining the avoidance of undesired critical points for the Lagrangian of RAAR. To some extent, promising empirical performance of non-convex ADMM conveys the impression that saddles exist in the Lagrangian function, which is not evident in the context of phase retrieval. This is a motivation of the current study. Recently, researchers have been cognizant of the importance of saddle structure in non-convex optimization research. Analysis of critical points in non-concave-convex problems leads to many interesting results in various applications. For instance, Lee et al. used a dynamical system approach to show that many gradient descent algorithms almost surely converge to local minimizers with random initialization, even though they can get stuck at critical points theoretically [29, 30, 31]. The terminology “local saddle” is a crucial concept in understanding the min-max algorithm employed in modern machine learning research, e.g., gradient descent-ascent algorithms in generative adversarial networks (GANs) [32] and multi-agent reinforcement learning [33]. With proper Hessian adjustment, [34] and [35] proposed novel saddle algorithms to escape undesired critical points and to reach local saddles of min-max problems almost surely with random initialization. Jin et al. [36] proposed one non-symmetric definition of local saddles to address one basic question, “what is a proper definition of local optima for the local saddle?” Later, Dai and Zhang gave saddle analysis on the constraint minimization problems [37].

Our study starts with one characterization of all the fixed point of RAAR algorithms in Theorem 2.3. These fixed points are critical points of (1). By varying \( \beta \), some of the
fixed points become “local saddles” of a concave-nonconvex function $F$,

$$\max_{\lambda} \min_z \left\{ F(z, \lambda; \beta) := \left( \frac{\beta}{2} \| A_\perp(z - \lambda) \|^2 - \frac{1}{2} \| \lambda \|^2 \right), \ z \in \mathcal{Z}, \ \lambda \in \mathbb{C}^N \right\}. \tag{5}$$

To characterize RAAR iterates, we investigate saddles in (5) lying in a high dimensional primal-dual space. Our study aims to answer a few intuitive questions, whether these local dimensional critical points on $\mathcal{Z}$ in the primal space can be lift to local saddles of (5) in a primal-dual space under some spectral gap condition, and how the ADMM iterates avoid or converge to these local saddles under a proper penalty parameter? The line of thought motivates the current study on local saddles of (5). Unfortunately, the definition of local saddles in [36] can not be employed to analyze the RAAR convergence, since the objective function in phase retrieval shares phase invariance, see Remark 3.3.

The main goal of the present work is to establish an optimization view to illustrate the convergence of RAAR, and show by analysis and numerics, under the framework for phase retrieval with coded diffraction patterns in [38], RAAR has a basin of attraction at a local saddle $(z_*^*, \lambda_*)$. For noiseless measurement, $z_* = A^* x_0$ is a strictly local minimizer of (1). In practice, numerical stagnation of RAAR on noiseless measurements disappears under sufficient large $\beta$ values. Specifically, Theorem 2.5 shows that RAAR is actually one ADMM to solve the constrained problem in (1). Based on this identification, Theorem 4.5 show that each limit of RAAR iterates can be viewed as a “local saddle” of $\text{max min } F$ in (5).

The rest of the paper is organized as follows. In section 2, we examine the fixed point condition of RAAR algorithm. By identifying RAAR as ADMM, we disclose the concave-non-convex function for the dynamics of RAAR, which provides a continuation viewpoint on RAAR iteration. In section 3, we present one proper definition for local saddles and show the existence of local saddles for oversampled coded diffraction patterns. In section 4, we show the convergence of RAAR to a local saddle under a sufficiently large parameter. Last, we provide experiments to illustrate the behavior of RAAR, (i) comparison experiments between RAAR and Douglas Rachford splitting proposed in [39]; (ii) applications of RAAR on coded diffraction patterns.

2. RAAR algorithms

2.1. Critical points

The following gives the first order optimality of the problem in (1). This is a special case of Prop. 3.2 with $\lambda = 0$. We skip the proof.

**Proposition 2.1.** Let $z_0 = b \odot u_0$ be a local minimizer of the problem in (1). Let $K_{z_0}^{-1} := \mathcal{R}(\text{diag}(\bar{u}_0)A_{\perp}^*A_{\perp}\text{diag}(u_0))$. Then the first-order optimality condition is

$$q_0 := z_0^{-1} \odot (A_{\perp}^*A_{\perp}z_0) \in \mathbb{R}^N, \tag{6}$$

and the second-order necessary condition is that for all $\xi \in \mathbb{R}^N$,

$$\xi^T(K_{z_0}^{-1} - \text{diag}(q_0))\xi \geq 0. \tag{7}$$
Remark 2.2. Once a local solution \( z \) is obtained, the unknown object of phase retrieval in (1) can be estimated by \( x = Az \). On the other hand, using \( I = A^*A + A^*_\perp A_\perp \), we can express the first order condition as

\[
 u^{-1} \odot (A^*A(b \odot u)) = b \odot (1 - q_0) \in \mathbb{R}^N.
\]  

Using \( \xi = e_i \) canonical vectors of \( \mathbb{R}^N \), we have a componentwise lower bound on \( 1 - q_0 \) from (7):

\[
 b^{-1} \odot ((1 - q_0) \geq \|Ae_i\|^2 \geq 0. \quad \text{In general, there exists many local minimizers on } \mathcal{Z}, \text{ satisfying (2) and (7).}
\]

2.2. Fixed point conditions of RAAR

We begin with fixed point conditions of \( \beta \)-RAAR iterations in (4). For each \( \beta \in (0, 1) \), introduce one auxiliary parameter \( \beta' \in (0, \infty) \) defined by

\[
 \beta = \frac{\beta'}{1 + \beta'}, \quad \text{i.e., } \beta = \frac{\beta}{1 - \beta}.
\]  

We shall show the reduction in the cardinality of fixed points under with a small penalty parameter.

Theorem 2.3. Consider the application of the \( \beta \)-RAAR algorithm on the constrained problem in (1). Write \( w \in \mathbb{C}^N \) in polar form \( w = u \odot |w| \). For each \( \beta \in (0, 1) \), let \( \beta' = \beta/(1 - \beta) \) and

\[
 c := (1 - \frac{1 - \beta}{\beta})b + \frac{1 - \beta}{\beta} |w| \in \mathbb{R}^N.
\]  

Then \( w \) is a fixed point of \( \beta \)-RAAR, if and only if \( w \) satisfies the phase condition

\[
 A^*A(b \odot u) = c \odot u,
\]  

and the magnitude condition,

\[
 |w| = \beta' c + (1 - \beta')b \geq 0, \quad \text{i.e., } c \geq (1 - \beta'^{-1})b.
\]  

In particular, for \( \beta \in [1/2, 1) \), we have \( c \geq 0 \) from (12). Observe that the inequality in (12) ensures the well-defined magnitude vector \( |w| \). Hence, the fixed points are critical points of (1).

Proof. Rearranging (4), we obtain the fixed point condition of RAAR,

\[
 ((1 - \beta)|w| - (1 - 2\beta)b) \odot \frac{w}{|w|} = \beta A^*A \left\{ (2b - |w|) \odot \frac{w}{|w|} \right\}.
\]  

Equivalently, taking the projections \( A^*A \) and \( I - A^*A \) on (13) yield

\[
 A^*A \left\{ (b - |w|) \odot \frac{w}{|w|} \right\} = 0,
\]  

\[
 (I - A^*A) \left\{ b \odot \frac{w}{|w|} \right\} = \beta'^{-1} \left\{ (b - |w|) \odot \frac{w}{|w|} \right\}.
\]
For the only-if part, let \( w \) be a fixed point of RAAR with (14), (15). With the definition of \( c \), (14) gives
\[
A^* A((c - b) \circ u) = \beta' A^* A(|w| - b) \circ u = 0,
\]
and (15) gives
\[
A^*_\bot A_\bot (c \circ \frac{w}{|w|}) = A^*_\bot A_\bot \left\{ b - \beta'^{-1}(b - |w|) \right\} \circ \frac{w}{|w|} = 0,
\]
which implies \( c \circ u \) in the range of \( A^* \). Together with (16), we have (11). Also, (12) is the result of the non-negativeness of \( |w| \) in (10).

To verify the if-part, we need to show that \( w \) constructed from a phase vector \( u \in \mathbb{C}^N \) satisfying (11) and a magnitude vector \( |w| \) satisfying (12) meets (14), (15). From (10), we have (14), i.e.,
\[
A^* A((b - |w|) \circ u) = A^* A \beta'(b - c) \circ u = \beta' \{ c \circ u - c \circ u \} = 0. \tag{18}
\]
With the aid of (14), (11), the fixed point condition in (15) is ensured by the computation:
\[
(I - A^* A) \left\{ (b - \beta'^{-1}(b - |w|)) \circ u \right\} = (I - A^* A) \{ c \circ u \} = 0. \tag{19}
\]
Finally, the condition in (11) is identical to the first optimality condition in (6). Hence, the fixed points must be critical points of (1). □

Theorem 2.3 indicates that each fixed point \( w \) can be re-parameterized by \( (u, \beta) \) satisfying (11) and (12). The condition in (12) always hold for \( \beta' \) sufficiently small.

Remark 2.4. The first order optimality in (6) yields that the phase condition in (11) is actually the critical point condition of \( u \in \mathcal{U}_* \) in (1). Fix one critical point \( u \in \mathcal{U}_* \) and let \( c \) be the corresponding vector given from (11). From Theorem 2.3, \( w \) given from the polar form \( w = u \circ |w| \) with (12) is a fixed point of \( \beta \)-RAAR, if \( \beta \) satisfies the condition in (12). To further examine (12), we parameterize the fixed point \( w \) by \( (u, \beta) \). Let \( b^{-1} \circ K^\bot b \) denote the threshold vector, where
\[
K := \Re(\text{diag}(\bar{u})A^*A\text{diag}(u)), \quad K^\bot := I - K.
\]
The fixed point condition in (12) indicates that \( (u, \beta_1) \) gives a fixed point of \( \beta_1 \)-RAAR with any \( \beta_1 \in (0, \beta) \). That is, the corresponding parameter \( (\beta')^{-1} \) must exceed the threshold vector,
\[
(\beta')^{-1} = \frac{1 - \beta}{\beta} \geq b^{-1} \circ (K^\bot b). \tag{20}
\]
Since \( \beta' = \beta/(1 - \beta) \) can be viewed as one penalty parameter in the associated Lagrangian in (23), we call (20) the penalty-threshold condition of RAAR fixed points. In general, the cardinality of RAAR fixed points decreases under a large parameter \( \beta \). See Fig. 1.
For $\beta = 1$, RAAR reduces to FDR, whose fixed point $w$ satisfies $\|A_{\perp}(b \odot w/|w|)\| = 0$ and thus $\|A(b \odot w/|w|)\| = \|b\|$. When phase retrieval has uniqueness property, $A(b \odot w/|w|)$ gives the reconstruction. On the other hand, for $\beta = 1/2$, (4) gives
\begin{equation}
S(w) = A^*A(b \odot \frac{w}{|w|}) + \frac{1}{2}(I - A^*A)w. \tag{21}
\end{equation}
Suppose a RAAR initialization is chosen from the range of $A^*$. The second term in (21) always vanishes and thus RAAR iterations reduce to alternating projection iterations (AP) in [2]. From this perspective, one can regard $\beta$-RAAR as one family of algorithms interpolating AP and FDR, varying $\beta$ from $1/2$ to $1$.

2.3. Alternative directions method of multipliers

Next, we present one relation between RAAR and the alternating direction method of multipliers (ADMM). The alternating direction method of multipliers was originally introduced in the 1970s [40, 41] and can be regarded as an approximation of the augmented Lagrangian method, whose primal update step is replaced by one iteration of the alternating minimization. Although ADMM is classified as one first-order method, practically ADMM could produce a solution with modest accuracy within a reasonable amount of time. Due to the algorithm simplicity, nowadays this approach is popular in many applications, in particular, applications of nonsmooth optimization. See [27, 42, 43] and the references therein.

Use the standard inner product
\[ \langle x, y \rangle := \Re(x^*y), \forall x, y \in \mathbb{C}^N. \]
To solve the problem in (1), introduce one auxiliary variable $y \in \mathbb{C}^N$ with one constraint $y = z$ and one associated multiplier $\lambda \in \mathbb{C}^N$, and form the Lagrangian function with some parameter $\beta' > 0$ in (9),
\[ \frac{\beta'}{2}\|A_{\perp}y\|^2 + \langle \lambda, y - z \rangle + \frac{1}{2}\|y - z\|^2, \quad y \in \mathbb{C}^N, z \in \mathcal{Z}. \tag{22} \]
Introducing one projection operator on \( Z \).

Local saddles of RAAR algorithms

From the initialization \( z_1 \in Z \) and \( \lambda_1 \in \mathbb{C}^N \) and generates the sequence \( \{(y_k, z_k, \lambda_k) : k = 1, 2, 3, \ldots\} \) with stepsizes \( s > 0 \), according to rules,

\[
y_{k+1} = \arg\min_y L(y, z_k, \lambda_k),
\]

\[
z_{k+1} = \arg\min_z L(y_{k+1}, z, \lambda_k),
\]

\[
\lambda_{k+1} = \lambda_k + s_\lambda L(y_{k+1}, z_{k+1}, \lambda)
\]

Introducing one projection operator on \( Z \), \( [w]_Z := w / |w| \odot b \) for \( w \in \mathbb{C}^N \). Algebraic computation yields

\[
y_{k+1} = (I + \beta A_1^* A_\perp)^{-1}(z_k - \lambda_k) = (I - \beta A_1^* A_\perp)(z_k - \lambda_k),
\]

\[
z_{k+1} = [y_{k+1} + \lambda_k]_Z,
\]

\[
\lambda_{k+1} = \lambda_k + s(y_{k+1} - z_{k+1}).
\]

From the \( y \)-update in (24), one reconstruction \( x \) for the unknown object in (1) can be computed by

\[
x = Ay = A(I - \beta A_1^* A_\perp)(z - \lambda) = A(z - \lambda).
\]

Theorem 2.5 indicates that RAAR is actually an ADMM with proper initialization applied to the problem in (1). In general, the step size \( s \) of the dual vector \( \lambda \) should be chosen properly to ensure the convergence. The following shows the relation between RAAR and the ADMM with \( s = 1 \). Hence, we shall focus \( s = 1 \) in this paper.

**Theorem 2.5.** Consider one \( \beta \)-RAAR iteration \( \{w_0, w_1, \ldots\} \) with nonzero initialization \( w_0 \in \mathbb{C}^N \). Let \( \lambda_1 = A_1^* A_\perp w_0 \), \( z_1 = [w_0]_Z \). Generate one ADMM sequence \( \{(y_{k+1}, z_{k+1}, \lambda_{k+1}) : k = 1, 2, \ldots\} \) with dual step size \( s = 1 \), according to (24), (25), (26) with initialization \( \lambda_1, z_1 \). Construct a sequence \( \{w'_k : k = 1, 2, \ldots\} \) from \( (y_k, z_k, \lambda_k) \),

\[
w'_k := y_{k+1} + \lambda_k = (I - \beta A_1^* A_\perp)z_k + \beta A_1^* A_\perp \lambda_k.
\]

Then the sequence \( \{w'_k : k = 1, 2, \ldots\} \) is exactly the \( \beta \)-RAAR sequence, i.e., \( w'_k = w_k \) for all \( k \geq 1 \).

**Proof.** Use induction. For \( k = 1 \), we have

\[
w'_1 = (I - \beta A_1^* A_\perp)z_1 + \beta A_1^* A_\perp \lambda_1 = (I - \beta A_1^* A_\perp)[w_0]_Z + \beta A_1^* A_\perp w_0 = w_1.
\]

Suppose \( w'_k = w_k \) for \( k \geq 1 \). From (28) and (25), we have

\[
A_1^* A_\perp w'_k = A_1^* A_\perp ((1 - \beta)z_k + \beta \lambda_k),
\]

and \( z_{k+1} = [w'_k]_Z \). From (26) and (24),

\[
A_1^* A_\perp \lambda_{k+1} = A_1^* A_\perp \{\beta \lambda_k + (1 - \beta)z_k\} - A_1^* A_\perp z_{k+1}.
\]
Local saddles of RAAR algorithms

Together with (29), (25) and (30), we complete the proof by the calculation,

\[ w'_{k+1} = (I - \beta A^*_\perp A_\perp)z_{k+1} + \beta A^*_\perp A_\perp \lambda_{k+1} \]

\[ = (I - \beta A^*_\perp A_\perp)z_{k+1} + \beta A^*_\perp A_\perp \{\beta \lambda_k + (1 - \beta)z_k\} - \beta A^*_\perp A_\perp z_{k+1} \]

\[ = (I - 2\beta A^*_\perp A_\perp)[w'_k]z + \beta A^*_\perp A_\perp w'_k \]

\[ = (I - 2\beta A^*_\perp A_\perp)[w_k]z + \beta A^*_\perp A_\perp w_k = w_{k+1}. \]  

\[ \square \]

Theorem 2.5 provides one max-min viewpoint to explore the dynamics of RAAR, which motivates the study in the next section. Indeed, after eliminating \( y \) in \( L \) in (22) via (24), we end up with a max-min problem of a concave-non-convex function \( F \),

\[ F(z, \lambda; \beta) := \left( \frac{\beta}{2} \|A_\perp(z - \lambda)\|^2 - \frac{1}{2} \|\lambda\|^2 \right), \quad z \in \mathcal{Z}, \quad \lambda \in \mathbb{C}^N. \]  

(35)

We can convert RAAR convergence to its fixed points into the convergence to saddles of the function \( F \) in (35) by one primal-dual algorithm. For notation simplicity, we shall omit \( \beta \) in the function \( F(z, \lambda; \beta) \), i.e., write \( F(z, \lambda) \), if no confusion occurs in the context.

3. Definition of local saddles for RAAR

When the objective function \( F \) of a max-min problem is not concave-convex, saddle points do not exist in general. For some smooth function \( F \), a point \((\lambda, z)\) is said to be a local max-min point in (44, 35), if \( z = z_s \) is a local minimizer and \( \lambda = \lambda_s \) is a local maximizer, i.e.,

\[ F(z_s, \lambda) \leq F(z_s, \lambda_s) \leq F(z, \lambda_s) \]  

(36)

for \((z, \lambda)\) near \((z_s, \lambda_s)\). The standard analysis can give the first-order and second-order characterizations. The existence of a saddle \((z_s, \lambda_s)\) in fact ensures the minimax equality. Indeed, since \( \min_z F(z, \lambda) \leq \min_z \max_\lambda F(z, \lambda) \), then \( \max_\lambda \min_z F(z, \lambda) \leq \min_z \max_\lambda F(z, \lambda) \). Together with

\[ \min_z \max_\lambda F(z, \lambda) \leq \max_\lambda \min_z F(z_s, \lambda_s) \leq \min F(z_s, \lambda_s) \leq \max_\lambda \min_z F(z, \lambda), \]  

(37)

we have that \( \min_z \max_\lambda F(z, \lambda) = \max_\lambda \min_z F(z, \lambda) \) for \((z, \lambda)\) holds near \((z_s, \lambda_s)\).

In this paper, we shall adopt the idea on “local max-min” proposed in [36] to emphasize the non-symmetric role of \( z, \lambda \), i.e., \((\lambda, z)\) is said to be a local max-min point, if for any \((\lambda, z)\) near \((\lambda_s, z_s)\) within a distance \( \delta > 0 \),

\[ \max_{z'} \{ F(z', \lambda) : \|z' - z_s\| \leq h(\delta) \} \leq F(z_s, \lambda_s) \leq F(z, \lambda_s) \]  

(38)

holds for some continuous function \( h : \mathbb{R} \to \mathbb{R} \) with \( h(0) = 0 \). That is, the minimizer \( z \) is driven by the dual vector \( \lambda \) maximizing the objective function \( F \). Since \( F \) in (35) is strictly concave in \( \lambda \) for \( \beta \in (0, 1) \), according to Prop. 18, 19 and 20 [36], the first-order condition is \( \nabla_z F(z_s, \lambda_s) = 0 \) and \( \nabla_\lambda F(z_s, \lambda_s) = 0 \), and the second-order
necessary/sufficient condition can be reduced to $\nabla_{zz} F(z_*, \lambda_*) \succeq 0$ and $\nabla_{zz} F(z_*, \lambda_*) \succ 0$, respectively. In short, thanks to the $\lambda$-concavity in $F$, we end up with the identical characterization for local-max-min points in [44, 35]. For simplicity, we shall also call these local max-min points as local saddles.

3.1. Local saddles

From Theorem 2.5, the RAAR convergence of $w_k$ to $w_*$ can be regarded as the convergence to a local saddle $(\lambda_k, z_k)$ of $F$ in (35). For each $\lambda_k, z_k$ is one approximate of the corresponding minimizer of $F(\lambda_k, z)$. From (36), we have the following optimality of $F$ in $\lambda$ and $z$, respectively. Since $0 < \beta < 1$, the strict concavity of $F$ in $\lambda$ ensures the uniqueness of $\lambda$ for any vector $z \in \mathcal{Z}$. We omit the proof of Prop. 3.1.

**Proposition 3.1.** Fix one $z \in \mathcal{Z}$. The maximizer $\lambda$ of $F(z; \lambda)$ in (32) satisfies

$$\lambda = -\beta' A_\perp^* A_\perp z. \quad (39)$$

Hence, $\lambda_\star = -\beta' A_\perp^* A_\perp z_\star$ holds for a saddle point $(z_\star, \lambda_\star)$.

**Proposition 3.2.** Let $q(z, \lambda) \in \mathbb{C}^N$ be a vector-valued function of $z$ and $\lambda$,

$$q(z, \lambda) := z^{-1} \odot A_\perp^* A_\perp (z - \lambda). \quad (40)$$

Fix one $\lambda$ in (32), and consider the $z$-minimization

$$\min_{z \in \mathcal{Z}} \left\{ F(z, \lambda) := \frac{\beta}{2} \|A_\perp (z - \lambda)\|^2 - \frac{1}{2} \|\lambda\|^2 \right\}. \quad (41)$$

When $z \in \mathcal{Z}$ is a local minimizer of $F(z, \lambda)$, then

$$q(z, \lambda) \in \mathbb{R}^N \quad (42)$$

and

$$\xi^\top (K_\perp^\top - \text{diag}(\Re(q))) \xi \geq 0, \quad \forall \xi \in \mathbb{R}^N. \quad (43)$$

**Proof.** Let $u = z/|z|$ be the phase vector of a local minimizer $z$. Consider the perturbation $z \rightarrow z \odot \exp(i\theta)$ in $\mathcal{Z}$ applied on $F(z, \lambda)$ with $\theta \in \mathbb{R}^N$, $\|\theta\|$ near $0$. Variation of $F$ can be expressed as one function of tangent vectors $\xi := b \odot \theta$, $H(\xi; z) = H(b \odot \theta; z) := \|A_\perp^* A_\perp (z \odot \exp(i\theta) - \lambda)\|^2$. With $z = b \odot u$, the Taylor expansion

$$H(\xi; z) = H(0; z) + \{2 \langle i(b \odot u) \odot \theta, A_\perp^* A_\perp (b \odot u - \lambda) \rangle - \langle (b \odot u) \odot \theta^2, A_\perp^* A_\perp (b \odot u - \lambda) \rangle + \|A_\perp (b \odot u \odot \theta)\|^2 \} + o(\|\theta\|^2)$$

$$= H(0; z) + 2 \langle i\xi, q \odot b \rangle + \{ -\langle \xi^2, q \rangle + \xi^\top K_\perp \xi \} + o(\|\xi\|^2)$$

implies that the first-order condition for a local minimizer $z$ is $q \in \mathbb{R}^N$, and the second-order condition is the positive semi-definite condition,

$$\langle \xi, (K_\perp - \text{diag}(q)) \xi \rangle \geq 0. \quad (44)$$

□
Unfortunately, the above optimality conditions for $F$ in (35) yields nonexistence of local saddles under the definition in (36)! Consider a noisy case, $\|A_\perp z\| > 0$ for all $z \in \mathcal{Z}$. No Nash equilibrium of $F$ in (35) can exist, since (43) and (39) cannot hold simultaneously at any stationary point of $F$. Indeed, suppose that $(z_*, \lambda_*)$ is a saddle point with $\|A_\perp z_*\| > 0$. The $\lambda$-optimality in (39) gives $\langle z_*, \lambda_* \rangle = -\beta^2 \|A_\perp z_*\|^2 < 0$. On the other hand, as $\xi = b$, (43) gives
\[
\xi^\top (K^\perp - \text{diag}(\Re(q))) \xi = \Re(\mathbf{z}^* A^*_\perp A_\perp \lambda) \geq 0.
\]
This inconsistency indicates that as $\lambda$ tends to $\lambda_*$, the corresponding local minimizer $z$ does not approach $z_*$ continuously. In the next subsection, we shall give a proper definition for the local Nash equilibrium applied on phase retrieval. Remark 3.3 indicates this difficulty always exists in the non convex-concave optimization with phase invariance.

**Remark 3.3 (phase-invariance).** Consider the problem
\[
\min_{\lambda} \max_{z \in \mathcal{Z}} \{-F(z, \lambda)\},
\]
where $F(z, \lambda) = F(\alpha z, \alpha \lambda)$ holds for any complex unit $\alpha$. Suppose that $(z, \lambda)$ is a local Nash equilibrium, then
\[
F(\alpha z, \lambda) \leq F(z, \lambda) \leq F(z, \alpha \lambda)
\]
for any complex unit $\alpha$ near 1 and $\lambda$ is a local maximizer. Then phase-invariance of $F$ implies
\[
F(\alpha z, \lambda) = F(z, \bar{\alpha} \lambda) \leq F(z, \lambda).
\]
Contradiction to (46) always occurs, if the above inequality in (47) is strict.

### 3.2. Cross sections

To alleviate the difficulty in Remark 3.3 we shall restrict the neighbourhood $U(z_*)$ by slicing the projected torus $A_\perp \mathcal{Z}$ into cross sections, such that (38) can hold locally at each critical point. Fix one $z_0 \in \mathcal{Z}$ and introduce the set $\mathcal{Z}'(z_0) := \{z_0 \odot \exp(i\theta) : \langle \theta, b^2 \rangle = 0, \theta \in \mathbb{R}^N\}$. Consider the optimization problem
\[
\min_{z \in \mathcal{Z}'(z_0)} \{\|A_\perp z\|^2\}.
\]
Note that one partition $\mathcal{Z} = \cup_{|\alpha| = 1} \{\mathcal{Z}'(z_0\alpha) : |\alpha| = 1\}$ indicates that for each $z \in \mathcal{Z}$, $z \in \mathcal{Z}'(\alpha z_0)$ holds for some complex unit $\alpha = \exp(i\rho)$, $\rho \in \mathbb{R}$. Indeed, let $1_N \in \mathbb{R}^N$ be a vector whose entries are all 1. Since $z, z_0$ both lie in $\mathcal{Z}$, then
\[
z \odot z_0^{-1} = \exp(i\delta) = \exp(i(\theta + \rho 1_N))\) for some $\delta, \theta \in \mathbb{R}^N$
\]
and $\rho := \|b\|^{-2} \langle \delta, b^2 \rangle$ with $0 = \langle \theta, b^2 \rangle$. To proceed, we need a few notations. At each $z \in \mathcal{Z}$, introduce a matrix $K_z^\perp$ and a tangent plane $\{iu \odot \xi : \xi \in \Xi, \ u = z/|z|\}$ with
\[
K_z^\perp := \Re \left( \text{diag} \left( \frac{z}{|z|} \right) A_\perp A_\perp \text{diag} \left( \frac{z}{|z|} \right) \right), \Xi := \{\xi : \xi \in \mathbb{R}^N, \langle \xi, b \rangle = 0\}.
\]
We shall drop the dependence on $z$ to simplify the notation, if no confusion occurs. Since the feasible set in $z$ is different from the setting in Prop. 3.2, we must investigate again the local $z$-optimality in $F(z, \lambda)$.

**Proposition 3.4.** Fix one $\lambda \in \mathbb{C}^N$. Consider the minimization problem

$$
\min_{z \in \mathcal{Z}(z_0)} \left\{ F(z, \lambda) := \frac{\beta}{2} \| A_\perp (z - \lambda) \|^2 - \frac{1}{2} \| \lambda \|^2 \right\}.
$$

Suppose $z = b \odot u$ is a local minimizer in (50). Then $\Im(q(z, \lambda)) = \rho 1_N$, $\rho \in \mathbb{R}$. The second-order necessary condition is that for all $\xi \in \Xi$, we have

$$
\langle \xi, (K_z - q(z, \lambda)) \xi \rangle \geq 0.
$$

In addition, a second-order sufficient condition is that for all nonzero $\xi \in \Xi$,

$$
\| \xi \|^{-2} \langle \xi, (K_z - q(z, \lambda)) \xi \rangle > 0.
$$

A local minimizer $z$ with (51) is called a **strictly local minimizer**.

**Proof.** Consider a perturbation $z \rightarrow z \odot \exp(i\theta)$ with $\theta \in \mathbb{R}^N$. Use arguments similar to the proof of Prop. 3.2. Since the objective function in (50) is continuously differentiable, with $\xi := b \odot \theta$, we have

$$
\langle \xi, i\bar{u} \odot A_\perp A_\perp (z - \lambda) \rangle = 0
$$

for all $\xi$ with

$$
\langle \xi, b \rangle = \langle \theta, b^2 \rangle = 0.
$$

Then (52) gives for some multiplier $\rho \in \mathbb{R}$,

$$
\Im(\bar{u} \odot A_\perp A_\perp (z - \lambda)) = \rho b, \text{ i.e., } \Im(q) = \rho 1_N \in \mathbb{R}^N.
$$

Note that $\xi \in \Xi$ and thus we have the second-order conditions. □

**Corollary 3.5.** Let $z_0 = b \odot u_0$ be a local minimizer of the problem in (48). Then the first-order condition is

$$
q_0 := z_0^{-1} \odot (A_\perp^* A_\perp z_0) \in \mathbb{R}^N,
$$

and the second-order necessary condition is that for all $\xi \in \Xi := \{ \xi \in \mathbb{R}^N; \langle b, \xi \rangle = 0 \}$,

$$
\xi^T (K_{z_0} - \text{diag}(q_0)) \xi \geq 0.
$$

Hence, $z_0$ is a strictly local minimizer on $\mathcal{Z}$, if

$$
\| \xi \|^{-2} \xi^T (K_{z_0} - \text{diag}(q_0)) \xi > 0, \ \xi \in \Xi, \ \| \xi \| > 0
$$

and (57) hold.

**Proof.** This is the special case of Prop. 3.4 with $\lambda = 0$. Note that $q_0 = q(z, 0)$ and we have $q_0 \in \mathbb{R}^N$ from (54) and

$$
\rho = \langle b^2, \rho 1_N \rangle = \langle b^2, \Im(q(z, \lambda)) \rangle = \Im(\| A_\perp z \|^2) = 0.
$$
and independently distributed on the unit circle. Let \( x \) be isometric with a proper choice of \( c \). Let \( x \) be a given rank \( \geq 1 \) object and at least one of \( \mu_j \), \( j = 1, \ldots, l \geq 2 \), be continuously and independently distributed on the unit circle. Let

\[
A^* = c_0 \left( \Phi \text{diag}\{\mu_1\} \right)
\]

\[
\vdots
\]

\[
\Phi \text{diag}\{\mu_l\}
\]

be isometric with a proper choice of \( c_0 \) and \( B := A \text{diag}(u_0) \), \( u_0 = |A^*x_0|^{-1} \odot (A^*x_0) \). Then with probability one,

\[
\lambda_2 = \max\{\|\Im(B^*v)\| : v \in \mathbb{C}^n, v \perp ix_0, \|v\| = 1\} < 1.
\]

**Theorem 3.7** (Theorem 6.3[16]). Let \( \Phi \) be the oversampled discrete Fourier transform. Let \( x_0 \) be a given rank \( \geq 2 \) object and at least one of \( \mu_j \), \( j = 1, \ldots, l \geq 2 \), be continuously and independently distributed on the unit circle. Let

\[
A^* = c_0 \left( \Phi \text{diag}\{\mu_1\} \right)
\]

\[
\vdots
\]

\[
\Phi \text{diag}\{\mu_l\}
\]

be isometric with a proper choice of \( c_0 \) and \( B := A \text{diag}(u_0) \), \( u_0 = |A^*x_0|^{-1} \odot (A^*x_0) \). Then with probability one,

\[
\lambda_2 = \max\{\|\Im(B^*v)\| : v \in \mathbb{C}^n, v \perp ix_0, \|v\| = 1\} < 1.
\]

**Theorem 3.8.** Let \( \Phi \) be the oversampled discrete Fourier transform. Let \( x_0 \) be a given rank \( \geq 2 \) object and at least one of \( \mu_j \), \( j = 1, \ldots, l \geq 2 \), be continuously and independently distributed on the unit circle. Let

\[
A^* = c_0 \left( \Phi \text{diag}\{\mu_1\} \right)
\]

\[
\vdots
\]

\[
\Phi \text{diag}\{\mu_l\}
\]
be isometric with a proper choice of \( c_0 \). Then with probability one, \( z = A^*x_0 \) is a strictly local minimizer of (24), and

\[
\min_{\xi} \{ \| \xi \|^2 - \langle \xi, (K^{-1}_z - \text{diag}(\Re(q(z, 0)))) \xi \rangle : \xi \in \mathbb{R}^N, \langle |z|, \xi \rangle = 0 \} \geq 1 - \lambda_2 > 0, \quad (65)
\]

where \( \lambda_2 \) is given in (63).

**Proof.** Note that (63) implies that \( \Re(B^*v) = \lambda_2 \xi \) holds for some unit vector \( \xi \in \mathbb{R}^N \). Then \( \xi \) is one left singular vector of

\[
\mathcal{B} := [\Re(B^*), \Im(B^*)] \in \mathbb{R}^{N \times 2N}
\]

with singular value \( \lambda_2 \), while the associated right singular vector is \( [\Im(v)^\top, \Re(v)^\top]^\top \). Let \( c = K|z| \) and \( z = Ax_0 \). The left and right singular vectors \( \mathcal{B} \) corresponding to singular value 1 are \( c \in \mathbb{R}^N \) and \( [\Im(ix_0)^\top, \Re(ix_0)]^\top \), respectively. Since \( \mathcal{B}^\top \mathcal{B} = \Re(B^*B) = \Re(\text{diag}(\Re(A^*A \Re(u))) = K_z \), then \( \xi, c \) are eigenvectors of \( K_z \). From theorem 3.7 with probability one we have

\[
1 > \lambda_2 \geq \max_{\xi} \{ \langle \xi, K_z \xi \rangle : \| \xi \| = 1, \langle |z|, \xi \rangle = 0 \}. \quad (67)
\]

By definition in (40), \( q(z, 0) = b^{-1} \odot (K_z^\top b) = 1 - b^{-1} \odot K_z b = 0 \). Finally, since \( K_z^\top = I - K_z \), (67) gives (65) and (51).

**4. RAAR Convergence analysis**

**4.1. Convergence of RAAR**

We shall derive one inequality stated in (82), which ensures the convergence of RAAR iterations \( \{w_k : k = 1, 2, \ldots \} \) in Prop. 4.2. In Theorem 4.5 we shall show that the condition in (82) holds near local saddles under a sufficient large penalty 1/\( \beta' \).

From the \( \lambda \)-iterations of ADMM in (26), we have

\[
\lambda_{k+1} = \lambda_k + (y_{k+1} - z_{k+1}), \quad \text{and} \quad w_k := \lambda_k + y_{k+1} = \lambda_{k+1} + z_{k+1}. \quad (68)
\]

The \( z \) iteration yields \( z_{k+1} = [w_k]_z = [z_{k+1} + \lambda_{k+1}]_z \) and \( z_{k+1} + \lambda_{k+1} \) shares the same phase with \( z_{k+1} \). We have a lower bound,

\[
\lambda_{k+1} \odot \frac{z_{k+1}}{|z_{k+1}|} \in (-|z_{k+1}|, \infty). \quad (69)
\]

With \( y_* = z_* \), \( \lambda_* = -\beta A_{\perp}^* A_{\perp} z_* \), and \( y_{k+1} = (I - \beta A_{\perp}^* A_{\perp})(z_k - \lambda_k) \), \( \beta' = \beta/(1 - \beta) \), introduce

\[
T(z_k, \lambda_k) := \beta' \| A_{\perp} (y_* - y_{k+1}) \|^2 + \| y_{k+1} - z_k \|^2
\]

\[
= \beta(1 - \beta) \| A_{\perp} ((z_k - z_*) - (\lambda_k - \lambda_*)) \|^2
\]

\[
+ \| A_{\perp} (-\beta (z_k - z_*) - (1 - \beta)(\lambda_k - \lambda_*)) \|^2 + \| A(\lambda_* - \lambda_k) \|^2
\]

\[
= \beta \| A_{\perp} (z_k - z_*) \|^2 + (1 - \beta) \| A_{\perp} (\lambda_k - \lambda_*)) \|^2 + \| A\lambda_k \|^2. \quad (70)
\]

We shall derive a few inequalities from the optimal condition of \( y \) and \( z \) in (22), respectively.
Proposition 4.1. Let $T$ be defined in (70). Then
\[
\|w_{k-1} - w_*\|^2 - \|w_k - w_*\|^2 \geq T(z_k, \lambda_k) + 2 \langle z_k - z_*, \lambda_k - \lambda_* \rangle \tag{71}
\]

Proof. Let $C$ be the cost function, $C(y) = \beta' ||A_\perp y||^2/2$, for $y \in \mathbb{C}^N$. We shall prove
\[
\frac{1}{2} \|z_k - z_*\|^2 \geq \langle \lambda_k - \lambda_* , y_{k+1} - y_* \rangle + \frac{1}{2}\|y_{k+1} - y_*\|^2 + \frac{1}{2}T(z_k, \lambda_k). \tag{72}
\]

To this end, we make two claims. First, the optimality of $y_{k+1}$ in (24) indicates that for all $y \in \mathbb{C}^N$,
\[
C(y) - C(y_{k+1}) = - \frac{1}{2} \|y - z_k\|^2 + \langle \lambda_k , (y_{k+1} - y) \rangle + \frac{1}{2} \|y - y_{k+1}\|^2 \tag{73}
\]
\[
+ \left( \frac{1}{2} \|y_{k+1} - z_k\|^2 + \frac{\beta'}{2} ||A_\perp (y - y_{k+1})||^2 \right). \tag{74}
\]

Second, the optimality of $y_*$ in $C(y)$ indicates
\[
C(y) - C(y_*) + \langle (y - y_*) , \lambda_* \rangle \geq 0. \tag{75}
\]

To verify (74), use the optimality $y_{k+1} = (I + \beta' A_\perp^* A_\perp)^{-1}(z_k - \lambda_k)$ in (24), which gives $\nabla_y L(y_{k+1}, z_k, \lambda_k) = 0$. The quadratic convexity of $L$ in $y$ gives (74), i.e.,
\[
L(y, z_k, \lambda_k) = C(y) + \langle \lambda_k, (y - z_k) \rangle + \frac{1}{2} \|y - z_k\|^2
\]
\[= C(y_{k+1}) + \langle \lambda_k , (y_{k+1} - z_k) \rangle + \frac{1}{2} \|y_{k+1} - z_k\|^2 + \left( \frac{\beta'}{2} ||A_\perp (y - y_{k+1})||^2 + \frac{1}{2} \|y - y_{k+1}\|^2 \right). \]

For (75), with $\lambda_* = -\beta' A_\perp^* A_\perp z_* = -\beta' A_\perp^* A_\perp y_*$, Taylor’s expansion of $C(y)$ at $y_*$ gives
\[
C(y) - C(y_*) = \beta' \langle A_\perp^* A_\perp y_*, y - y_* \rangle + \frac{\beta'}{2} ||A_\perp (y_* - y)||^2 \geq \langle -\lambda_*, y - y_* \rangle \tag{76}
\]

With $y = y_* = z_*$ in (74) and $y = y_{k+1}$ in (75), (74)+(75) gives (72).

Next, from (72), we have two identities,
\[
(y_{k+1} - z_*) = w_k - w_* - (\lambda_k - \lambda_*), \tag{77}
\]
\[(z_k - z_*) = w_{k-1} - w_* - (\lambda_k - \lambda_*). \tag{78}
\]

The difference of the squares of (77) and (78) gives
\[
- \|z_* - z_k\|^2 + \|z_* - y_{k+1}\|^2
\]
\[= \|w_k - w_*\|^2 - \|w_{k-1} - w_*\|^2 - 2 \langle w_k - w_{k-1}, \lambda_k - \lambda_* \rangle \tag{79}
\]
\[= \|w_k - w_*\|^2 - \|w_{k-1} - w_*\|^2 + 2 \langle z_k - z_*, \lambda_k - \lambda_* \rangle
\]
\[- 2 \langle \lambda_k - \lambda_* , y_{k+1} - z_* \rangle, \tag{80}
\]

where the last equality is given by the difference of (77) and (78). The proof of (71) is completed by (80) and (72).

Note that for each fixed point $w_* := z_* + \lambda_*$ of RAAR, $\alpha w_*$ is also a fixed point of RAAR with any complex unit $\alpha$. □
Local saddles of RAAR algorithms

Proposition 4.2. For \( z, \lambda \in \mathbb{C}^N \), let \( \alpha \) be the corresponding global phase factor between \( w \) and \( w_* \),

\[
\arg \min_{\alpha} \{ \| w - \alpha w_* \| : |\alpha| = 1 \}, \ w = z + \lambda, \ w_* := z_* + \lambda_*.
\] (81)

Suppose that there exists some constant \( c_0 > 1 \) such that the following inequality

\[
T(z, \lambda) \geq 2c_0 \langle \alpha z_* - z, \lambda - \alpha \lambda_* \rangle
\] (82)

holds for \((z, \lambda) = (z_k, \lambda_k)\) for \( k \geq k_0 \). Then any limit point \((z', \lambda')\) of RAAR satisfies

\[
A_{\perp}(z' - \alpha z_* ) = 0, \ \lambda - \alpha \lambda_* = 0 \quad \text{for some complex unit } \alpha.
\]

Proof. Recall \( w_{k-1} = z_k + \lambda_k \) and \( w_* = z_* + \lambda_* \). Let \( \alpha_k \) be some global factor in

\[
(83)
\]

\[
(84)
\]

\[
(85)
\]

\[
(86)
\]

\[
(87)
\]

Hence,

\[
(1 - c_0^{-1})(k_1 - k_0) \left\{ \frac{1}{k = k_0, \ldots, k_1-1} \min_{\lambda} T(z_k, \lambda) \right\} \leq \| w_{k_0-1} - \alpha_{k_0-1} w_* \|^2. \] (88)

Let \( k_1 \to \infty \). Since the left-hand side is bounded above and \( 1 - c_0^{-1} > 0 \), then

\[
\liminf_{k \to \infty} T(z_k, \lambda_k) = 0.
\]

Let \((z', \lambda')\) be a limiting point and \( \alpha \) be the limiting phase factor. For \( \beta \in (0, 1), \) from

\[
A\lambda' = 0, \ A_{\perp}(-\beta(z' - \alpha z_* ) - (1-\beta)(\lambda' - \alpha \lambda_* )) = 0 = A_{\perp}((z' - \alpha z_* ) - (\lambda' - \alpha \lambda_* )). \] (89)

The second part of (89) gives \( A_{\perp} \lambda' = \alpha A_{\perp} \lambda_* \). Thus \( \lambda' = \alpha \lambda_* \) and \( A_{\perp} z' = \alpha A_{\perp} z_* \).

Remark 4.3. When (82) holds eventually, then (88) indicates that \( T(z_k, \lambda_k) \) sub-linearly converges to 0, i.e., (70) indicates sub-linear convergence of RAAR, \( O((k_0 - k_1)^{-1}) \). This is consistent with sub-linear convergence in FDR numerical experiments in (70).
4.2. Justification of (82) from local saddles

Next we shall verify that the convergence condition in (82) holds, i.e.,

\[ \beta \|A_\perp (z - z_*)\|^2 + (1 - \beta) \|A_\perp (\lambda - \lambda_*)\|^2 + \|A\lambda\|^2 > 2 \langle z_* - z, \lambda - \lambda_* \rangle. \quad (90) \]

if the positive definite condition (92) holds at \(z_*\) and \((z, \lambda)\) is close to \((z_*, \lambda_*)\). For the sake of simplicity, we shall omit the global factors \(\alpha\) in front of \(z\) and \(\lambda\), if no confusion occurs.

**Remark 4.4.** With \(\beta \in (0, 1)\), \(T(z, \lambda)\) can quantize one distance between \((z, \lambda)\) and \((z_*, \lambda_*)\). That is, for \(\epsilon > 0\), from (70), \(T(z, \lambda) < \epsilon\) implies

\[
\max \left\{ \beta \|A_\perp (z - z_*)\|^2, (1 - \beta) \|A_\perp (\lambda - \lambda_*)\|^2, \|A\lambda\|^2 \right\} \leq \epsilon. 
\]

Thus, \(\|A_\perp (z - z_*)\|^2 \leq \epsilon / \beta, \quad \|\lambda - \lambda_*\|^2 \leq \epsilon / (1 - \beta).\)

**Theorem 4.5.** Let \(z_*\) be a strictly local minimizer in (1). Then we can find \(\beta \in (0, 1)\) satisfying

\[
(1 - \beta) \left\langle \xi, K_{zz}^+ \xi \right\rangle > 2 \left\langle \xi, \text{diag}(b^{-1} \circ (K_{zz}^+) b) \xi \right\rangle \quad \text{for any unit vector } \xi \in \Xi. \quad (92)
\]

Consider \(\beta\)-RAAR with this \(\beta \in (0, 1)\). Let \(\lambda_* = -\beta' A_\perp^* A_\perp z_*\) and \(w_* = z_* + \lambda_*\). Then there is some constant \(\epsilon > 0\), such that (82) holds for all \((z, \lambda)\) with \(\|w - w_*\|^2 < \epsilon\), where a proper complex unit is applied on \(w_*\) according to (87).

**Proof.** First, the existence of \(\beta\) for (92) is ensured by Prop. 3.6. The RAAR iterations satisfying (25) (26) (68) (69) indicate the decomposition \(w = z + \lambda\) with \(z \in \mathcal{Z}\) and \(\lambda \circ z^{-1} \in \mathbb{R}^N\). Let \(u = z / |z|, \ u_* = z_*/|z_*|\) and \(q_0 = b^{-1} \circ K_\perp b\). Then \(u \circ \lambda \in \mathbb{R}^N\), \(u_* \circ \lambda_* \in \mathbb{R}^N\) and

\[
(z_*)^{-1} \circ \lambda_* = -\beta' b^{-1} \circ K_\perp b = -\beta' q_0. 
\]

Using continuity arguments on (92), we have with \(u' = (z + z_*)/|z + z_*|, \xi = (-i)u' \circ (z_* - z) \in \mathbb{R}^N\),

\[
(1 - \beta) \left\langle \xi, u' \circ (A_\perp^* A_\perp (u' \circ \xi)) \right\rangle > - (\beta')^{-1} \left\langle \xi, \left(\frac{\lambda_*}{z_*} + \frac{\lambda}{z}\right) \circ \xi \right\rangle 
\]

for \((\lambda, z)\) sufficiently close to \((\lambda_*, z_*\)). Observe that as \(z \to z_*\), we have \(\langle \xi, b \rangle = 0\). Note that \(T(z, \lambda)\) has an upper bound \(2(1 - c_0^{-1})^{-1} \|w_{k_0 - 1} - w_*\|^2 / 2\) from (88). According to Remark 4.4 (94) holds, if \(\|w_{k_0 - 1} - w_*\| < \epsilon\) holds for some \(\epsilon\) sufficiently small. With (94), algebraic computation gives (82). Indeed,

\[
2c_0 \langle \lambda - \lambda_*, z_* - z \rangle = 2c_0 \langle \lambda \circ z^{-1}, z \circ (z_* - z) \rangle - 2c_0 \langle \lambda_* \circ z_*^{-1}, z_* \circ (z_* - z) \rangle = -c_0 \langle \lambda \circ z^{-1} + \lambda_* \circ z_*^{-1}, |(z - z_*)|^2 \rangle \leq \beta \langle \xi, u' \circ (A_\perp^* A_\perp (u' \circ \xi)) \rangle = \beta \|A_\perp (z - z_*)\|^2 \leq T(z, \lambda).
\]

Thus, \(\|w_{k_0} - w_*\| < \epsilon\) gives the closeness condition for the sequential vector \((z, \lambda)\). □
5. Numerical experiments

5.1. Gaussian-DRS

The $\beta$-RAAR algorithm is not the only algorithm, which can screen out some undesired local saddles via varying penalty parameters. Recently, [39] proposed Gaussian-Douglas-Rachford Splitting to solve phase retrieval via minimizing a loss function $\|z - b\|^2$ subject to $z$ in the range of $A^*$. Let $x$ be the unknown object. Let $1_\mathcal{F}(y)$ be the indicator function of the range $\mathcal{F}$ of $A^*$. Then $A^*x \in \mathcal{F}$. Similar to RAAR, the algorithm can be formulated as ADMM with a penalty parameter $\rho > 0$ to reach a local max-min point of the Lagrangian function

$$\max_{\lambda} \min_{y, z \in \mathbb{C}^N} \left\{ \frac{1}{2} \|z - b\|^2 + \langle \lambda, z - y \rangle + \frac{\rho}{2} \|z - y\|^2 + 1_\mathcal{F}(y) \right\}. \tag{95}$$

The ADMM scheme consists of repeating the following three updates to reach a fixed point $(y, z, \lambda)$:

- $y \leftarrow A^*A(z + \rho^{-1}\lambda)$;
- $z \leftarrow (1 + \rho)^{-1}(\|w\| \odot b + \rho w)$ where $w = y - \rho^{-1}\lambda$;
- $\lambda \leftarrow \lambda + \rho(z - y)$.

Similar to the RAAR reconstruction in (27), once a fixed point of this ADMM is obtained, the object $x$ can be computed by $x = Ay = A(z + \rho^{-1}\lambda)$ from the $y$-update. Introduce $P = A^*A$ and $P^\perp = I - P$. After eliminating $y$, the local max-min problem reduces to

$$\max_{\lambda} \min_{z} \left\{ \frac{1}{2} \|z - b\|^2 + \frac{\rho}{2} \left\|P^\perp(z + \frac{\lambda}{\rho})\right\|^2 - \left\|\frac{\lambda}{\rho}\right\|^2 \right\}. \tag{96}$$

Algebraic computations on ADMM scheme yield the fixed-point condition of DRS.

**Proposition 5.1.** Denote $\mu := \lambda/\rho$. Let $(z, \mu)$ be a fixed point of $\rho$-DRS. Then

$$z + \rho \mu = [z - \mu]_Z = [z]_Z, \quad P\mu = 0 \text{ and } P^\perp z = 0. \tag{97}$$

We skip the proof of (97). Note that the condition implies that the vector $z$ shares the same phase vector $u := z/|z|$ with $z - \mu$, and $[z]_Z$ has the $(P, P^\perp)$-decomposition $[z]_Z = z + \rho \mu$, where $P\mu = 0$ and $P^\perp z = 0$. Hence, $\mu/|z| \in \mathbb{R}^N$ is a real vectors with bounds, $-\rho^{-1} \leq z^{-1} \odot \mu \leq 1$.

Next we derive conditions for local saddles of $L$ in (96).

**Proposition 5.2.** Let $(z, \mu)$ be a local saddle of $L$ in (96). Then the first-order optimal condition is

$$z + \rho \mu = [z]_Z, \quad P\mu = 0, \quad P^\perp z = 0. \tag{98}$$

When $\rho > 0$, the concavity of $L$ in $\lambda$ is obvious. Let $K := \Re(\text{diag}(\bar{u}) P \text{diag}(u))$ and $u = z/|z|$. The second-order necessary condition of $z$ in (96) is

$$(\rho + 1)I - \frac{b}{|z|} \succeq \rho K. \tag{99}$$
Proof. The optimality of $\mu$ in (96) is
\[ P^\perp(z + \mu) = \mu, \text{ which implies } P^\perp z = 0 \text{ and } P \mu = 0. \] (100)
From the derivative of $L$ with respect to $z$, the $z$-optimality is
\[ \frac{z}{|z|} \odot (|z| - b) + \rho P^\perp(z + \mu) = 0, \text{ i.e., } z - [z]_Z + \rho P^\perp(z + \mu) = 0. \]
Together with (100), we have $z + \rho \mu = [z]_Z$. Next, we derive the second-order necessary condition of $z$. Consider a perturbation $z \rightarrow z + \epsilon$ with $\epsilon \in \mathbb{C}^N$. From $|z + \epsilon| = |z| \left(1 + \Re(\frac{\epsilon}{|\epsilon|}) + \frac{1}{2} \Im(\frac{\epsilon}{|\epsilon|})^2\right) + o(\epsilon^2)$, and
\[ \|z + \epsilon| - b\|^2 - \|z| - b\|^2 \]
\[ = 2 \left\langle z - b \odot \frac{z}{|z|}, \epsilon \right\rangle + \|\epsilon\|^2 - \left\langle \frac{b}{|z|}, \Im(\epsilon \odot \bar{u})^2 \right\rangle + o(\|\epsilon\|^2), \]
we have the second-order condition of $L$,
\[ \rho \left\langle \epsilon, P^\perp \epsilon \right\rangle + \|\epsilon\|^2 - \left\langle \frac{b}{|z|}, \Im(\epsilon \odot \bar{u})^2 \right\rangle \geq 0. \] (103)
Taking $\epsilon = ic \odot u$ for $c \in \mathbb{R}^N$ yields
\[ \|c\|^2 \geq \frac{1}{\rho + 1} \left\langle \frac{b}{|z|}, c^2 \right\rangle + \frac{\rho}{\rho + 1} \langle c, Kc \rangle. \] (104)

Next, we show that a local saddle $(z, \mu)$ is always a fixed point of DRS.

**Proposition 5.3.** If $(\mu, z)$ is a local max-min point in (96), then $z$ is a fixed point of DRS.

**Proof.** At each stationary point $z$ of DRS, we have $|z| = Kb$ and $[z] = b \odot u$ of DRS. Taking $c = e_i$ in (104) yields
\[ (\rho + 1)Kb = (\rho + 1)|z| \geq b, \text{ i.e., } \bar{u} \odot (z - \mu) = Kb - \mu \odot \bar{u} \geq 0, \]
which implies $[z - \mu]_Z = [z]_Z$. Together with the $(P, P^\perp)$-decomposition, $[z]_Z = z + \rho \mu$, we have the fixed point condition. $\square$

From the second-order condition in (99), we expect that DRS with smaller $\rho$ yields a stronger screening-out ability. The next remark illustrates the screening-out similarity between RAAR and DRS in the case $\rho$ close to 0.

**Remark 5.4.** Roughly, for $\rho$ close to 0, a local saddle at some phase vector $u$ of $\beta$-RAAR would be a local saddle at the same phase vector $u$ of $\rho$-DRS, if $\rho$ and $\beta$ satisfy $\beta^{-1} = \rho + 1$. Indeed, the second-order necessary condition for RAAR function in (57) is given by
\[ K^\perp - \text{diag}(\Re(q)) = -\frac{\beta}{1 - \beta} I - K + \frac{1}{1 - \beta} \left(\frac{Kb}{b}\right) \succeq 0. \] (105)
That is, for any nonzero $\xi \in \mathbb{R}^N$,
\[
\left\langle \frac{Kb}{b}, \xi^2 \right\rangle \geq \beta \|\xi\|^2 + (1 - \beta)\xi^\top K\xi.
\] (106)

On the other hand, for DRS, replacing $c$ with $\pm (Kb/b)^{1/2} \odot \xi$ and $|z| = Kb$ in (104) gives
\[
\left\langle \frac{Kb}{b}, \xi^2 \right\rangle \geq \frac{1}{\rho + 1}\|\xi\|^2 + \frac{\rho}{\rho + 1} \left\langle \left(\frac{Kb}{b}\right)^{1/2} \odot \xi, K\left(\frac{Kb}{b}\right)^{1/2} \odot \xi \right\rangle.
\] (107)

Comparing with (106), (104) is almost identical to (106) under $\beta^{-1} = \rho + 1$, if $\rho$ is close to 0 and we ignore the difference of the second terms of their right hand side.

5.1.1. Simulation of Gaussian matrices

We provide one simulation to present the screening-out effect for $\beta$-RAAR and $\rho$-DRS. Generate Gaussian matrices $A$ with size $n \times N$, $n = 100$, $N/n = 3, 3.5, 4, 4.5$ and 5, respectively. For simplicity, generate noise-free data $b = |A^*x_0|$ from some $x_0$. Apply $\beta$-RAAR and $\rho$-DRS to reconstruct phase retrieval solutions under a set of parameters $\beta, \rho$, respectively. Here, we test
\[
\rho = 1, 1/2, 1/3, 1/4, \ldots, 1/10 \quad \text{and} \quad \beta = 1/2, 2/3, 3/4, \ldots, 10/11.
\] (108)

Figures 2 show the success rate of reaching a global solution for each parameter value (among 40 trials with random initializations). For $\rho$ close to 0 or $\beta$ close to 1, $\beta$-RAAR with
\[
\beta = \frac{\rho^{-1}}{1 + \rho^{-1}}
\]
gives a similar empirical performance as $\rho$-DRS, although RAAR performs slightly better. For instance, as $\beta \geq 0.8$ or $\rho \leq 1/4$, with success rates higher than 70%, $\beta$-RAAR and $\rho$-DRS algorithms both can reconstruct a global solution in the case with $N/n \geq 4$. These empirical results are consistent with the theoretical analysis in Remark 5.4.
5.2. Coded diffraction patterns

The following experiments present convergence behavior of RAAR on coded diffraction patterns. Consider $1\frac{1}{2}$ coded diffraction patterns with oversampling, i.e., one coded pattern and one uncoded pattern as used in [16]. For test images $x_0$, we use the Randomly Phased Phantom(RPP) $x_0 = p \odot \mu_0$, where $\mu_0 := e^{i\phi}$ and $\phi$ are i.i.d. uniform random variables over $[0, 2\pi]$. The size is $128 \times 128$, including the margins. Here, we randomize the original phantom $p$ (in the left of Fig. 3) in order to make its reconstruction more challenging. A random object such as RPP is more difficult to recover than a deterministic object.

Theorem 3.8 states the existence and strictly local minimizer and Theorem 4.5 indicates that the existence of a local saddle relies on a sufficiently large penalty parameter. The following experiment validates RAAR convergence to the local saddle under proper selection on $\beta$.

Empirically, $\beta$-RAAR with large $\beta$ can easily diverge, but $\beta$-RAAR with small $\beta$ can easily get stuck (not necessarily converged) near distinct critical points on $\mathcal{Z}$. To demonstrate the effectiveness of RAAR, we shall make two adjustments on application of $\beta$-RAAR. First, to alleviate the stagnation at far critical solutions, we employ the null vector method [8], which is one spectral initialization, to generate an initialization of RAAR. See the middle and right subfigures in Fig. 3 for the initialization. Second, to reach a local saddle within 600 RAAR iterations, we vary the parameter $\beta$ along a $\beta$-path, starting from some initial value and then decreases to 0.5, shown in Fig. 4. Each path consists of two phases: (i) $\beta$ remains one constant value selecting from $0.95, 0.9, 0.8, 0.7$ and $0.6$ within the first 300 iterations; (ii) $\beta$ decreases to 0.5 piecewise linearly within the second 300 iterations.

Conduct four experiments to examine $\beta$-RAAR along five $\beta$-paths:

(a) Noiseless data, $b = |A^*x_0|$ with $A$ defined in (62). Use the null vector method $x_{null}$
as one initial vector for $\beta$-RAAR, i.e.,
\[
z_1 = [A^* x_{null}]_Z, \quad \lambda_1 = A^* x_{null} - z_1.
\] (109)

(b) Noiseless data. RAAR with random initialization.

c) Noisy data. RAAR with null vector initialization as in (109).

d) Noisy data. RAAR with random initialization.

In (c) and (d), the source of noise is the counting statistics [45], i.e., each entry of the squared measurement $b^2$ follows a Poisson distribution,
\[
b^2 \sim \text{Poisson}(|A^* x_0|^2), \quad \kappa > 0.
\] (110)

In the RPP experiment, the parameter $\kappa$ is chosen so that the noise level is $\|b - |A^* x_0||/\|b\| \approx 0.18$.

5.2.1. Performance metrics Results in the case (a,b) and (c,d) are reported in figure 5 and figure 6 respectively. Each row shows the performance metrics of $\beta$-RAAR iterations $\{w_k\}$. Here, $z, \lambda$ are computed from $w$ according to
\[
\{z_{k+1} := [w_k]_Z, \lambda_{k+1} := w_k - z_{k+1}\}.
\] (111)

From (27), the reconstruction of the object $x$ is estimated by $A(z_{k+1} - \lambda_{k+1})$.

- Residual: $\|A_{\perp} z\|/\|b\|$. 
- Norm of derivative: The Wirtinger derivative of the objective $F$ in the $\lambda$-direction is
\[
-\{\beta A_{\perp}^* (A_{\perp} (z - \lambda)) + \lambda\}.
\] (112)

When RAAR converges to one local saddle, the derivative norm would be 0. The norm
\[
\mathbb{D}_\lambda(z, \lambda) := (\|A_{\perp}((1 - \beta)\lambda + \beta z)\|^2 + \|A\lambda\|^2)^{1/2}
\]
Local saddles of RAAR algorithms

can be employed to examine the quality of convergence. (Empirically, the derivative in the $z$-direction $D_z$ has a behavior similar to the one of $D_\lambda$. For simplicity, we do not report $D_z$ here.)

- Inequality ratio $T(z_k, \lambda_k)$. In the noiseless setting, $T(z, \lambda)$ is positive, as $(z, \lambda)$ approaches a local saddle $(z_\ast, \lambda_\ast)$ with $\|A_\perp z_\ast\| = 0$. Hence, a positive ratio

$$T(z, \lambda) := 1 + (\beta \|A_\perp z\|^2 + (1 - \beta)\|A_\perp \lambda\|^2 + \|A\lambda\|^2)^{-1}(2(z, \lambda))$$

can be used as one indicator that RAAR iterates enter the attraction basin of $(z_\ast, \lambda_\ast)$.

5.2.2. Results on noiseless measurement In Fig. 5, the left column and the right column show the metric performance in the cases (a) and (b).

- The left column shows the result of the 600 RAAR iterations along five $\beta$ curves with null initialization, i.e., case (a). The null initialization is illustrated in the middle of Fig. 3. Based on the residual and derivative metrics, RAAR converges to the global solutions for all five $\beta$-paths. The $T$ become positive after 100 iterations, which indicates the closeness of the null initialization to the attraction basin. In particular, $T$ reaches 1 in the early iterations of the case $\beta = 0.95$.

- The right column shows the case (b). The initialization difference between case (a) and case (b) reflects the influence of undesired local saddles. We observe two distinct convergence behaviors. First, in the case of $\beta = 0.6, \beta = 0.7$ and $\beta = 0.8$, based on the metrics of the derivative norm and the residual, RAAR fails to converge within the first 300 iterations. As $\beta$ decreases in the second 300 iterations, the iterates tend to different local saddles. Second, for the $\beta$ paths starting with $\beta = 0.9$ or $\beta = 0.95$, RAAR successfully converge to global solutions. Their $T$-values are negative in the early 50 iterations, but quickly turn to be positive after 100 iterations. Fig. 7 demonstrates the reconstruction.

5.2.3. Results on noisy measurements The left column in Fig. 6 demonstrates the metric performance of RAAR in the noisy case (c). The null initialization shown in the right of Fig. 3 is used to reduce the chance of getting stuck at far local saddles. The reconstructed objects after the first 300 RAAR iterations are shown in the top row of in Fig. 8. Even though these reconstructions are very similar to the RPP, the metric $D_\lambda$ indicates that these RAAR with $\beta \geq 0.7$ fail to converge within the first 300 iterations. Hence, we decrease $\beta$ in the second 300 iterations to obtain local saddles. Observe that the derivative norm in all cases decays to 0. Actually, by examining the correlation of reconstructed objects after 600 RAAR iterations, we verify that these five reconstructed objects are identical up to a phase factor.

The right column in Fig. 6 shows the metric performance of the noisy case (d). Five $\beta$-RAAR tend to different residual values in the first 300 iterations. For large $\beta$, i.e., $\beta = 0.95, \beta = 0.9$ and $\beta = 0.8$, RAAR produce rather successful reconstructed objects.
shown the bottom row of in Fig. 8. These RAAR do not converge within the first 300 iterations. Hence, we reduce the $\beta$ value during the second 300 iterations. By examining the correlation of reconstructed objects, we verify that three final reconstructions are all identical to the final reconstruction in (c). For small $\beta$, RAAR could get stuck at poor solutions, e.g., $\beta = 0.7$ and $\beta = 0.6$. Indeed, after the second 300 iterations, these two RAAR converge to non-global local solutions with larger residual values. The above experiment results suggest that RAAR starting with large $\beta$ typically performs better than RAAR starting with small $\beta$ in the lack of spectral methods. In numerical simulations, we demonstrate the effectiveness of RAAR on coded diffraction patterns, where $\beta$ travels from a large value to 0.5.

5.3. Conclusion and outlook

In this paper, we examine the RAAR convergence from a viewpoint of local saddles of a concave-non-convex max-min problem. We show that the global solution is a strictly local minimizer in oversampled coded diffraction patterns, which ensures the existence of local saddles. Convergence to each local saddle of the RAAR Lagrangian function requires a sufficient large penalty parameter, which explains the avoidance of undesired local solutions under RAAR with a moderate penalty parameter.

ADMM is a popular algorithm in handling various constraints. The current paper does not introduce any further assumption on unknown objects, except for the condition $x \in \mathbb{C}^n$ in (1). Stable recovery from incomplete measurements is actually possible, provided that additional assumptions of unknown objects are used. For instance, when unknown objects can be characterized by piecewise-smooth functions with small total variation seminorm, the recovery can be obtained from incomplete Fourier measurements with the aid of total variation regularization [46]. Another interesting work [47] demonstrates the number of measurements ensuring stable recovery of a sparse object under independent measurement vectors. From the above perspective, one interesting future work is the saddle analysis of ADMM associated with these additional object assumptions.

5.4. Acknowledgements

The author would like to thank Albert Fannjiang for helpful discussions.

[1] 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. *Signal Processing Magazine, IEEE*, 32(3):87–109, 2015.

[2] P. Chen, A. Fannjiang, and G Liu. Phase retrieval with one or two coded diffraction patterns by alternating projection with the null initialization. *Journal of Fourier Analysis and Applications*, pages 1–40, 2017.

[3] R. W. Gerchberg. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237, 1972.

[4] Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. Phase retrieval using alternating minimization. *arxiv:1306.0160*, 2013.
Local saddles of RAAR algorithms

Figure 5. Left and right columns show the residue metric and the derivative metric of RAAR in case (a) and case (b), respectively.

[5] Yuxin Chen and Emmanuel J. Candès. Solving random quadratic systems of equations is nearly as easy as solving linear systems. arxiv:1505.05114, 2015.

[6] Emmanuel J. Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. IEEE Transactions on Information Theory, 61(4):1985–2007, apr 2015.

[7] Huishuai Zhang and Yingbin Liang. Reshaped Wirtinger flow for solving quadratic system of equations. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 29, pages 2622–2630. Curran Associates, Inc., 2016.

[8] P. Chen, A. Fannjiang, and G Liu. Phase retrieval by linear algebra. SIAM J. Matrix Anal. Appl., 38(3):864–868, 2017.

[9] Wangyu Luo, Wael Alghamdi, and Yue M. Lu. Optimal spectral initialization for signal recovery
Local saddles of RAAR algorithms

Figure 6. Left and right columns show the residue metric and the derivative metric of RAAR in case(c) and case (d), respectively.

Figure 7. The row from left to right show the reconstruction of 300 RAAR iterations in the case (b), corresponding to $\beta$-paths with $\beta = 0.95, 0.9, 0.8, 0.7$ and $0.6$, respectively.

with applications to phase retrieval. arXiv:1811.04420, 2018.

[10] Yue M. Lu and Gen Li. Phase transitions of spectral initialization for high-dimensional nonconvex estimation. arxiv:1702.06435, 2017.

[11] Marco Mondelli and Andrea Montanari. Fundamental limits of weak recovery with applications to phase retrieval. Foundations of Computational Mathematics, 19(3):703–773, Jun 2019.

[12] John C Duchi and Feng Ruan. Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval. Information and Inference: A Journal of the IMA, 8(3):471–529, 2019.

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

[14] J. R. Fienup. Phase retrieval algorithms: a personal tour. Applied Optics, 52(1):45–56, 2013.

[15] Heinz H. Bauschke, Patrick L. Combettes, and D. Russell Luke. Hybrid projection–reflection
method for phase retrieval. *J. Opt. Soc. Am. A*, 20(6):1025–1034, Jun 2003.

[16] Pengwen Chen and Albert Fannjiang. Fourier phase retrieval with a single mask by Douglas–Rachford algorithms. *Applied and Computational Harmonic Analysis*, 44(3):665 – 699, 2018.

[17] Zaiwen Wen, Chao Yang, Xin Liu, and Stefano Marchesini. Alternating direction methods for classical and ptychographic phase retrieval. *Inverse Problems*, 28(11):115010, oct 2012.

[18] D Russell Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems*, 21(1):37–50, nov 2004.

[19] Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, Apr 1992.

[20] Bingsheng He and Xiaoming Yuan. On the convergence rate of Douglas-Rachford operator splitting method. *Mathematical Programming*, 153(2):715–722, 2015.

[21] Ji Li and Tie Zhou. On relaxed averaged alternating reflections (RAAR) algorithm for phase retrieval with structured illumination. *Inverse Problems*, 33(2):025012, jan 2017.

[22] Dimitri P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods (Optimization and Neural Computation Series)*. Athena Scientific, 1 edition, 1996.

[23] Mingyi Hong, Zhi-Quan Luo, and Meisam Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM Journal on Optimization*, 26(1):337–364, 2016.

[24] Guoyin Li and Ting Kei Pong. Global convergence of splitting methods for nonconvex composite optimization. *SIAM Journal on Optimization*, 25(4):2434–2460, 2015.

[25] Yu Wang, Wotao Yin, and Jinchuan Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78(1):29–63, 2019.

[26] D.P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Computer science and applied mathematics. Elsevier Science, 2014.

[27] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011.

[28] J. Sun, Q. Qu, and J. Wright. A geometric analysis of phase retrieval. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2379–2383, July 2016.

[29] Jason D. Lee, Max Simchowitz, Michael I. Jordan, and Benjamin Recht. Gradient descent only

---

**Figure 8.** The columns from left to right show the reconstruction of 300 RAAR iterations in the case (c,d), corresponding to $\beta$-paths with $\beta = 0.95, 0.9, 0.8, 0.7$ and $0.6$, respectively. The top row is the case (c) and the bottom row is the case (d).
Local saddles of RAAR algorithms

converges to minimizers. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, 29th Annual Conference on Learning Theory, volume 49 of Proceedings of Machine Learning Research, pages 1246–1257, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR.

[30] Simon S Du, Chi Jin, Jason D Lee, Michael I Jordan, Aarti Singh, and Barnabas Poczos. Gradient descent can take exponential time to escape saddle points. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 30, pages 1067–1077. Curran Associates, Inc., 2017.

[31] Chi Jin, Rong Ge, Praneeth Netrapalli, Sham M Kakade, and Michael I Jordan. How to escape saddle points efficiently. In 34th International Conference on Machine Learning, ICML 2017, pages 2727–2752. International Machine Learning Society (IMLS), 2017.

[32] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. Communications of the ACM, 63(11):139–144, 2020.

[33] Shayegan Omidshafiei, Jason Pazis, Christopher Amato, Jonathan P How, and John Vian. Deep decentralized multi-task multi-agent reinforcement learning under partial observability. arXiv preprint arXiv:1703.06182, 2017.

[34] Leonard Adolphs, Hadi Daneshmand, Aurelien Lucchi, and Thomas Hofmann. Local saddle point optimization: A curvature exploitation approach. In The 22nd International Conference on Artificial Intelligence and Statistics, pages 486–495. PMLR, 2019.

[35] Constantinos Daskalakis and Ioannis Panageas. The limit points of (optimistic) gradient descent in min-max optimization. Advances in Neural Information Processing Systems, 31:9236–9246, 2018.

[36] Chi Jin, Praneeth Netrapalli, and Michael Jordan. What is local optimality in nonconvex-nonconcave minimax optimization? In International Conference on Machine Learning, pages 4880–4889. PMLR, 2020.

[37] Yu-Hong Dai and Liwei Zhang. Optimality conditions for constrained minimax optimization. CSIAM Transactions on Applied Mathematics, 1(2):296–315, 2020.

[38] Albert Fannjiang. Absolute uniqueness of phase retrieval with random illumination. Inverse Problems, 28(7):075008, jun 2012.

[39] Albert Fannjiang and Zheqing Zhang. Fixed point analysis of douglas–rachford splitting for ptychography and phase retrieval. SIAM Journal on Imaging Sciences, 13(2):609–650, 2020.

[40] R. Glowinski and A. Marroco. Sur l’approximation, par ´el´ement finis d’ordre un, et la r´esolution, par p´enalisation-dualit´e d’une classe de probl`emes de dirichlet non lin´eaires. ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique, 9(R2):41–76, 1975.

[41] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Computers Mathematics with Applications, 2(1):17–40, 1976.

[42] Junfeng Yang, Yin Zhang, and Wotao Yin. An efficient tvl1 algorithm for deblurring multichannel images corrupted by impulse noise. SIAM Journal on Scientific Computing, 31(4):2842–2865, 2009.

[43] Bhabesh Deka and Sumit Datta. Compressed sensing magnetic resonance image reconstruction algorithms. Springer Series on Bio-and Neurosystems, 2019.

[44] A. Cherukuri, B. Gharesifard, and J. Cortés. Saddle-point dynamics: Conditions for asymptotic stability of saddle points. SIAM Journal on Control and Optimization, 55(1):486–511, 2017.

[45] P Thibault and M Guizar-Sicairos. Maximum-likelihood refinement for coherent diffractive imaging. New Journal of Physics, 14(6):063004, jun 2012.

[46] Huibin Chang, Yifei Lou, Michael K. Ng, and Tieyong Zeng. Phase retrieval from incomplete magnitude information via total variation regularization. SIAM J Sci Comput, 38(6):A3672–A3695, 2016.
[47] Y. C. Eldar and S. Mendelson. Phase retrieval: Stability and recovery guarantees. *Applied and Computational Harmonic Analysis*, 36(3):473–494, 2014.