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

$$\boldsymbol{y} = |\boldsymbol{a} \circledast \boldsymbol{x}| = |\boldsymbol{A}\boldsymbol{x}| .$$

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 $\boldsymbol{A}$ is generic, the general phase retrieval [4], [5], [6] is $\mathcal{O}(mn)$ per iteration cost. In comparison, the benign structure of the convolutional model allows us to design much more efficient methods with $\mathcal{O}(m \log m)$ 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 $\boldsymbol{a} = [a_1, \cdots, a_m]^\top$ is complex Gaussian

$$a_k = \frac{1}{\sqrt{2}}\left(X_k + \mathrm{i} Y_k\right), \quad X_k \sim \mathcal{N}(0,1),\ Y_k \sim \mathcal{N}(0,1),$$

and we solve the problem by minimizing a weighted[1] nonconvex and nonsmooth objective

$$f(\boldsymbol{z}) = \frac{1}{2m}\left\|\boldsymbol{b} \odot (\boldsymbol{y} - |\boldsymbol{A}\boldsymbol{z}|)\right\|^2, \tag{.1}$$

where the weights $\boldsymbol{b} = \zeta_{\sigma^2}^{1/2}(\boldsymbol{y})$ that

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

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

$$\exp\left(\mathrm{i}\phi(u)\right) \doteq \begin{cases} u/|u| & \text{if } |u| \neq 0, \\ 1 & \text{otherwise.} \end{cases}$$

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

$$\frac{\partial}{\partial \boldsymbol{z}} f(\boldsymbol{z}) \doteq \frac{1}{m}\boldsymbol{A}^* \operatorname{diag}(\boldsymbol{b})\left[\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y} \odot \exp\left(\mathrm{i}\phi(\boldsymbol{A}\boldsymbol{z})\right)\right],$$

where $\odot$ denotes the Hadamard product. Thus for the kth iterate, the gradient descent step takes the form

$$\boldsymbol{z}_{k+1} = \boldsymbol{z}_k - \tau \frac{\partial}{\partial \boldsymbol{z}_k} f(\boldsymbol{z}_k),$$

where τ is the stepsize. If we define

$$\operatorname{dist}(\boldsymbol{z}, \boldsymbol{x}) \doteq \inf_{\theta \in [0, 2\pi)} \left\|\boldsymbol{x} e^{\mathrm{i}\theta} - \boldsymbol{z}\right\|,$$

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

Theorem 0.1: Whenever $m \geq C_0 n \log^{31} n$, spectral method [9], [10] produces an initialization $\boldsymbol{z}_0$ that satisfies

$$\operatorname{dist}(\boldsymbol{z}_0, \mathcal{X}) \leq c_0 \log^{-6} n \left\|\boldsymbol{x}\right\|$$

with probability at least $1 - c_1 \min\left\{\sqrt{m}^{-c_2}, \left(m^{1/4}\right)^{-c_3 \log^{3/4} n}\right\}$. Starting from the initialization $\boldsymbol{z}_0$, whenever $m \geq C_1 \frac{\|\boldsymbol{C_x}\|^2}{\|\boldsymbol{x}\|^2} \max\left\{\log^{17} n, n \log^4 n\right\}$, with $\sigma^2 = 0.51$ and stepsize $\tau = 2.02$, we have for every $k \geq 1$

$$\operatorname{dist}(\boldsymbol{z}_k, \mathcal{X}) \leq (1-\delta)^k \operatorname{dist}(\boldsymbol{z}_0, \mathcal{X}), \tag{.2}$$

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

In Theorem 0.1, a dependence of the sample complexity m on $\|\boldsymbol{C_x}\|$ 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 $\boldsymbol{x}$, avoiding the need to analyze high-order moments of the highly structured and inhomogeneous random process $|\boldsymbol{A}\boldsymbol{z}|$.

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.

[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 $\|\boldsymbol{C_z}\|$ over the space causes tremendous difficulties for concentration with $m \geq \Omega(n \operatorname{poly} \log n)$ samples.

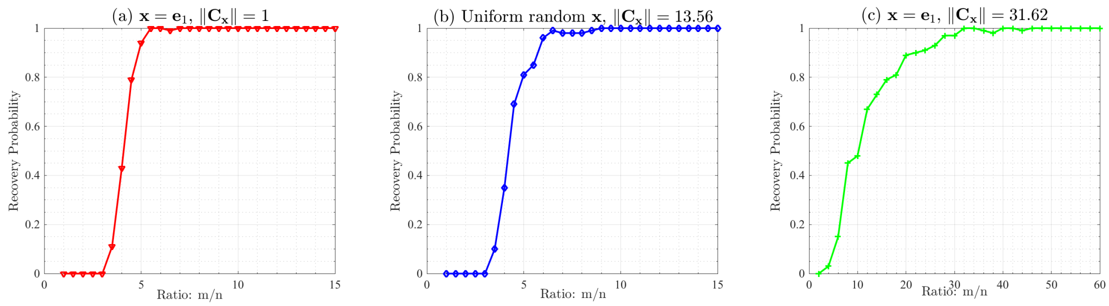

Fig. 1: Phase transition for signal $\boldsymbol{x} \in \mathbb{C}^n$ with different $\|\boldsymbol{C_x}\|$. We normalize the signal with $\|\boldsymbol{x}\| = 1$, fix $n = 1000$ and vary the ratio m/n. (a) shows the case when $\boldsymbol{x}$ is a standard basis vector; (b) shows the case when $\boldsymbol{x}$ is uniformly generated on the complex sphere $\mathbb{CS}^{n-1}$, where $\|\boldsymbol{C_x}\| \sim \mathcal{O}(\|\boldsymbol{x}\|)$; (c) shows the case when $\boldsymbol{x} = \frac{1}{\sqrt{n}}\mathbf{1}_n$, such that $\|\boldsymbol{C_x}\| = \sqrt{n}\,\|\boldsymbol{x}\|$.

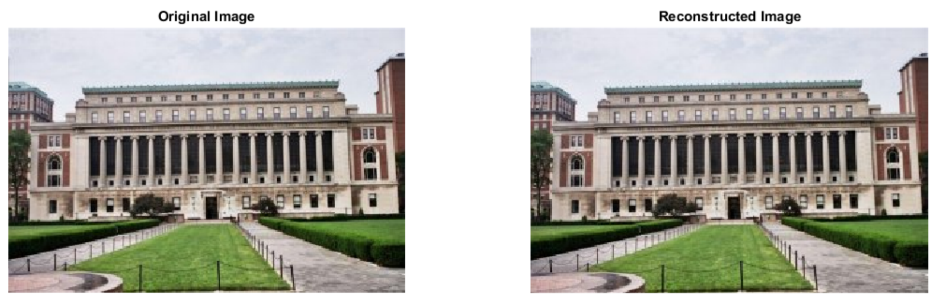

Fig. 2: Experiment on real images. The image is of size 200×300, we vectorize the image and use $m = 5n \log n$ samples for reconstruction. The kernel $\boldsymbol{a} \in \mathbb{C}^m$ is randomly generated as complex Gaussian. We run power method for 100 iterations for initialization, and stop the algorithm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

References

[1] P. Walk, H. Becker, and P. Jung, "Ofdm channel estimation via phase retrieval," in *Asilomar 2015*, 2015.

[2] R. M. Gagliardi and S. Karp, "Optical communications," *New York, Wiley-Interscience, 1976. 445 p.*, vol. 1, 1976.

[3] M. Stojanovic, J. A. Catipovic, and J. G. Proakis, "Phase-coherent digital communications for underwater acoustic channels," *IEEE Journal of Oceanic Engineering*, vol. 19, no. 1, pp. 100–111, 1994.

[4] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, 2013.

[5] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[6] 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*, vol. 32, pp. 87–109, May 2015.

[7] J. Sun, Q. Qu, and J. Wright, "When are nonconvex problems not scary?," *arXiv preprint arXiv:1510.06096*, 2015.

[8] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retreival," *arXiv preprint arXiv:1602.06664*, 2016.

[9] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Advances in Neural Information Processing Systems*, pp. 2796–2804, 2013.

[10] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *Information Theory, IEEE Transactions on*, vol. 61, pp. 1985–2007, April 2015.

[11] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," *arXiv preprint arXiv:0906.4835*, 2009.

[12] V. De la Pena and E. Giné, *Decoupling: from dependence to independence*. Springer, 1999.

[13] F. Krahmer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property," *Communications on Pure and Applied Mathematics*, vol. 67, no. 11, pp. 1877–1904, 2014.

[14] I. Waldspurger, "Phase retrieval with random gaussian sensing vectors by alternating projections," *arXiv preprint arXiv:1609.03088*, 2016.

[15] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *arXiv preprint arXiv:1505.05114*, 2015.