Algorithms for phase retrieval problems
Part III

Irène Waldspurger

CNRS et CEREMADE (Université Paris Dauphine)
Équipe MOKAPLAN (INRIA)

12 décembre 2017

Journées de géométrie algorithmique
Aussois

Reminder

Reconstruct $x \in \mathbb{C}^n$ from $(|\langle x, f_k\rangle|)_{1\leq k\leq m}$?

(The $f_1, \dots, f_m$ are known elements of $\mathbb{C}^n$.)

Reminder

Reconstruct $x \in \mathbb{C}^n$ from $(|\langle x, f_k\rangle|)_{1\le k\le m}$?

(The $f_1, \dots, f_m$ are known elements of $\mathbb{C}^n$.)
They have been chosen at random, with a normal distribution :

$$f_k \overset{i.i.d.}{\sim} \mathcal{N}_{\mathbb{C}}(0, \mathrm{Id}_n), \quad k = 1, \dots, m.$$

Reminder

Reconstruct $x \in \mathbb{C}^n$ from $(|\langle x, f_k\rangle|)_{1\le k\le m}$?

(The $f_1, \dots, f_m$ are known elements of $\mathbb{C}^n$.)
They have been chosen at random, with a normal distribution :

$$f_k \overset{i.i.d.}{\sim} \mathcal{N}_{\mathbb{C}}(0, \mathrm{Id}_n), \quad k = 1, \dots, m.$$

Provided that, for some $\alpha > 0$,

$$m \ge \alpha n,$$

the phase retrieval problem is well-posed with high probability.

Reminder

Reconstruct $x \in \mathbb{C}^n$ from $(|\langle x, f_k \rangle|)_{1 \le k \le m}$?

(The $f_1, \dots, f_m$ are known elements of $\mathbb{C}^n$.)
They have been chosen at random, with a normal distribution :

$$f_k \overset{i.i.d.}{\sim} \mathcal{N}_{\mathbb{C}}(0, \mathrm{Id}_n), \quad k = 1, \dots, m.$$

Provided that, for some $\alpha > 0$,

$$m \ge \alpha n,$$

the phase retrieval problem is well-posed with high probability.

It is however difficult to solve, because it is non-convex.

Yesterday afternoon : convexification methods

Principle

- Approximate the non-convex problem by a convex one.
- Show that the non-convex problem and its convex approximation have the same solution.

Results

- Algorithms with strong theoretical guarantees.
- Solving *PhaseLift* is slow, but solving *PhaseMax* is faster.

Today : non-convex methods

Very broadly speaking,

- Ignore the non-convexity of the problem ; solve as if it was convex.
- Hope that it works.

Practitioners traditionally used this kind of methods.
But there was not much theoretical understanding.

In the last three years, these methods have been improved, and a theoretical analysis has been done.

Outline

1. Presentation of the main non-convex methods
2. Analysis of “improved” non-convex methods
3. Analysis of “non-improved” non-convex methods ?

Main non-convex methods

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

Main non-convex methods

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

Alternating projections

$$\text{Find } x \in \mathbb{C}^n \quad \text{s.t. } |\langle x, f_k \rangle| = b_k, \quad \forall k \leq m$$

Instead of directly reconstructing x, we focus on reconstructing

$$y = (\langle x, f_k \rangle)_{1 \leq k \leq m}.$$

(Equivalent because $(f_k)_{k \leq m}$ is a generating family.)

All we know about y is that it has the following properties.

(Property 1) $|y_k| = b_k, \forall k \leq m.$
(Property 2) $y \in Range\,(z \to (\langle z, f_k \rangle)_{1 \leq k \leq m}).$

Alternating projections

Find y that satisfies (Property 1) and (Property 2) ?

Let E_1 be the set of points that satisfy (Property 1).
Let E_2 be the set of points that satisfy (Property 2).

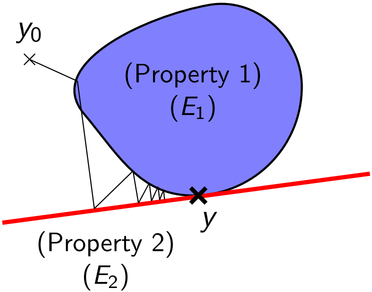
$$\text{Find } y \in E_1 \cap E_2 \,?$$

Alternating projections

Find y that satisfies (Property 1) and (Property 2) ?

Let E_1 be the set of points that satisfy (Property 1).
Let E_2 be the set of points that satisfy (Property 2).

$$\text{Find } y \in E_1 \cap E_2 \text{ ?}$$

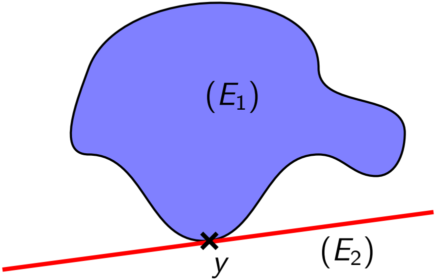


If the sets are convex, a possible algorithm is to

- start from any point y_0 ;
- alternately project it onto E_1 and E_2.

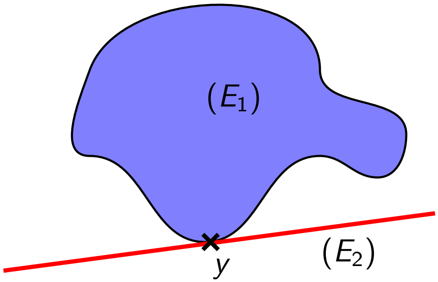
Alternating projections

The sets are not convex, but we use the same algorithm.



Alternating projections

The sets are not convex, but we use the same algorithm.

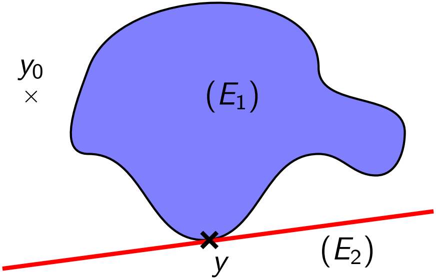


Algorithm

1. Choose any y_0.
2. Project it on E_1.
3. Project it on E_2.
4. Repeat Steps 2-3 several times.

Alternating projections

The sets are not convex, but we use the same algorithm.

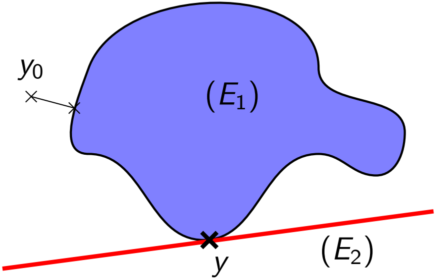


Algorithm

1. Choose any y_0.
2. Project it on E_1.
3. Project it on E_2.
4. Repeat Steps 2-3 several times.

Alternating projections

The sets are not convex, but we use the same algorithm.

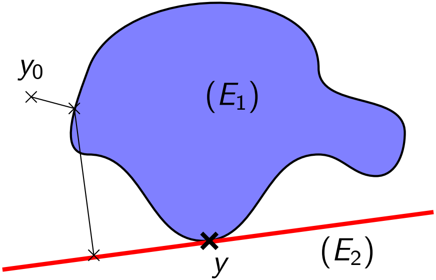


Algorithm

1. Choose any y_0.
2. Project it on E_1.
3. Project it on E_2.
4. Repeat Steps 2-3 several times.

Alternating projections

The sets are not convex, but we use the same algorithm.

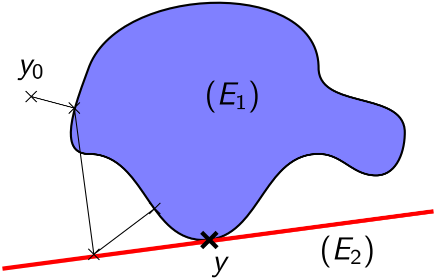


Algorithm

1. Choose any y_0.
2. Project it on E_1.
3. Project it on E_2.
4. Repeat Steps 2-3 several times.

Alternating projections

The sets are not convex, but we use the same algorithm.

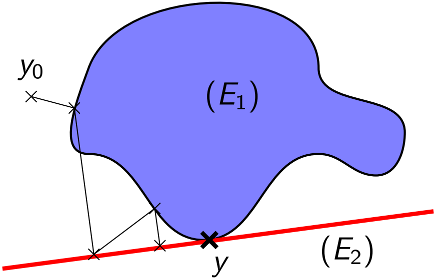


Algorithm

1. Choose any y_0.
2. Project it on E_1.
3. Project it on E_2.
4. Repeat Steps 2-3 several times.

Alternating projections

The sets are not convex, but we use the same algorithm.

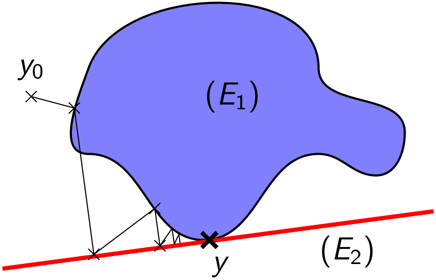


Algorithm

1. Choose any y_0.
2. Project it on E_1.
3. Project it on E_2.
4. Repeat Steps 2-3 several times.

Alternating projections

The sets are not convex, but we use the same algorithm.



Algorithm

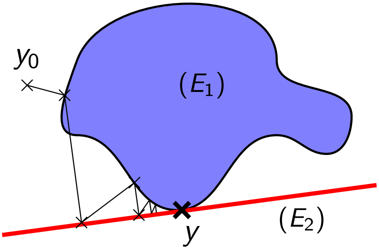
1. Choose any y_0.
2. Project it on E_1.
3. Project it on E_2.
4. Repeat Steps 2-3 several times.

Alternating projections

- Simple algorithm, easy to implement.
- Easily incorporates available additional information if any.
- Fast.
  $\rightarrow$ $O(n + n^2 \log(1/\epsilon))$ operations per iteration.

Alternating projections

- Simple algorithm, easy to implement.
- Easily incorporates available additional information if any.
- Fast.
  $\rightarrow O(n + n^2 \log(1/\epsilon))$ operations per iteration.
- But non-convexity can a priori make it fail.



(Success) ("Bad critical point")

Gradient descent

Find $x \in \mathbb{C}^n$
s.t. $|\langle x, f_k\rangle| = b_k, \quad \forall k \leq m$

Choose a reasonable objective function, like :

$$Obj : \begin{array}{ccc} \mathbb{C}^n & \to & \mathbb{R} \\ z & \to & \frac{1}{m}\sum_{k=1}^{m}\left(|\langle z, f_k\rangle|^2 - b_k^2\right)^2 . \end{array}$$

The solution is the only global minimum of Obj.

Gradient descent

$$\begin{aligned} &\text{Find } x \in \mathbb{C}^n \\ &\text{s.t. } |\langle x, f_k \rangle| = b_k, \quad \forall k \leq m \end{aligned}$$

Choose a reasonable objective function, like :

$$\begin{array}{rccc} Obj: & \mathbb{C}^n & \to & \mathbb{R} \\ & z & \to & \frac{1}{m}\sum_{k=1}^m \left(|\langle z, f_k\rangle|^2 - b_k^2\right)^2. \end{array}$$

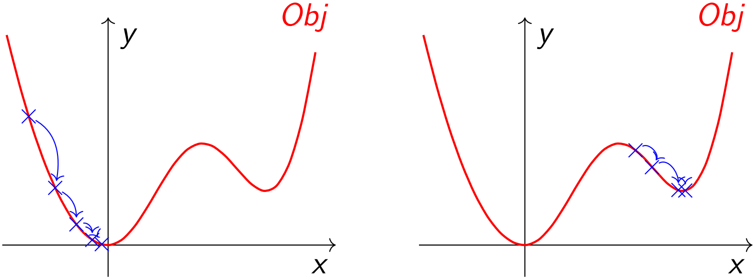
The solution is the only global minimum of Obj.

Algorithm

- Run gradient descent on Obj.
- Hope that it finds the global minimum.

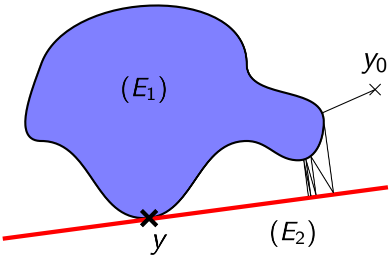
Gradient descent

- Simple algorithm, easy to implement.
- Fast
  $\rightarrow O(n^2)$ operations per iteration.
- But non-convexity can a priori make it fail.



(Success) ("Bad critical point")

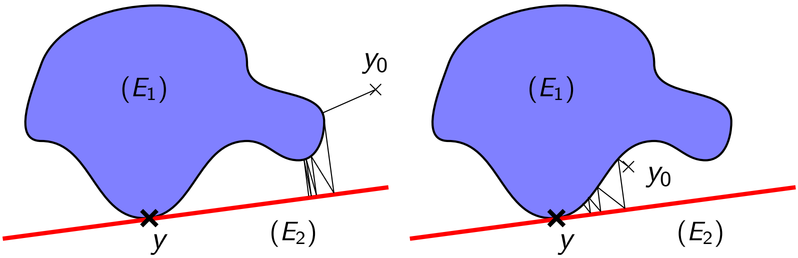
How to avoid bad critical points ?



(Bad critical point)

[Netrapalli, Jain, and Sanghavi, 2013]

How to avoid bad critical points ?



(Bad critical point)

This would not have happened if y_0 had been close enough to y.

[Netrapalli, Jain, and Sanghavi, 2013]

Compute x_0 close enough to x_{true} ?

It is possible, using the randomness of $f_1, \dots, f_m$.

Idea : by concentration inequalities,

$$\frac{1}{m}\sum_{k=1}^{m} |\langle x_{true}, f_k\rangle|^2 f_k f_k^* \approx \mathbb{E}\left(|\langle x_{true}, f\rangle|^2 ff^*\right)$$

where $f \sim \mathcal{N}_{\mathbb{C}}(0, \mathrm{Id}_n)$.

Compute x_0 close enough to x_{true} ?

It is possible, using the randomness of $f_1, \dots, f_m$.

Idea : by concentration inequalities,

$$\frac{1}{m}\sum_{k=1}^{m} |\langle x_{true}, f_k\rangle|^2 f_k f_k^* \approx \mathbb{E}\left(|\langle x_{true}, f\rangle|^2 ff^*\right) = \mathrm{Id}_n + x_{true}x_{true}^*,$$

where $f \sim \mathcal{N}_{\mathbb{C}}(0, \mathrm{Id}_n)$.

Compute x_0 close enough to x_{true} ?

It is possible, using the randomness of $f_1, \dots, f_m$.

Idea : by concentration inequalities,

$$\frac{1}{m}\sum_{k=1}^{m} |\langle x_{true}, f_k\rangle|^2 f_k f_k^* \approx \mathbb{E}\left(|\langle x_{true}, f\rangle|^2 f f^*\right) = \mathrm{Id}_n + x_{true} x_{true}^*,$$

where $f \sim \mathcal{N}_{\mathbb{C}}(0, \mathrm{Id}_n)$.

Set

$$x_0 = \text{main eigenvector}\left(\tfrac{1}{m}\sum_{k=1}^{m} |\langle x_{true}, f_k\rangle|^2 f_k f_k^*\right).$$

Compute x_0 close enough to x_{true} ?

$$x_0 = \text{main eigenvector} \left(\frac{1}{m} \sum_{k=1}^{m} |\langle x_{true}, f_k \rangle|^2 f_k f_k^* \right) .$$

This exact definition is not optimal.

One issue : the indexes k for which $|\langle x_{true}, f_k \rangle|$ is large can induce large unwanted deviations of the eigenvector.

One solution : generalize to

$$x_0 = \text{main eigenvector} \left(\frac{1}{m} \sum_{k=1}^{m} \sigma(|\langle x_{true}, f_k \rangle|) f_k f_k^* \right) ,$$

with σ better than the square (e.g. $\sigma = (s \to s^2 1_{|s| \leq 3})$).

Theorem (Spectral initialization works)

Let $\delta > 0$ be fixed.
There exist $\alpha, \gamma > 0$ such that, when

$$m \geq \alpha n,$$

then, if we define x_0 as in the previous slide,

$$||x_0 - x_{true}||_2 \leq \delta ||x_{true}||_2,$$

with probability at least $1 - O(e^{-\gamma m})$.

[Chen and Candès, 2015]
[Chen, Fannjiang, and Liu, 2015]
[Mondelli and Montanari, 2017]

Consider one of the following algorithms :

- Gradient descent with a (specific) smooth objective ;
- Gradient descent with a (specific) non-smooth objective ;
- Alternating projections.

Theorem (With a good initialization, it works)

There exists $\alpha, \gamma > 0$ and $\eta \in]0;1[$ such that, if

$$m \geq \alpha n,$$

when the algorithm is initialized with the previous x_0, then its estimate x_t after t steps satisfies

$$||x_t - x_{true}||_2 \leq \eta^t ||x_{true}||_2,$$

with probability at least $1 - O(e^{-\gamma m})$.

[Candès, Li, and Soltanolkotabi, 2015]
[Chen and Candès, 2015] [Zhang and Liang, 2016]
[Wang, Giannakis, and Eldar, 2017] [Waldspurger, 2017]

Idea of proof for smooth gradient descent

$$\begin{aligned} &\text{Find } x \in \mathbb{C}^n \\ &\text{s.t. } |\langle x, f_k \rangle| = b_k, \quad \forall k \leq m \end{aligned}$$

The smooth objective function is

$$\begin{array}{rccc} Obj: & \mathbb{C}^n & \to & \mathbb{R} \\ & z & \to & \frac{1}{m}\sum_{k=1}^{m}\left(|\langle z, f_k\rangle|^2 - b_k^2\right)^2. \end{array}$$

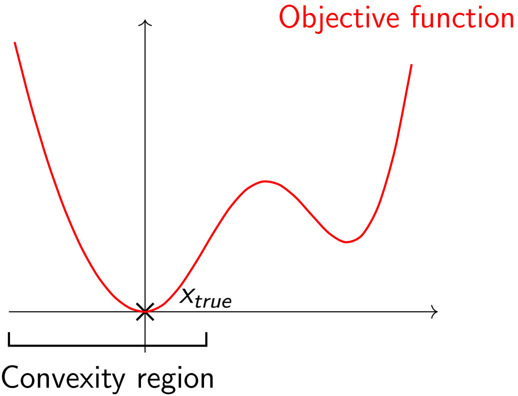
Wirtinger Flow algorithm : $\forall t \in \mathbb{N}$, $x_{t+1} = x_t - \mu \nabla\, Obj(x_t)$.

Idea of proof

Intuition

The objective function is (more or less) locally convex around x_{true}.
[White, Sanghavi, and Ward, 2017]

Being close to x_{true}, the initial point x_0 belongs to the convexity region.

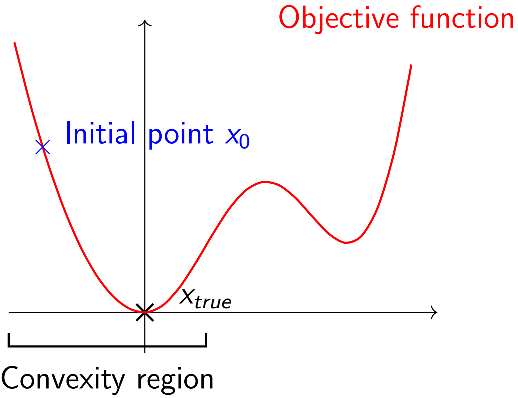


Idea of proof

Intuition

The objective function is (more or less) locally convex around x_{true}. [White, Sanghavi, and Ward, 2017]

Being close to x_{true}, the initial point x_0 belongs to the convexity region.

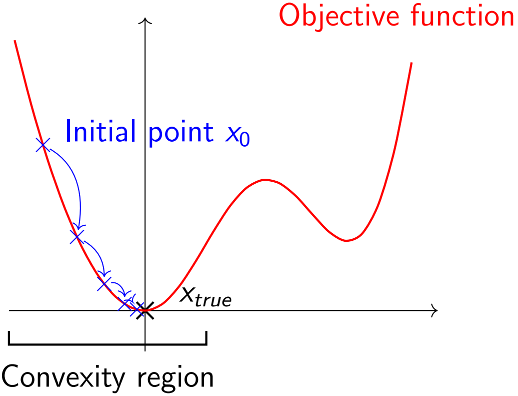


Idea of proof

Intuition

The objective function is (more or less) locally convex around x_{true}. [White, Sanghavi, and Ward, 2017]

Being close to x_{true}, the initial point x_0 belongs to the convexity region.



Idea of proof

To simplify, we assume $||x_{true}|| = 1$.

We have seen that, with high probability,

$$||x_0 - x_{true}||_2 \leq \frac{1}{8}.$$

Idea of proof

To simplify, we assume $||x_{true}|| = 1$.

We have seen that, with high probability,

$$||x_0 - x_{true}||_2 \leq \frac{1}{8}.$$

We will show that, for all $z \in B(x_{true}, 1/8)$,

$$\Big|\Big|\big(z - \mu\nabla\, Obj(z)\big) - x_{true}\Big|\Big|_2 \leq \eta||z - x_{true}||_2,$$

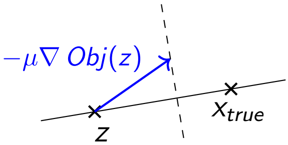
for some fixed constant $\eta < 1$.

Idea of proof

Show that, for all $z \in B(x, 1/8)$,

$$\left|\left|\left(z - \mu\nabla\, Obj(z)\right) - x_{true}\right|\right|_2 \leq \eta ||z - x_{true}||_2,$$

for some fixed constant $\eta < 1$?



If, in addition, we can control $||\mu\nabla\, Obj(z)||_2$, it is enough to show

$$\mathrm{Re}\,\langle x_{true} - z, -\mu\nabla\, Obj(z)\rangle \geq \epsilon ||z - x_{true}||_2^2.$$

Idea of proof

Show that, for all $z \in B(x, 1/8)$,

$$\mathrm{Re}\,\langle z - x_{true}, \nabla\, Obj(z)\rangle \geq \epsilon ||z - x_{true}||_2^2,$$

that is, for all $h \in B(0, 1/8)$,

$$\mathrm{Re}\,\langle h, \nabla\, Obj(x_{true} + h)\rangle \geq \epsilon ||h||_2^2.$$

Idea of proof

Show that, for all $z \in B(x, 1/8)$,

$$\mathrm{Re}\,\langle z - x_{true}, \nabla\, Obj(z)\rangle \geq \epsilon ||z - x_{true}||_2^2,$$

that is, for all $h \in B(0, 1/8)$,

$$\mathrm{Re}\,\langle h, \nabla\, Obj(x_{true} + h)\rangle \geq \epsilon ||h||_2^2.$$

We can compute

$$\begin{array}{rccc} \nabla\, Obj : & \mathbb{C}^n & \to & \mathbb{C}^n \\ & z & \to & \frac{4}{m}\sum_{k=1}^{m}\left(|\langle z, f_k\rangle|^2 - |\langle x_{true}, f_k\rangle|^2\right)\overline{\langle z, f_k\rangle} f_k. \end{array}$$

Idea of proof

Show that, for all $z \in B(x, 1/8)$,

$$\mathrm{Re}\,\langle z - x_{true}, \nabla\, Obj(z)\rangle \geq \epsilon ||z - x_{true}||_2^2,$$

that is, for all $h \in B(0, 1/8)$,

$$\mathrm{Re}\,\langle h, \nabla\, Obj(x_{true} + h)\rangle \geq \epsilon ||h||_2^2.$$

We can compute

$$\mathrm{Re}\,\langle h, \nabla\, Obj(x_{true} + h)\rangle = \frac{4}{m}\sum_{k=1}^{m}\Big(2\,\mathrm{Re}\,(\overline{\langle x_{true}, f_k\rangle}\langle h, f_k\rangle)^2 + 3\,\mathrm{Re}\,(\overline{\langle x_{true}, f_k\rangle}\langle h, f_k\rangle)|\langle h, f_k\rangle|^2 + |\langle h, f_k\rangle|^4\Big)$$

Idea of proof

For a fixed h, with high probability,

$$\mathrm{Re}\,\langle h, \nabla\, Obj(x_{true} + h)\rangle = \frac{4}{m}\sum_{k=1}^{m}\Big(2\,\mathrm{Re}\,(\overline{\langle x_{true}, f_k\rangle}\langle h, f_k\rangle)^2 + 3\,\mathrm{Re}\,(\overline{\langle x_{true}, f_k\rangle}\langle h, f_k\rangle)|\langle h, f_k\rangle|^2 + |\langle h, f_k\rangle|^4\Big)$$

$\approx$ its expectation

(Concentration inequalities)

Idea of proof

For a fixed h, with high probability,

$$\mathrm{Re}\,\langle h, \nabla\, Obj(x_{true} + h)\rangle = \frac{4}{m}\sum_{k=1}^{m}\Big(2\,\mathrm{Re}\,(\overline{\langle x_{true}, f_k\rangle}\langle h, f_k\rangle)^2 + 3\,\mathrm{Re}\,(\overline{\langle x_{true}, f_k\rangle}\langle h, f_k\rangle)|\langle h, f_k\rangle|^2 + |\langle h, f_k\rangle|^4\Big)$$

$$\approx \text{its expectation}$$

$$= 4\left(3|\langle x_{true}, h\rangle|^2 + ||h||^2 + 3\langle x_{true}, h\rangle||h||^2 + 2||h||^4\right)$$

(Concentration inequalities)

Idea of proof

For a fixed h, with high probability,

$$\mathrm{Re}\,\langle h, \nabla\, Obj(x_{true} + h)\rangle = \frac{4}{m}\sum_{k=1}^{m}\Big(2\,\mathrm{Re}\,(\overline{\langle x_{true}, f_k\rangle}\langle h, f_k\rangle)^2 + 3\,\mathrm{Re}\,(\overline{\langle x_{true}, f_k\rangle}\langle h, f_k\rangle)|\langle h, f_k\rangle|^2 + |\langle h, f_k\rangle|^4\Big)$$

$$\approx \text{ its expectation}$$

(Concentration inequalities)

$$= 4\left(3|\langle x_{true}, h\rangle|^2 + ||h||^2 + 3\langle x_{true}, h\rangle||h||^2 + 2||h||^4\right)$$

$$\geq \frac{5}{2}||h||^2$$

Idea of proof

For a fixed h, with high probability,

$$\mathrm{Re}\,\langle h, \nabla\, Obj(x_{true} + h)\rangle = \frac{4}{m}\sum_{k=1}^{m}\Big(2\,\mathrm{Re}\,(\overline{\langle x_{true}, f_k\rangle}\langle h, f_k\rangle)^2$$

$$+3\,\mathrm{Re}\,(\overline{\langle x_{true}, f_k\rangle}\langle h, f_k\rangle)|\langle h, f_k\rangle|^2 + |\langle h, f_k\rangle|^4\Big)$$

$$\approx \text{ its expectation}$$

(Concentration inequalities)

$$= 4\left(3|\langle x_{true}, h\rangle|^2 + ||h||^2 + 3\langle x_{true}, h\rangle ||h||^2 + 2||h||^4\right)$$

$$\geq \frac{5}{2}||h||^2$$

It holds for all $h \in B(0, 1/8)$ by a union bound argument.

Related literature

The same method

Good initialization + gradient descent

has been used to develop algorithms for related non-convex problems ("low-rank matrix recovery problems").

Examples

- Matrix sensing
  [Zhao, Wang, and Liu, 2015]
- Matrix completion
  [Jain, Netrapalli, and Sanghavi, 2013]
- Sparse PCA
  [Chen and Wainwright, 2015]
- ...

How important is the initialization ?

The previous proof strongly relied on the use of a carefully chosen initial point.

Is it an artifact of the proof technique, or is it really necessary to carefully choose the initial point ?

For some related problems, it has been shown that non-convex algorithms can succeed regardless of their initial point in certain regimes :

- Matrix sensing
  [Bhojanapalli, Neyshabur, and Srebro, 2016]
- Matrix completion
  [Ge, Lee, and Ma, 2016]
- Phase synchronization
  [Boumal, 2016]

Non-convex algorithms with arbitrary initialization

Theorem (Sun, Qu, and Wright [2017])

There exist $\alpha, \gamma > 0$ such that, when

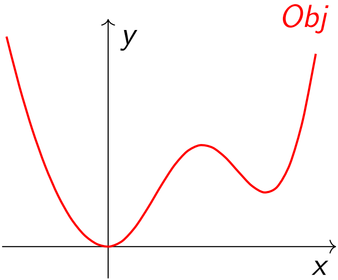
$$m \geq \alpha n \log^3(n),$$

then, with probability at least $1 - \frac{\gamma}{m}$,
non-convex gradient descent with the same smooth objective as previously returns a sequence $(x_t)_{t\in\mathbb{N}}$ such that

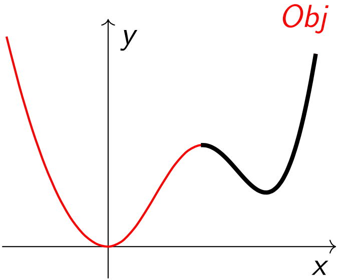
$$x_t \overset{t\to+\infty}{\longrightarrow} x_{true},$$

except possibly for x_0 in a set with zero Lebesgue measure.

Why is it possible ?

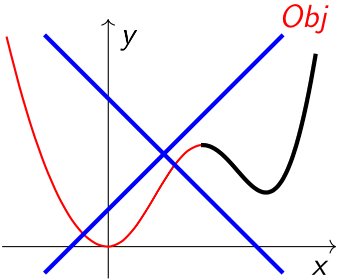


Why is it possible ?



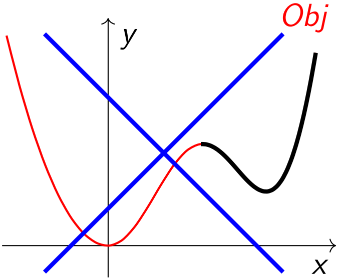
For this non-convex function, the set of "bad initial points" has non-zero Lebesgue measure.

Why is it possible ?

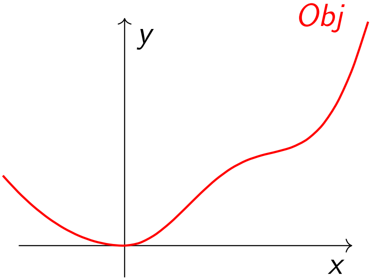


For this non-convex function, the set of “bad initial points” has non-zero Lebesgue measure.

Why is it possible ?

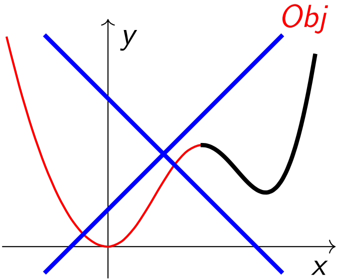


For this non-convex function, the set of "bad initial points" has non-zero Lebesgue measure.

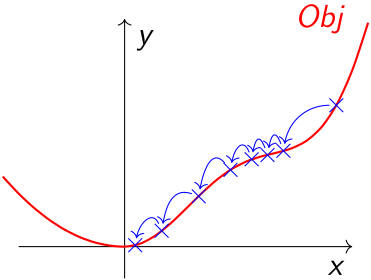


This non-convex function has no bad initial point.

Why is it possible ?



For this non-convex function, the set of “bad initial points” has non-zero Lebesgue measure.



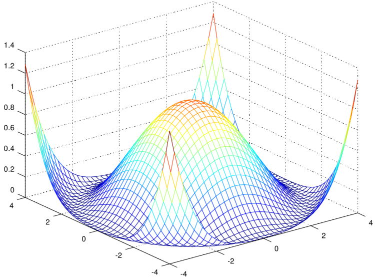
This non-convex function has no bad initial point.

Idea of proof

Principle : show that there is no point in which gradient descent can get stuck, unless it starts from a zero measure set.

Show that for any z that is not the solution :

- Either $\nabla Obj(z) \neq 0$: z is not a critical point.
- Or z is a critical point, but an unstable critical point.



Idea of proof

When is a critical point z unstable?
$\rightarrow$ At least when the Hessian $\nabla^2(z)$ has a (strictly) negative eigenvalue.
[Lee, Simchowitz, Jordan, and Recht, 2016]

Show that for all z that is not the solution,

$$\nabla\, Obj(z) \neq 0 \quad \text{or} \quad \lambda_{min}(\nabla^2\, Obj(z)) < 0?$$

Idea of proof

$$\nabla \, Obj(z) \neq 0 \quad \text{or} \quad \lambda_{min}(\nabla^2 \, Obj(z)) < 0?$$

Split $\mathbb{C}^n$ in zones :

- Zone 1 : when $||z||$ is small or $\langle x_{true}, z \rangle \approx 0$,

$$\nabla^2 \, Obj(z).(x_{true}, x_{true}) < 0.$$

- Zone 2 : when $||z||$ is large,

$$\langle \nabla \, Obj(z), z \rangle \neq 0.$$

- Zone 3 : when $||z||$ is medium, and $\langle x_{true}, z \rangle \not\approx 0$,

$$\langle \nabla \, Obj(z), z - x_{true} \rangle \neq 0.$$

Idea of proof

Zone 1 : show that when $||z||$ is small or $\langle x_{true}, z\rangle \approx 0$,

$$\nabla^2 Obj(z).(x_{true}, x_{true}) < 0?$$

Same principle as before

- Write the expression of $\nabla^2 Obj(z).(x_{true}, x_{true})$.
- Compute its expectation, and show that it is negative.
- With concentration inequalities, show that $\nabla^2 Obj(z).(x_{true}, x_{true})$ is close to its expectation.

Does it work for other algorithms ?

For alternating projections, one can show that bad critical points (more or less) disappear, with high probability, when

$$m \geq \alpha n^2.$$

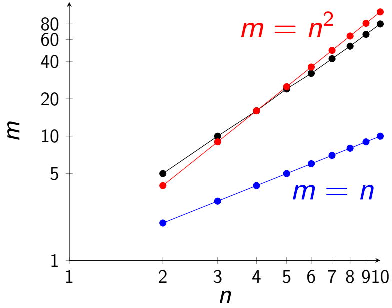
This is much worse than for smooth gradient descent.

Does it work for other algorithms ?

For alternating projections, one can show that bad critical points (more or less) disappear, with high probability, when

$$m \geq \alpha n^2.$$

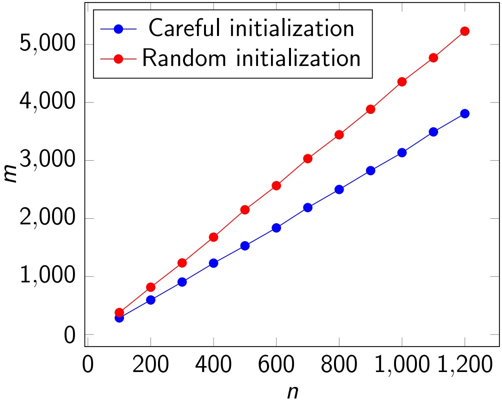
This is much worse than for smooth gradient descent.



(Dim m at which no bad critical points exists with proba $1/2$.)

Alternating projections with random initialization

Nevertheless, starting from a random initial point, alternating projections seem to succeed even when $m = O(n)$.

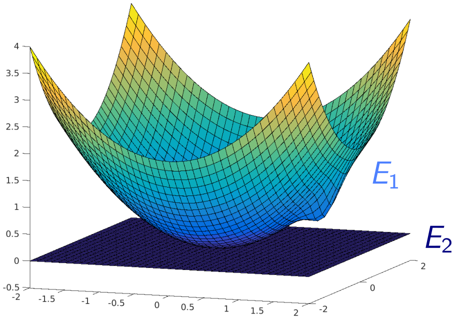


Value of m for which success probability is 50%.

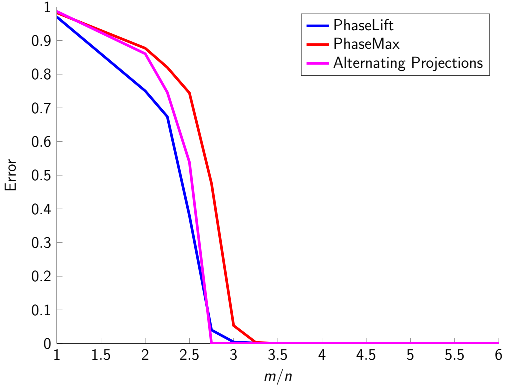
Apparently, there are bad critical points, but their attraction basin is small.

⇒ If the initialization is chosen at random, the probability to land in one of these attraction basins is small.

Tentative illustration in 3D



Numerical results



Median error as a function of m/n for $n = 64$.

Summary

Today, we have discussed non-convex methods.

- Almost the same theoretical guarantees as convexification techniques.
- Simpler and faster to implement.
- Theoretical analysis is more involved.

Open questions

- Better understanding of the importance (or not) of the initialization method ?
  Why don't all algorithms behave the same with this respect ?
- Incorporate the structure of x in the reconstruction algorithms ?
  [Soltanolkotabi, 2017]
- Extend these algorithms to non-random measurement vectors $f_1, \dots, f_m$?