Phase Retrieval using the Iterative Regularized Least-Squares Algorithm

Radu Balan

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

August 16, 2017
Workshop on ”Phaseless Imaging in Theory and Practice: Realistic Models, Fast Algorithms, and Recovery Guarantees”
IMA, University of Minnesota, Minneapolis, MN
This material is based upon work supported by the National Science Foundation under Grant No. DMS-1413249, by ARO under Contract No. W911NF-16-1-0008, and by LTS under grant H9823013D00560049. "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."

Joint work with Naveed Haghani (UMD).
Hilbert space $H = \mathbb{C}^n$, $\hat{H} = H / T^1$, frame $\mathcal{F} = \{f_1, \cdots, f_m\} \subset \mathbb{C}^n$ and measurements

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

The frame is said *phase retrievable* (or that it gives phase retrieval) if $\hat{x} \mapsto (|\langle x, f_k \rangle|)_{1 \leq k \leq m}$ 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 $y = (y_k)_k$ up to a global phase factor.

Our problem today: A reconstruction algorithm.
General Purpose Algorithms
Unstructured Frames. Unstructured Data

1. Iterative Algorithms:
   - Gerchberg-Saxton [Gerchberg&all]
   - Wirtinger flow - gradient descent [CLS14]
   - IRLS [B13]

2. Rank 1 Tensor Recovery:
   - PhaseLift; PhaseCut [CSV12]; [WdAM12]
   - Higher-Order Tensor Recovery [B09]
Specialized Algorithms
Structured Frames and/or Structured Data

1. Structured Frames:
   - Fourier Frames: 4n-4 [BH13]; Masking DFT [CLS13];
     STFT/Spectograms [B.][Eldar&all][Hayes&all]; Alternating Projections
     [GriffinLim][Fannjiang]; Hybrid I-O [Fienup82]
   - Polarization: 3-term [ABFM12], masking [BCM]
   - Shift-Invariant Spaces: Bandlimited [Thakur11]; Filterbanks/Circulant
     Matrices [IVW2]; Other spaces [Chen&all]
   - X-Ray Crystallography – over 100 years old, lots of Nobel prizes ...

2. Special Signals:
   - Sparse general case: GESPAR[SBE14];
   - Specialized: sparse [IVW1]; speech [ARF03]

... and others – ”phase retrieval” in title: 2680 papers
Our algorithm (IRLS and variants) belongs to the class of \textit{Graduation Methods}, or \textit{Homotopic Continuations}.

Idea:
Graduation Method. Homotopic Continuation
First Motivation

Our algorithm (IRLS and variants) belongs to the class of *Graduation Methods*, or *Homotopic Continuations*.

Idea:
Our target is to optimize a complicated (possibly non-convex) optimization criterion $J(x)$, $\arg\min_{x \in D} J(x)$. 
Graduation Method. Homotopic Continuation

First Motivation

Our algorithm (IRLS and variants) belongs to the class of *Graduation Methods*, or *Homotopic Continuations*.

Idea:
Our target is to optimize a complicated (possibly non-convex) optimization criterion \( J(x) \), \( \text{argmin}_{x \in D} J(x) \).
However we know how to optimize a closely related criterion \( J_0(x) \), \( \text{argmin}_{x \in D_0} J_0(x) \).
Graduation Method. Homotopic Continuation
First Motivation

Our algorithm (IRLS and variants) belongs to the class of Graduation Methods, or Homotopic Continuations.

Idea:
Our target is to optimize a complicated (possibly non-convex) optimization criterion \( J(x) \), \( \text{argmin}_{x \in D} J(x) \).
However we know how to optimize a closely related criterion \( J_0(x) \), \( \text{argmin}_{x \in D_0} J_0(x) \).
Then we introduce a monotonic sequence \( 0 \leq t_n \leq 1 \) with \( t_0 = 1 \) and \( t_n \to 0 \) and solve iteratively

\[
x^{n+1} = \text{argmin}_{x \in D_n} F(t_n, J(x), J_0(x))
\]

using \( x^n \) as starting point. Here \( F \) is a continue function so that \( F(1, J(x), J_0(x)) = J_0(x) \) and \( F(0, J(x), J_0(x)) = J(x) \).
Graduation Method. Homotopic Continuation

First Motivation

M.C. Escher (1937) - Metamorphosis I
online at: http://www.mcescher.com/gallery/
Graduation Method. Homotopic Continuation
Second Motivation: LARS Algorithm

Least Angle Regression (LARS) [EHJT04] designed to solve LASSO, or variants:

$$\arg\min_x ||y - Ax||_2^2 + \lambda ||x||_1$$
Graduation Method. Homotopic Continuation
Second Motivation: LARS Algorithm

Least Angle Regression (LARS) [EHJT04] designed to solve LASSO, or variants:

$$\arg\min_x \|y - Ax\|_2^2 + \lambda \|x\|_1$$

It is proved the optimizer $x_{opt} = x(\lambda)$ is a continuous and piecewise differentiable function of $\lambda$ (linear, in the case of LASSO).
Graduation Method. Homotopic Continuation
Second Motivation: LARS Algorithm

Least Angle Regression (LARS) [EHJT04] designed to solve LASSO, or variants:

$$\text{argmin}_x \| y - Ax \|^2_2 + \lambda \| x \|^1_1$$

It is proved the optimizer $x_{opt} = x(\lambda)$ is a continuous and piecewise differentiable function of $\lambda$ (linear, in the case of LASSO).

Method: Start with $\lambda = \lambda_0 = \frac{2}{\|A^T y\|_2}$ and the optimal solution is $x^0 = 0$. 

Radu Balan (UMD)
Graduation Method. Homotopic Continuation
Second Motivation: LARS Algorithm

Least Angle Regression (LARS) [EHJT04] designed to solve LASSO, or variants:

$$\text{argmin}_x \|y - Ax\|_2^2 + \lambda \|x\|_1$$

It is proved the optimizer $x_{opt} = x(\lambda)$ is a continuous and piecewise differentiable function of $\lambda$ (linear, in the case of LASSO).

Method: Start with $\lambda = \lambda_0 = \frac{2}{\|A^T y\|_2}$ and the optimal solution is $x^0 = 0$.

Then LARS finds monotonically decreasing $\lambda$ values where the slope (and support) of $x(\lambda)$ changes. The algorithm ends at the desired value of $\lambda = \lambda_\infty$ (see also Hierarchical Decompositions of Tadmor&all).
The ultimate goal is to find the global minimum of the following functional:

\[
I(x) = \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2
\]

over \( x \in \mathbb{C}^n \), given the set of real numbers \( y_1, \cdots, y_m \) and frame vectors \( f_1, \cdots, f_m \in \mathbb{C}^n \). The problem is hard because the criterion is non-convex (it is a quartic multivariate polynomial).
Homotopy Method

Our Main Problem

The ultimate goal is to find the global minimum of the following functional:

\[ I(x) = \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 \]

over \( x \in \mathbb{C}^n \), given the set of real numbers \( y_1, \ldots, y_m \) and frame vectors \( f_1, \ldots, f_m \in \mathbb{C}^n \). The problem is hard because the criterion is non-convex (it is a quartic multivariate polynomial). Denote \( \hat{x} \) this global optimum, and assume it is unique up to a global phase.
The ultimate goal is to find the global minimum of the following functional:

$$I(x) = \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2$$

over $x \in \mathbb{C}^n$, given the set of real numbers $y_1, \cdots, y_m$ and frame vectors $f_1, \cdots, f_m \in \mathbb{C}^n$. The problem is hard because the criterion is non-convex (it is a quartic multivariate polynomial). Denote $\hat{x}$ this global optimum, and assume it is unique up to a global phase. Let $J(x; \lambda)$ denote the regularized form:

$$J(x; \lambda) = \frac{1}{4} \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 + \frac{\lambda}{2} \|x\|^2$$
Homotopy Method
Quartic Criteria: The Convex Regime

\[ J(x; \lambda) = \frac{1}{4} \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 + \frac{\lambda}{2} \|x\|^2 \]
Homotopy Method
Quartic Criteria: The Convex Regime

$$J(x; \lambda) = \frac{1}{4} \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 + \frac{\lambda}{2} \|x\|^2$$

Since

$$J(x; \lambda) = \frac{1}{4} \sum_{k=1}^{m} |\langle x, f_k \rangle|^4 + \frac{1}{2} \langle (\lambda I - R_0)x, x \rangle + \frac{1}{4} \sum_{k=1}^{m} y_k^2$$

with $R_0 = \sum_{k=1}^{m} y_k f_k f_k^*$, it follows for $\lambda > \lambda_0 = \lambda_{\text{max}}(R_0)$ the criterion is strongly convex and $x = 0$ is the global minimum.
The Phase Retrieval Problem

Existing Algorithms

The Homotopy Method

Numerical Results

Homotopy Method
Quartic Criteria: The Convex Regime

\[ J(x; \lambda) = \frac{1}{4} \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 + \frac{\lambda}{2} \|x\|^2 \]

Since

\[ J(x; \lambda) = \frac{1}{4} \sum_{k=1}^{m} |\langle x, f_k \rangle|^4 + \frac{1}{2} \langle (\lambda I - R_0)x, x \rangle + \frac{1}{4} \sum_{k=1}^{m} y_k^2 \]

with \( R_0 = \sum_{k=1}^{m} y_k f_k f_k^* \), it follows for \( \lambda > \lambda_0 = \lambda_{max}(R_0) \) the criterion is strongly convex and \( x = 0 \) is the global minimum.

A good candidate for the homotopy method is to start with \( J(x; \lambda_0 - \varepsilon) \) whose global minimum is along the principal eigenvector of \( R_0 \), and then decrease \( \lambda \) until desired value (e.g. 0).
At each $\lambda \geq 0$ consider the set of critical points: $\nabla_x J(x; \lambda) = 0$. To illustrate the method, restrict to the real case. The characteristic equation (of critical points) is given by:

$$\sum_{k=1}^{m} (|\langle x, f_k \rangle|^2 - y_k)\langle x, f_k \rangle f_k + \lambda x = 0$$

or

$$R(x)x + (\lambda I - R_0)x = 0$$

(3.1)

where $R_0 = \sum_{k=1}^{m} y_k f_k f_k^*$ and $R(x) = \sum_{k=1}^{m} |\langle x, f_k \rangle|^2 f_k F_k^*$. 
At each $\lambda \geq 0$ consider the set of critical points: $\nabla_x J(x; \lambda) = 0$. To illustrate the method, restrict to the real case. The characteristic equation (of critical points) is given by:

$$
\sum_{k=1}^{m} (|\langle x, f_k \rangle|^2 - y_k)\langle x, f_k \rangle f_k + \lambda x = 0
$$

or

$$
R(x)x + (\lambda I - R_0)x = 0 \quad (3.1)
$$

where $R_0 = \sum_{k=1}^{m} y_k f_k f_k^*$ and $R(x) = \sum_{k=1}^{m} |\langle x, f_k \rangle|^2 f_k F_k^*$.

Note, (3.1) is a system of cubic equations in $n$ variables. Assume the number of roots is always finite (true, unless a degenerate case). Then the number of critical points is at most $3^n$. 
Homotopy Method
Bifurcation Diagrams

An example of $\lambda$-dependent characteristic roots:

Figure: Plot of $x_1 = x_1(\lambda)$ in a low-dimensional case $n = 3$, $m = 5$. $\lambda_0 = 55.84$
Homotopy Method

Bifurcation Diagrams

An example of $\lambda$-dependent characteristic roots:

![Bifurcation Diagram](image)

**Figure:** Plot of $x_1 = x_1(\lambda)$ in a low-dimensional case $n = 3$, $m = 5$. $\lambda_0 = 55.84$

$$y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 2 & 1 \\ 1 & -1 & 1.01 \end{bmatrix} \begin{bmatrix} 1 \\ 1.5 \\ -1 \end{bmatrix}^2$$
Homotopy Method
Bifurcation Diagrams

Strategy: Start with

\[ J(x; \lambda) = \frac{1}{4} \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 + \frac{\lambda}{2} \|x\|^2 \]

at \((\lambda_0 - \varepsilon, s(\varepsilon)e_0)\) and then continually track the critical point branch, while decreasing the criterion.
Homotopy Method
Bifurcation Diagrams

**Strategy:** Start with

\[ J(x; \lambda) = \frac{1}{4} \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 + \frac{\lambda}{2} \|x\|^2 \]

at \((\lambda_0 - \epsilon, s(\epsilon)e_0)\) and then continually track the critical point branch, while decreasing the criterion. Thus:

\[ \frac{1}{4} \sum_{k=1}^{m} |\langle x, f_k \rangle|^4 + \frac{1}{2} \langle (\lambda I - R_0)x, x \rangle + \frac{1}{4} \sum_{k=1}^{m} y_k^2 = J(x; \lambda) \leq J(0, \lambda_0) = \frac{1}{4} \sum_{k=1}^{m} y_k^2 \]

Thus

\[ \sum_{k=1}^{m} |\langle x, f_k \rangle|^4 \leq 2 \langle (R_0 - \lambda I)x, x \rangle \]

Let \(a_{24} = \min_{\|x\|_2 = 1} \|Tx\|_4 > 0\). We obtain:

\[ \|x\| \leq \frac{\sqrt{\|R_0\| - \lambda}}{a_{24}^2} \]
Consider a parametrization of the characteristic curves $(\lambda = \lambda(t), x = x(t))$:

$$\nabla_x J(x(t); \lambda(t)) = 0 \iff R(x(t))x(t) + (\lambda(t)I - R_0)x(t) = 0$$

Differentiate to obtain:

$$\begin{bmatrix} x & : & H(x, \lambda) \end{bmatrix} \begin{bmatrix} \frac{d\lambda}{dt} \\ \frac{dx}{dt} \end{bmatrix} = 0 \quad (\text{Diff.Sys.})$$

with the Hessian $H(x, \lambda) = 3R(x) + \lambda I - R_0$. 
Consider a parametrization of the characteristic curves 
\((\lambda = \lambda(t), x = x(t)):\)
\[
\nabla_x J(x(t); \lambda(t)) = 0 \iff R(x(t))x(t) + (\lambda(t)I - R_0)x(t) = 0
\]
Differentiate to obtain:
\[
\begin{bmatrix}
x \\
\lambda
\end{bmatrix} \cdot H(x, \lambda) \begin{bmatrix}
\frac{d\lambda}{dt} \\
\frac{dx}{dt}
\end{bmatrix} = 0 \quad (\text{Diff. Sys.})
\]
with the Hessian \(H(x, \lambda) = 3R(x) + \lambda I - R_0\).
If Hessian nonsingular, we can parametrize \(x = x(\lambda)\) and
\[
\frac{dx}{d\lambda} = - (H(x, \lambda))^{-1} x.
\]
The IRLS Algorithm

The Iterative Regularized Least-Squares Algorithm attempts to find the global minimum of the non-convex problem

$$\arg\min_x \sum_{k=1}^m |y_k - |\langle x, f_k \rangle|^2|^2 + 2\lambda_\infty \|x\|^2$$
The IRLS Algorithm

IRLS Algorithm

The Iterative Regularized Least-Squares Algorithm attempts to find the global minimum of the non-convex problem

$$\arg\min_x \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 + 2\lambda_\infty \|x\|^2_2$$

using a sequence of iterative least-squares problems:

$$x^{(t+1)} = \arg\min_x \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 + 2\lambda_t \|x\|^2_2 + \mu_t \|x - x^{(t)}\|^2$$
The IRLS Algorithm

The Iterative Regularized Least-Squares Algorithm attempts to find the global minimum of the non-convex problem

\[
\arg\min_x \sum_{k=1}^m |y_k - |\langle x, f_k \rangle|^2|^2 + 2\lambda_\infty \|x\|_2^2
\]

using a sequence of iterative least-squares problems:

\[
x^{(t+1)} = \arg\min_x \sum_{k=1}^m |y_k - |\langle x, f_k \rangle|^2|^2 + 2\lambda_t \|x\|_2^2 + \mu_t \|x - x^{(t)}\|_2^2
\]

together with a polarization relaxation:

\[
|\langle x, f_k \rangle|^2 \approx \frac{1}{2} (\langle x, f_k \rangle \langle f_k, x^{(t)} \rangle + \langle x^{(t)}, f_k \rangle \langle f_k, x \rangle)
\]
The IRLS Algorithm
Main Optimization

The optimization problem:

$$x^{(t+1)} = \arg\min_x \sum_{k=1}^{m} \left| y_k - \frac{1}{2} \left( \langle x, f_k \rangle \langle f_k, x^{(t)} \rangle + \langle x^{(t)}, f_k \rangle \langle f_k, x \rangle \right) \right|^2 +$$

$$+ \lambda_t \|x\|_2^2 + \mu_t \|x - x^{(t)}\|_2^2 + \lambda_t \|x^{(t)}\|_2^2$$

$$= \arg\min_x J(x, x^{(t)}; \lambda, \mu)$$
The IRLS Algorithm

Main Optimization

The optimization problem:

\[
\begin{align*}
x^{(t+1)} &= \arg\min_x \sum_{k=1}^{m} \left| y_k - \frac{1}{2} (\langle x, f_k \rangle \langle f_k, x^{(t)} \rangle + \langle x^{(t)}, f_k \rangle \langle f_k, x \rangle) \right|^2 + \\
&\quad + \lambda_t \| x \|^2 + \mu_t \| x - x^{(t)} \|^2 + \lambda_t \| x^{(t)} \|^2 \\
&= \arg\min_x J(x, x^{(t)}; \lambda, \mu)
\end{align*}
\]

Note:

J(x, x^{(t)}; \lambda, \mu) is quadratic in x \Rightarrow hence a least-squares problem!
The IRLS Algorithm
Main Optimization

The optimization problem:

\[ x^{(t+1)} = \arg\min_x \sum_{k=1}^m \left| y_k - \frac{1}{2} (\langle x, f_k \rangle \langle f_k, x^{(t)} \rangle + \langle x^{(t)}, f_k \rangle \langle f_k, x \rangle) \right|^2 + \]

\[ + \lambda_t \| x \|^2 + \mu_t \| x - x^{(t)} \|^2 + \lambda_t \| x^{(t)} \|^2 \]

\[ = \arg\min_x J(x, x^{(t)}; \lambda, \mu) \]

Note:

- \( J(x, .; ., .) \) is quadratic in \( x \) \( \Rightarrow \) hence a least-squares problem!
The IRLS Algorithm
Main Optimization

The optimization problem:

\[ x^{(t+1)} = \arg\min_{x} \sum_{k=1}^{m} \left| y_k - \frac{1}{2}(\langle x, f_k \rangle \langle f_k, x^{(t)} \rangle + \langle x^{(t)}, f_k \rangle \langle f_k, x \rangle) \right|^2 + \lambda_t \| x \|^2_2 + \mu_t \| x - x^{(t)} \|^2_2 + \lambda_t \| x^{(t)} \|^2_2 = \arg\min_{x} J(x, x^{(t)}; \lambda, \mu) \]

Note:

- \( J(x, :, :, :) \) is quadratic in \( x \) \( \Rightarrow \) hence a least-squares problem!
- \( J(x, x; \lambda, \mu) = \sum_{k=1}^{m} \left| y_k - |\langle x, f_k \rangle|^2 \right|^2 + 2\lambda\| x \|^2_2 \Rightarrow \) Fixed points of IRLS are local minima of the original problem.
The IRLS Algorithm

Second Motivation: Relaxation of Constraints

Another motivation: seek $X = xx^*$ that solves

$$
\min_{X \geq 0, \text{rank}(X) = 1} \sum_{k=1}^{m} |y_k - \langle X, f_k f_k^* \rangle_{HS}|^2 + 2\lambda \text{trace}(X).
$$
The IRLS Algorithm
Second Motivation: Relaxation of Constraints

Another motivation: seek $X = xx^*$ that solves

$$\min_{X \geq 0, \text{rank}(X) = 1} \sum_{k=1}^{m} |y_k - \langle X, f_k f_k^* \rangle_{HS}|^2 + 2\lambda \text{trace}(X).$$

PhaseLift algorithm removes the condition $\text{rank}(X) = 1$ and shows (for large $\lambda$) this produces the desired result with high probability.
The IRLS Algorithm
Second Motivation: Relaxation of Constraints

Another motivation: seek $X = xx^*$ that solves

$$\min_{X \geq 0, \text{rank}(X)=1} \sum_{k=1}^{m} |y_k - \langle X, f_k f_k^* \rangle_{HS}|^2 + 2\lambda \text{trace}(X).$$

PhaseLift algorithm removes the condition $\text{rank}(X) = 1$ and shows (for large $\lambda$) this produces the desired result with high probability.

Another way to relax the problem is to search for $X$ in a larger space. The IRLS is essentially equivalent to optimize a convex functional of $X$ on the larger space

$$S^{1,1} = \{ T = T^* \in \mathbb{C}^{n \times n}, \ T \text{ has at most one positive eigenvalue and at most one negative eigenvalue}\}.$$
Consider the following three convex criteria:

\[
J_1(X; \lambda, \mu) = \sum_{k=1}^{m} |y_k - \langle X, f_k f_k^* \rangle_{HS}|^2 + 2(\lambda + \mu)\|X\|_1 - 2\mu \text{trace}(X)
\]

\[
J_2(X; \lambda, \mu) = \sum_{k=1}^{m} |y_k - \langle X, f_k f_k^* \rangle_{HS}|^2 + 2\lambda \text{eig}_{\text{max}}(X) - (2\lambda + 4\mu)\text{eig}_{\text{min}}(X)
\]

\[
J_3(X; \lambda, \mu) = \sum_{k=1}^{m} |y_k - \langle X, f_k f_k^* \rangle_{HS}|^2 + 2\lambda\|X\|_1 - 4\mu \text{eig}_{\text{min}}(X)
\]

which coincide on \( S^{1,1} \).
Consider the following three convex criteria:

\[
J_1(X; \lambda, \mu) = \sum_{k=1}^{m} \left| y_k - \langle X, f_k f_k^* \rangle_{HS} \right|^2 + 2(\lambda + \mu)\|X\|_1 - 2\mu \text{trace}(X)
\]

\[
J_2(X; \lambda, \mu) = \sum_{k=1}^{m} \left| y_k - \langle X, f_k f_k^* \rangle_{HS} \right|^2 + 2\lambda \text{eig}_{\text{max}}(X) - (2\lambda + 4\mu)\text{eig}_{\text{min}}(X)
\]

\[
J_3(X; \lambda, \mu) = \sum_{k=1}^{m} \left| y_k - \langle X, f_k f_k^* \rangle_{HS} \right|^2 + 2\lambda\|X\|_1 - 4\mu \text{eig}_{\text{min}}(X)
\]

which coincide on $S^{1,1}$. Consider the optimization problem

\[
(J_{\text{opt}}, X) = \min_{X \in S^{1,1}} J_k(X; \lambda, \mu), \quad 1 \leq k \leq 3
\]
The IRLS Algorithm
Second Formulation -2

The following are true:

1. Optimization in $S^{1,1}$:

   $$\min_{X \in S^{1,1}} J_k(X; \lambda, \mu) = \min_{u, v \in \mathbb{C}^n} J(u, v; \lambda, \mu)$$

   If $\hat{X}$ and $(\hat{u}, \hat{v})$ denote optimizers so that $\text{imag}(\langle \hat{u}, \hat{v} \rangle) = 0$, then
   $$\hat{X} = \frac{1}{2}(\hat{u}\hat{v}^* + \hat{v}\hat{u}^*).$$
The IRLS Algorithm
Second Formulation -2

The following are true:

1. Optimization in $S^{1,1}$:

$$\min_{X \in S^{1,1}} J_k(X; \lambda, \mu) = \min_{u, v \in \mathbb{C}^n} J(u, v; \lambda, \mu)$$

If $\hat{X}$ and $(\hat{u}, \hat{v})$ denote optimizers so that $\text{imag}(\langle \hat{u}, \hat{v} \rangle) = 0$, then

$$\hat{X} = \frac{1}{2}(\hat{u}\hat{v}^* + \hat{v}\hat{u}^*).$$

2. Optimization in $S^{1,0}$:

$$\min_{X \in S^{1,0}} J_k(X; \lambda, \mu) = \min_{x \in \mathbb{C}^n} J(x, x; \lambda, \mu)$$

If $\hat{X}$ and $\hat{x}$ denote optimizers, then $\hat{X} = \hat{x}\hat{x}^*$. $S^{1,0} = \{xx^*\}$. 
The IRLS Algorithm

Initialization

For $\lambda \geq \text{eig}_{\text{max}}(R(y))$, where $R(y) = \sum_{k=1}^{m} y_k f_k f_k^*$,

$$J(x; \lambda) = \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 + 2\lambda \|x\|_2^2$$

is convex. The unique global minimum is $x^0 = 0$. 

$\epsilon > 0$ is a parameter that depends on the frame set as well as the spectral gap of $R(y)$.
The IRLS Algorithm

Initialization

For $\lambda \geq \text{eig}_{\text{max}}(R(y))$, where $R(y) = \sum_{k=1}^{m} y_k f_k f_k^*$, $J(x; \lambda) = \sum_{k=1}^{m} |y_k - |\langle x, f_k \rangle|^2|^2 + 2\lambda\|x\|^2_2$ is convex. The unique global minimum is $x^0 = 0$.

Initialization Procedure:

- Solve the principal eigenpair $(e, \text{eig}_{\text{max}})$ of matrix $R(y)$ using e.g. the power method;
- Set

$$\lambda_0 = (1 - \varepsilon)\text{eig}_{\text{max}}, \ x^0 = \sqrt{\frac{(1 - \varepsilon)\text{eig}_{\text{max}}}{\sum_{k=1}^{m} |\langle e, f_k \rangle|^4}} e.$$ 

Here $\varepsilon > 0$ is a parameter that depends on the frame set as well as the spectral gap of $R(y)$.
- Set $\mu_0 = \lambda_0$ and $t = 0$. 

Radu Balan (UMD)
The IRLS Algorithm

Iterations

Repeat the following steps until stopping:

- **Optimization**: Solve the least-square problem:

  \[
  x^{(t+1)} = \text{argmin}_x \sum_{k=1}^{m} \left| y_k - \frac{1}{2} \left( \langle x, f_k \rangle \langle f_k, x^{(t)} \rangle + \langle x^{(t)}, f_k \rangle \langle f_k, x \rangle \right) \right|^2 + \\
  + \lambda_t \| x \|^2 + \mu_t \| x - x^{(t)} \|^2 + \lambda_t \| x^{(t)} \|^2 \\
  = \text{argmin}_x \ J(x, x^{(t)}; \lambda, \mu)
  \]

- **Update**: \( \lambda_{t+1} = \gamma \lambda_t, \mu_{t+1} = \max(\gamma \mu_t, \mu_{\text{min}}), \ t = t + 1. \) Here \( \gamma \) is the learning rate, and \( \mu_{\text{min}} \) is related to performance.

Radu Balan (UMD)
The IRLS Algorithm

Performance

Let $y_k = |\langle x, f_k \rangle|^2 + \nu_k$. Assume the algorithm is stopped at some $T$ so that

$$J(x^{(T)}, x^{(T-1)}; \lambda, \mu) \leq J(x, x; \lambda, \mu).$$

Denote $\hat{X} = \frac{1}{2}(x^{(T)}x^{(T-1)*} + x^{(T-1)}x^{(T)*})$ and $\hat{x}\hat{x}^* = P_+(\hat{X})$. 

Let $y_k = |\langle x, f_k \rangle|^2 + \nu_k$. Assume the algorithm is stopped at some $T$ so that

$$J(x^{(T)}, x^{(T-1)}; \lambda, \mu) \leq J(x, x; \lambda, \mu).$$

Denote $\hat{X} = \frac{1}{2} (x^{(T)} x^{(T-1)*} + x^{(T-1)} x^{(T)*})$ and $\hat{x}\hat{x}^* = P_+(\hat{X})$. Then the following hold true:

1. **Matrix norm error:**

   $$\|\hat{X} - xx^*\|_1 \leq \frac{\lambda}{C_0} + \sqrt{C_0} \|\nu\|$$

2. **Natural distance:**

   $$D(\hat{x}, x)^2 = \|\hat{X} - xx^*\|_1 + |\text{eig}_{\min}(\hat{X})| \leq \frac{\lambda}{C_0} + \sqrt{C_0} \|\nu\| + \frac{\|\nu\|^2}{4\mu} + \frac{\lambda \|x\|^2}{2\mu}$$

where $C_0$ is a frame dependent constant (lower Lipschitz constant in $S^{1,1}$).
Numerical Simulations

Setup

The algorithm requires $O(m)$ memory. Simulations with $m = Rn$ (complex case) with $n = 1000$ and $R \in \{4, 6, 8, 12\}$. Frame vectors corresponding to masked (windowed) DFT:

$$f_{jn+k} = \frac{1}{\sqrt{Rn}} \left( w_j e^{2\pi i k(l-1)/n} \right)_{0 \leq l \leq n-1}, \quad 1 \leq j \leq R, 1 \leq k \leq n$$

$$\begin{bmatrix}
    f_1 & f_2 & \cdots & f_m \\
\end{bmatrix} = \begin{bmatrix}
    \text{Diag}(w^1) & \cdots & \text{Diag}(w^R) \\
\end{bmatrix}\begin{bmatrix}
    DFT_n & 0 & 0 \\
    0 & \ddots & 0 \\
    0 & 0 & DFT_n \\
\end{bmatrix}$$

Parameters: $\varepsilon = 0.1$, $\gamma = 0.95$, $\mu^{min} = \frac{\mu^0}{10}$. Power method tolerance: $10^{-8}$
Conjugate gradient tolerance: $10^{-14}$. 
Numerical Simulations

MSE Plots
Numerical Simulations

MSE Plots

- **Bias/Variance/MSE/CRLB vs SNR**
  - Various SNR values are shown, with different markers indicating different conditions.

- **MSE vs SNR**
  - A graph showing the Mean Squared Error (MSE) with respect to Signal-to-Noise Ratio (SNR).
  - Different lines represent varying conditions, possibly different sample sizes or conditions.

- **Variance vs SNR**
  - A graph showing variance with respect to SNR.
  - Similar to previous plots, different lines could indicate different conditions.

- **Bias vs SNR**
  - A graph showing bias with respect to SNR.
  - Again, different lines could represent different conditions or parameters.

These graphs illustrate the performance metrics for the Homotopy Method under varying conditions, highlighting how bias, variance, MSE, and CRLB change with respect to SNR.
Numerical Simulations

Performance

**Iterations/Realization vs SNR**

- $\gamma = 0.6$
- $\gamma = 0.8$
- $\gamma = 1.2$

**Iterations/Realization vs SNR**

- $\gamma = 0.90$
- $\gamma = 0.99$
- $\gamma = 0.95$
Numerical Simulations

Performance

**Iterations/Realization vs SNR**

- $R = 6$
- $R = 8$
- $R = 12$

SNR (dB) vs Iterations/Realization

**Error vs Iteration, Input 1, Weight 1, SNR 0**

SNR (dB) vs Iteration

**Iterations/Realization vs SNR**

- Gamma = 0.90
- Gamma = 0.99
- Gamma = 0.95

SNR (dB) vs Iterations/Realization

**Error vs Iteration, Input 1, Weight 1, SNR 30**

SNR (dB) vs Iteration

Radu Balan (UMD)

Homotopy Method

16 Aug. 2017
Numerical Simulations

Performance - 2

<table>
<thead>
<tr>
<th>SNR</th>
<th>$10 \cdot \log_{10}(\text{Bias})$</th>
<th>$10 \cdot \log_{10}(\text{Variance})$</th>
<th>$10 \cdot \log_{10}(\text{MSE})$</th>
<th>CRLB</th>
</tr>
</thead>
<tbody>
<tr>
<td>-30</td>
<td>32.13</td>
<td>43.78</td>
<td>44.07</td>
<td>58.39</td>
</tr>
<tr>
<td>-20</td>
<td>32.39</td>
<td>35.29</td>
<td>40.69</td>
<td>45.39</td>
</tr>
<tr>
<td>-10</td>
<td>32.27</td>
<td>35.56</td>
<td>37.23</td>
<td>39.39</td>
</tr>
<tr>
<td>0</td>
<td>22.17</td>
<td>30.24</td>
<td>30.67</td>
<td>29.39</td>
</tr>
<tr>
<td>10</td>
<td>-2.21</td>
<td>19.16</td>
<td>19.19</td>
<td>19.39</td>
</tr>
<tr>
<td>20</td>
<td>-10.88</td>
<td>9.05</td>
<td>9.05</td>
<td>9.39</td>
</tr>
<tr>
<td>30</td>
<td>-30.63</td>
<td>-0.96</td>
<td>-0.96</td>
<td>-0.61</td>
</tr>
</tbody>
</table>
Numerical Simulations
Performance - 2

### Bias/Variance/MSE/CRLB vs SNR

<table>
<thead>
<tr>
<th>SNR</th>
<th>$10 \cdot \log_{10}(Bias)$</th>
<th>$10 \cdot \log_{10}(Variance)$</th>
<th>$10 \cdot \log_{10}(MSE)$</th>
<th>CRLB</th>
</tr>
</thead>
<tbody>
<tr>
<td>-30</td>
<td>32.13</td>
<td>43.78</td>
<td>44.07</td>
<td>58.39</td>
</tr>
<tr>
<td>-20</td>
<td>32.39</td>
<td>35.29</td>
<td>40.69</td>
<td>45.39</td>
</tr>
<tr>
<td>-10</td>
<td>32.27</td>
<td>35.56</td>
<td>37.23</td>
<td>39.39</td>
</tr>
<tr>
<td>0</td>
<td>22.17</td>
<td>30.24</td>
<td>30.34</td>
<td>29.39</td>
</tr>
<tr>
<td>10</td>
<td>-2.21</td>
<td>19.16</td>
<td>19.19</td>
<td>19.39</td>
</tr>
<tr>
<td>20</td>
<td>-18.88</td>
<td>9.05</td>
<td>9.05</td>
<td>9.39</td>
</tr>
<tr>
<td>30</td>
<td>-30.63</td>
<td>-0.96</td>
<td>-0.96</td>
<td>-0.61</td>
</tr>
</tbody>
</table>

### MSE vs SNR

- Red: Gamma = 0.30
- Green: Gamma = 0.99
- Blue: Gamma = 0.95

16 Aug. 2017
References


E. Candès, X. Li, *Solving Quadratic Equations Via PhaseLift When There Are As Many Equations As Unknowns*, available online arXiv:1208.6247

E. Candès, X. Li, M. Soltanolkotabi, *Phase Retrieval from Coded Diffraction Patterns*,


The Phase Retrieval Problem

Existing Algorithms

The Homotopy Method

Numerical Results


Existing Algorithms

The Homotopy Method

Numerical Results


