Globally Convergent Levenberg-Marquardt Method For Phase Retrieval

Zaiwen Wen

Beijing International Center For Mathematical Research
Peking University

Thanks: Chao Ma, Xin Liu
Outline

1. Introduction
2. The WF Algorithm
3. The LM Method
4. Numerical Results
Phase and magnitude

Phase carries more information than magnitude
X-ray crystallography (Thanks: Candes)

Method for determining atomic structure within a crystal

principle

10 Nobel Prizes in X-ray crystallography, and counting...
Missing phase problem (Thanks: Candes)

Detectors record **intensities** of diffracted rays \(\implies\) **phaseless data only!**

Fraunhofer diffraction \(\implies\) intensity of electrical \(\approx\) Fourier transform

\[
|\hat{x}(f_1, f_2)|^2 = \left| \int x(t_1, t_2) e^{-i2\pi(f_1t_1 + f_2t_2)} dt_1 dt_2 \right|
\]

Electrical field \(\hat{x} = |\hat{x}| e^{i\phi}\) with intensity \(|\hat{x}|^2\)

**Phase retrieval problem (inversion)**

How can we recover the phase (or signal \(x(t_1, t_2)\)) from \(|\hat{x}(f_1, f_2)|\)
Ptychographic Phase Retrieval (Thanks: Chao)

Given $|\mathcal{F}(Q_i \psi)|$ for $i = 1, \ldots, k$, can we recover $\psi$?

Ptychographic imaging along with advances in detectors and computing have resulted in X-ray microscopes with increased spatial resolution without the need for lenses.
Outline

1 Introduction
2 The WF Algorithm
3 The LM Method
4 Numerical Results
Discretized model for phase retrieval problem

Solve the equations:

\[ y_r = |\langle a_r, x \rangle|^2, \quad r = 1, 2, ..., m. \]  (1)

- **Gaussian model:**
  \[ a_r \in \mathbb{C}^n \text{ i.i.d.} \sim \mathcal{N}(0, I/2) + i\mathcal{N}(0, I/2). \]

- **Coded Diffraction model:**
  \[ y_r = \left| \sum_{t=0}^{n-1} x[t] \bar{d}_l(t) e^{-i2\pi kt/n} \right|^2, \quad r = (l, k), \quad 0 \leq k \leq n - 1, \quad 1 \leq l \leq L. \]
Nonlinear least square problem:

\[
\min_{z \in \mathbb{C}^n} f(z) = \frac{1}{4m} \sum_{k=1}^{m} \left( y_k - |\langle a_k, z \rangle|^2 \right)^2
\]

- Pro: operates over vectors and not matrices
- Con: \( f \) is nonconvex, many local minima
Wirtinger flow: Candes, Li and Soltanolkotabi ('14)

Strategies:
- Start from a sufficiently accurate initialization
- Make use of Wirtinger derivative

\[
\begin{align*}
  f(z) &= \frac{1}{4m} \sum_{k=1}^{m} (y_k - |\langle a_k, z \rangle|^2)^2 \\
  \nabla f(z) &= \frac{1}{m} \sum_{k=1}^{m} (|\langle a_k, z \rangle|^2 - y_k)(a_k a_k^*)z
\end{align*}
\]

- Careful iterations to avoid local minima
Algorithm: Gaussian model

**Spectral Initialization:**
1. Input measurements \(\{a_r\}\) and observation \(\{y_r\}\) \((r = 1, 2, \ldots, m)\).
2. Calculate \(z_0\) to be the leading eigenvector of \(Y = \frac{1}{m} \sum_{r=1}^{m} y_r a_r a_r^*\).
3. Normalize \(z_0\) such that \(\|z_0\|^2 = n \frac{\sum_r y_r}{\sum_r \|a_r\|^2}\).

**Iteration via Wirtinger derivatives:** for \(\tau = 0, 1, \ldots\)

\[
z_{\tau+1} = z_\tau - \frac{\mu_{\tau+1}}{\|z_0\|^2} \nabla f(z_\tau)
\]
Convergence property: Gaussian model

distance (up to global phase)

\[ \text{dist}(z, x) = \arg \min_{\pi \in [0, 2\pi]} \|z - e^{i\phi} x\| \]

**Theorem**

**Convergence for Gaussian model (C. Li and Soltanolkotabi ('14))**

- number of samples \( m \gtrsim n \log n \)
- Step size \( \mu \leq c/n \) (\( c > 0 \))

Then with probability at least \( 1 - 10e^{-\gamma n} - 8/n^2 - me^{-1.5n} \), we have \( \text{dist}(z_0, x) \leq \frac{1}{8}\|x\| \) and after \( \tau \) iteration

\[ \text{dist}(z_\tau, x) \leq \frac{1}{8}(1 - \frac{\mu}{4})^{\tau/2}\|x\|. \]

Here \( \gamma \) is a positive constant.
Outline

1. Introduction
2. The WF Algorithm
3. The LM Method
4. Numerical Results
Nonlinear least square problem

\[
\min_{z \in \mathbb{C}^n} \quad f(z) = \frac{1}{4m} \sum_{k=1}^{m} (y_k - |\langle a_k, z \rangle|^2)^2
\]

Using Wirtinger derivative:

\[
z := \begin{bmatrix} z \\ \bar{z} \end{bmatrix};
\]

\[
g(z) := \nabla_{c} f(z) = \frac{1}{m} \sum_{r=1}^{m} (|a_r^T z|^2 - y_r) \begin{bmatrix} (a_r a_r^T) z \\ (\bar{a}_r a_r^T) \bar{z} \end{bmatrix};
\]

\[
J(z) := \frac{1}{\sqrt{m}} \sum_{r=1}^{m} \begin{bmatrix} |a_1^* z| a_1, & |a_2^* z| a_2, & \cdots, & |a_m^* z| a_m \\ |a_1^* \bar{z}| \bar{a}_1, & |a_2^* \bar{z}| \bar{a}_2, & \cdots, & |a_m^* \bar{z}| \bar{a}_m \end{bmatrix}^T;
\]

\[
\Psi(z) := J(z)^T J(z) = \frac{1}{m} \sum_{r=1}^{m} \begin{bmatrix} |a_r^T z|^2 a_r a_r^T & (a_r^T z)^2 a_r a_r^T \\ (\bar{a}_r^T z)^2 \bar{a}_r a_r^T & |a_r^T z|^2 \bar{a}_r a_r^T \end{bmatrix}.
\]
The Modified LM method for Phase Retrieval

Levenberg-Marquardt Iteration:

\[ z_{k+1} = z_k - (\Psi(z_k) + \mu_k I)^{-1} g(z_k) \]

Algorithm

1 **Input:** Measurements \( \{a_r\} \), observations \( \{y_r\} \). Set \( \epsilon \geq 0 \).

2 Construct \( z_0 \) using the spectral initialization algorithms.

3 **While** \( \|g(z_k)\| \geq \epsilon \) **do**

   - Compute \( s_k \) by solving equation

     \[ \Psi_{z_k}^{\mu_k} s_k = (\Psi(z_k) + \mu_k I) s_k = -g(z_k). \]

   until

     \[ \|\Psi_{z_k}^{\mu_k} s_k + g(z_k)\| \leq \eta_k \|g(z_k)\|. \]

   - Set \( z_{k+1} = z_k + s_k \) and \( k := k + 1 \).

3 **Output:** \( z_k \).
If the measurements follow the Gaussian model, the LM equation is solved accurately ($\eta_k = 0$ for all $k$), and the following conditions hold:

- $m \geq cn \log n$, where $c$ is sufficiently large;
- If $f(z_k) \geq \frac{\|z_k\|^2}{900n}$, let $\mu_k = 70000n \sqrt{nf(z_k)}$; if else, let $\mu_k = \sqrt{f(z_k)}$.

Then, with probability at least $1 - 15e^{-\gamma n} - 8/n^2 - me^{-1.5n}$, we have $\text{dist}(z_0, x) \leq (1/8)\|x\|$, and

$$\text{dist}(z_{k+1}, x) \leq c_1 \text{dist}(z_k, x),$$

Meanwhile, once $f(z_s) < \frac{\|z_s\|^2}{900n}$, for any $k \geq s$ we have

$$\text{dist}(z_{k+1}, x) < c_2 \text{dist}(z_k, x)^2.$$
Convergence of the Gaussian Model

In the theorem above,

\[ c_1 := \begin{cases} 
(1 - \frac{\|x\|}{4\mu_k}), & \text{if } f(z_k) \geq \frac{1}{900n} \|z_k\|^2; \\
\frac{4.28 + 5.56 \sqrt{n}}{9.89 \sqrt{n}}, & \text{otherwise.}
\end{cases} \]

and

\[ c_2 = \frac{4.28 + 5.56 \sqrt{n}}{\|x\|}. \]
Key to proof

**Lower bound of GN matrix’s second smallest eigenvalue**

For any $y, z \in \mathbb{C}^n$, $Im(y^* z) = 0$, we have:

$$y^* \Psi(z)y \geq \|y\|^2 \|z\|^2,$$

holds with high probability.

$$Im(y^* z) = 0 \quad \Rightarrow \quad \|(\Psi^\mu_z)^{-1} y\| \leq \frac{2}{\|z\|^2 + \mu} \|y\|.$$
Local error bound property

\[ \frac{1}{4} \text{dist}(z, x)^2 \leq f(z) \leq 8.04 \text{dist}(z, x)^2 + 6.06n \text{dist}(z, x)^4, \]

holds for any \( z \) satisfying \( \text{dist}(z, x) \leq \frac{1}{8} \).

Regularity condition

\[ \mu(z) h^* \left( \Psi_z^\mu \right)^{-1} g(z) \geq \frac{1}{16} \|h\|^2 + \frac{1}{64100n\|h\|} \|g(z)\|^2 \]

holds for any \( z = x + h, \|h\| \leq \frac{1}{8} \), and \( f(z) \geq \frac{\|z\|^2}{900n} \).
Theorem

Convergence of the inexact LM method for the Gaussian model:
- \( m \gtrsim n \log n \);
- \( \mu_k \) takes the same value as in the exact LM method for the Gaussian model;
- \( \eta_k \leq \frac{(1-c_1)\mu_k}{25.55n\|z_k\|} \) if \( f(z_k) \geq \frac{\|z_k\|^2}{900n} \); otherwise \( \eta_k \leq \frac{(4.33\sqrt{n}-4.28)\mu_k\|g_k\|}{372.54n^2\|z_k\|^3} \).

Then, with probability at least \( 1 - 15e^{-\gamma n} - 8/n^2 - me^{-1.5n} \), we have \( \text{dist}(z_0, x) \leq \frac{1}{8}\|x\| \), and

\[
\text{dist}(z_{k+1}, x) \leq \frac{1 + c_1}{2} \text{dist}(z_k, x), \quad \text{for all } k = 0, 1, \ldots
\]

\[
\text{dist}(z_{k+1}, x) \leq \frac{9.89 \sqrt{n} + c_2\|x\|}{2\|x\|} \text{dist}(z_k, x)^2, \quad \text{for all } f(z_k) < \frac{\|z_k\|^2}{900n}.
\]

Here \( c_1 \) and \( c_2 \) take the same values as in the exact algorithm for the Gaussian model.
Solving the LM Equation: PCG

Solve

\[(\Psi_k + \mu_k I)u = g_k\]

by Pre-conditioned Conjugate Gradient Method:

\[M^{-1}(\Psi_k + \mu_k I)u = M^{-1}g_k, \quad M = \Phi_k + \mu_k I.\]

\[\Phi(z) := \begin{bmatrix} zz^* & 2zz^T \\ 2\bar{z}z^* & \bar{z}z^T \end{bmatrix} + \|z\|^2 I_{2n}\]

- small condition number
- **Easy to inverse:** \(M = (\mu_k + \|z_k\|^2)I + M_1\), where \(M_1\) is rank-2 matrix.
Solving the LM Equation: PCG

- small condition number.

**Lemma**

Consider solving the equation \((\Phi_z^\mu)^{-1}\Psi_z^\mu s = (\Phi_z^\mu)^{-1}g(z)\) by the CG method from \(s_0 := -(\Phi_z^\mu)^{-1}g(z)\). Let \(s_*\) be the solution of the system. Define \(V := \{x : x = [x^*, x^T]^*, x \in \mathbb{C}^n\}\). Then, \(V\) is an invariant subspace of \((\Phi_z^\mu)^{-1}\Psi_z^\mu\), and \(s_0, s_* \in V\). Meanwhile, choosing \(\mu_k = Kn \sqrt{f(z)}\), then the eigenvalues of \((\Phi_z^\mu)^{-1}\Psi_z^\mu\) on \(V\) satisfy:

\[
1 - \frac{57}{K \sqrt{n}} \leq \lambda \leq 1 + \frac{57}{K \sqrt{n}}.
\]
Solving the LM Equation: PCG

- Easy to inverse.

Calculate by Sherman-Morrison-Woodbury theorem:

$$(\Phi_z^{\mu})^{-1} = aI_{2n} + b \begin{bmatrix} z & z^* \end{bmatrix} [z^*, z^T] + c \begin{bmatrix} z & -z \end{bmatrix} [z^*, -z^T]$$

where

$$a = \frac{1}{\|z\|^2 + \mu}, \quad b = -\frac{3}{2(\|z\|^2 + \mu)(4\|z\|^2 + \mu)}, \quad c = \frac{1}{2(\|z\|^2 + \mu)\mu}. $$
Coded diffraction model

- **Initialization via resampled Wirtinger Flow:**
  1. Input measurements \( \{a_r\} \) and observation \( \{y_r\} (r = 1, 2, ..., m) \).
  2. Divide the measurements and observations equally into \( B + 1 \) groups of size \( m' \). The measurements and observations in group \( b \) are denoted as \( a_r^{(b)} \) and \( y_r^{(b)} \) for \( b = 0, 1, ..., B \).
  3. Obtain \( u_0 \) by conducting the spectral initialization on group 0.
  4. For \( b = 0 \) to \( B - 1 \), perform the following update:

\[
    u_{b+1} = u_b - \frac{\mu}{\|u_0\|^2} \left( \frac{1}{m'} \sum_{r=1}^{m'} \left( |z^* a_r^{(b+1)}|^2 - y_r^{(b+1)} \right) (a_r^{(b+1)}(a_r^{(b+1)})^*) z \right)
\]

  5. Set \( z_0 = u_B \).

- **WF: same as the Gaussian model:**
  for \( \tau = 0, 1, \ldots \)

\[
    z_{\tau+1} = z_\tau - \frac{\mu_{\tau+1}}{\|z_0\|^2} \nabla f(z_\tau)
\]
Convergence of WF: coded diffraction model

Theorem

Convergence for CD model (C. Li and Soltanolkotabi ('14))

- \( L \gtrsim (\log n)^4 \)
- Step size \( \mu \leq c (c > 0) \)

Then with probability at least \( 1 - (2L + 1)/n^3 - 1/n^2 \), we have

\[ \text{dist}(z_0, x) \leq \frac{1}{8\sqrt{n}} \|x\| \quad \text{and after } t \text{ iteration} \]

\[ \text{dist}(z_\tau, x) \leq \frac{1}{8\sqrt{n}} (1 - \frac{\mu}{3})^{\tau/2} \|x\|. \]
Convergence of LM: Coded Diffraction Model

**Theorem**

If the measurements follow the coded diffraction model, the LM equation is solved accurately, and the following conditions hold:

- $L \geq c (\log n)^4$, where $c$ is sufficiently large;
- If $f(z_k) \geq \frac{\|z_k\|^2}{3260n}$, let $\mu_k = 35000n \sqrt{f(z_k)}$; if else, let $\mu_k = \sqrt{f(z_k)}$.

Then, with probability at least $1 - (2L + 1)/n^3 - 1/n^2$, we have

$$\text{dist}(z_0, x) \leq (1/8 \sqrt{n})\|x\|,$$

and

$$\text{dist}(z_{k+1}, x) \leq c_1 \text{dist}(z_k, x). \quad (2)$$

Meanwhile, once $f(z_s) < \frac{\|z_s\|^2}{3260n}$, for any $k \geq s$ we have

$$\text{dist}(z_{k+1}, x)\text{dist}(z_k, x)^2 < c_2 \text{dist}(z_k, x)^2. \quad (3)$$
Outline

1. Introduction
2. The WF Algorithm
3. The LM Method
4. Numerical Results
Consider the following two kinds of signals:

- **Random low-pass signals:**

  \[ x[t] = \sum_{k=-(M/2-1)}^{M/2} (X_k + iY_k)e^{2\pi i(k-1)(t-1)/n}, \]

  with \( M=n/8 \) and \( X_k \) and \( Y_k \) are i.i.d. \( \mathcal{N}(0, 1) \).

- **Random Guassian signals:** where \( x \in \mathbb{C}^n \) is a random complex Gaussian vector with i.i.d. entries of the form

  \[ X[t] = X + iY, \]

  with \( X \) and \( Y \) distributed as \( \mathcal{N}(0, 1/2) \).

Set \( n = 512 \), solve the problem by WF method, accurate LM method and accurate LM method.
Success rate: Gauss signal

Figure: Empirical probability of success based on 100 random trials for different signal/measurement models, different algorithms, and a varied number of measurements. For Gauss signal.
Success rate: Low-pass signal

Figure: Empirical probability of success based on 100 random trials for different signal/measurement models, different algorithms, and a varied number of measurements. For low-pass signal.
### Time Cost and Number of Iterations

<table>
<thead>
<tr>
<th>Signal model</th>
<th>Gaussian signal</th>
<th>Low-pass signal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gaussian</td>
<td>Coded diffraction</td>
</tr>
<tr>
<td>iter CPU</td>
<td>iter CPU</td>
<td>iter CPU</td>
</tr>
<tr>
<td>ALM</td>
<td>8.39 20.40s</td>
<td>6.41 0.53s</td>
</tr>
<tr>
<td>ILM</td>
<td>10.07 2.47s</td>
<td>6.58 0.06s</td>
</tr>
<tr>
<td>WF</td>
<td>300.79 7.14s</td>
<td>171.79 0.21s</td>
</tr>
</tbody>
</table>

**Table:** The averaged number of iterations and CPU time used by the LM and WF methods to achieve $10^{-5}$ accuracy for different signals and models, where $m/n=6$ for Gaussian model and $L=10$ for coded diffraction model.
Figure: Relationship between relative error and number of iterations. $m/n = 6$ for Gaussian model and $L = 10$ for CD model.
Relative Error vs CPU Time

Figure: Relationship between relative error and CPU time. $m/n = 6$ for Gaussian model and $L = 10$ for CD model.
**Figure:** Turret of Palace Museum. Image size is 352×1000 pixels. For the ALM method, the CPU time is 9464.57s for 25 iterations, with a final relative error to be $2.42 \times 10^{-16}$; for the ILM algorithm, the CPU time is 1507.16s for 25 iterations, and the final relative error is $2.44 \times 10^{-16}$; for WF algorithm, the CPU time is 1617.01s for 300 iterations, while the final relative error is $4.89 \times 10^{-16}$
Figure: The Milky Way Galaxy. Image size is 1080×1920 pixels. For the ALM method, the CPU time is 20240.64s, with a final relative error to be $2.44 \times 10^{-16}$; for the ILM algorithm, the CPU time is 4733.43s, and the final relative error is $2.42 \times 10^{-16}$; for WF algorithm, the CPU time is 5211.35s, while the final relative error is $4.91 \times 10^{-16}$
**Table:** The number of iterations and CPU time used (Not including the CPU time of initialization) by the LM and WF algorithms to reduce the relative error to a certain criterion in natural image recovering.

<table>
<thead>
<tr>
<th>Image</th>
<th>Turret of Palace Museum</th>
<th>The Milky Way Galaxy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Criterion</strong></td>
<td><strong>10^-5</strong></td>
<td><strong>10^-10</strong></td>
</tr>
<tr>
<td></td>
<td><strong>iter</strong></td>
<td><strong>CPU</strong></td>
</tr>
<tr>
<td>ALM</td>
<td>5.3</td>
<td>2244s</td>
</tr>
<tr>
<td>ILM</td>
<td>6.3</td>
<td>349s</td>
</tr>
<tr>
<td>WF</td>
<td>136.3</td>
<td>685s</td>
</tr>
</tbody>
</table>
Numerical Result: Natural Image

Figure: Relation between relative error and number of iterations for natural images recovery.
Figure: Relation between relative error and CPU time used for natural images recovery.