An Undersampled Phase Retrieval Algorithm via Gradient Iteration

Qiang Li and Lei Huang
College of Information Engineering, Shenzhen University, Shenzhen 518060, China

Wei Liu
Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield S1 4ET, United Kingdom

9 July 2018

1.Introduction

- The problem of estimating an N-dimensional complex signal $\boldsymbol{x}$ from M magnitude-only linear measurements $\boldsymbol{y}$ is called phase retrieval.
- A basic phase retrieval model with intensity measurements is

$$y_i = |(\boldsymbol{A}\boldsymbol{x})_i|^2 + n_i, \quad i = 1, \cdots, M \tag{1}$$

  where $|\cdot|$ is the element-wise magnitude, y_i and complex measurement matrix $\boldsymbol{A} \in \mathbb{C}^{M \times N}$ are known beforehand and $\boldsymbol{n} = [n_1, \cdots, n_M]^T$ denotes noise.
- Phase retrieval is an inherently non-convex ill-posed inverse problem. Normally, to recover a signal with large probability, the number of measurements needs to be greater than the dimensions of incident signal.
- In practice, the undersampled phase retrieval problem is often encountered, which refers to the case of $M < N$.
- In this work, we consider undersampled phase retrieval and assume that the incident signal is sparse.

2. Phase Retrieval by Majorization-Minimization (MM) Technique

- The undersampled phase retrieval problem as the following optimization model

$$\min_{\boldsymbol{x}} \sum_{i=1}^{M} \left(\sqrt{y_i} - |(\boldsymbol{A}\boldsymbol{x})_i|\right)^2 + \lambda \left\|\boldsymbol{x}\right\|_1 \tag{2}$$

where the parameter $\lambda > 0$ is a regularization penalty factor and $\left\|\boldsymbol{x}\right\|_1$ denotes ℓ_1 norm of vector $\boldsymbol{x}$, which is used to regularize the ill-posed phase retrieval problem and promote sparsity in $\boldsymbol{x}$.

- Employing the MM technique, in [13], an efficient C-PRIME method was proposed to solve a convex surrogate problem instead. The surrogate optimization problem is convex with regard to $\boldsymbol{x}$ and equivalent to the following problem

$$\boldsymbol{x} = \arg\min_{\boldsymbol{x}} \left[C \left\|\boldsymbol{x} - \boldsymbol{c}\right\|_2^2 + \lambda \left\|\boldsymbol{x}\right\|_1 \right] \tag{3}$$

where C is a constant satisfying $C \geq \lambda_{\text{max}}(\boldsymbol{A}^H\boldsymbol{A})$ and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix.

- In (3), the vector $\boldsymbol{c}$ is a constant independent of the variable $\boldsymbol{x}$:

$$\boldsymbol{c} = \boldsymbol{x}^{k-1} - \frac{1}{C}\boldsymbol{A}^{H}\left(\boldsymbol{A}\boldsymbol{x}^{k-1} - \sqrt{\boldsymbol{y}} \odot e^{j\,\mathrm{ang}(\boldsymbol{A}\boldsymbol{x}^{k-1})}\right) \tag{4}$$

- The C-PRIME method solves the surrogate optimization problem in (3) with a simple closed-form solution at the k iteration

$$\boldsymbol{x}^{k} = e^{j\,\mathrm{ang}(\boldsymbol{c})} \odot \max\left\{|\boldsymbol{c}| - \frac{\lambda}{2C}, 0\right\} \tag{5}$$

where $\mathrm{ang}(\cdot)$ denotes the phase angle and $\odot$ denotes the Hadamard (element-wise) product of two vectors.

3. Proposed Phase Retrieval based on the Gradient Framework

- The optimization in (3) can be considered the following general formulation

$$\boldsymbol{x} = \arg\min_{\boldsymbol{x}} \left[F(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x}) \right] \tag{6}$$

  where f is a smooth convex function and g is a continuous convex function which is possibly nonsmooth.

- For the optimization problem (3), let $f(\boldsymbol{x}) = C \|\boldsymbol{x} - \boldsymbol{c}\|_2^2$ and $g(\boldsymbol{x}) = \lambda \|\boldsymbol{x}\|_1$.

- We considered a given quantity $\boldsymbol{\eta}$ which may or may not be equal to $\boldsymbol{x}_{k-1}$. According to Taylor series expansion and the proximal regularization theorem [14], for the given point $\boldsymbol{\eta}$, a quadratic approximation of $F(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})$ can be written as

$$Q_L(\boldsymbol{x}, \boldsymbol{\eta}) = f(\boldsymbol{\eta}) + \langle \boldsymbol{x} - \boldsymbol{\eta}, \nabla f(\boldsymbol{\eta}) \rangle + \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{\eta}\|^2 + g(\boldsymbol{x}) \tag{7}$$

  where L plays the role of a step and $\nabla f(\cdot)$ is the complex gradient vector.

- Then, we have

$$\boldsymbol{x}^k = \arg\min_{\boldsymbol{x}} \left\{ Q_L(\boldsymbol{x}, \boldsymbol{\eta}) \right\} \tag{8}$$

- Discarding the constant term about $\boldsymbol{\eta}$, the optimization function (8) is simplified as

$$\boldsymbol{x}^k = \arg\min \left\{ g\left(\boldsymbol{x}\right) + \frac{L}{2} \left\| \boldsymbol{x} - \left(\boldsymbol{\eta} - \frac{1}{L} \nabla f\left(\boldsymbol{\eta}\right) \right) \right\|^2 \right\} \tag{9}$$

- We know that $f\left(\boldsymbol{x}\right) = C \left\| \boldsymbol{x} - \boldsymbol{c} \right\|_2^2$ and then can get

$$\nabla f\left(\boldsymbol{\eta}\right) = 2C\left(\boldsymbol{\eta} - \boldsymbol{c}\right) \tag{10}$$

- After that, $\boldsymbol{x}^k$ can be represented as

$$\boldsymbol{x}^k = \arg\min \left\{ \lambda \left\| \boldsymbol{x} \right\|_1 + \frac{L}{2} \left\| \boldsymbol{x} - \left[\boldsymbol{\eta} - \frac{2C}{L}\left(\boldsymbol{\eta} - \boldsymbol{c}\right) \right] \right\|^2 \right\} \tag{11}$$

- Furthermore, according to the soft thresholding method [16], we have

$$\boldsymbol{x}^k = e^{j\,\mathrm{ang}(\boldsymbol{b})} \odot \max\left\{ \left|\boldsymbol{b}\right| - \frac{\lambda}{L}, 0 \right\} \tag{12}$$

where

$$\boldsymbol{b} = \boldsymbol{\eta} - \frac{2C}{L}\left(\boldsymbol{\eta} - \boldsymbol{c}\right) \tag{13}$$

- Then, if $\boldsymbol{\eta} = \boldsymbol{x}^{k-1}$, substituting (4) into (13) and simplifying it, we have

$$\boldsymbol{b} = \boldsymbol{x}^{k-1} - \frac{2}{L}\boldsymbol{A}^H\left(\boldsymbol{A}\boldsymbol{x}^{k-1} - \sqrt{\boldsymbol{y}} \odot e^{j\,\mathrm{ang}(\boldsymbol{A}\boldsymbol{x}^{k-1})}\right) \tag{14}$$

- It is interesting that we obtain the same solution of the problem (3) as the C-PRIME algorithm in the case of $L = 2C$ but from a totally different gradient theorem.
- Moreover, the C-PRIME method can be regarded as a special case of the proposed G-PRIME in the case of $\boldsymbol{\eta} = \boldsymbol{x}^{k-1}$.

4. Design Examples

- The measurement matrix $\boldsymbol{A}$ is standard complex Gaussian distributed, corrupted with real-valued additive white Gaussian noise and the original complex signal is generated randomly.
- The length N of the original complex signal is set as 128 with sparsity level $P = 8$ and the number of measurements is 120.
- The signal-to-noise ratio is SNR=25dB. The parameter C and regularization penalty factor λ in all tested methods are set as $C = \lambda_{\text{max}}(\boldsymbol{A}^H\boldsymbol{A})$ and $\lambda = 0.1$, respectively.
- We assign step size $L = 2C$ for our proposed G-PRIME algorithm unless specified otherwise.
- For the ISTA-PRIME algorithm [15], the iterative step size μ should satisfy $\mu \in (0, 1/||\boldsymbol{A}^H\boldsymbol{A}||]$.

5. Results

- Magnitude recovery ability: for the proposed G-PRIME algorithm, the magnitude curves are shown in Fig. 1 at the 200th iteration.
- It is observed that the nonzero values in the recovered signal are almost the same as those in the original signal, which proves that the G-PRIME algorithm can recover the magnitude information successfully.

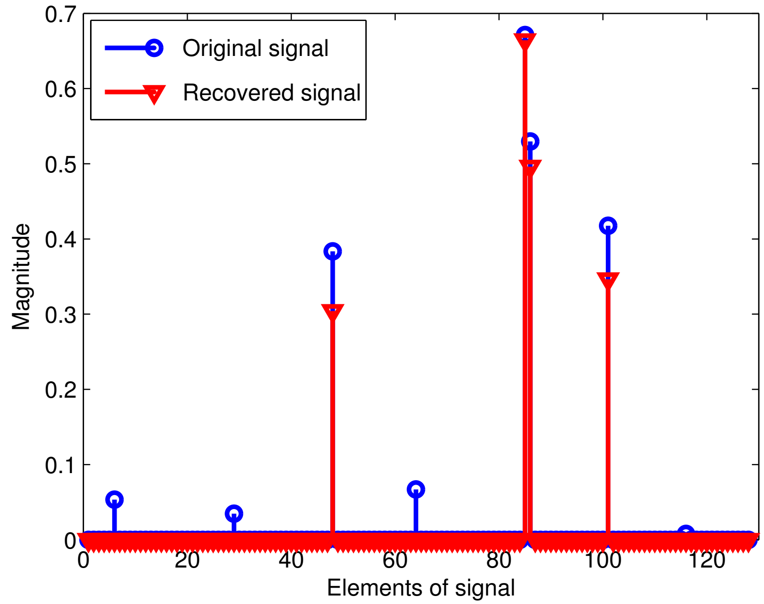

Figure 1: Magnitudes of original and recovered signals.

- Phase recovery ability: Fig. 2 plots the recovered signal at iteration $k = 1, 20, 200$.

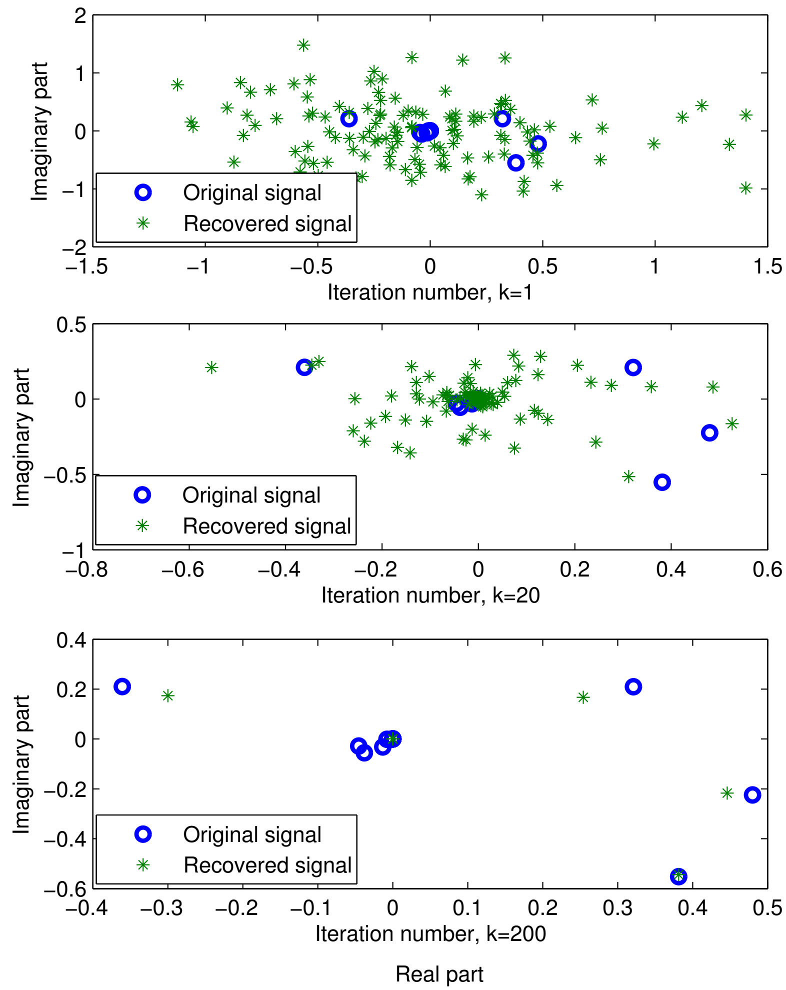

Figure 2: Original and recovered signals.

- The recovered signal is a random complex vector at the first iteration and the position of recovered signal is already close to that of the original signal after 200 iterations. The phase recovery ability of the G-PRIME algorithm is also excellent.

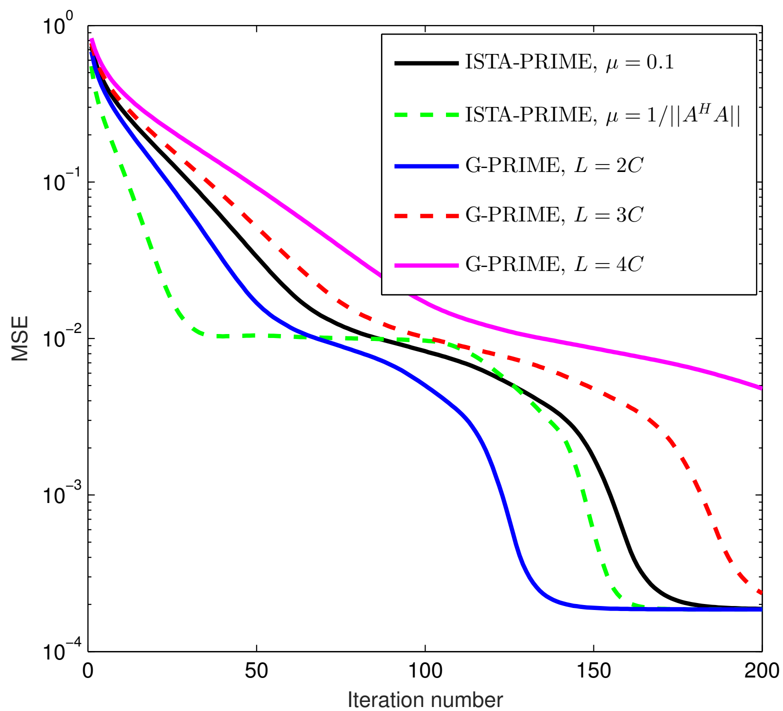

Figure 3: MSE versus iteration number.

- The MSE performance of the ISTA-PRIME ($\mu = 0.1, 1/||\boldsymbol{A}^H\boldsymbol{A}||$) [15], C-PRIME [13] and the proposed G-PRIME algorithms are shown in Fig. 3. As mentioned above, the G-PRIME algorithm has the same solution as the C-PRIME algorithm in the case of $L = 2C$. So the MSE curve of C-PRIME is not shown in Fig. 3.

- It is obvious that the G-PRIME $(L = 2C)$, C-PRIME and ISTA-PRIME algorithms have the same steady-state value 2×10^{-4}.
- Furthermore, the G-PRIME $(L = 2C)$ and C-PRIME algorithms converge when the iteration number is close to 150 and the ISTA-PRIME algorithm approach convergence when the iteration number reaches 170-190.

Conclusions:

- Inspired by the C-PRIME technique, a gradient-based PRIME algorithm is proposed to solve a quadratic approximation of the original problem.
- The C-PRIME method can be regarded as a special case of the proposed G-PRIME algorithm.
- Numerical results have confirmed that the proposed algorithm has excellent phase recovery ability.