Nonlinear Analysis with Frames. Part III: Algorithms

Radu Balan

Department of Mathematics, AMSC, CSCAMM and NWC
University of Maryland, College Park, MD

July 28-30, 2015
Modern Harmonic Analysis and Applications
Summer Graduate Program
University of Maryland, College Park, MD 20742
Thanks to our sponsors:

"This material is based upon work supported by the National Science Foundation under Grants No. DMS-1413249, DMS-1501640. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Problem Formulation</td>
</tr>
<tr>
<td>2</td>
<td>PhaseLift</td>
</tr>
<tr>
<td>3</td>
<td>IRLS</td>
</tr>
</tbody>
</table>
Table of Contents

1. Problem Formulation
2. PhaseLift
3. IRLS
Problem Formulation

The phase retrieval problem

- Let $H = \mathbb{C}^n$. The quotient space $\hat{H} = \mathbb{C}^n/T^1$, with classes induced by $x \sim y$ if there is real $\varphi$ with $x = e^{i\varphi}y$.
- Frame $\mathcal{F} = \{f_1, \cdots, f_m\} \subset \mathbb{C}^n$ and nonlinear maps
  \[
  \alpha : \hat{H} \to \mathbb{R}^m, \quad \alpha(x) = (|\langle x, f_k \rangle|)_{1 \leq k \leq m}.
  \]
  \[
  \beta : \hat{H} \to \mathbb{R}^m, \quad \beta(x) = \left(|\langle x, f_k \rangle|^2\right)_{1 \leq k \leq m}.
  \]

The frame is said phase retrievable (or that it gives phase retrieval) if $\alpha$ (or $\beta$) is injective.
Problem Formulation
The phase retrieval problem

- Let $H = \mathbb{C}^n$. The quotient space $\hat{H} = \mathbb{C}^n / T^1$, with classes induced by $x \sim y$ if there is real $\varphi$ with $x = e^{i\varphi} y$.
- Frame $\mathcal{F} = \{f_1, \cdots, f_m\} \subset \mathbb{C}^n$ and nonlinear maps

  \[ \alpha : \hat{H} \rightarrow \mathbb{R}^m, \quad \alpha(x) = (|\langle x, f_k \rangle|)_{1 \leq k \leq m}. \]

  \[ \beta : \hat{H} \rightarrow \mathbb{R}^m, \quad \beta(x) = (|\langle x, f_k \rangle|^2)_{1 \leq k \leq m}. \]

The frame is said phase retrievable (or that it gives phase retrieval) if $\alpha$ (or $\beta$) is injective.

- The general phase retrieval problem a.k.a. phaseless reconstruction: Decide when a given frame is phase retrievable, and, if so, find an algorithm to recover $x$ from $y = \alpha(x)$ (or from $y = \beta(x)$) up to a global phase factor.
Our Problem Today: Assume $\mathcal{F}$ is phase retrievable. Want reconstruction algorithms.
Problem Formulation

Algorithms

Our Problem Today: Assume $\mathcal{F}$ is phase retrievable. Want reconstruction algorithms.

- Recursive Projections: Gerchberg-Saxton
- Matrix Estimation: PhaseLift (Candes, Strohmer, Voroninski’12, CandesLi)
- Signal Estimation: Iterative Regularized Least Squares (IRLS), Wirtinger Flow (Candes’14)
- Algorithms for special frames: Reconstruction via Polarization (Alexeev, Bandeira, Fickus, Mixon; Bodmann, Hammen), Fourier transform (Lim & co MIT; Bates’82; Bal’09; PhaseLift with Masking; 4n-4 by Bodmann, Hammen), Shift Invariant Frames (Iwen, Viswanathan, Wang), High Redundancy (BBCE’09)
- Algorithms for special signals: sparse signals (e.g. Iwen, Viswanathan, Wang)
Our Problem Today: Assume $\mathcal{F}$ is phase retrievable. Want reconstruction algorithms.

- Recursive Projections: Gerchberg-Saxton
- Matrix Estimation: PhaseLift (Candes, Strohmer, Voroninski’12, CandesLi)
- Signal Estimation: Iterative Regularized Least Squares (IRLS), Wirtinger Flow (Candes’14)
- Algorithms for special frames: Reconstruction via Polarization (Alexeev,Bandeira,Fickus, Mixon; Bodmann,Hammen), Fourier transform (Lim&co MIT; Bates’82; Bal’09; PhaseLift with Masking; 4n-4 by Bodmann,Hammen), Shift Invariant Frames (Iwen, Viswanathan,Wang), High Redundancy (BBCE’09)
- Algorithms for special signals: sparse signals (... e.g. Iwen,Viswanathan,Wang)
Our Problem Today: Assume $\mathcal{F}$ is phase retrievable. Want reconstruction algorithms.

- Recursive Projections: Gerchberg-Saxton
- Matrix Estimation: PhaseLift (Candes, Strohmer, Voroninski’12, CandesLi)
- Signal Estimation: Iterative Regularized Least Squares (IRLS), Wirtinger Flow (Candes’14)
- Algorithms for special frames: Reconstruction via Polarization (Alexeev, Bandeira, Fickus, Mixon; Bodmann, Hammen), Fourier transform (Lim & co MIT; Bates’82; Bal’09; PhaseLift with Masking; 4n-4 by Bodmann, Hammen), Shift Invariant Frames (Iwen, Viswanathan, Wang), High Redundancy (BBCE’09)
- Algorithms for special signals: sparse signals (... e.g. Iwen, Viswanathan, Wang)
PhaseLift
The Idea

Consider the noiseless case $y = \beta(x)$. The main idea is embodied in the following feasibility problem:

Find $X$
subject to:
$X = X^* \geq 0$
$A(X) = y$  \hspace{1cm} (Feas)
$\text{rank}(X) = 1$
PhaseLift
The Idea

Consider the noiseless case $y = \beta(x)$. The main idea is embodied in the following feasibility problem:

Find $X$
subject to:
$X = X^* \geq 0$ (Feas)
$\mathbb{A}(X) = y$
$\text{rank}(X) = 1$

Alternatively, since there is a unique rank 1 that satisfies this problem:

Min $\text{rank}(X)$
subject to:
$X = X^* \geq 0$ (L0)
$\mathbb{A}(X) = y$

Except for $\text{rank}(X)$ the optimization problem would be convex.
IDEA: Replace \( \text{rank}(X) \) by \( \text{trace}(X) \) as in the \textit{Matrix Completion} problem.
IDEA: Replace $\text{rank}(X)$ by $\text{trace}(X)$ as in the Matrix Completion problem. Once a solution $X$ is found, the vector $x$ can be easily obtained from the factorization: $X = xx^*$. 
PhaseLift

The Algorithm

\[(\text{PhaseLift}) \quad \min_{A(X) = y, X = X^* \geq 0} \text{trace}(X)\]

which is a convex optimization problem (a semi-definite program: SDP).
PhaseLift
The Algorithm

\[(\text{PhaseLift})\quad \min_{\mathbf{A}(\mathbf{X}) = \mathbf{y}, \mathbf{X} = \mathbf{X}^* \geq 0} \text{trace}(\mathbf{X})\]

which is a convex optimization problem (a semi-definite program: SDP).

**Theorem (Candés-Li 2014)**

Assume each vector $f_k$ is drawn independently from $\mathcal{N}(0, I_n/2) + i\mathcal{N}(0, I_n/2)$, or each vector is drawn independently from the uniform distribution on the complex sphere of radius $\sqrt{n}$. Then there are universal constants $c_0, c_1, \gamma > 0$ so that for $m \geq c_0 n$, for every $\mathbf{x} \in \mathbb{C}^n$ the problem (PhaseLift) has the same solution as (L0) with probability at least $1 - c_1 e^{-\gamma n}$. 

Radu Balan (UMD)
Phase Retrieval
July 28-30, 2015
The Algorithm

\[
\text{(PhaseLift)} \quad \min_{\mathbf{A}(\mathbf{X}) = \mathbf{y}, \mathbf{X} = \mathbf{X}^* \geq 0} \text{trace}(\mathbf{X})
\]

which is a convex optimization problem (a semi-definite program: SDP).

**Theorem (Candés-Li 2014)**

Assume each vector \( f_k \) is drawn independently from \( \mathcal{N}(0, I_n/2) + i\mathcal{N}(0, I_n/2) \), or each vector is drawn independently from the uniform distribution on the complex sphere of radius \( \sqrt{n} \). Then there are universal constants \( c_0, c_1, \gamma > 0 \) so that for \( m \geq c_0 n \), for every \( \mathbf{x} \in \mathbb{C}^n \) the problem (PhaseLift) has the same solution as (L0) with probability at least \( 1 - c_1 e^{-\gamma n} \).

Hand & Demanet (2013) showed (PhaseLift) is in essence a feasibility problem.
Consider the measurement model in the presence of noise

\[ y = \beta(x) + \nu \]
Consider the measurement model in the presence of noise

\[ y = \beta(x) + \nu \]

Modify the optimization problem:

\[ \min_{X = X^* \geq 0} \| A(X) - y \|_1 \quad \text{(PL2)} \]
Modified Phase Lift algorithm is robust to noise:

**Theorem (Candés-Li 2014)**

Consider the same stochastic process for the random frame $\mathcal{F}$. There is a universal constant $C_0 > 0$ so that for all $x \in \mathbb{C}^n$ the solution to (PL2) obeys

$$\|X - xx^*\|_2 \leq C_0 \frac{\|\nu\|_1}{m}$$

For the Gaussian model this holds with the same probability as in the noiseless case, whereas the probability of failure is exponentially small in $n$ in the uniform model. The principal eigenvector $x^0$ of $X$ (normalized by the square root of the principal eigenvalue) obeys

$$D_2(x^0, x) \leq C_0 \min(\|x\|_2, \frac{\|\nu\|_1}{m\|x\|_2}).$$
Table of Contents

1. Problem Formulation
2. PhaseLift
3. IRLS
Iterative Regularized Least-Squares

The Idea

Consider the measurement process

\[ y_k = |\langle x, f_k \rangle|^2 + \nu_k, \quad 1 \leq k \leq m \]

The Least-Squares criterion:

\[
\min_{x \in \mathbb{C}^n} \sum_{k=1}^{m} \left( |\langle x, f_k \rangle|^2 - y_k \right)^2
\]

can be understood as the Maximum Likelihood Estimator (MLE) when the noise vector \( \nu \in \mathbb{R}^m \) is normal distributed with zero mean and covariance \( \sigma^2 I_m \). However the optimization problem is not convex and has many local minima.
Consider the following optimization criterion:

\[ J(u, v; \lambda, \mu) = \sum_{k=1}^{m} \left| \frac{1}{2} \left( \langle u, f_k \rangle \langle f_k, v \rangle + \langle v, f_k \rangle \langle f_k, u \rangle \right) - y_k \right|^2 + \lambda \| u \|^2 + \mu \| u - v \|^2 + \lambda \| v \|^2 \]
Consider the following optimization criterion:

\[ J(u, v; \lambda, \mu) = \sum_{k=1}^{m} \left| \frac{1}{2} \left( \langle u, f_k \rangle \langle f_k, v \rangle + \langle v, f_k \rangle \langle f_k, u \rangle \right) - y_k \right|^2 + \lambda \| u \|^2 + \mu \| u - v \|^2 + \lambda \| v \|^2 \]

The Iterative Regularized Least-Squares (IRLS) algorithm is based on minimization:

\[ x^{t+1} = \arg\min_u J(u, x^t; \lambda_t, \mu_t) \]
Iterative Regularized Least-Squares
The Algorithm: Initialization

Step 1. Initialization. Compute the principal eigenvector of $R_y = \sum_{k=1}^{m} y_k f_k f_k^*$ using e.g. the power method. Let $(e_1, a_1)$ be the eigen-pair with $e_1 \in \mathbb{C}^n$ and $a_1 \in \mathbb{R}$. If $a_1 \leq 0$ then set $x = 0$ and exit. Otherwise initialize:

$$x^0 = \sqrt{\frac{(1 - \rho)a_1}{\sum_{k=1}^{m} |\langle e_1, f_k \rangle|^4}} e_1$$  \hspace{1cm} (3.1)

$$\lambda_0 = \rho a_1$$  \hspace{1cm} (3.2)

$$\mu_0 = \rho a_1$$  \hspace{1cm} (3.3)

$$t = 0$$  \hspace{1cm} (3.4)
Iterative Regularized Least-Squares

The Algorithm: Iterations

**Step 2. Iteration.** Perform:

2.1 Solve the least-square problem:

\[ x^{t+1} = \arg\min_u J(u, x^t; \lambda_t, \mu_t) \]

using the conjugate gradient method.

2.2 Update:

\[ \lambda_{t+1} = \gamma \lambda_t, \quad \mu_t = \max(\gamma \mu_t, \mu_{\text{min}}), \quad t = t + 1 \]
Iterative Regularized Least-Squares
The Algorithm: Stopping

Step 3. Stopping. Repeat Step 2 until:

- The error criterion is achieved: \( J(x^t, x^t; 0, 0) < \varepsilon \); or
- The desired signal-to-noise-ratio is reached: \( \frac{\|x^t\|^2}{J(x^t, x^t; 0, 0)} > \text{snr} \); or
- The maximum number of iterations is reached: \( t > T \).
Iterative Regularized Least-Squares
The Algorithm: Stopping

Step 3. Stopping. Repeat Step 2 until:

- The error criterion is achieved: \( J(x^t, x^t; 0, 0) < \varepsilon \); or
- The desired signal-to-noise-ratio is reached: \( \frac{\|x^t\|^2}{J(x^t, x^t; 0, 0)} > snr \); or
- The maximum number of iterations is reached: \( t > T \).

The final estimate can be \( x^T \), or the best estimate obtained in the iteration path: \( x^{est} = x^{t_0} \), where \( t_0 = \arg\min_t J(x^t, x^t; 0, 0) \).
Iterative Regularized Least-Squares
Performance Bounds

**Theorem (B. 2013)**

Fix $0 \neq z_0 \in \mathbb{C}^n$. Assume the frame $\mathcal{F}$ is so that $\ker A \cap S^{2,1} = \{0\}$. Then there is a constant $A_3 > 0$ that depends of $\mathcal{F}$ so that for every $x \in \Omega_{z_0}$ and $\nu \in \mathbb{C}^n$ that produce $y = \beta(x) + \nu$ if there are $u, \nu \in \mathbb{C}^n$ so that $J(u, \nu; \lambda, \mu) < J(x, x; \lambda, \mu)$ then

$$\| [u, \nu] - xx^* \|_1 \leq \frac{4\lambda}{A_3} + \frac{2\| \nu \|_2}{\sqrt{A_3}}$$

(3.5)

Moreover, let $[u, \nu] = a_+ e_+ e^*_+ + a_- e_- e^*_-$ be its spectral factorization with $a_+ \geq 0 \geq a_-$ and $\| e_+ \| = \| e_- \| = 1$. Set $\tilde{x} = \sqrt{a_+} e_+$. Then

$$D_2(x, \tilde{x})^2 \leq \frac{4\lambda}{A_3} + \frac{2\| \nu \|_2}{\sqrt{A_3}} + \frac{\| \nu \|_2^2}{4\mu} + \frac{\lambda \| x \|_2^2}{2\mu}$$

(3.6)
Iterative Regularized Least-Squares
Performance Bounds

**Theorem (B. 2013)**

Fix \( 0 \neq z_0 \in \mathbb{C}^n \). Assume the frame \( \mathcal{F} \) is so that \( \ker \mathbf{A} \cap S^{2,1} = \{0\} \). Then there is a constant \( A_3 > 0 \) that depends of \( \mathcal{F} \) so that for every \( x \in \Omega z_0 \) and \( \nu \in \mathbb{C}^n \) that produce \( y = \beta(x) + \nu \) if there are \( u, \nu \in \mathbb{C}^n \) so that \( J(u, \nu; \lambda, \mu) < J(x, x; \lambda, \mu) \) then

\[
\|[[u, \nu]] - xx^*\|_1 \leq \frac{4\lambda}{A_3} + \frac{2\|\nu\|_2}{\sqrt{A_3}} \tag{3.5}
\]

Moreover, let \( [[[u, \nu]] = a_+e_+e^*_+ + a_-e_-e_-e^* \) be its spectral factorization with \( a_+ \geq 0 \geq a_- \) and \( \|e_+\| = \|e_-\| = 1 \). Set \( \tilde{x} = \sqrt{a_+e_+} \). Then

\[
D_2(x, \tilde{x})^2 \leq \frac{4\lambda}{A_3} + \frac{2\|\nu\|_2}{\sqrt{A_3}} + \frac{\|\nu\|_2^2}{4\mu} + \frac{\lambda\|x\|_2^2}{2\mu} \tag{3.6}
\]
Thank you!

Questions?
References


