A Local Analysis of Block Coordinate Descent for Gaussian Phase Retrieval

David Barmherzig and Ju Sun

Stanford University

davidbar@stanford.edu,sunju@stanford.edu

1. Introduction

- Converge of ADMM is guaranteed for convex functions
- Also works well on nonconvex functions empirically but not well-understood
- We analyze phase retrieval as a model nonconvex function to understand local convergence

2. Phase Retrieval and ADMM

The Gaussian Phase Retrieval (GPR) problem: Recover $\boldsymbol{x} \in \mathbb{R}^n$ from the measurements $y_k = |\boldsymbol{a}_k^* \boldsymbol{x}|, k = 1, \cdots, m$ ($\boldsymbol{a}_k$'s are i.i.d Gaussian vectors).
A biconvex least-squares formulation:

$$\begin{aligned} &\text{minimize}_{\boldsymbol{z},\boldsymbol{w}\in\mathbb{R}^n} \quad f(\boldsymbol{z},\boldsymbol{w}) \doteq \frac{1}{4m}\sum_{k=1}^{m}\left(y_k^2 - \boldsymbol{a}_k^T \boldsymbol{z} \boldsymbol{a}_k^T \boldsymbol{w}\right)^2 \\ &\text{subject to} \quad \boldsymbol{z} = \boldsymbol{w}. \end{aligned} \tag{2.1}$$

ADMM updates for Eq. (2.1):

$$\begin{aligned} \boldsymbol{z}^{(k+1)} &= \arg\min_{\boldsymbol{z}} \mathcal{L}\left(\boldsymbol{z}, \boldsymbol{w}^{(i)}, \boldsymbol{\lambda}^{(k)}\right), \\ \boldsymbol{w}^{(k+1)} &= \arg\min_{\boldsymbol{w}} \mathcal{L}\left(\boldsymbol{z}^{(k+1)}, \boldsymbol{w}, \boldsymbol{\lambda}^{(k)}\right), \\ \boldsymbol{\lambda}^{(k+1)} &= \boldsymbol{\lambda}^{(k)} + \rho\left(\boldsymbol{z}^{(k+1)} - \boldsymbol{w}^{(k+1)}\right), \end{aligned} \tag{2.2}$$

where $\rho > 0$ and $\mathcal{L}$ is the augmented Lagrangian:

$$\mathcal{L}(\boldsymbol{z},\boldsymbol{w},\boldsymbol{\lambda}) \doteq f(\boldsymbol{z},\boldsymbol{w}) + \langle \boldsymbol{\lambda}, \boldsymbol{z} - \boldsymbol{w}\rangle + \frac{\rho}{2}\|\boldsymbol{z}-\boldsymbol{w}\|^2. \tag{2.3}$$

3. A natural reduction to block coordinate descent

Lemma 3.1. *A triple $(\boldsymbol{z},\boldsymbol{w},\boldsymbol{\lambda})$ is a critical point of $\mathcal{L}(\boldsymbol{z},\boldsymbol{w},\boldsymbol{\lambda})$ if and only if*

$$\partial_{\boldsymbol{z}} f = \partial_{\boldsymbol{w}} f = \boldsymbol{0}, \quad \boldsymbol{z} = \boldsymbol{w}, \quad \boxed{\boldsymbol{\lambda} = \boldsymbol{0}}. \tag{3.1}$$

Motivates fixing $\boldsymbol{\lambda} = \boldsymbol{0}$ so that ADMM becomes block coordinate descent (BCD):

$$\begin{aligned} \boldsymbol{z}^{(k+1)} &= \arg\min_{\boldsymbol{z}} f(\boldsymbol{z}, \boldsymbol{w}^{(k)}) + \frac{\rho}{2}\|\boldsymbol{z} - \boldsymbol{w}^{(k)}\|^2, \\ \boldsymbol{w}^{(k+1)} &= \arg\min_{\boldsymbol{w}} f(\boldsymbol{z}^{(k+1)}, w) + \frac{\rho}{2}\|\boldsymbol{z}^{(k+1)} - \boldsymbol{w}\|^2. \end{aligned} \tag{3.2}$$

4. Linear Convergence

Consider the expected objective function:

$$g(\boldsymbol{z},\boldsymbol{w}) \doteq \mathbb{E}\left[f(\boldsymbol{z},\boldsymbol{w})\right] + \frac{\rho}{2}\|\boldsymbol{z}-\boldsymbol{w}\|^2 \tag{4.1}$$

$$= \frac{3}{2}\|\boldsymbol{x}\|^4 + \left(\boldsymbol{w}^T\boldsymbol{z}\right)^2 + \frac{1}{2}\|\boldsymbol{z}\|^2\|\boldsymbol{w}\|^2 - 2\boldsymbol{x}^T\boldsymbol{z}\boldsymbol{x}^T\boldsymbol{w} - \|\boldsymbol{x}\|^2\boldsymbol{w}^T\boldsymbol{z} + \frac{\rho}{2}\|\boldsymbol{z}-\boldsymbol{w}\|^2. \tag{4.2}$$

Using a spectral initialization (i.e., see [1]), we may assume $\left(\boldsymbol{z}^{(0)}, \boldsymbol{w}^{(0)}\right)$ lies within the set:

$$N_{\boldsymbol{x}} \doteq \left\{(\boldsymbol{z},\boldsymbol{w}) : \|\boldsymbol{z}-\boldsymbol{x}\| \le \frac{1}{8}\|\boldsymbol{x}\| \text{ and } \|\boldsymbol{w}-\boldsymbol{x}\| \le \frac{1}{8}\|\boldsymbol{x}\|\right\}. \tag{4.3}$$

Lemma 4.1 (Local strong convexity). *Suppose $\rho \ge \|\boldsymbol{x}\|^2$. For all $(\boldsymbol{z},\boldsymbol{w}) \in N_{\boldsymbol{x}}$,*

$$\nabla^2 g(\boldsymbol{z},\boldsymbol{w}) \succeq \frac{1}{3}\|\boldsymbol{x}\|^2\boldsymbol{I}. \tag{4.4}$$

Lemma 4.2 (No-escape). *Suppose $\rho \ge \frac{27}{8}\|\boldsymbol{x}\|^2$. The BCD iterate sequence $\left\{\left(\boldsymbol{z}^{(k)}, \boldsymbol{w}^{(k)}\right)\right\}$ stays in $N_{\boldsymbol{x}}$.*

Lemma 4.3 (Locally block Lipschitz). *$g(\boldsymbol{z},\boldsymbol{w})$ is block Lipschitz on $N_{\boldsymbol{x}}$, i.e., for all $(\boldsymbol{z},\boldsymbol{w}) \in N_{\boldsymbol{x}}$ and all $\boldsymbol{h}_{\boldsymbol{z}}, \boldsymbol{h}_{\boldsymbol{w}} \in \mathbb{R}^n$,*

$$\|\nabla_{\boldsymbol{z}} g(\boldsymbol{z}+\boldsymbol{h}_{\boldsymbol{z}},\boldsymbol{w}) - \nabla_{\boldsymbol{z}} g(\boldsymbol{z},\boldsymbol{w})\| \le (4\|\boldsymbol{x}\|^2+\rho)\|\boldsymbol{h}_{\boldsymbol{z}}\|, \tag{4.5}$$

$$\|\nabla_{\boldsymbol{w}} g(\boldsymbol{z},\boldsymbol{w}+\boldsymbol{h}_{\boldsymbol{w}}) - \nabla_{\boldsymbol{z}} g(\boldsymbol{z},\boldsymbol{w})\| \le (4\|\boldsymbol{x}\|^2+\rho)\|\boldsymbol{h}_{\boldsymbol{w}}\|. \tag{4.6}$$

With local convexity in place, we establish a linear convergence rate:

Theorem 4.4. *Suppose $\rho \ge \frac{27}{8}\|\boldsymbol{x}\|^2$ and $(\boldsymbol{z}^{(0)}, \boldsymbol{w}^{(0)}) \in N_{\boldsymbol{x}}$. Then, the BCD iterate sequence converges linearly to the point $(\boldsymbol{x},\boldsymbol{x})$ with the rate given by:*

$$\left\|\left(\boldsymbol{z}^{(k)}, \boldsymbol{w}^{k}\right) - (\boldsymbol{x},\boldsymbol{x})\right\| \le \left(1 - \frac{\|\boldsymbol{x}\|^2}{12\|\boldsymbol{x}\|^2 + 3\rho}\right)^{k/2} \sqrt{\frac{6}{\|\boldsymbol{x}\|}\left[g\left(\boldsymbol{z}^{(0)}, \boldsymbol{w}^{(0)}\right) - g(\boldsymbol{x},\boldsymbol{x})\right]}. \tag{4.7}$$

5. Acknowledgments

The authors would like to foremostly thank their research advisor, Professor Emmanuel J. Candès, for introducing and guiding this research. D.B. is also very grateful to Professor Walter Murray and Professor Gordon Wetzstein for many helpful discussions, etc.

References

[1] Mondelli, M. and Montanari, A. Fundamental Limits of Weak Recovery with Applications to Phase Retrieval. *arXiv preprint*, arXiv:1708.05932.