Local convergence of the method of alternating projections

Dominikus Noll

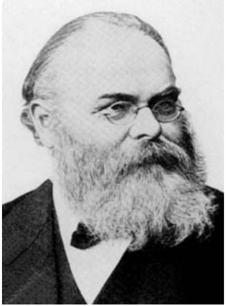
INSTITUT de MATHEMATIQUES de TOULOUSE

Université Paul Sabatier

Alternating projections invented by Hermann Amandus Schwarz in 1869

H. A. Schwarz. Über einen Grenzübergang durch alternirendes Verfahren.

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



- Used to solve Dirichlet problem rigorously
- First domain decomposition method ever
- Modern re-interpretation by P.-L. Lions 1978, 1988-89.

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.

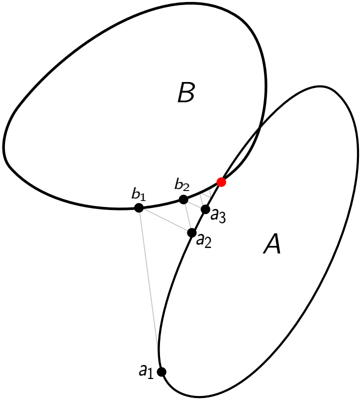
R. Courant, D. Hilbert. Verfahren der mathematischen Physik, Band 2 1930s.

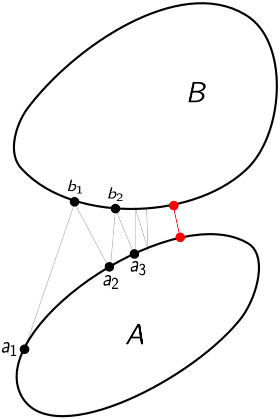
J. von Neumann. Functional Operators II. Lecture Notes 1950

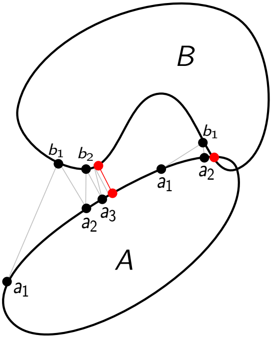
- Presents no citations
- Claims that original version in 1933 has it already

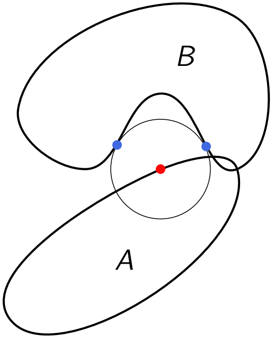
Convex Alternating Projections

- Schwarz, Sobolev, v. Neumann : Subspaces
- 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
b1
b2
a3
a2
A
a1


B
b1
b2
a3
a2
A
a1


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

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

Non-convex Alternating Projections :

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

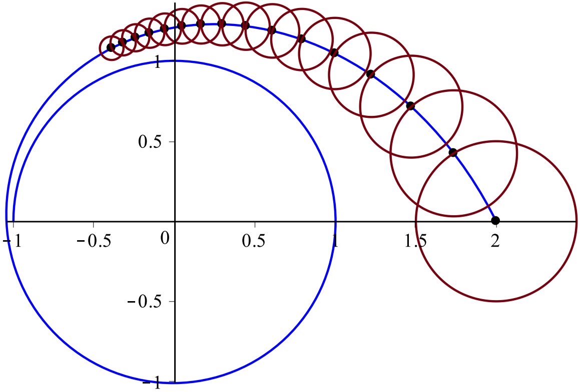
Failure of convergence

Theorem

(Combettes, Trussell 1990). *Let A, B be closed. Suppose the sequence of alternating projections a_k, b_k is bounded and satisfies $a_k - b_k \to 0$. Then the set of accumulation points of a_k, b_k is either singleton or a compact continuum. Every accumulation point is a solution of the feasibility problem.*

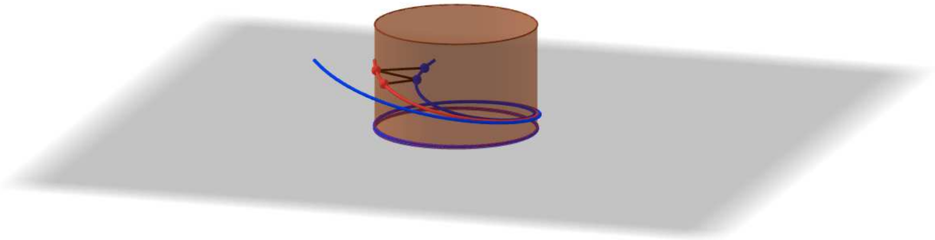
Theorem

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



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

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



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

Bauschke, Noll (2014, Archiv der Mathematik)

Are there applications?

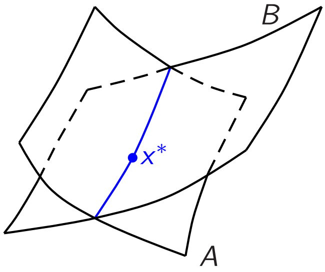
Non-convex Alternating Projections used in :

- 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)
- EM-algorithm for Gaussian laws
- Phase retrieval

Local convergence

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

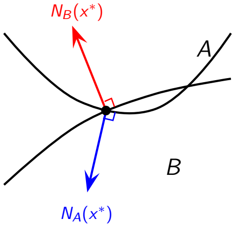
$$T_A(x^*) + T_B(x^*) = \mathbb{R}^n$$

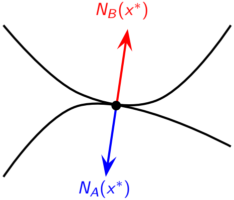
Theorem

(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.

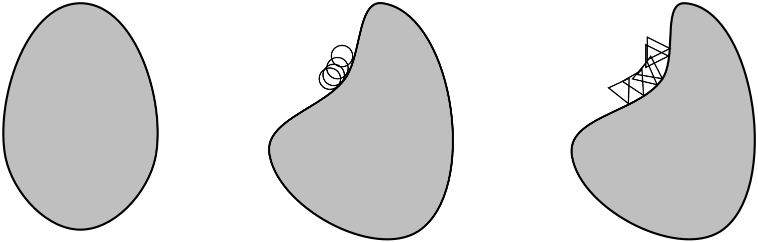




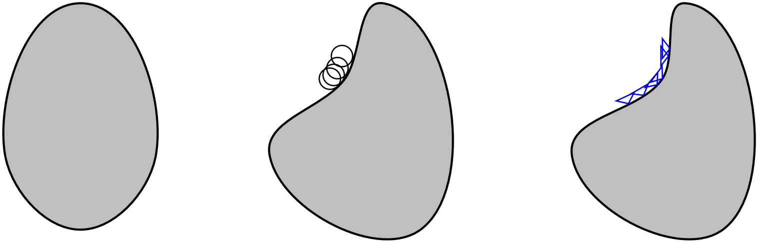
A.S. Lewis, J. Malick (2008). Alternating projections on manifolds.
Math. Oper. Res.

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

- 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.

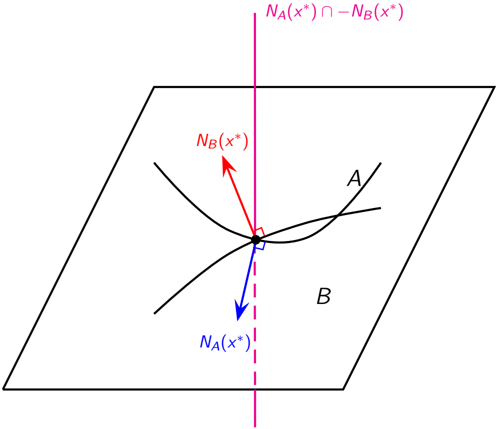


convex prox-regular super-regular



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.



Use restricted normal cones instead :

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

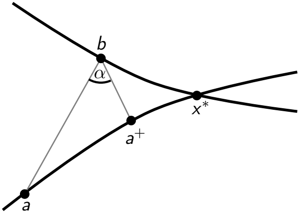
Transversality at x^* becomes :

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

Works better, but still not good enough.

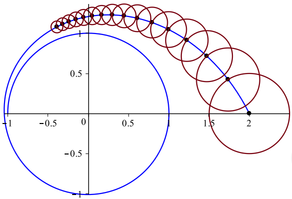
Definition

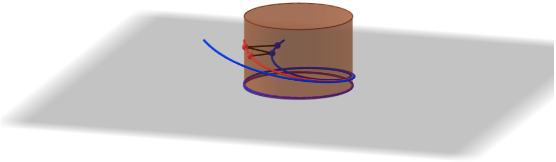
(Noll, Rondepierre 2013). Transversality is when α stays away from 0° in neighborhood of x^*.



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 ?





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

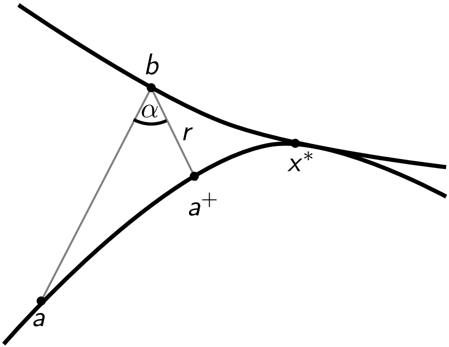
Regularity missing

intersection tangential

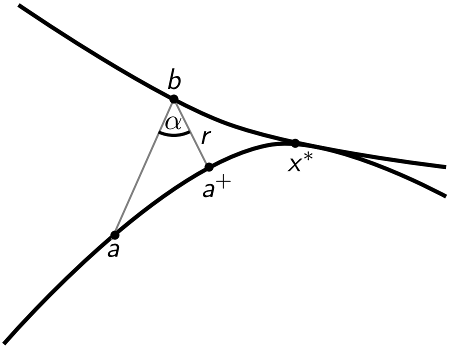
Regularity OK

How to deal with tangential intersection ?

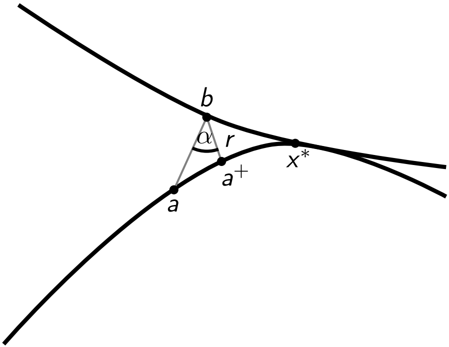
Noll, Rondepierre 2013 :



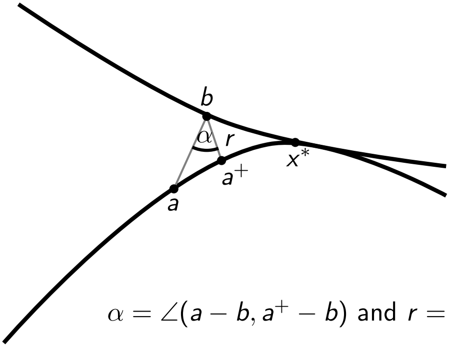
Tangential intersection :



Tangential intersection :



Tangential intersection :



$\alpha = \angle(a - b, a^+ - b)$ 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 $a \xrightarrow{P_B} b \xrightarrow{P_A} a^+$ in U 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).

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^*.*

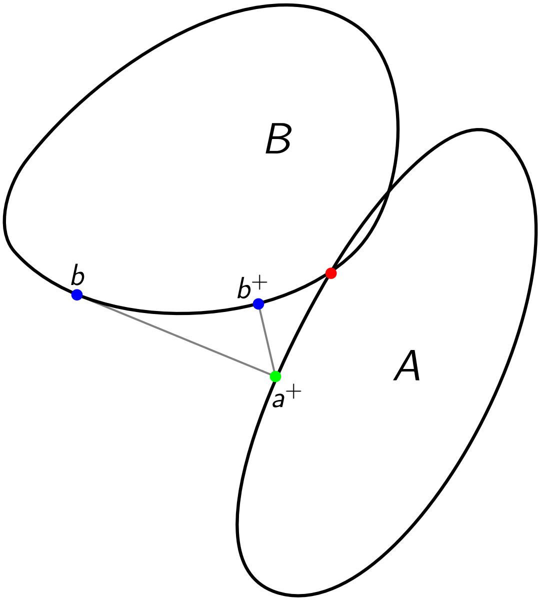
Semi-analytic set :

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

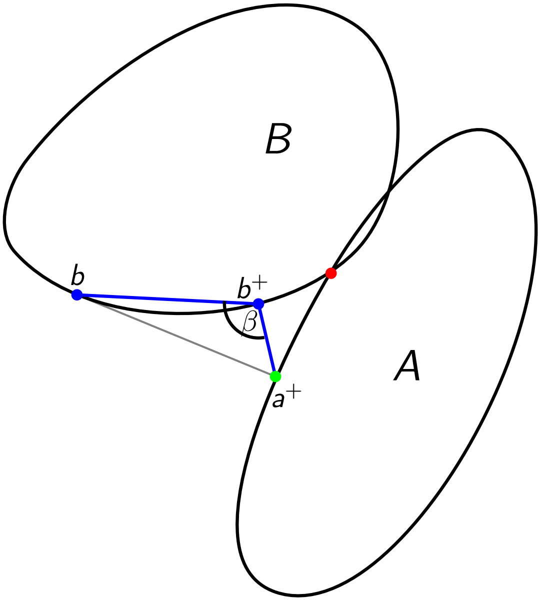
with real-analytic functions ϕ_{ij}, ψ_{ij}.

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

How about Hölder regularity ?

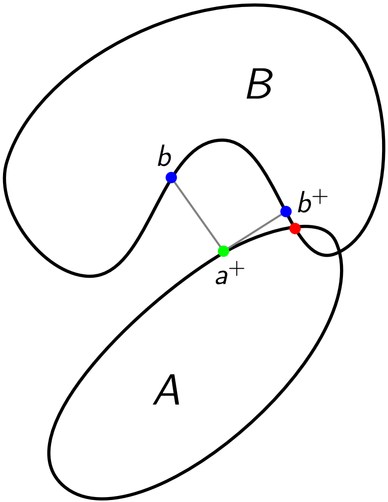


B convex

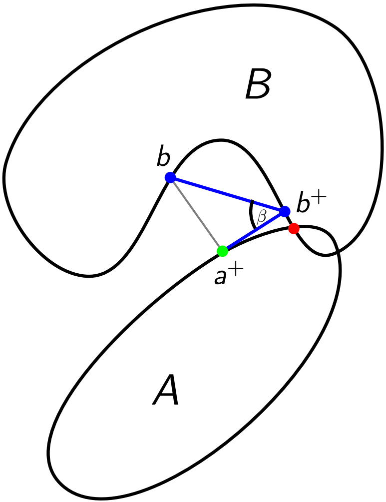


B convex

$\beta \geq 90°$

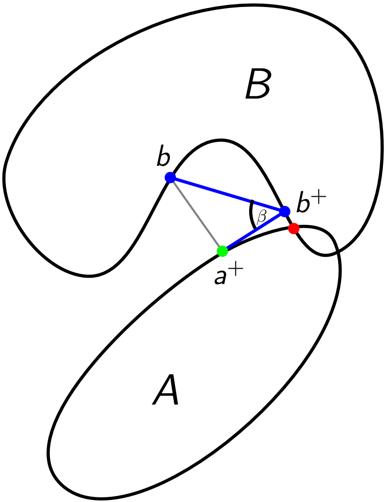


B non-convex



B non-convex

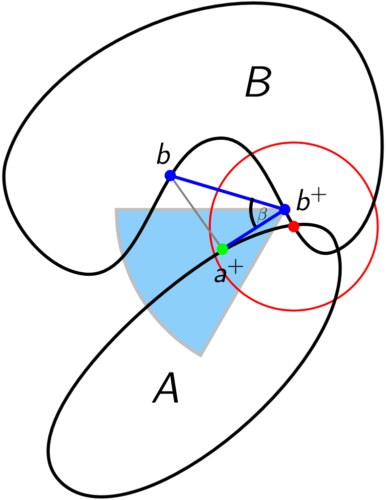
$\beta < 90^\circ$ possible



B non-convex

B super-regular :

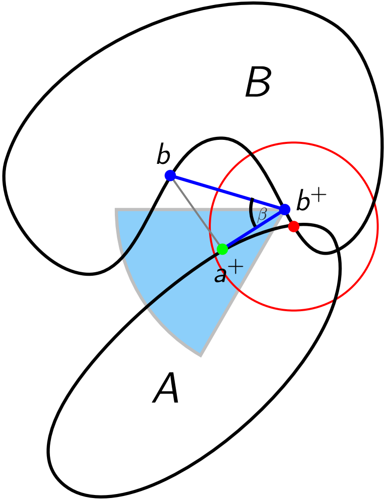
β not too small



B non-convex

B superregular :

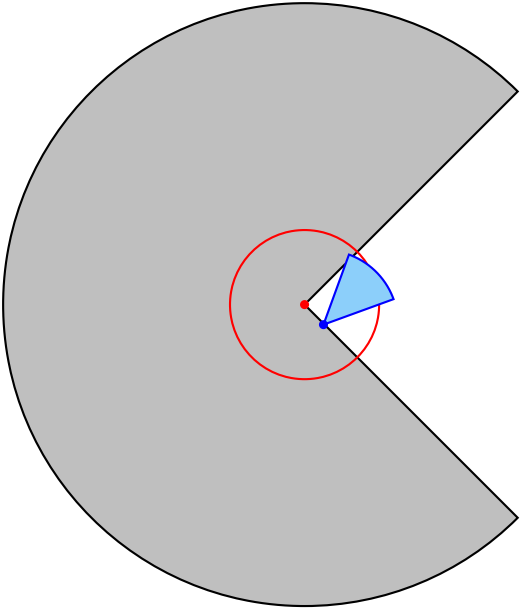
β not too small

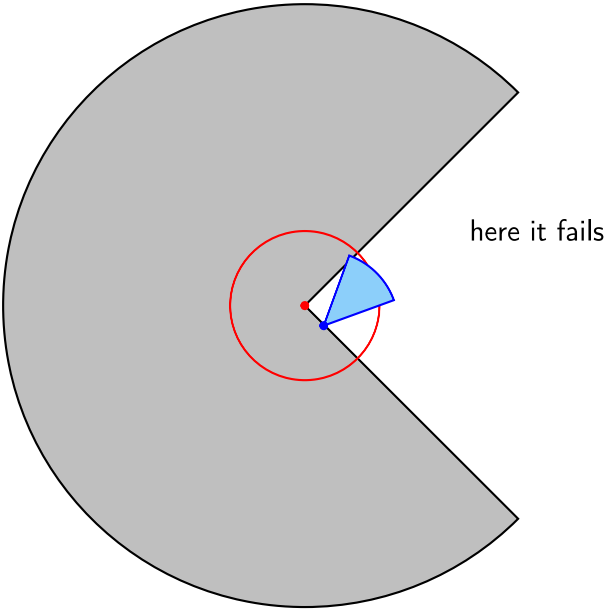


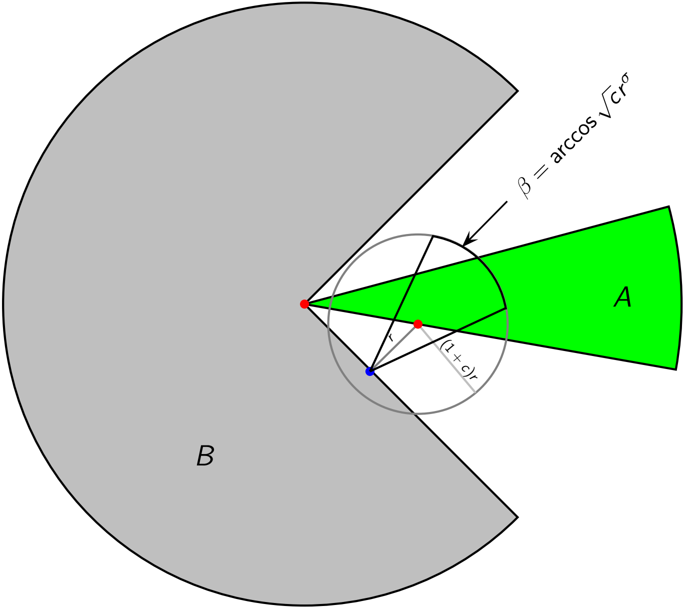
B non-convex

B superregular :

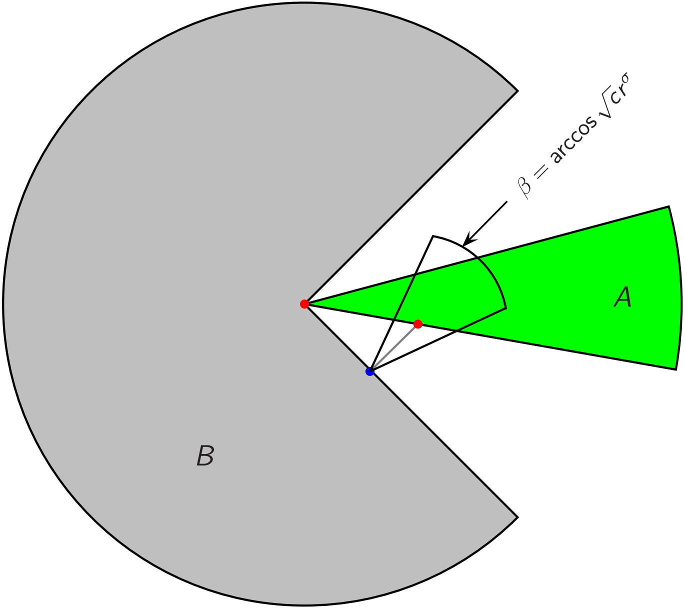
β not too small

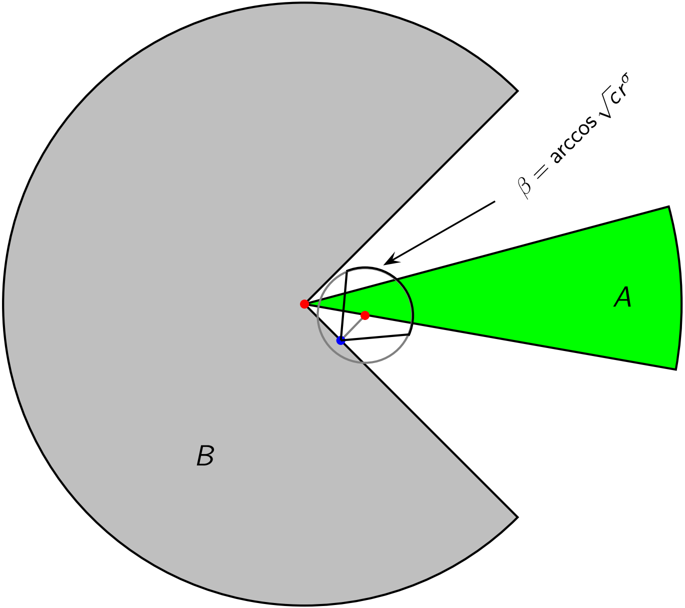



here it fails

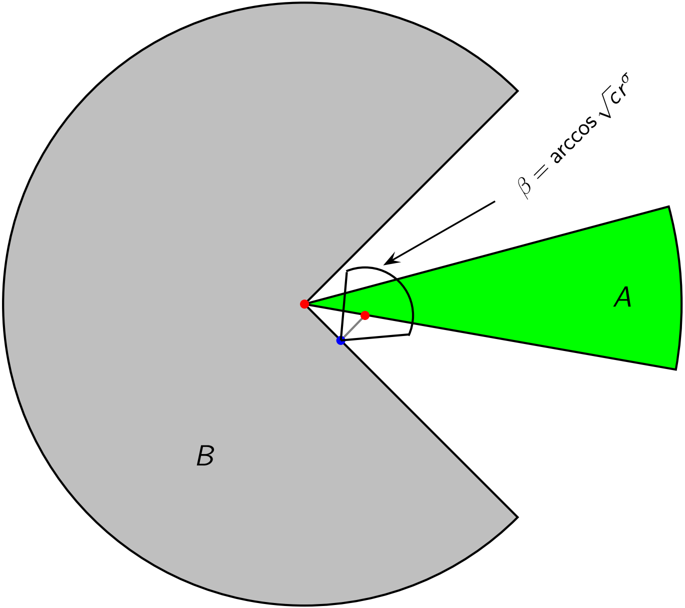


$\beta = \arccos\sqrt{cr^\sigma}$
A
B
r
$(1+c)r$


$\beta = \arccos\sqrt{cr^{\sigma}}$
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.

Corollary

(Noll, Rondepierre 2013). *Suppose* A, B *are 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*

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

Application : Phase retrieval

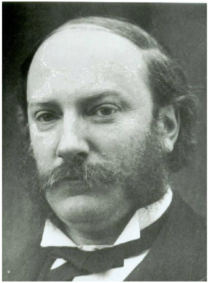
Phase retrieval

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

- Retrieve unknown phase $\widehat{x}(f)/|\widehat{x}(f)|$, hence the name.
- Have to add prior information like known support of x in time domain : $x(t) = 0$ for $t \notin S$.
- Or additional measurements (Fourier amplitude from a second Fourier plane ; Gerchberg-Saxton 1972).

Phase retrieval in interferometry (optics) :

- First mentioned in a letter by Lord Rayleigh to A. Michelson in 1892.
- Impossibility to solve without prior information clearly stated.
- First numerical scheme : Gerchberg-Saxton algorithm 1972

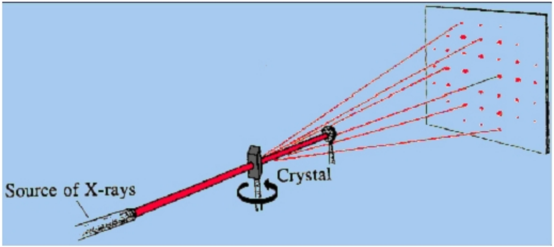


Lord Rayleigh
(early 1900s)

A. Michelson
(1907)

Some history

- Max von Laue (1912) proposes to use X-rays to visualise crystal structure via diffraction.
- David Sayre (1952) shows that non-periodic x can in principle also be retrieved from $|\widehat{x}|$ if $a = |\widehat{x}|$ is sampled twice the Nyquist rate in every dimension.
  - Deplores lack of methods to do it.
  - Hence participates in development of 1st `fortran` compiler.
- R.W. Gerchberg - O.W. Saxton (1972). 1st algorithm to retrieve x from $|\widehat{x}|$.
- J. Miao, P. Charalambous, J. Kirz, D. Sayre, Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens, *Nature* 400, 342-344, 1999. 45 years later Sayre is back !
- 2014. Individual proteins and nano-crystals can be visualized by CDI.



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 amplitude $\widehat{y}(f) := a(f)\dfrac{\widehat{x}(f)}{|\widehat{x}(f)|}$.
2. Take inverse Fourier transform y of $\widehat{y}$, and «correct» domain by putting $x^+(t) = \begin{cases} y(t) & \text{for } t \in S \\ 0 & \text{for } t \notin S \end{cases}$.
3. Replace x by x^+ and loop on.

$\Longrightarrow$ Optic, astronomy, crystallography, nano-materials, ...

$\Longrightarrow$ Cited 2643 times. Fourth most used algorithm ever

$\Longrightarrow$ No convergence proof since 1972.
We give the first.

Theorem

(Noll, Rondepierre 2013). *Suppose the phase retrieval problem has a solution $x^* \in A \cap B$. Suppose the physical domain constraint is represented by a sub-analytic set B. Then the Gerchberg-Saxton error reduction method 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 :

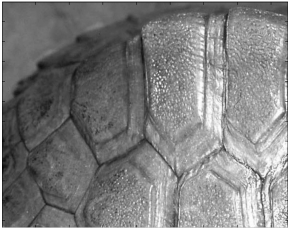
$$A = \{x \in \mathbb{C}^N : |\widehat{x}(f)| = a(f) \text{ for all } f\}$$

$$B = \{y \in \mathbb{C}^N : y(t) = 0 \text{ for all } t \notin S\}$$

$$P_A(x) = (a\widehat{x}/|\widehat{x}|)^{\widetilde{}} \qquad P_B(y) = y \cdot \mathbf{1}_S$$

□

Fourier phase and amplitude

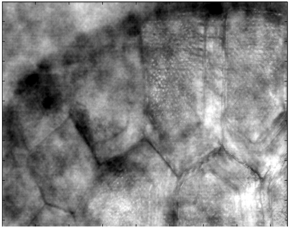


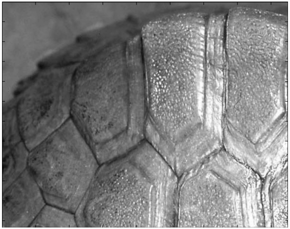
$r_t e^{i\phi_t}$

$r_w e^{i\phi_w}$

$r_t e^{i\phi_w}$

$r_w e^{\phi_t}$



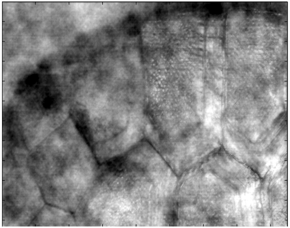


$r_t e^{i\phi_t}$

$r_w e^{i\phi_w}$

$r_t e^{i\phi_w}$

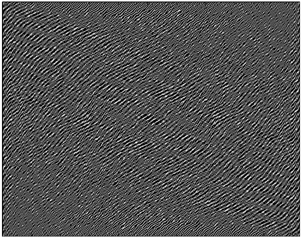
$r_w e^{\phi_t}$



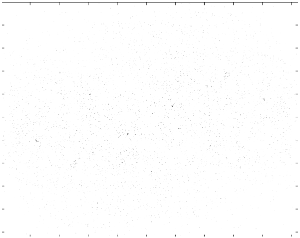
Consequences :

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

original (unknown)

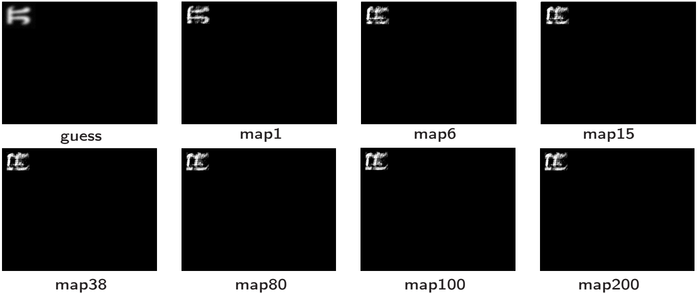


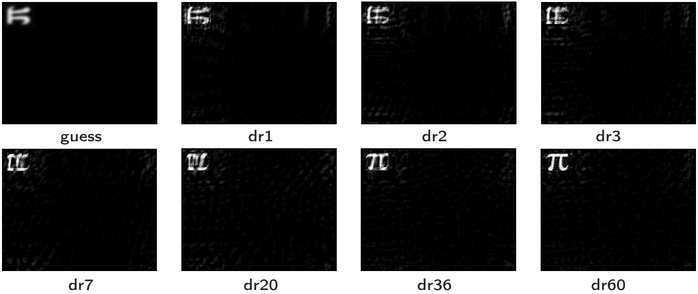
Fourier phase (unknown)



Fourier amplitude (known)

estimated support (prior)


guess
map1
map6
map15
map38
map80
map100
map200

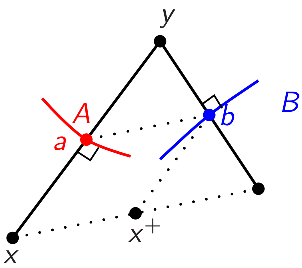

guess
dr1
dr2
dr3
dr7
dr20
dr36
dr60

- 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 amplitude $a = |\widehat{x_0}|$ is known exactly.
- $A = \{x \in \mathbb{C}^{1024 \times 1024} : |\widehat{x}(f)| = a(f) \text{ for all frequencies } f\}$.
- $B = \{y \in \mathbb{C}^{1024 \times 1024} : y(t) = 0 \text{ for all pixels } t \text{ not in mask}\}$.
- Mask is gray region around the PI-symbol. Prior assumption is that values outside that mask equal 0.
- MAP does not fully succeed within reasonable time.
- Douglas-Rachford recovers phase quite nicely.

J. Douglas, H.H. Rachford. On the numerical solution of heat conduction problems in two and three dimensions. *TAMS* 82 (1956), 421 – 439.

P.-L. Lions, B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J.Num. Anal.* 16 (1979), 946–979.

J.R. Fienup. Phase retrieval algorithms : a comparison. *Applied Optics*, 1982. HIO = Hybrid-Input-Output



$a \in P_A(x)$
$y = 2a - x \in R_A(x)$
$b \in P_B(y)$
$x^+ = x + b - a$

reflect-reflect-average

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. *Set-valued Var. Anal.*

D. Noll, A. Rondepierre (2013). On local convergence of the method of alternating projections. Foundations of Computational Mathematics, 2015.

Convergence for non-convex Douglas-Rachford

R. Hesse, D.R. Luke (2013). Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems, *SIAM Journal on Optimization* 23(4), 2397–2419.

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

H.M. Phan (2014). Linear convergence of the Douglas-Rachford method for two closed sets. *Optimzation* 2015.

Pointer to Actual News

Comparison of resolution of CDI with :

Chemistry Nobel Price 2014 : Fluorescence Microscopy

Stefan Hell
*1962

William Moerner
*1953

Eric Betzig
*1960

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

CDI : $10\text{nm} = 10^{-8}\text{m}$ (organic)
$2\text{ nm} = 2 \cdot 10^{-9}\text{m}$ (anorganic)

Thanks for your attention !

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

Structured low-rank matrix approximation

Given a structured matrix $x \in \mathcal{S}$, solve the problem

$$(P) \quad \begin{array}{ll} \text{minimize} & \|x' - x\|_F \\ \text{subject to} & x' \in \mathcal{S} \\ & \mathrm{rank}(x') \leq r \end{array}$$

- $\mathcal{S} =$ Hankel matrices (denoising of time series)
- $\mathcal{S} =$ Toeplitz matrices (spectral estimation problems)
- $\mathcal{S} =$ positive semidefinite matrices
- $\mathcal{S} =$ stable matrices

Use non-convex alternating projections :

$$A = \{x : x \in \mathcal{S}\} \qquad P_A = \text{easy ???}$$

$$B = \{x : \mathrm{rank}(x) \leq r\} \qquad P_B = \text{truncated SVD}$$

Can now prove local convergence to $x^* \in \{x : \mathrm{rank}(x) \leq r, x \in \mathcal{S}\}$. Need not be solution of (P)

Sparse affine feasibility

$$\begin{array}{ll} \text{minimize} & \|x\|_0 = \text{number of non-zero entries in } x \\ \text{subject to} & Ax = b \\ & x \in \mathbb{R}^n \end{array}$$

Use non-convex alternating projections :

$$A = \{x \in \mathbb{R}^n : \|x\|_0 \leq k\} = \bigcup_{\mathrm{card}(I) \leq k} \underbrace{\mathrm{span}\{e_i : i \in I\}}_{=:A_I}$$

$$P_A(x) = \bigcup_{I_{\text{active}}} P_{A_I}(x)$$

$$B = \{x \in \mathbb{R}^n : Ax = b\} \qquad P_B(x) = x - A^{\dagger}(Ax - b)$$

Packings in Grassmannian manifolds

$\mathbb{G}(k, \mathbb{C}^d)$ = all k-dimensional subspaces of $\mathbb{C}^d$

Represent $\underline{S} \in \mathbb{G}(k, \mathbb{C}^d)$ by unitary $S \in \mathbb{C}^{k \times d}$: $S^* S = I_k$, range$(S) = \underline{S}$

For two subspaces S, T do SVD :

$$S^* T = UCV^*$$

then $c_{kk} = \cos\theta_k$ the principal angles between $\underline{S}, \underline{T}$. Leads to distances between $\underline{S}$ and $\underline{T}$:

- Chordal distance : $\sqrt{\sin^2\theta_1 + \cdots + \sin^2\theta_k} = \left(k - \|S^* T\|_F^2\right)^{1/2}$
- Spectral distance : $\min_i \sin\theta_i = \left(k - \|S^* T\|_{2,2}^2\right)^{1/2}$
- Fubini-Study distance : $\arccos\left(\Pi_j \cos\theta_j\right)$
- Geodesic distance : $\sqrt{\theta_1^2 + \cdots + \theta_k^2}$

Packing for the chordal distance :

$$\text{pack}(S_1, \ldots, S_N) := \min_{m \neq n} d_{\mathrm{chord}}(S_m, S_n) = \min_{m \neq n} \left(k - \|S_m^* S_n\|_F^2\right)^{1/2}$$

True problem : $\max\limits_{\{S_1, \ldots, S_N\}} \mathrm{pack}_{\mathrm{chord}}(S_1, \ldots, S_N)$

Instead feasibility problem : Given $\rho > 0$, want $\{S_1, \ldots, S_N\}$ such that $\mathrm{pack}_{\mathrm{chord}}(S_1, \ldots, S_N) \geq \rho$

$$\min_{m \neq n} \left(k - \|S_m^* S_n\|_F^2\right)^{1/2} \geq \rho \equiv \max_{m \neq n} \|S_m^* S_n\|_F \leq \mu := \sqrt{k - \rho^2}$$

Put $S := [S_1 S_2 \dots S_N]$

Gramian $G = S^* S \in \mathbb{C}^{kN \times kN} \succeq 0$, $G_{mn} = S_m^* S_n$

1. G is Hermitian
2. Each diagonal block of G is identity I_k
3. $G \succeq 0$
4. $\mathrm{rank}(G) \leq d$
5. $\mathrm{trace}(G) = kN$

Conversely, any G with these properties can be factored $G = S^* S$ and $S = [S_1 \dots S_N]$ gives then rise to a configuration of N subspaces in the Grassmannian $\mathbb{G}(k, \mathbb{C}^d)$.

Now alternating projections :

Structural constraint (convex) :

$$A = \{H \in \mathbb{C}^{kN \times kN} : H = H^*, H_{nn} = I_k, \|H_{mn}\|_F \leq \mu\}$$

Spectral constraint (non-convex) :

$$B = \{G \in \mathbb{C}^{kN \times kN} : G \succeq 0, \mathrm{rank}(G) \leq d, \mathrm{trace}(G) = kN\}$$

Any solution $G \in A \cap B$ gives a packing of size N of the Grassmannian with chordal distance packing index $\geq \rho$.

Compute projection on A (easy) :

$$H = P_A(G) : H_{mn} = \begin{cases} G_{mn} & \|G_{mn}\|_F \leq \mu \\ \mu G_{mn}/\|G_{mn}\|_F & \text{else} \end{cases}$$

Compute projection on B (more involved but possible) :

Let $H = \sum_{j=1}^{kN} \lambda_j u_j u_j^*$ be spectral decomposition, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{kN}$. Then

$$G = \sum_{j=1}^{d} (\lambda_j - \gamma)_+ u_j u_j^* \in P_B(H)$$

provided γ is chosen such that $\sum_{j=1}^{d} (\lambda_j - \gamma)_+ = kN$.