Convolutional Phase Retrieval via Gradient Descent

Qing Qu†, Yuqian Zhang†, Yonina Eldar*, John Wright†,
† Department of Electrical Engineering, Columbia University, New York, 10027
* Department of Electrical Engineering, Israel Institute of Technology, Haifa, 32000

We consider the problem of recovering an unknown signal \( x \in \mathbb{C}^n \) from measurements \( y = |a \circ x| \), where \( a \in \mathbb{C}^m \) \((m \geq n)\) is a given kernel, \( \circ \) denotes the cyclic convolution. Let \( C_a \in \mathbb{C}^{m \times m} \) be the circulant matrix generated by \( a \), and let \( A \in \mathbb{C}^{m \times n} \) denote the first \( n \) columns of \( C_a \). The observation model can also be written in the matrix-vector form as

\[
y = |a \circ x| = |Ax|.
\]

This problem is motivated by applications such as channel estimation [1], noncoherent optical communication [2], and underwater acoustic communication [3]. In these scenarios, the phase measurements can be very noisy and unreliable, while their magnitudes are often much easier to obtain. On the other hand, we know that if \( A \) is generic, the general phase retrieval [4], [5], [6] is \( O(mn) \) per iteration cost. In comparison, the benign structure of the convolutional model allows us to design much more efficient methods with \( O(m \log m) \) memory and computational cost, by using the fast Fourier transform for matrix-vector products.

In this work, we consider a generic random model in which the kernel \( a = [a_1, \ldots, a_m] \) is complex Gaussian

\[
a_k = \frac{1}{\sqrt{2}} (X_k + iY_k), \quad X_k \sim \mathcal{N}(0, 1), \quad Y_k \sim \mathcal{N}(0, 1),
\]

and we solve the problem by minimizing a weighted\(^1\) nonconvex and nonsmooth objective

\[
f(z) = \frac{1}{2m} \|b \circ (y - |Az|)\|^2, \quad (1)
\]

where the weights \( b = c^{1/2} \delta(y) \) that

\[
\zeta_\sigma^2(t) = 1 - 2\pi^2 \sigma^2 \xi_\sigma^2(t), \quad \xi_\sigma^2(t) = \frac{1}{2\sigma^2} \exp \left( -\frac{|t|^2}{2\sigma^2} \right),
\]

and \( \sigma^2 > 1/2 \) is a numerical constant.

We analyze a local\(^2\) (generalized) gradient descent method. The same as [9], [10], the algorithm is initialized via a spectral method. Although the objective \((1)\) is nonsmooth, by defining the phase of \( u \in \mathbb{C} \) as

\[
\exp (i\phi(u)) = \begin{cases} u/|u| & \text{if } |u| \neq 0, \\ 1 & \text{otherwise}. \end{cases}
\]

\(^1\)The introduction of weights is purely for the ease of analysis.
\(^2\)It would be nicer to characterize the global geometry of the problem as we did in [7], [8]. However, the nonhomogeneity of \( \|C_a\| \) over the space causes tremendous difficulties for concentration with \( m \geq \Omega(n \log \log n) \) samples.

the generalized Wirtinger gradient [11] of \((1)\) can be uniquely specified as

\[
\frac{\partial}{\partial z} f(z) \doteq \frac{1}{m} A^* \text{diag} (b) [Az - y \circ \exp (i\phi(Az))],
\]

where \( \circ \) denotes the Hadamard product. Thus for the \( k \)th iterate, the gradient descent step takes the form

\[
z_{k+1} = z_k - \tau \frac{\partial}{\partial z_k} f(z_k),
\]

where \( \tau \) is the stepsize. If we define

\[
dist(z, x) = \inf_{\theta \in [0, 2\pi]} \|xe^{i\theta} - z\|,
\]

then we show that gradient descent converges linearly in a small region close to the optimal by the following theorem.

**Theorem 0.1:** Whenever \( m \geq C_0n \log^3 n \), spectral method [9], [10] produces an initialization \( z_0 \) that satisfies

\[
dist(z_0, \mathcal{X}) \leq c_0 \log^{-6} n \|x\|
\]

with probability at least \( 1 - c_1 \min \left\{ \sqrt{m/c_2}, \left( \frac{m^{1/4}}{c_3 \log^{3/4} n} \right) \right\} \). Starting from the initialization \( z_0 \), whenever \( m \geq C_1 \max \{ \log^7 n, n \log^4 n \} \), with \( \sigma^2 = 0.51 \) and stepsize \( \tau = 2.02 \), we have for every \( k \geq 1 \)

\[
dist(z_k, \mathcal{X}) \leq (1 - \delta)^k \dist(z_0, \mathcal{X}), \quad (2)
\]

holds for some small constant \( \delta \in (0, 1) \) with probability at least \( 1 - c_4 m^{-c_5} \). Here, \( c_0, c_1, c_2, c_3, c_4, c_5, C_0 \) and \( C_1 \) are some positive numerical constants.

In Theorem 0.1, a dependence of the sample complexity \( m \) on \( \|C_a\| \) seems necessary; please see experiments in Fig. 1 for the demonstration. Our proof is based on ideas from decoupling theory [12], the restricted isometry property of random circulant matrices [13], and a new analysis of alternating projection method due to [14]. More specifically, instead of using restricted strong convexity as [10], [15] to show that the iterates contract, our analysis is largely inspired by the recent work of [14]. This argument controls the bulk effect of phase errors uniformly in a neighborhood around the ground truth signal \( x \), avoiding the need to analyze high-order moments of the highly structured and inhomogeneous random process \( |Az| \).

Finally, Fig. 2 demonstrates the proposed method on a real image. As we observe from Fig. 1 and Fig. 2, the sample complexity in Theorem 0.1 loose by at least a few log factors. Improving this is a direction for future work.
Fig. 1: Phase transition for signal $x \in \mathbb{C}^n$ with different $\|Cx\|$. We normalize the signal with $\|x\| = 1$, fix $n = 1000$ and vary the ratio $m/n$. (a) shows the case when $x$ is a standard basis vector; (b) shows the case when $x$ is uniformly generated on the complex sphere $\mathbb{S}^{n-1}$, where $\|Cx\| \sim O(\|x\|)$; (c) shows the case when $x = \frac{1}{\sqrt{n}} 1_n$, such that $\|Cx\| = \sqrt{n} \|x\|$.

Fig. 2: Experiment on real images. The image is of size $200 \times 300$, we vectorize the image and use $m = 5n \log n$ samples for reconstruction. The kernel $a \in \mathbb{C}^m$ is randomly generated as complex Gaussian. We run power method for 100 iterations for initialization, and stop the algorithm once the error is smaller than $1 \times 10^{-4}$. It takes 197.08s to reconstruct all the RGB channels. Methods using general Gaussian measurements $A \in \mathbb{C}^{m \times n}$ could easily run out of memory on a personal computer for problems of this size.

References