Alternating projections with applications to Gerchberg-Saxton error reduction

Dominikus Noll

INSTITUT de MATHEMATIQUES de TOULOUSE

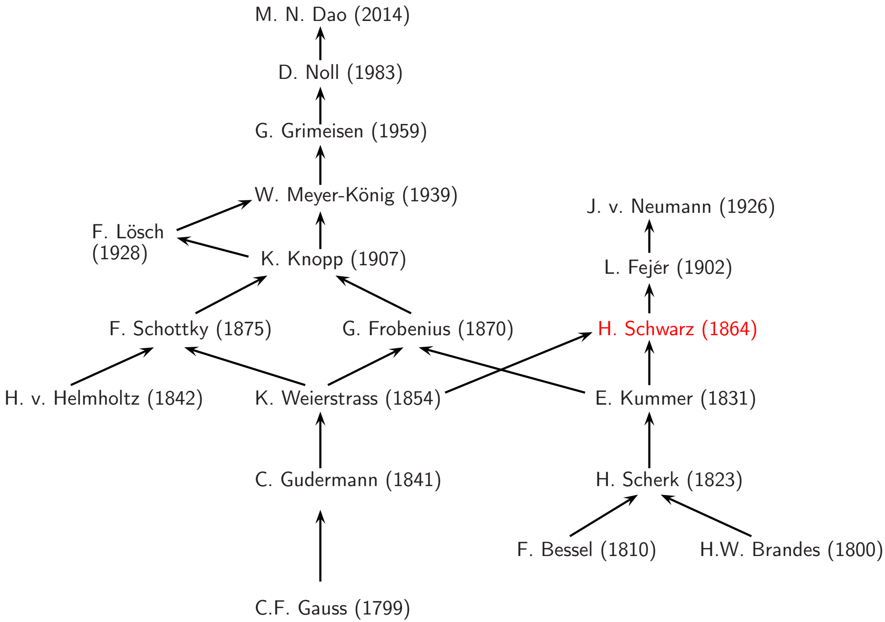
Université Paul Sabatier

Alternating projections invented by Hermann Schwarz in 1869

H. Schwarz. Über einen Grenzübergang durch alternirendes Verfahren.

Vierteljahresschrift der Naturforschenden Gesellschaft in Zürich, 15(1870), 272–286.

- Rigorous solution to Dirichlet problem 1870 - 1901
- First domain decomposition method ever
- Modern re-interpretation as MAP by P.-L. Lions 1978, 1988-89.


M. N. Dao (2014)
D. Noll (1983)
G. Grimeisen (1959)
W. Meyer-König (1939)
F. Lösch (1928)
K. Knopp (1907)
J. v. Neumann (1926)
L. Fejér (1902)
F. Schottky (1875)
G. Frobenius (1870)
H. Schwarz (1864)
H. v. Helmholtz (1842)
K. Weierstrass (1854)
E. Kummer (1831)
C. Gudermann (1841)
H. Scherk (1823)
F. Bessel (1810)
H.W. Brandes (1800)
C.F. Gauss (1799)

Method erroneously attributed to J. von Neumann :

S.L. Sobolev. L'algorithme de Schwarz dans la théorie de l'élasticité. Comptes Rendus de l'Académie des Sciences de l'URSS, IV :243–246, 1936.

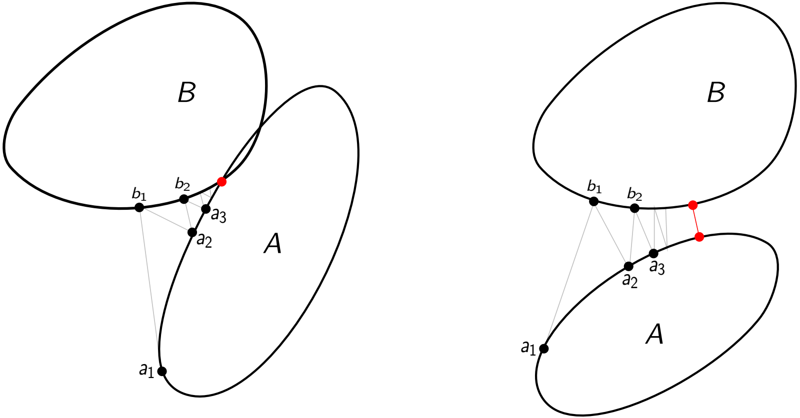
R. Courant, D. Hilbert. Verfahren der mathematischen Physik, Band 2 1930s.

J. von Neumann. Functional Operators II. Lecture Notes 1950

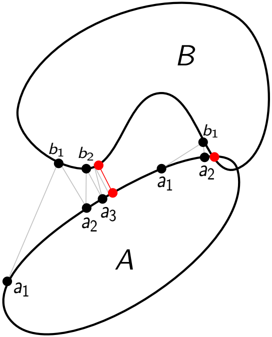
- Presents no citations
- Claims that original version in 1933 has it already

Convex Alternating Projections

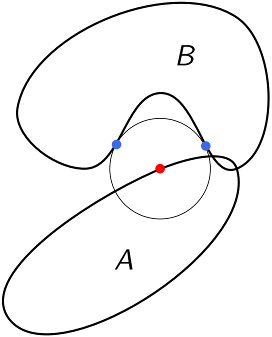
- Schwarz, Sobolev, v. Neumann : Subspaces
- W. Cheney, A.A. Goldstein. Proximity maps for convex sets. 1959.
- L.M. Bregman : Weak convergence for convex sets in Hilbert space. 1965
- H.H. Bauschke : Convex case essentially settled 1993.
- H. Hundal. Norm convergence may fail, 2002.



B
A
b_1
b_2
a_1
a_2
a_3
B
A
b_1
b_2
a_1
a_2
a_3



B
b_1
b_2
b_1
a_1
a_2
a_2
a_3
A
a_1



B
A

Given : closed sets A, B in $\mathbb{R}^n$

$$A \cap B \neq \emptyset$$

Want : solution x of feasibility problem

$$x \in A \cap B$$

Method :

$$b_1 \in P_B(a_1),\ a_2 \in P_A(b_1),\ b_2 \in P_B(a_2),\ a_3 \in P_A(b_2),\ \ldots$$

or

$$a_1 \xrightarrow{P_B} b_1 \xrightarrow{P_A} a_2 \xrightarrow{P_B} b_2 \xrightarrow{P_A} \ldots$$

Given : closed sets A, B in $\mathbb{R}^n$

$$A \cap B = \emptyset$$

Want : generalized solution

$$a^* \in P_A(b^*), b^* \in P_B(a^*), \|a^* - b^*\| = \mathrm{dist}(A, B)$$

Expect : convergence to gap

$$a_k \to a^*, b_k \to b^*$$

Non-convex Alternating Projections :

- Are there applications ?
- Conditions for local convergence ?
- Conditions for global convergence ?
- May convergence fail ?

Are there applications?

Non-convex Alternating Projections used in :

- Color plane interpolation (de-mosaicking)
- De-noising of time-series (Cadzow's basic algorithm, Singular Spectrum Analysis)
- Inverse eigenvalue problems
- Pole placement (control)
- Synthesis of low-order feedback controllers (control)
- Road profile design (western Canada)
- Recovery of lost image blocks in JPEG and MPEG images
- Sparse affine feasibility (for error correction in linear codes)
- Packings in Grassmannian manifolds (wireless communication)

- Brain signal localization
- sudokus
- Nearest correlation matrix
- Robust matrix completion
- Bi-proportional matrix scaling (matrix ranking - statistics)
- EM-algorithm for Gaussian laws.
- Phase retrieval
  - optics
  - X-ray crystrallography
  - astronomy
  - speech proccessing
  - computational biology
  - blind deconvolution

Failure of convergence

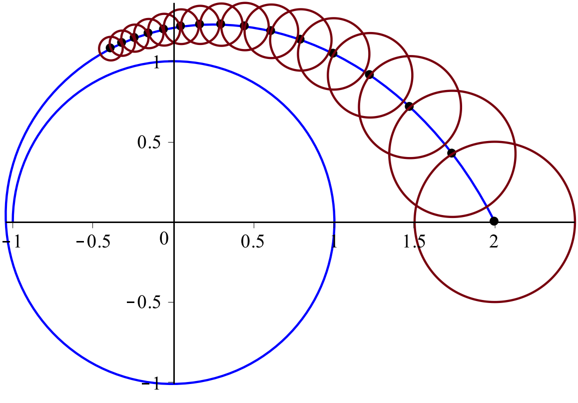
Proposition

Let A, B be closed. Suppose the sequence of alternating projections a_k, b_k is bounded and satisfies $a_k - b_k \to 0$, $a_k - a_{k-1} \to 0$. Then the set of accumulation points of a_k, b_k is either singleton or a compact continuum.

Proof. By A. Ostrowski's Theorem. □

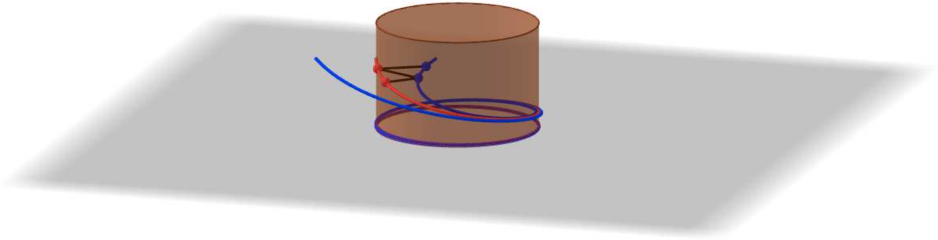
Example

(Bauschke, Noll 2013). The case of a non-trivial compact continuum may indeed occur.



$$A = \{1^{\text{st}}, 3^{\text{rd}}, 5^{\text{th}}, 7^{\text{th}}, \ldots\} \cup C, \quad B = \{2^{\text{nd}}, 4^{\text{th}}, 6^{\text{th}}, \ldots\} \cup C.$$

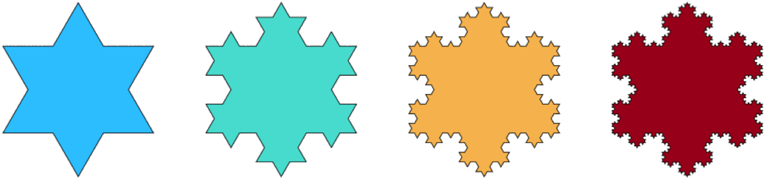
$$A \cap B = C = \{z : |z| = 1\}$$



$$B = \{(\cos t, \sin t, s) : 0 \leq s \leq 1, 0 \leq t \leq 2\pi\}$$
$$A = \{(\cos t(1+e^{-t}), \sin t(1+e^{-t}), e^{-2t}) : 0 \leq t \leq \infty\}$$

Bauschke, Noll (2014, Archiv der Mathematik)

Bauschke, Noll (2014 unpublished)



Koch Snowflake

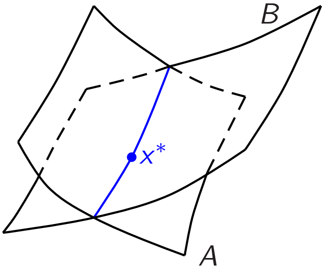
Proposition

(Bauschke, Noll 2014 unpublished). *Let $F \subset \mathbb{R}^n$ be compact connected and locally connected and contained in a $(n-1)$-dimensional C^1 manifold ($\sim$ a Peano set). Then there exist $A, B \subset \mathbb{R}^n$ closed bounded with $A \cap B = F$ such that every alternating sequence a_k, b_k has all points of F as accumulation points.*

Convergence with transversality

Theorem

(A.S. Lewis, J. Malick 2008). *Let A, B be C^2-manifolds in $\mathbb{R}^n$ intersecting transversally at $x^* \in A \cap B$. Then there exists a neighborhood U of x^* such that every alternating sequence a_k, b_k which enters U converges to some $a^* \in A \cap B$ with R-linear speed.*



Transversality

$$T_A(x^*) + T_B(x^*) = \mathbb{R}^n$$

Theorem

(A.S. Lewis, R. Luke, J. Malick 2009). *Suppose*

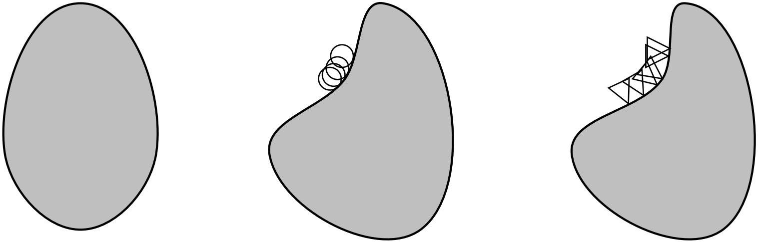
1. *There exists* $x^* \in A \cap B$ *such that* $N_A(x^*) \cap -N_B(x^*) = \{0\}$ *(replaces transversality).*
2. B *is super-regular (replaces convexity).*

Then there exists a neighborhood U *of* x^* *such that every alternating sequence* a_k, b_k *which enters* U *converges to some* $a^* \in A \cap B$ *with R-linear speed.*

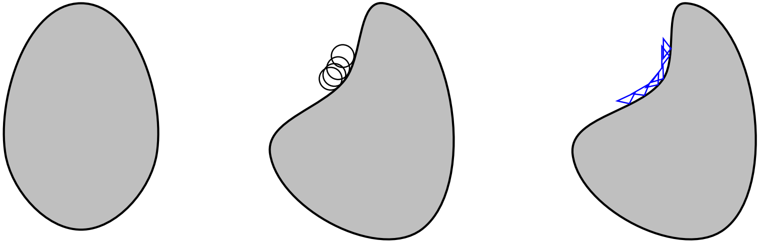
A.S. Lewis, J. Malick (2008). Alternating projections on manifolds. *Math. Oper. Res.* 33 :2008, 216–234.

A.S. Lewis, R. Luke, J. Malick (2009). Local linear convergence for alternating and averaged non-convex projections. *Foundations Comp. Math.* 9 :2009, 485–513.

- Transversality too restrictive. Two non-parallel lines in $\mathbb{R}^2$ intersect transversally, but no longer in $\mathbb{R}^3$
- Same for $N_A(x^*) \cap -N_B(x^*) \subset \{0\}$.
- Need an additional regularity hypothesis called super-regularity.



convex prox-regular super-regular



convex prox-regular super-regular

H.H. Bauschke, D.R. Luke, H.M. Phan, X. Wang (2013). Restricted normal cones and the method of alternating projections.
Set-Valued Var. Anal. 21 :2013, 431 – 473.

Use restricted normal cones instead :

$$N_A^B(x^*) = \text{ normals to } A \text{ at } x^* \text{ pointing into } B$$

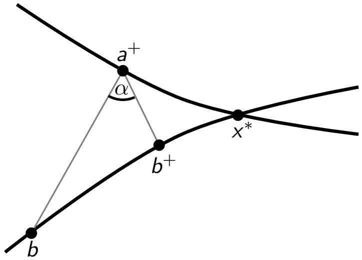
Transversality at x^* becomes :

$$N_A^B(x^*) \cap -N_B^A(x^*) \subset \{0\}$$

Works better, but still not good enough.

Definition

(Noll, Rondepierre 2013). Transversality means α stays away from 0° in neighborhood of x^*.

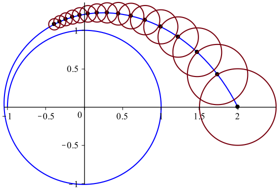


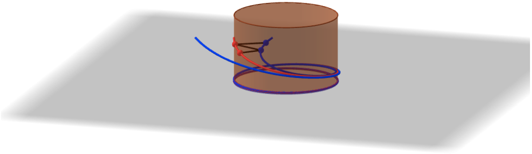
Could say : A, B intersect *at an angle* at x^*

What happens when the intersection is tangentiel ?

- Is failure of convergence due to the lack of regularity ?
- or is it because the intersection is (too) tangential ?

Remember convex case :
Linear convergence when intersection *at an angle*.
Slow convergence when intersection *tangential*.





$\alpha \approx 180^\circ$ (transversal)

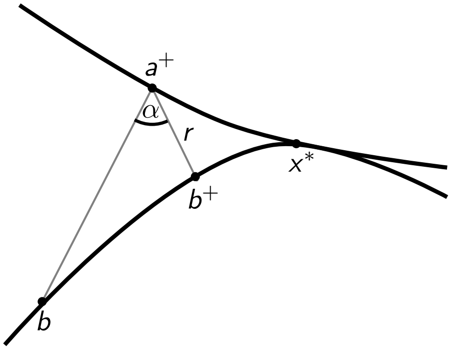
Regularity missing

intersection too tangential

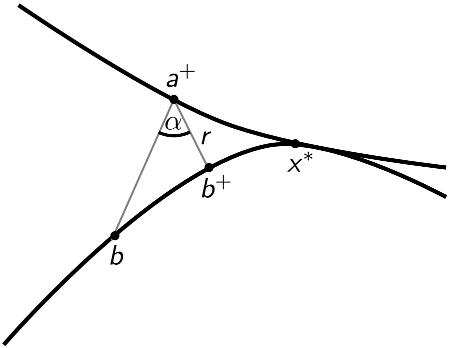
Regularity OK

How to deal with tangential intersection ?

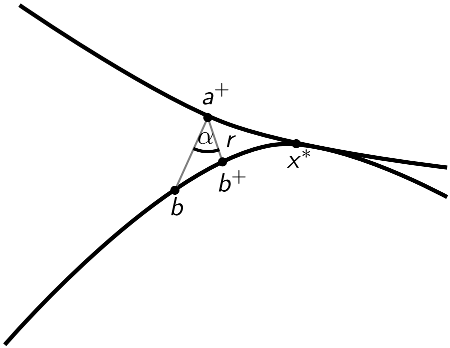
Noll, Rondepierre 2013 :



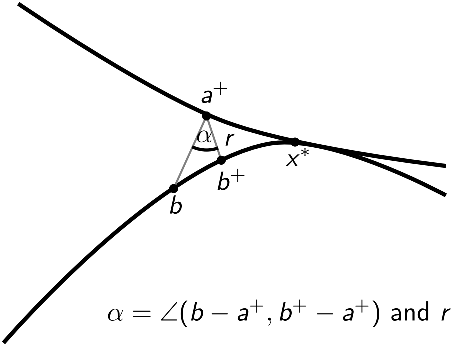
Tangential intersection :



Tangential intersection :



Tangential intersection :



$\alpha = \angle(b - a^+, b^+ - a^+)$ and $r = \|b^+ - a^+\|$

both shrink to 0 as we approach x^*

Definition

(Noll, Rondepierre 2013). The sets A, B satisfy the angle condition at $x^* \in A \cap B$ if there exists $\gamma > 0$, $\omega \in [0, 2)$ and a neighborhood U of x^* such that for every building block $b \xrightarrow{P_A} a^+ \xrightarrow{P_B} b^+$ in U and $r = \|b^+ - a^+\|$ we have

$$\frac{\sin^2 \alpha}{r^\omega} \geq \gamma$$

- Tangential intersection means α and r both shrink to 0.
- Angle condition means they shrink in controlled fashion. Angle does not shrink too fast.
- Special case $\omega = 0$ gives back transversality (angle does not shrink, but distance r does).

Definition

(Noll 2020). A gap (A^*, B^*, r^*) satisfies the angle condition with constant $\gamma > 0$ and exponent $\omega \in [0, 2)$ if there exist neighborhoods U of B^* and V of A^* such that, for every building block $b \to a^+ \to b^+$ with $r = \|a^+ - b^+\| > r^*$ and $a^+ \in V$, $b^+ \in U$, the estimate

$$\frac{1 - \cos\alpha}{(r - r^*)^\omega} \geq \gamma$$

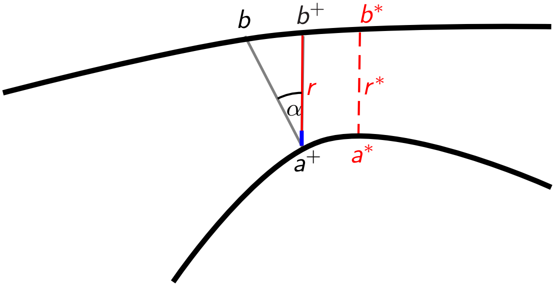
holds for the angle $\alpha = \angle(b - a^+, b^+ - a^+)$.

What is a gap between A, B?

Definition

We call (A^*, B^*, r^*) a gap if for every $a^* \in A^*$ there exists $b^* \in B^* \cap P_B(a^*)$ with $\|a^* - b^*\| = r^*$, and vice versa for every $b^* \in B^*$ there exists $a^* \in A^* \cap P_A(b^*)$ with $\|a^* - b^*\| = r^*$.

Angle condition extends to gaps :



$$r = \|b^+ - a^+\|$$

$$\frac{1-\cos\alpha}{(r-r^*)^{\omega}} \geq \gamma \qquad \text{or} \qquad \frac{\sin\alpha}{(r-r^*)^{\omega/2}} \geq \gamma'$$

Theorem

(Noll, Rondepierre 2013). *Suppose there exists $x^* \in A \cap B$ such that*

1. *A, B satisfy the ω-angle condition at x^*.*
2. *B is $\omega/2$-Hölder regular at x^* with respect to A.*

Then there exists a neighborhood U of x^* such that every alternating sequence a_k, b_k which enters U converges to some point $a^* \in A \cap B$. The speed of convergence is

$$\|a_k - a^*\| = \mathcal{O}\left(k^{-\frac{2-\omega}{2\omega}}\right), \qquad \|b_k - a^*\| = \mathcal{O}\left(k^{-\frac{2-\omega}{2\omega}}\right)$$

Special case $\omega = 0$ gives R-linear convergence

Theorem

(Noll, Rondepierre 2013). *Suppose A, B are sub-analytic sets and $x^* \in A \cap B$. Then there exists $\omega \in [0, 2)$ such that A, B intersect with ω-angle condition at x^*.*

Semi-analytic set :

$$A = \bigcup_{i=1}^{N} \bigcap_{j=1}^{M} \{x \in \mathbb{R}^n : \phi_{ij}(x) = 0, \psi_{ij}(x) > 0\}$$

with real-analytic functions ϕ_{ij}, ψ_{ij}.

$$A \text{ sub-analytic } \iff \forall a \in A \ \exists r > 0 \ \exists \mathcal{A} \text{ bounded semi-analytic} \\ A \cap B(a, r) = \{x : (x, y) \in \mathcal{A}\}$$

Proposition

(Noll, Rondepierre 2013, Noll 2020). *Suppose $f = i_A + \frac{1}{2}d_B^2$ satisfies the Łojasiewicz inequality with exponent $\theta \in [\frac{1}{2}, 1)$. Then A, B satisfy the ω-angle condition with $\omega = 4\theta - 2$.*

- $\omega \in [0, 2)$
- $\omega = 0$ transversality
- Gets weaker as ω increases.
- For $\omega \geq 2$ too weak.
- Works also with gaps.

Corollary

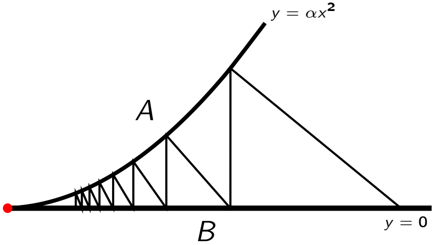
(Noll, Rondepierre 2013). *Rates for* $r^* = 0$. *Suppose* $f = i_A + \frac{1}{2}d_B^2$ *has Łojasiawicz exponent* θ *and* A, B *are* $(2\theta - 1)$*-Hölder regular.*

- *For* $\theta \in (\frac{1}{2}, 1)$ *convergence rate is* $\|b_k - b^*\| = O(k^{-\frac{1-\theta}{2\theta-1}})$.
- *For* $\theta = \frac{1}{2}$ *speed is R-linear.*

Corollary

(Noll 2020). *Rates for* $r^* > 0$. *Suppose* $f = i_A + \frac{1}{2}d_B^2$ *has Łojasiewicz exponent* θ *and the gap is* $(2\theta - 1)$*-Hölder regular.*

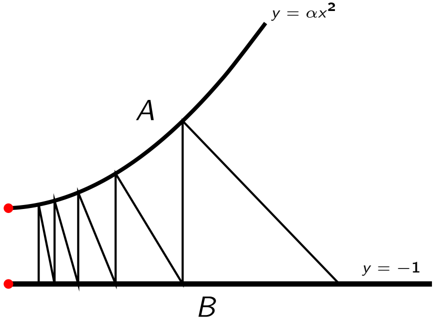
- *For* $\theta \in (\frac{3}{4}, 1)$ *convergence rate is* $\|b_k - b^*\| = O(k^{-\frac{1-\theta}{2\theta-3/2}})$.
- *For* $\theta = \frac{3}{4}$ *convergence is R-linear.*
- *For* $\theta \in (\frac{1}{2}, \frac{3}{4})$ *convergence is R-linear with rate* $\frac{1}{2} + \epsilon$ *for any* $\epsilon > 0$.
- *For* $\theta = \frac{1}{2}$ *convergence is finite.*



$$x_+ = P_B(P_A(x))$$

$$x_+ + 2\alpha x_+^3 = x$$

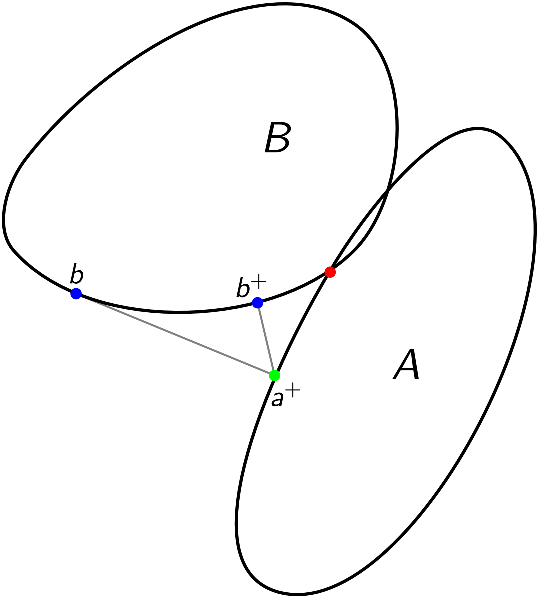
$$\frac{x_+}{x} \to 1$$



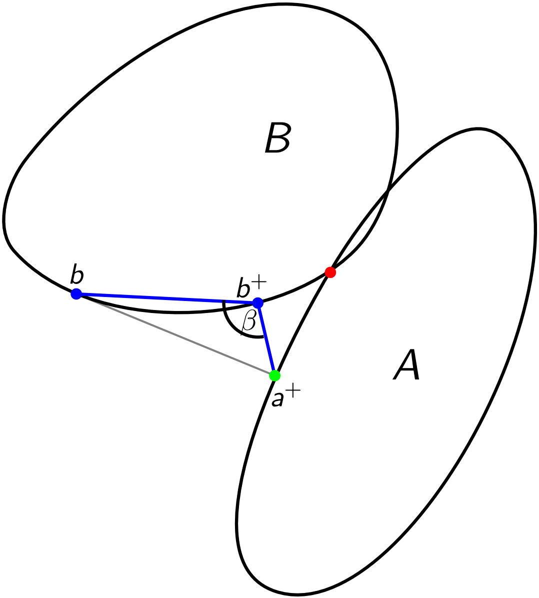
$$x_+ + 2\alpha x_+(1 + \alpha x_+^2) = x$$

$$\frac{x_+}{x} \to \frac{1}{1+2\alpha}$$

How about Hölder regularity ?

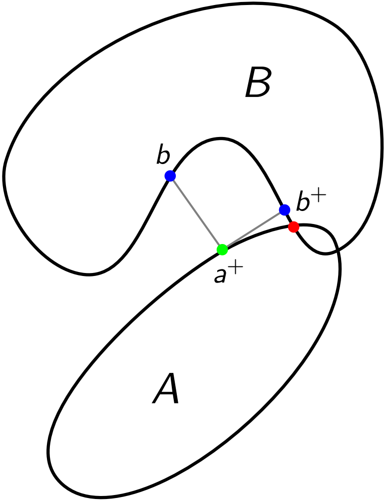


B convex

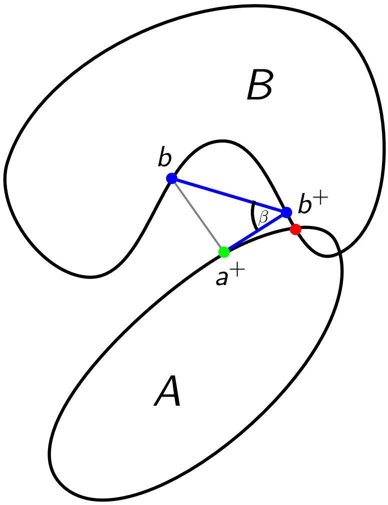


B convex

$\beta \geq 90^\circ$

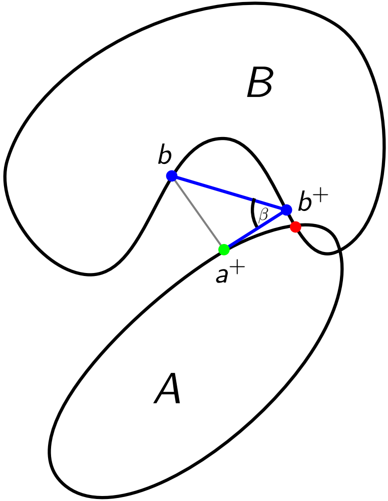


B non-convex



B non-convex

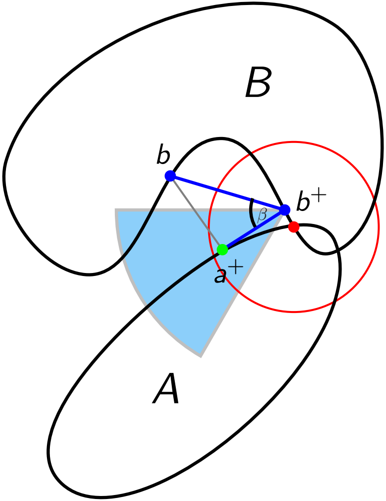
$\beta < 90°$ possible



B non-convex

B super-regular :

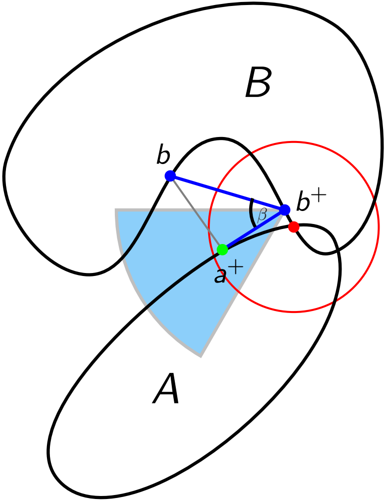
β not too small



B non-convex

B superregular :

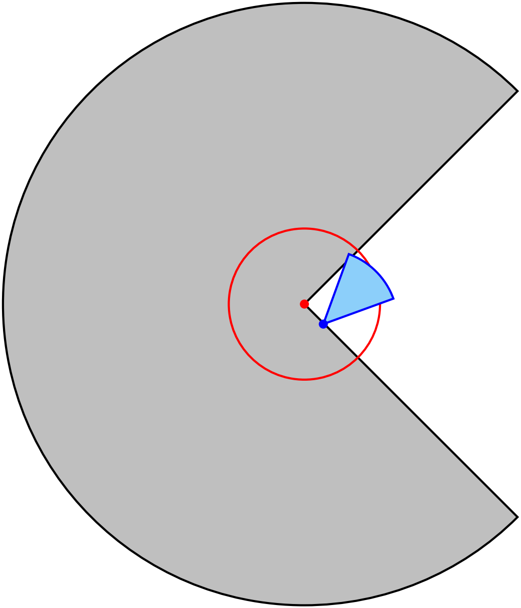
β not too small

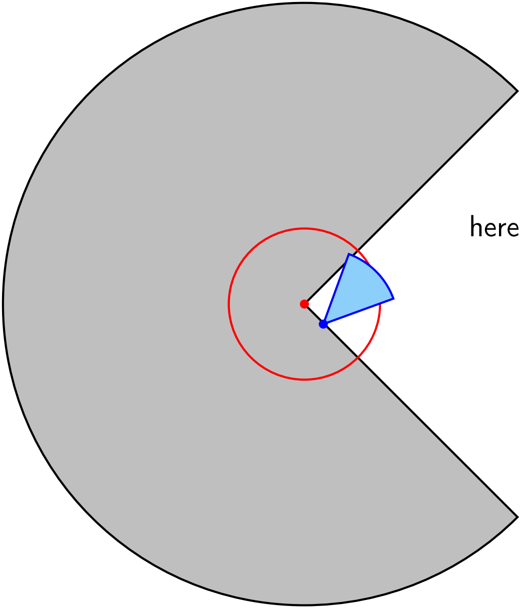


B non-convex

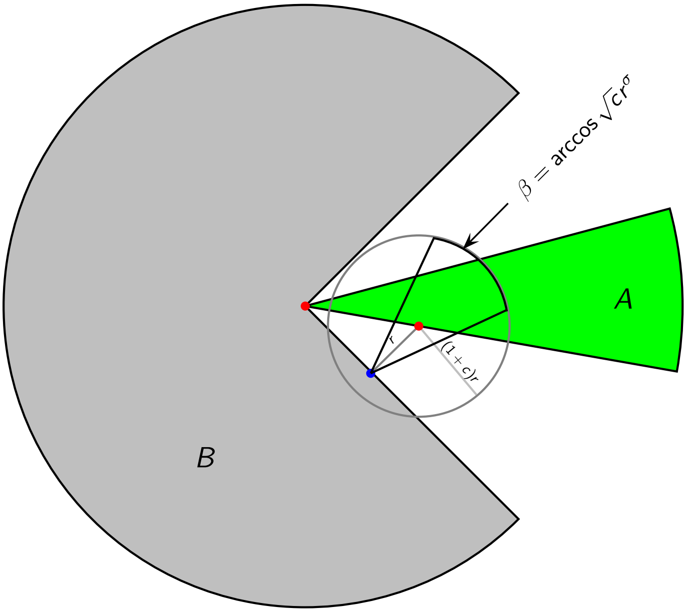
B superregular :

β not too small

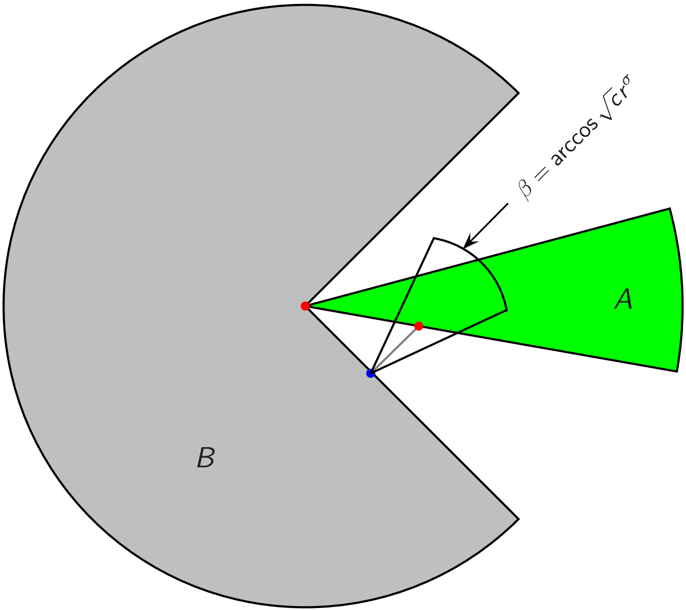


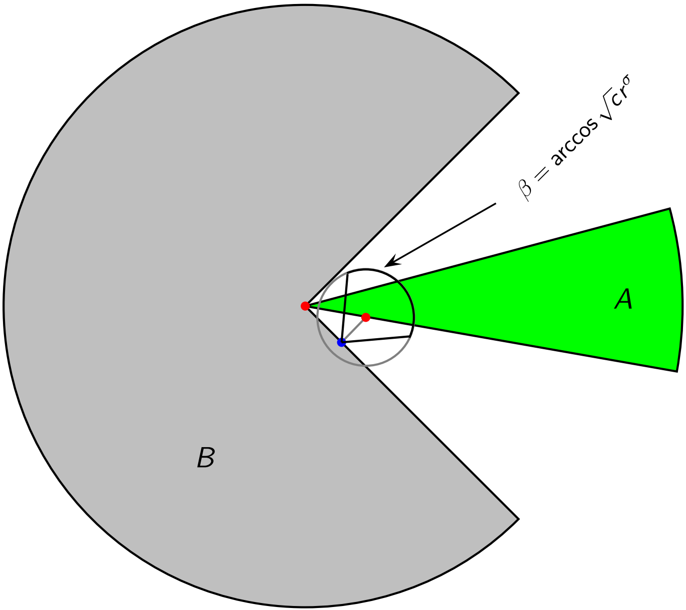


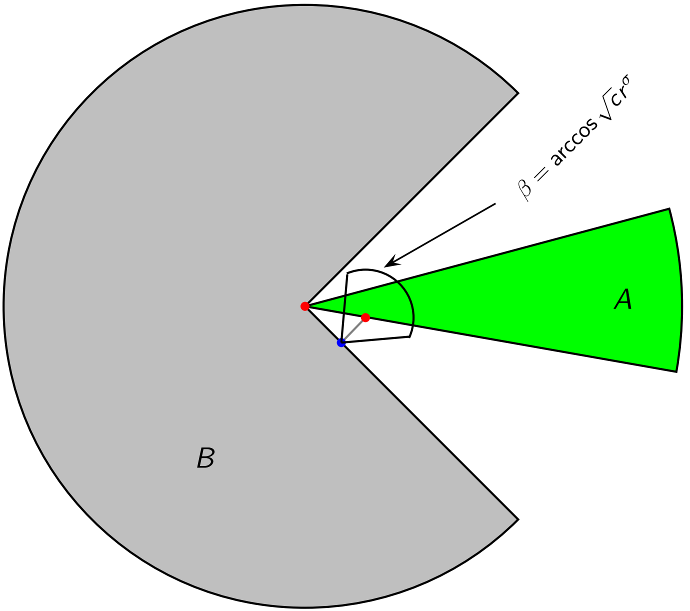
here super-regularity fails



$\beta = \arccos \sqrt{cr^{\sigma}}$
A
B
r
$(1+c)r$


$\beta = \arccos \sqrt{cr^\sigma}$
A
B


$\beta = \arccos\sqrt{cr^{\sigma}}$
A
B


$\beta = \arccos \sqrt{cr^{\sigma}}$
A
B

Consequence : Our notion of Hölder regularity still in business for packman. Can enter into corners.

- **Super-regular :** Aperture increases as distance decreases.

$$\cos\beta = f(\delta) \to 0 \text{ as } \delta \to 0, \delta = \mathrm{dist}(b^+, b^*)$$

- σ-**Hölder regular** $(\sigma \in [0,1))$: Aperture increase quantified.

$$\cos\beta = \sqrt{c} \cdot r^\sigma, r = \|a^+ - b^+\|$$

  What makes the difference :

  - relate to r not δ ;
  - require only for $b \in B(a^+, (1+c)r)$

- For $\sigma = 0$ weaker than existing notions of regularity.

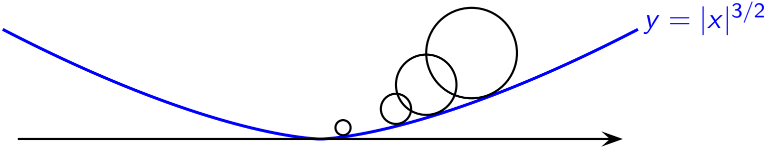
For gaps :

- $\cos\beta = \sqrt{c}(r - r^*)^\sigma$;
- But still for $b \in B(a^+, (1+c)r)$.

Proposition

Slowly shrinking reach implies Hölder regularity. In particular, for $r^* = 0$ *prox-regularity implies* σ*-Hölder regularity for all* $\sigma \in [0,1)$ *and arbitrarily small* $c > 0$*. For* $r^* > 0$ *need reach* $\geq r^*$*.*

Vanishing reach :



Curvature radius $R_x \sim \frac{4}{3}|x|^{1/2}$

Corollary

(Noll, Rondepierre 2013). *Suppose A, B are closed sub-analytic, and B is Hölder regular with respect to A. Suppose the alternating sequence a_k, b_k is bounded and satisfies $a_k - b_k \to 0$. Then there exists $\omega \in [0, 2)$ such that it converges to a point $a^* \in A \cap B$ with speed*

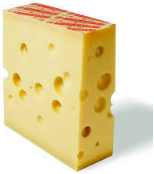
$$\|a_k - a^*\| = \mathcal{O}\left(k^{-\frac{2-\omega}{2\omega}}\right), \qquad \|b_k - a^*\| = \mathcal{O}\left(k^{-\frac{2-\omega}{2\omega}}\right)$$

Corollary

(Noll 2020). *Suppose A, B are closed sub-analytic, and B is prox-regular. Let a_k, b_k be a bounded alternating sequence with gap $r^* < \mathrm{reach}(B^*)$. Then $\|b_k - b^*\| = O(k^{-\rho})$ for some $\rho > 0$.*

Open problems :

- Prove convergence for sub-analytic A, B without regularity hypothesis.
- Construct A, B bounded prox-regular, $A \cap B \neq \emptyset$, $a_k - a_{k-1} \to 0$, $a_k - b_k \to 0$, such that A^*, B^* are not singleton. (One of the sets must fail to be sub-analytic).
- Find case where $b_k \to b^*$ but a_k fail to converge, possibly with $a_k - a_{k-1} \to 0$. Then A must have a crater.



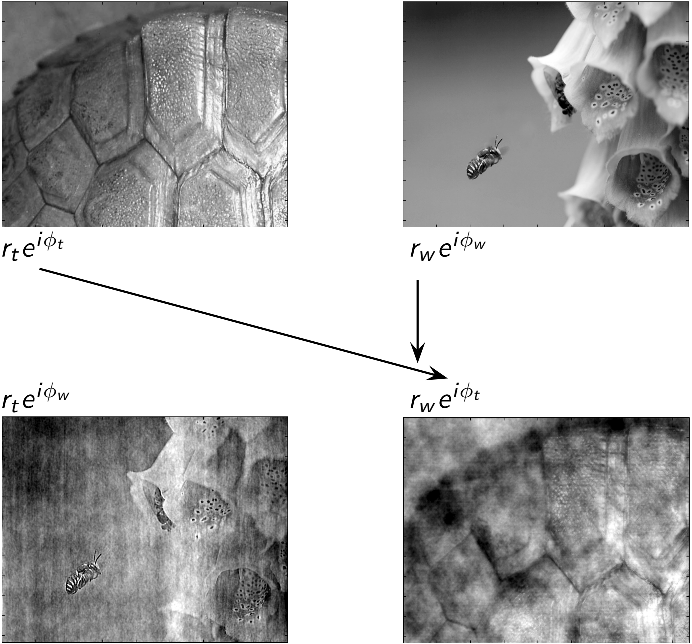
Application : Phase retrieval

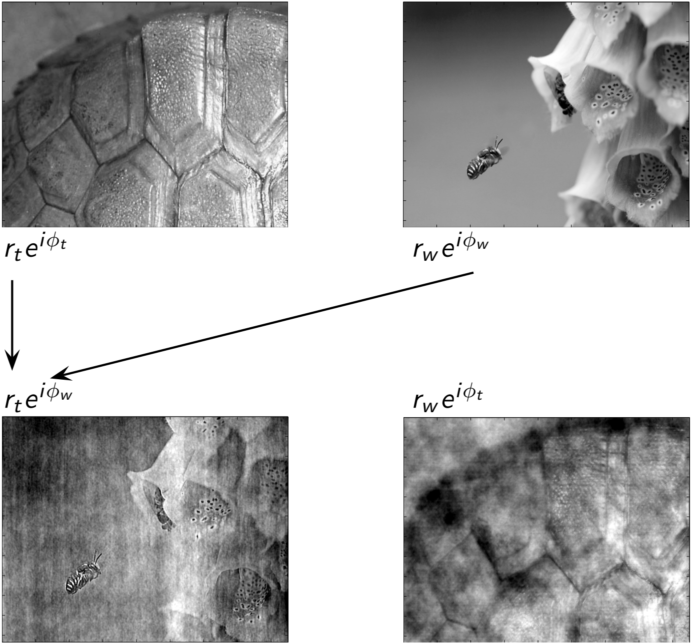
Phase retrieval

Reconstruct unknown signal $x(t)$, $t = 0, \ldots, N-1$ from known Fourier magnitude $m(f) = |\widehat{x}(f)|$, $f = 0, \ldots, N-1$.

- Retrieve unknown phase $\widehat{x}(f)/|\widehat{x}(f)|$, hence the name.
- Underdetermined and ill-posed. Solution set leaves ambiguity :

$$B = \{y \in \mathbb{C}^N : |\widehat{y}(f)| = m(f), f = 0, \ldots, N-1\}$$


$r_t e^{i\phi_t}$
$r_w e^{i\phi_w}$
$r_t e^{i\phi_w}$
$r_w e^{i\phi_t}$


$r_t e^{i\phi_t}$
$r_w e^{i\phi_w}$
$r_t e^{i\phi_w}$
$r_w e^{i\phi_t}$

Consequences :

- Phase of Fourier transform $\widehat{x}/|\widehat{x}|$ gives the essential information about x.
- Magnitude of Fourier transform $|\widehat{x}|$ does not help to localize image x.
- Example : shift in time domain changes phase but not magnitude.
- Hence phase retrieval must be difficult. And it is !

Some history

- Max von Laue (1912) proposes to use X-rays to visualise crystal structure via diffraction.

- David Sayre (1952) argues non-periodic x could in principle also be retrieved from $|\widehat{x}|$ if $m(f) = |\widehat{x}(f)|$ were sampled twice the Nyquist rate in every dimension.

- R.W. Gerchberg - O.W. Saxton (1972). First algorithm.

- J. Miao, P. Charalambous, J. Kirz, D. Sayre, Extending the methodology of X-ray crystallography to imaging micrometre-sized non-crystalline specimens, *Nature* 400, 342-44, 1999.

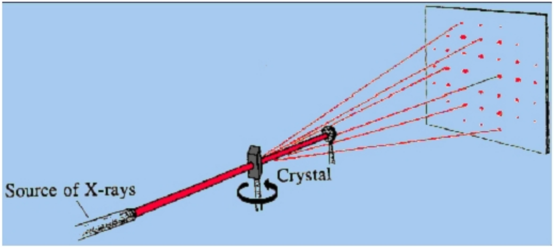
- 2017. Individual proteins and nano-crystals can be visualized by CDI.

  Coherent Diffraction Imaging : $10\text{nm} = 10^{-8}\text{m}$ (organic)

  $2\text{ nm} = 2 \cdot 10^{-9}\text{m}$ (inorganic)

  Fluorescence Microscopy : $1\mu\text{m} = 10^{-6}\text{m}$ (organic)

  Electron microscopy : $0.1\text{ nm} = 10^{-10}\text{m}$ (inorganic)



Max von Laue
(photo 1929)

David Sayre
(photo 1972)

W. O. Saxton
(photo 2012)

Gerchberg-Saxton error reduction (1972)

1. Given current estimate x compute $\widehat{x}$ and *correct* Fourier magnitude $\widehat{y}(f) := m(f)\dfrac{\widehat{x}(f)}{|\widehat{x}(f)|}$.
2. Take inverse Fourier transform y of $\widehat{y}$, and *correct* physical domain constraint by putting $x^+ = P_A(y)$
3. Replace x by x^+ and loop on.

- Add prior information $x \in A$ (= prior or pattern)
- $y \in B$ (= possible phase retrievals)
- Fourier magnitude correction $y = P_B(x)$. Physical domain correction $x = P_A(y)$.

$\Longrightarrow$ No convergence proof since 1972. We gave the first in 2013.

Challenge : not enough samples to retrieve phase $y^* \in B$

- (Unique solution). Ideally pattern A such that $A \cap B = \{y^*\}$, i.e., pattern removes ambiguity.
- (Converge to particular solution). If not possible want at least $x_k \in A$, $y_k \in B$ to converge to $x^* = y^* \in A \cap B$ which has pattern *and* is phase retrieval.
- (Unique generalized solution). In case of noise : Want pattern A to single out gap $(\{x^*\}, \{y^*\}, r^*)$ to determine phase retrieval y^*.
- (Converge to generalized solution). Due to noise : accept $x_k \to x^* \in A$, $y_k \to y^* \in B$ with $P_A(y^*) = x^*$, $P_B(x^*) = y^*$. That is, x^* with that pattern has phase retrieval y^*, and among elements with pattern x^* is closest to y^*.

Typical priors or pattern $x \in A$

- Historically first instance : 2^{nd} Fourier plane

$$A = \{x : |x(t)| = \widetilde{m}(t), t = 0, \ldots, N-1\}$$

- Localization in physical domain (known support)

$$A = \{x : x(t) = 0 \text{ for } t \notin S\}$$

- Sparsity $n \ll N$

$$A = \{x : x(t) \neq 0 \text{ for at most } n \text{ of the } x(t)\}$$

- Sparse phase $n \ll N$

$$A = \{x : \arg(\widehat{x}(f)) \neq 0 \text{ for at most } n \text{ frequencies } f\}$$

Theorem

(Noll, Rondepierre 2013). *Let the phase retrieval problem have an exact solution $x^* \in A \cap B$. Suppose the physical domain constraint is represented by a sub-analytic set A. Then Gerchberg-Saxton error reduction converges in a neighborhood of x^* with speed of convergence $\mathcal{O}\left(k^{-\frac{2-\omega}{2\omega}}\right)$ for some $\omega \in (0, 2)$.*

Proof. Equivalent to non-convex alternating projections between A and :

$$B = \{y \in \mathbb{C}^N : |\widehat{y}(f)| = m(f) \text{ for } f = 0, \ldots, N-1\}$$

B is sub-analytic and prox-regular. $\square$

Definition

The physical domain prior set A allows a guess better than 0 if there exists $x_0 \in A$ with $\text{dist}(x_0, B) < \|m\| = \text{dist}(0, B)$.

Theorem

(Noll, 2020). *Suppose A is closed sub-analytic and allows a guess $x_0 \in A$ better that 0. If Gerchberg-Saxton $x_k \in A$, $y_k \in B$ is started from that guess, then $y_k \to y^* \in B$ with speed $\|y_k - y^*\| = O(k^{-\rho})$ for some $\rho > 0$. That is, there is a unique phase retrieval y^*, and all accumulation points $x^* \in A$ of the x_k admit the same y^* as their phase retrieval.*

Proof. The reach of B is $\|m\|$. Hence started from x_0 iterates stay within reach of B. $\square$

Historically first case. Phase measurements in second Fourier plane (Gerchberg-Saxton 1972). (Ptychography).

$$B = \{y \in \mathbb{C}^N : |\widehat{y}(f)| = m(f)\}, \quad A = \{x \in \mathbb{C}^N : |x(t)| = \widetilde{m}(t)\}$$

$$\|m\| = \|\widetilde{m}\|$$

Theorem

(Noll 2020). *Suppose there exists an initial guess better than* 0. *If started from that guess, Gerchberg-Saxton converges* $x_k \to x^* \in A$, $y_k \to y^* \in B$ *with speed* $\|x_k - x^*\| = O(k^{-\rho})$, $\|y_k - y^*\| = O(k^{-\rho})$ *for some* $\rho > 0$, *and the limits give generalized solution*

$$|x^*| = \widetilde{m}, \quad |\widehat{y}^*| = m, \quad \widehat{y}^* = m \cdot \frac{\widehat{x}^*}{|\widehat{x}^*|}, \quad x^* = \widetilde{m} \cdot \frac{y^*}{|y^*|}$$

Sparsity as prior. Fix $n \ll N$.

$$B = \{y \in \mathbb{C}^N : |\widehat{y}(f)| = m(f), f = 0, \dots, N-1\}$$
$$A = \{x \in \mathbb{C}^N : \text{at most } n \text{ of the } x(t) \text{ are non-zero}\}$$

Projection $x \in P_A(y)$:

$$|y(t_0)| \le |y(t_1)| \le \cdots \le |y(t_{N-1})| \qquad \text{(permutation of } 0, \dots, N-1)$$

Put

$$\underbrace{x(t_0) = 0, \dots, x(t_{N-n-1}) = 0}_{\text{zero the first } N-n}, \underbrace{x(t_{N-n}) = y(t_{N-n}), \dots, x(t_{N-1}) = y(t_{N-1})}_{\text{keep } n \text{ largest}}$$

Possible ambiguity if
$|y(t_{N-n-s})| = \cdots = |y(t_{N-n-1})| = \cdots = |y(t_{N-n-1+r})|$

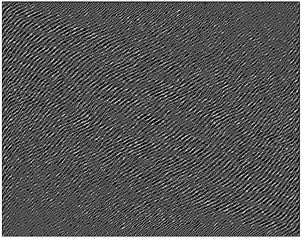
Theorem

(Noll 2020). *Let* x_k, y_k *be the Gerchberg-Saxton sequence for the sparsity prior started from a guess better than 0. Then* $y_k \to y^*$ *with speed* $\|y_k - y^*\| = O(k^{-\rho})$ *for some* $\rho > 0$*. The* x_k *admit a finite set of accumulation points, all having* y^* *as their phase retrieval. If* $x_k - x_{k+1} \to 0$*, then the* x_k *converge to a unique sparse prior with that same speed.*

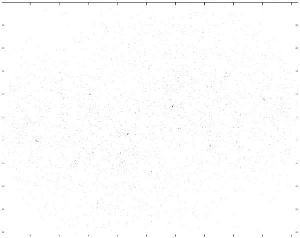
Proof. Sparsity set A is sub-analytic. $\square$

Fourier phase and magnitude

original (unknown)

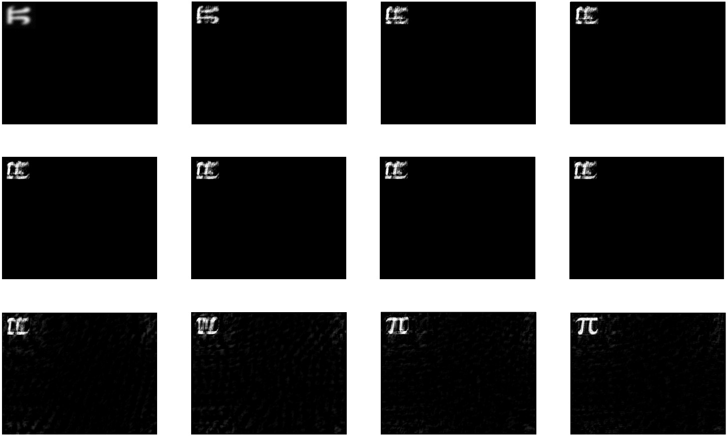


Fourier phase (unknown)



Fourier magnitude (known)

estimated support (prior)



- Ideal image x_0 is PI-image enlarged to size 1024×1024 by 0-padding.
- 0 is black, 256 is white.
- Initial guess is blurred and noisy version of the PI-image which is then rotated 90°.
- Fourier magnitude $m = |\widehat{x_0}|$ is known exactly.
- $B = \{y \in \mathbb{C}^{1024\times 1024} : |\widehat{y}(f)| = m(f)$ for all frequencies $f\}$.
- $A = \{x \in \mathbb{C}^{1024\times 1024} : x(t) = 0$ for all pixels t not in mask$\}$.
- Mask is gray region around the PI-symbol. Prior assumption is that values outside that mask equal 0.
- MAP slow convergence.

Convergence for cactus sets and counterexamples

H.H. Bauschke, D. Noll (2013). On cluster points of alternating projections. *Serdica Math. J.* 39 (2013), 355-364.

H.H. Bauschke, D. Noll (2014). On the local convergence of the Douglas–Rachford algorithm. *Archiv der Mathematik* 102, 589 – 600.

Counterexamples for Gerchberg-Saxton, Fienup variants

D. Noll (2020). Alternating projections with applications to Gerchberg-Saxton error reduction. *SVVA*.

Convergence for non-convex alternating projections

A.S. Lewis, J. Malick (2008). Alternating projections on manifolds. *Math. Oper. Res.*

A.S. Lewis, R. Luke, J. Malick (2009). Local linear convergence for alternating and averaged non-convex projections. *Foundations Comp. Math.*

H.H. Bauschke, D.R. Luke, H.M. Phan, X. Wang (2013). Restricted normal cones and the method of alternating projections. *SVVA*.

D. Noll, A. Rondepierre (2013). On local convergence of the method of alternating projections. *Foundations of Comp. Math.* 2015

Z. Zhu, X. Li (2018). Convergence analysis of alternating projection method for nonconvex sets. `arXiv` 2019.

D. Noll (2020). Alternating projections with applications to Gerchberg-Saxton error reduction. *SVVA* 2021.

Łojasiewicz inequality

J. Bolte, A. Daniilidis, A.S. Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Opt.* 17(4) :2007, 1205-1223.

H. Attouch, J. Bolte, P. Redont, A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems : an approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.*, 35(2) :2010, 438–457.

Structured low-rank matrix approximation

Cadzow's basic denoising algorithm

- Noisy time series $\widetilde{c}_0, \widetilde{c}_1, \ldots, \widetilde{c}_{n-1}$
- Encode as Toeplitz matrix

$$\widetilde{T} = \begin{bmatrix} \widetilde{c}_0 & \widetilde{c}_1 & \ldots & & \widetilde{c}_{n-1} \\ \widetilde{c}_{n-1} & \widetilde{c}_0 & \widetilde{c}_1 & \ldots & \widetilde{c}_{n-2} \\ \ldots & & \ddots & \ddots & \\ \widetilde{c}_2 & & & \ddots & \widetilde{c}_1 \\ \widetilde{c}_1 & \widetilde{c}_2 & \ldots & & \widetilde{c}_0 \end{bmatrix}$$

- Find Toeplitz matrix T with $\mathrm{rank}(T) \leq r$ as close as possible to $\widetilde{T}$. Decoding T gives denoised signal $c_0, \ldots, c_{n-1}$.
- Heuristic : apply MAP to $A = \{T \in \mathbb{C}^{n\times n} : T \text{ Toeplitz}\}$, $B = \{R \in \mathbb{C}^{n\times n} : \mathrm{rank}(R) \leq r\}$, starting at $\widetilde{T}$

Generalized Cadzow method

$A =$ closed sub-analytic set of structured matrices $T \in \mathbb{C}^{n \times m}$

$B = \{R \in \mathbb{C}^{n \times m} : \mathrm{rank}(R) \leq r\}$ (non-convex)

$$\boxed{\text{Find } T \in A \text{ with rank} \leq r}$$

Apply non-convex MAP in Frobenius norm $\| \cdot \|_F$

- How to compute $R \in P_B(T)$?
- $T = U \Sigma V^\top$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_{\min(n,m)})$
- $\Sigma_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots)$; $R = U \Sigma_r V^\top$.
- Finite number of possibilities if σ_r has multiplicity.
- Reach of B at $R \in B$ is σ_r.

Theorem

(Noll 2020). *Let A be closed subanalytic matrix structure, R_k, T_k a bounded Cadzow alternating sequence with gap r^*. Suppose $r^* < \sigma_r^*$ for the accumulation points R^* of the R_k. Then $\|R_k - R^*\| = O(k^{-\rho})$ for some $\rho > 0$. All accumulation points T^* of the T_k have structure A and admit the same R^* as their low-rank approximation.*

Proof. $R \in B \iff$ all $(r+1) \times (r+1)$ minors of R vanish. Therefore set of polynomial equations (semi-algebraic). Known as determinantal variety of dimension $r(n+m-r)$. $\square$

Theorem

(Noll 2020). *Suppose the matrix structure set A is closed subanalytic and prox-regular. Let T_k, R_k be a bounded Cadzow sequence with gap r^*, where $r^* < \mathrm{reach}(A)$. Then the T_k converge to a structured matrix $T^* \in A$ with speed $\|T_k - T^*\| = O(k^{-\rho})$ for some $\rho > 0$. The sequence R_k has a finite set of accumulation points R^*, and each of these R^* is a low-rank approximation of the same T^*.*

- For convex A convergence from arbitrary starting point.
- Many convergence claims in literature - evoking transversality.
- All wrong.
- Transversality cannot be checked - need not be satisfied.
- Boundedness hypothesis needed.

Example

$B = 2 \times 2$ of rank ≤ 1. $A = \left\{ \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} + t \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} : t \in \mathbb{R} \right\}$

$A \cap B = \left\{ \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} \right\}$. Convergence of MAP sublinear.

Application : EM-algorithm

A.P. Dempster, N.M. Laird, D.B. Rubin. Maximum likelihood from incomplete data via the EM-algorithm.
J. Royal Stat. Soc. Series B, vol. 39, no. 1 (1977), 1 – 38.

$\Longrightarrow$ Cited 38230 times since 1977
$\Longrightarrow$ However, convergence proof incorrect.
$\Longrightarrow$ Since then only proofs for specific situations.

Our approach gives the first local convergence proof for Gaussian laws when parameter set is not convex.

H.H. Bauschke, J.M. Borwein, A. Celler, D. Noll. An EM-algorithm for dynamic SPEC tomography. IEEE Transactions on Medical Imaging 18, no. 3, 1999, 252 - 261.

Thanks for your attention !