Phase retrieval with alternating projections for random sensing vectors

Irène Waldspurger
CNRS and CEREMADE, Université Paris Dauphine, MOKAPLAN, Inria

August 15, 2017
Phaseless imaging in theory and practice
IMA, University of Minnesota
Reconstruct $x_0 \in \mathbb{C}^n$ from $|Ax_0| \in (\mathbb{R}^+)^m$?

We denote by
- $n$ the dimension of $x_0$;
- $m \geq n$ the number of measurements;
- $A \in \mathbb{C}^{m \times n}$ the sensing matrix.
Reconstruct $x_0 \in \mathbb{C}^n$ from $|Ax_0| \in (\mathbb{R}^+)^m$?

We denote by
- $n$ the dimension of $x_0$;
- $m \geq n$ the number of measurements;
- $A \in \mathbb{C}^{m \times n}$ the sensing matrix.

Ideally, we would like to rigorously understand when and why available algorithms succeed in solving problems of this form.
The focus is on establishing correctness guarantees for a realistic algorithm.

For the sensing matrix, on the other hand, we assume a very simple model.

→ The sensing matrix has independent Gaussian entries.

\[ A_{ij} \sim_{i.i.d.} \mathcal{N}_\mathbb{C}(0, 1). \]

[This is not realistic, but handling more general \( A \) is difficult.]
The focus is on establishing correctness guarantees for a realistic algorithm.

For the sensing matrix, on the other hand, we assume a very simple model.
- The sensing matrix has independent Gaussian entries.

\[ A_{ij} \sim^{i.i.d.} \mathcal{N}_\mathbb{C}(0, 1). \]

[This is not realistic, but handling more general \( A \) is difficult.]

Let us first review the main algorithms for which, in this setting, we can establish reconstruction guarantees.
Convexification methods

First algorithms for which correctness guarantees were proved.

[Candès, Eldar, Strohmer, and Voroninski, 2011]
[Chai, Moscoso, and Papanicolaou, 2011]
[Candès and Li, 2014]
[Waldspurger, d’Aspremont, and Mallat, 2015]

Principle: “lift” the problem to a matricial space by a suitable change of variables.
Convexification methods

First algorithms for which correctness guarantees were proved.

[Candès, Eldar, Strohmer, and Voroninski, 2011]
[Chai, Moscoso, and Papanicolaou, 2011]
[Candès and Li, 2014]
[Waldspurger, d’Aspremont, and Mallat, 2015]

Principle: “lift” the problem to a matricial space by a suitable change of variables.

Theorem (Candès and Li [2014])

If $m \geq Cn$, for $C$ large, convexification methods recover $x_0$ with high probability.
Convexification methods

Now $n^2$ variables. $\rightarrow$ High computational cost.
$\rightarrow$ Difficult to use in practice.
(Although progress has been done. [Yurtsever et al., 2017])

Convexification without lifting is possible.
But still seems slower than non-convexified methods.

[Goldstein and Studer, 2016; Bahmani and Romberg, 2017]
Convexification methods

Now $n^2$ variables. \rightarrow High computational cost.
\rightarrow Difficult to use in practice.
(Although progress has been done. [Yurtsever et al., 2017])

Convexification without lifting is possible.
But still seems slower than non-convexified methods.

[Goldstein and Studer, 2016; Bahmani and Romberg, 2017]

\rightarrow More realistic algorithms?
Local search with careful initialization

First step: initialization

Find an approximation of $x_0$ via a “spectral method”.

Second step: local search

Typically, gradient descent over a non-convex cost, like

$$F(x) \overset{\text{def}}{=} \sum_{k=1}^{m} \left( |(Ax)_i|^2 - |(Ax_0)_i|^2 \right)^2.$$ 

[Netrapalli, Jain, and Sanghavi, 2013]
[Candes, Li, and Soltanolkotabi, 2015]
[Chen and Candès, 2015; Zhang and Liang, 2016]
[Wang, Giannakis, and Eldar, 2017]
Local search with careful initialization

Same guarantees as convexification techniques: work when

\[ m \geq Cn. \]

Much faster than convexification techniques.
→ Easy to use, even in high dimension.

However, these methods were introduced for theoretical reasons. They are not traditional methods.
Local search with careful initialization

Same guarantees as convexification techniques: work when

\[ m \geq Cn. \]

Much faster than convexification techniques.
→ Easy to use, even in high dimension.

However, these methods were introduced for theoretical reasons. They are not traditional methods.

→ Can we show similar guarantees for traditional methods?
Traditional methods?

- Alternating projections [Gerchberg and Saxton, 1972]
- Hybrid Input Output [Fienup, 1982]
Traditional methods?

- Alternating projections [Gerchberg and Saxton, 1972]
- Hybrid Input Output [Fienup, 1982]
Traditional methods?

- Alternating projections [Gerchberg and Saxton, 1972]
- Hybrid Input Output [Fienup, 1982]

Main result

Alternating projections also work with high probability when

\[ m \geq Cn, \]

under the condition that they are initialized with a spectral method.
Outline

1. Proof sketch
2. Do we really need the careful initialization?
Reconstruct $x_0 \in \mathbb{C}^n$ from $|Ax_0|$?

Alternating projections

Idea: focus on the reconstruction of $y_0 = Ax_0$.

Reconstruct $y_0 \in \mathbb{C}^m$ s.t. $y_0 \in \text{Range}(A)$ and $|y_0| = |Ax_0|$.

Natural heuristic: (Gerchberg and Saxton [1972])
Reconstruct $x_0 \in \mathbb{C}^n$ from $|Ax_0|$?

Alternating projections

Idea: focus on the reconstruction of $y_0 = Ax_0$.

Reconstruct $y_0 \in \mathbb{C}^m$ s.t.

- $y_0 \in \text{Range}(A)$
- $|y_0| = |Ax_0|$.

Natural heuristic: (Gerchberg and Saxton [1972])

- Choose an initial guess for $y_0$. 

Alternating projections

Idea: focus on the reconstruction of $y_0 = Ax_0$.

\[
\text{Reconstruct } y_0 \in \mathbb{C}^m \text{ s.t. } y_0 \in \text{Range}(A) \text{ and } |y_0| = |Ax_0|.
\]

Natural heuristic: (Gerchberg and Saxton [1972])

- Choose an initial guess for $y_0$.
- Project onto $\text{Range}(A)$.
Reconstruct $x_0 \in \mathbb{C}^n$ from $|Ax_0|$?

Alternating projections

Idea: focus on the reconstruction of $y_0 = Ax_0$.

Reconstruct $y_0 \in \mathbb{C}^m$ s.t. $y_0 \in \text{Range}(A)$ and $|y_0| = |Ax_0|$.

Natural heuristic: (Gerchberg and Saxton [1972])

- Choose an initial guess for $y_0$.
- Project onto $\text{Range}(A)$.
- Project onto $\{y \text{ s.t. } |y| = |Ax_0|\}$. 
Convergence with careful initialization

Reconstruct $x_0 \in \mathbb{C}^n$ from $|Ax_0|$?

Alternating projections

Idea: focus on the reconstruction of $y_0 = Ax_0$.

Reconstruct $y_0 \in \mathbb{C}^m$ s.t. $y_0 \in \text{Range}(A)$ and $|y_0| = |Ax_0|$.

Natural heuristic: (Gerchberg and Saxton [1972])

- Choose an initial guess for $y_0$.
- Project onto $\text{Range}(A)$.
- Project onto $\{y \text{ s.t. } |y| = |Ax_0|\}$.

Repeat the double projection
Alternating projections

Idea : focus on the reconstruction of $y_0 = Ax_0$.

Reconstruct $y_0 \in \mathbb{C}^m$ s.t. $y_0 \in \text{Range}(A)$ and $|y_0| = |Ax_0|$.

Natural heuristic : (Gerchberg and Saxton [1972])

- Choose an initial guess for $y_0$.
- Project onto $\text{Range}(A)$.
- Project onto $\{y \text{ s.t. } |y| = |Ax_0|\}$.
- Hope it converges to $y_0$.

Repeat the double projection
Theorem (Waldspurger [2016])

We initialize the algorithm by

\[ y_{\text{init}} = Ax_{\text{init}}, \]

for \( x_{\text{init}} \) given by the spectral method [Chen and Candès, 2015].

If \( m \geq Cn \), for \( C > 0 \) large enough, then the output of the algorithm after \( T \) steps satisfies

\[ ||y_{\text{output}} - y_0|| \leq \delta^T ||y_0||, \]

for some \( 0 < \delta < 1 \), with probability \( 1 - e^{-O(m)} \).
Theorem (Waldspurger [2016])

We initialize the algorithm by

\[ y_{\text{init}} = Ax_{\text{init}}, \]

for \( x_{\text{init}} \) given by the spectral method [Chen and Candès, 2015].

If \( m \geq Cn \), for \( C > 0 \) large enough, then the output of the algorithm after \( T \) steps satisfies

\[ \| y_{\text{output}} - y_0 \| \leq \delta^T \| y_0 \|, \]

for some \( 0 < \delta < 1 \), with probability \( 1 - e^{-O(m)} \).

So \( y_0 \) and \( x_0 \) can be recovered with arbitrary precision.
Proof principle, following the literature

Step 1: the initial point is close to the solution.
With high probability,
\[ \| y_{init} - y_0 \|_2 < \epsilon. \]

Step 2: local contraction.
With high probability, for some \( \delta < 1 \),
\[ \forall y \in B(y_0, \epsilon), \quad \| P(y) - y_0 \|_2 \leq \delta \| y - y_0 \|_2, \]
where \( P \) is the double projection operator.

Conclusion
\[ \| y_{output} - y_0 \| = \| P^T(y_{output}) - y_0 \| \leq \delta^T \| y_{init} - y_0 \|. \]

\( \| y_0 \| = 1 \)
Proof principle, following the literature

Step 1: the initial point is close to the solution.
With high probability,
\[ \| \mathbf{y}_{\text{init}} - \mathbf{y}_0 \|_2 < \epsilon. \]

Step 2: local contraction.
With high probability, for some \( \delta < 1 \),
\[ \forall \mathbf{y} \in B(\mathbf{y}_0, \epsilon), \quad \| P(\mathbf{y}) - \mathbf{y}_0 \|_2 \leq \delta \| \mathbf{y} - \mathbf{y}_0 \|_2, \]
where \( P \) is the double projection operator.

Conclusion
\[ \| \mathbf{y}_{\text{output}} - \mathbf{y}_0 \| = \| P^T(\mathbf{y}_{\text{output}}) - \mathbf{y}_0 \| \leq \delta^T \| \mathbf{y}_{\text{init}} - \mathbf{y}_0 \|. \]
Proof principle, following the literature

Step 1: the initial point is close to the solution.
With high probability,
\[ \| y_{init} - y_0 \|_2 < \epsilon. \]

Step 2: local contraction.
With high probability, for some \( \delta < 1 \),
\[ \forall y \in B(y_0, \epsilon), \quad \| P(y) - y_0 \|_2 \leq \delta \| y - y_0 \|_2, \]
where \( P \) is the double projection operator.

Conclusion
\[ \| y_{output} - y_0 \| = \| P^T(y_{output}) - y_0 \| \leq \delta^T \| y_{init} - y_0 \|. \]
Convergence with careful initialization

We want to show:

\[ \forall y \in B(y_0, \epsilon), \quad ||P(y) - y_0||_2 \leq \delta ||y - y_0||_2 \]

(small, but independent from \( n, m \))

(Uniform over \( y \))

Previous works

1. [Noll and Rondepierre, 2016]
   More general measurements, but \( \epsilon \) strongly depends on \( n \).

2. [Netrapalli, Jain, and Sanghavi, 2013]
   Proof for \( y \) fixed, not uniformly over \( y \).
   \[ \rightarrow \] Convergence guarantees for a resampled version of the algorithm.

3. [Soltanolkotabi, 2014]
   Proof for \( \epsilon \) that depends on \( n \).
   \[ \rightarrow \] Convergence guarantees for a complex initialization procedure.
We want to show:

\[ \forall y \in B(y_0, \epsilon), \quad \| P(y) - y_0 \|_2 \leq \delta \| y - y_0 \|_2 \]

Explicit expression of \( P \)

\[ P(y) = |y_0| \times \text{phase}(A A^\dagger y), \]

Projection onto Range(\( A \))

Projection onto the phase manifold

with \( \text{phase}(\mathbf{v}) \stackrel{\text{def}}{=} \left( \begin{array}{c} v_1/|v_1| \\ \vdots \\ v_m/|v_m| \end{array} \right) \).
We write $y = y_0 + d$, with $\|d\| < \epsilon$.

\[
P(y) = |y_0| \times \text{phase}(AA^\dagger y)
\approx y_0 + \text{Im} \left( \cdot \frac{AA^\dagger d}{\text{phase}(y_0)} \right).
\]

**Intuition:** $AA^\dagger d$ behaves like a random vector, with Gaussian coordinates, independent from $y_0$, so

\[
\left\| \text{Im} \left( \cdot \frac{AA^\dagger d}{\text{phase}(y_0)} \right) \right\| \approx \frac{1}{\sqrt{2}} \left\| \cdot \frac{AA^\dagger d}{\text{phase}(y_0)} \right\| \lesssim \frac{1}{\sqrt{2}} \|d\|.
\]
We write \( y = y_0 + d \), with \( ||d|| < \epsilon \).

\[
P(y) = |y_0| \times \text{phase}(AA^\dagger y)
\approx y_0 + \text{Im} \left( \frac{AA^\dagger d}{\text{phase}(y_0)} \right).
\]

First order

\[
\Rightarrow \quad ||P(y) - y_0|| \lesssim \frac{1}{\sqrt{2}} ||y - y_0||.
\]

This is our contraction property.
Warning: we have considered the first order expansion of a function that is not even continuous.

→ We have specific error terms to control.
Convergence with random initialization?

Pairs \((n, m)\) for which success probability is 50%.
Convergence with random initialization?

Pairs \((n, m)\) for which success probability is 50%.
Convergence with random initialization?

Pairs \((n, m)\) for which success probability is 50%.

→ The initialization method does not seem to change much.
It seems that alternating projections with random initialization also work with high probability in the regime

\[ m \geq Cn. \]

Why?
It seems that alternating projections with random initialization also work with high probability in the regime

\[ m \geq Cn. \]

Why?

First idea

Alternating projections converge to a “stagnation point”.

Show that there is no stagnation point other than \( x_0 \)? This property holds for several other non-convex methods. [Sun, Qu, and Wright, 2016; Ge, Lee, and Ma, 2016] [Bhojanapalli, Neyshabur, and Srebro, 2016; Boumal, 2016]
It seems that alternating projections with random initialization also work with high probability in the regime

\[ m \geq Cn. \]

Why?

First idea

Alternating projections converge to a "stagnation point".

Show that there is no stagnation point other than \( x_0 \)?

This property holds for several other non-convex methods. [Sun, Qu, and Wright, 2016; Ge, Lee, and Ma, 2016]
[Bhojanapalli, Neyshabur, and Srebro, 2016; Boumal, 2016]

No. It seems that bad stagnation points disappear only in the regime \( m = O(n^2) \).
Question 1

Why does the alternating projections method behave differently from other non-convex methods?

→ Non-continuity of the double projection operator?

Question 2

Can we show that the algorithm succeeds, with random initialization, despite the presence of stagnation points?
Question 1

Why does the alternating projections method behave differently from other non-convex methods?

→ Non-continuity of the double projection operator?

Question 2

Can we show that the algorithm succeeds, with random initialization, despite the presence of stagnation points?

Thank you.


