Analysis (and geometry) of alternating projection algorithms

Jérôme MALICK

CNRS (Lab. Jean Kuntzmann, Grenoble)

Séminaire de l'équipe ACSIOM, Montpellier – March 2010

based on joint work with Adrian Lewis and Russell Luke

Outline

Outline

Projection, distance and convexity

In a Euclidean space $(\mathbb{R}^n, \|\cdot\|)$

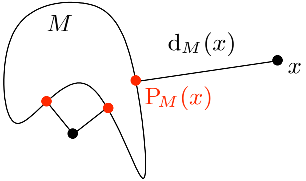
For closed $M \subset \mathbb{R}^n$, the distance of x from M

$$\mathrm{d}_M(x) = \min\{\|x - y\| : \; y \in M\}$$

and the projection of x onto M

$$\mathrm{P}_M(x) = \operatorname{argmin}\{\|x - y\| : \; y \in M\}$$

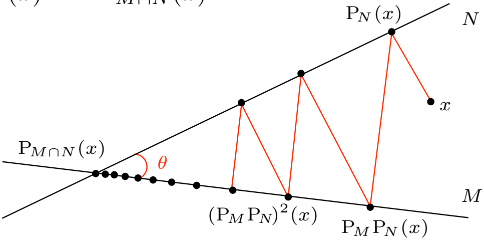
If M is convex, $\mathrm{P}_M(x)$ is singleton. Otherwise, it is not for some x for sure!



Alternating projections on subspaces

For affine subspaces M and N, von Neumann '33 studied:

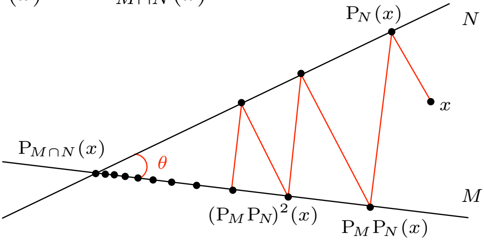
$$(\mathrm{P}_M\,\mathrm{P}_M)^n(x) \longrightarrow \mathrm{P}_{M\cap N}(x)$$



Alternating projections on subspaces

For affine subspaces M and N, von Neumann '33 studied:

$$(\mathrm{P}_M\,\mathrm{P}_M)^n(x) \longrightarrow \mathrm{P}_{M\cap N}(x)$$



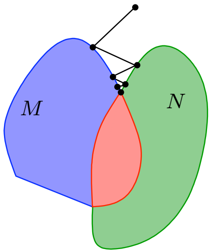
Convergence is linear at rate $(\cos\theta)^2$, indeed:

$$\|(\mathrm{P}_M\,\mathrm{P}_M)^n(x) - \mathrm{P}_{M\cap N}(x)\| \leqslant (\cos\theta)^{2n-1}\|x\|$$

where θ is the angle between M and N (Aronszajn '50)

Alternating convex projections

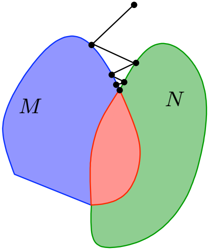
Alternating projections naturally extends to closed convex sets M and N



Bregman '65 proves: $(\mathrm{P}_M\,\mathrm{P}_M)^n(x) \longrightarrow M \cap N$

Alternating convex projections

Alternating projections naturally extends to closed convex sets M and N



Bregman '65 proves: $(\mathrm{P}_M\,\mathrm{P}_M)^n(x) \longrightarrow M \cap N$

Convergence is linear providing

$$M \cap \operatorname{int} N \neq \varnothing \qquad \text{(more generally } \operatorname{ri} M \cap \operatorname{ri} N \neq \varnothing\text{)}$$

(Polyak *et al* '67, Bauschke-Borwein '93)

Algorithm simple but powerful

To find a point $x \in M \cap N$, with M and N closed convex sets on $\mathbb{R}^n$

Algorithm simple but powerful

To find a point $x \in M \cap N$, with M and N closed convex sets on $\mathbb{R}^n$

Alternating convex projections is a basic algorithm

$\longrightarrow$ many enhancements, among them:

- in Hilbert, complex spaces...
- several sets, averaged, cyclic projections...
- relaxations, regularization (under/over-relaxed, AAR, Dykstra,...)

Algorithm simple but powerful

To find a point $x \in M \cap N$, with M and N closed convex sets on $\mathbb{R}^n$

Alternating convex projections is a basic algorithm

$\longrightarrow$ many enhancements, among them:

- in Hilbert, complex spaces...
- several sets, averaged, cyclic projections...
- relaxations, regularization (under/over-relaxed, AAR, Dykstra,...)

Alternating convex projections is very simple

$\longrightarrow$ very popular in practice !

Applications: statistics, finance, engineering sciences, image processing. . .

Algorithm simple but powerful

To find a point $x \in M \cap N$, with M and N closed convex sets on $\mathbb{R}^n$

Alternating convex projections is a basic algorithm

$\longrightarrow$ many enhancements, among them:

- in Hilbert, complex spaces...
- several sets, averaged, cyclic projections...
- relaxations, regularization (under/over-relaxed, AAR, Dykstra,...)

Alternating convex projections is very simple

$\longrightarrow$ very popular in practice !

Applications: statistics, finance, engineering sciences, image processing...

Ex1: Hounsfield '73 Nobel Prize winner, inventor of the EMI scanner

Algorithm simple but powerful

To find a point $x \in M \cap N$, with M and N closed convex sets on $\mathbb{R}^n$

Alternating convex projections is a basic algorithm

$\longrightarrow$ many enhancements, among them:

- in Hilbert, complex spaces...
- several sets, averaged, cyclic projections...
- relaxations, regularization (under/over-relaxed, AAR, Dykstra,...)

Alternating convex projections is very simple

$\longrightarrow$ very popular in practice !

Applications: statistics, finance, engineering sciences, image processing...

Ex1: Hounsfield '73 Nobel Prize winner, inventor of the EMI scanner

Ex2: Xu-Zikatanov '02 draw connection alternating projection with methods of PDE's (domain decomposition or multigrid methods)

Algorithm simple but powerful

To find a point $x \in M \cap N$, with M and N closed convex sets on $\mathbb{R}^n$

Alternating convex projections is a basic algorithm

$\longrightarrow$ many enhancements, among them:

- in Hilbert, complex spaces...
- several sets, averaged, cyclic projections...
- relaxations, regularization (under/over-relaxed, AAR, Dykstra,...)

Alternating convex projections is very simple

$\longrightarrow$ very popular in practice !

Applications: statistics, finance, engineering sciences, image processing...

Ex1: Hounsfield '73 Nobel Prize winner, inventor of the EMI scanner

Ex2: Xu-Zikatanov '02 draw connection alternating projection with methods of PDE's (domain decomposition or multigrid methods)

Ex3: Combettes et al. '10 "on the effectiveness of projection methods for convex feasibility problems with linear inequality constraints"

Example in Finance

For symmetric matrix C, computing the nearest correlation matrix: computing the projection of C onto the intersection of $\mathcal{S}_n^+$ the semidefinite positive matrices, and the matrices with ones the diagonal

Example in Finance

For symmetric matrix C, computing the nearest correlation matrix: computing the projection of C onto the intersection of $\mathcal{S}_n^+$ the semidefinite positive matrices, and the matrices with ones the diagonal

Used as calibration for evaluating extreme risks (Stress testing)

Example in Finance

For symmetric matrix C, computing the nearest correlation matrix: computing the projection of C onto the intersection of $\mathcal{S}_n^+$ the semidefinite positive matrices, and the matrices with ones the diagonal

Used as calibration for evaluating extreme risks (Stress testing)

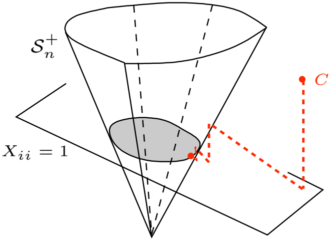
How to compute the nearest correlation matrix:

Example in Finance

For symmetric matrix C, computing the nearest correlation matrix: computing the projection of C onto the intersection of $\mathcal{S}_n^+$ the semidefinite positive matrices, and the matrices with ones the diagonal

Used as calibration for evaluating extreme risks (Stress testing)

How to compute the nearest correlation matrix:



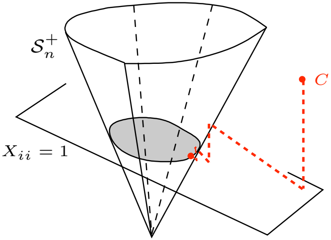
1. alternating projection (Higham '02) (+ Dykstra correction)

Example in Finance

For symmetric matrix C, computing the nearest correlation matrix: computing the projection of C onto the intersection of $\mathcal{S}_n^+$ the semidefinite positive matrices, and the matrices with ones the diagonal

Used as calibration for evaluating extreme risks (Stress testing)

How to compute the nearest correlation matrix:



1. alternating projection (Higham '02) (+ Dykstra correction)
2. Lagrangian duality (Malick '04) $\longrightarrow$ efficient algorithm

Nonconvex heuristic

alternating convex projections is a good method...

Nonconvex heuristic

alternating convex projections is a good method...
...alternating nonconvex projections is also a popular heuristic !

Nonconvex heuristic

alternating convex projections is a good method...
...alternating nonconvex projections is also a popular heuristic !

Examples

- Optics : phase retrieval of images (Combettes *et al* '02)
  Simple version : given $a_j \in \mathbb{C}^k$, find $x \in \mathbb{C}^k$, so

  $$|\langle a_j, x\rangle| = b_j \quad (j = 1, \ldots, m)$$

  with alternative projections onto

  $$\begin{aligned} M &= \{(x,z) \in \mathbb{C}^n \times \mathbb{C}^m : \; Ax = z\} \\ N &= \{(x,z) : \; |z_j| = b_j, \; (j = 1, \ldots, m)\} \end{aligned}$$

Nonconvex heuristic

alternating convex projections is a good method...
...alternating nonconvex projections is also a popular heuristic !

Examples

- Optics : phase retrieval of images (Combettes *et al* '02)
  Simple version : given $a_j \in \mathbb{C}^k$, find $x \in \mathbb{C}^k$, so

  $$|\langle a_j, x\rangle| = b_j \quad (j = 1, \ldots, m)$$

  with alternative projections onto

  $$\begin{aligned} M &= \{(x,z) \in \mathbb{C}^n \times \mathbb{C}^m : \; Ax = z\} \\ N &= \{(x,z) : \; |z_j| = b_j, \; (j = 1, \ldots, m)\} \end{aligned}$$

- Control : low-order control design (eg Grigoriadis-Beran '00)

  $$\begin{aligned} \text{affine } M &\subset \{n\text{-by-}n \text{ symmetric matrices}\} \\ N &= \{\text{positive semidefinite matrices of rank } r\} \end{aligned}$$

Numerical illustration

Find a 100-by-110 matrix X of rank 4, satisfying 450 equations

$$A(X) = b$$

(simple analogue of the low-rank control problem)

Numerical illustration

Find a 100-by-110 matrix X of rank 4, satisfying 450 equations

$$A(X) = b$$

(simple analogue of the low-rank control problem)

$\longrightarrow$ alternating projections onto

$$\begin{aligned} M &= \{X : \; A(X) = b\} && \text{(by inverting } AA^\top\text{)} \\ N &= \{X : \; \mathrm{rank}(X) = 4\} && \text{(via singular value decomposition)} \end{aligned}$$

Numerical illustration

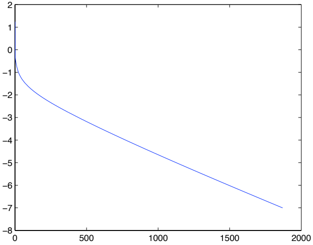
Find a 100-by-110 matrix X of rank 4, satisfying 450 equations

$$A(X) = b$$

(simple analogue of the low-rank control problem)

$\longrightarrow$ alternating projections onto

$$M = \{X : A(X) = b\} \qquad \text{(by inverting } AA^\top\text{)}$$
$$N = \{X : \operatorname{rank}(X) = 4\} \qquad \text{(via singular value decomposition)}$$



Alternating nonconvex projections

Why does alternating nonconvex projections work ?

Alternating nonconvex projections

Why does alternating nonconvex projections work ?

Projecting onto nonconvex sets ??
How does linear convergence appear ?

Alternating nonconvex projections

Why does alternating nonconvex projections work ?

Projecting onto nonconvex sets ??
How does linear convergence appear ?

Few answers:

- 1st results in Combettes-Trussel '90 (with convex-like techniques)
- linear cvg in special cases (Orsi '06 for a matrix analysis pb),...
- or of special algos (Attouch-Bolte-Redont-Soubeyrand '08, Luke '09)
- Lewis-Malick '07, Lewis-Luke-Malick '08 – whose ingredients are:
  - geometry of the intersection and of the algorithm
  - interpretation with nonsmooth analysis

Alternating nonconvex projections

Why does alternating nonconvex projections work ?

Projecting onto nonconvex sets ??
How does linear convergence appear ?

Few answers:

- 1st results in Combettes-Trussel '90 (with convex-like techniques)
- linear cvg in special cases (Orsi '06 for a matrix analysis pb),...
- or of special algos (Attouch-Bolte-Redont-Soubeyrand '08, Luke '09)
- Lewis-Malick '07, Lewis-Luke-Malick '08 – whose ingredients are:
  - geometry of the intersection and of the algorithm
  - interpretation with nonsmooth analysis

In this talk:

1. easy-to-compute nonconvex projections
2. convergence of the algorithm through nice geometry

Outline

Easy nonconvex projections

For closed nonconvex $M \subset \mathbb{R}^n$, the projection $\mathrm{P}_M(x)$ is somewhere nonsingleton. But projection may still be easy...

Easy nonconvex projections

For closed nonconvex $M \subset \mathbb{R}^n$, the projection $\mathrm{P}_M(x)$ is somewhere nonsingleton. But projection may still be easy...

Examples

- Single quadratic constraint

$$M = \left\{x \in \mathbb{R}^n : \quad x^\top A\,x + b^\top x = c\right\}$$

  Projection is analogous to trust-region subproblems, solvable with a special Newton method

Easy nonconvex projections

For closed nonconvex $M \subset \mathbb{R}^n$, the projection $\mathrm{P}_M(x)$ is somewhere nonsingleton. But projection may still be easy...

Examples

- Single quadratic constraint

$$M = \left\{ x \in \mathbb{R}^n : \quad x^\top A\, x + b^\top x = c \right\}$$

  Projection is analogous to trust-region subproblems, solvable with a special Newton method

- Rank constraint

$$M = \left\{ X \in \mathbb{R}^{n \times m} : \quad \mathrm{rank}(X) = r \right\}$$

  To project, find a singular value decomposition $X = UDV$ and zero all but the first r largest singular values in D

Spectral sets

For permutation-invariant $K \subset \mathbb{R}^n$, the spectral set of symmetric matrices

$$\lambda^{-1}(K) = \{X \in \mathcal{S}_n : \quad (\lambda_1(X), \ldots, \lambda_n(X)) \in K\}$$

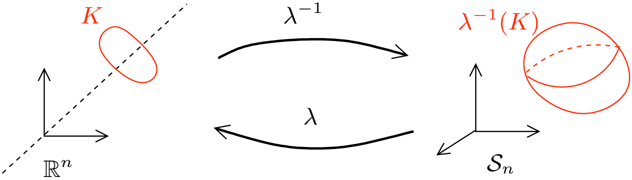
(analogous definition for sets of matrices described by singular values)

Spectral sets

For permutation-invariant $K \subset \mathbb{R}^n$, the spectral set of symmetric matrices

$$\lambda^{-1}(K) = \{X \in \mathcal{S}_n : \quad (\lambda_1(X), \dots, \lambda_n(X)) \in K\}$$

(analogous definition for sets of matrices described by singular values)

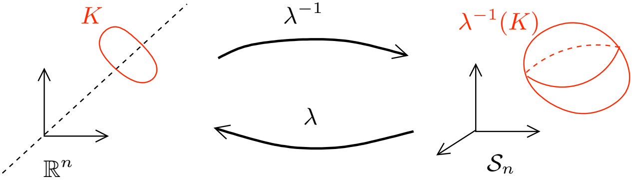


Spectral sets

For permutation-invariant $K \subset \mathbb{R}^n$, the spectral set of symmetric matrices

$$\lambda^{-1}(K) = \{X \in \mathcal{S}_n : \ (\lambda_1(X), \ldots, \lambda_n(X)) \in K\}$$

(analogous definition for sets of matrices described by singular values)



Examples

- $K = \mathbb{R}^n_+$ gives the positive semidefinite cone $\mathcal{S}^+_n$
- $K = \Sigma_n \cdot x$ gives an isospectral set (given x)
- $K = \{x : \ \mathrm{Card}\,(\mathrm{argmax}\{x_i\}) = r\}$ gives $\{X : \ \lambda_{\max}(X)$ with $r\}$

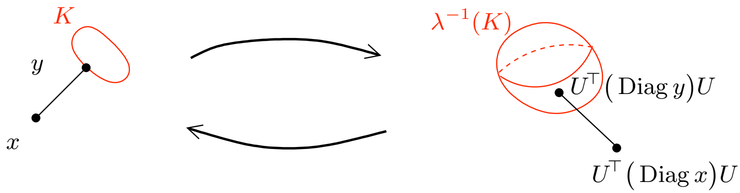
Easy spectral projections

The following result (Lewis-Malick '07) generalizes previous (partial) results about projections onto some spectral sets (eg Higham '88, Oustry '02)

Theorem (projection onto spectral sets)

If $y \in \mathrm{P}_K(x)$ *and* U *orthogonal, then*

$$U^\top(\mathrm{Diag}\, y)U \;\in\; \mathrm{P}_{\lambda^{-1}(K)}(U^\top(\mathrm{Diag}\, x)U)$$



Prox-regular spectral sets

Let's take the chance to say more on spectral sets

Transfer of structure: if K is invariant by permutation of entries

Prox-regular spectral sets

Let's take the chance to say more on spectral sets

Transfer of structure: if K is invariant by permutation of entries

1. Lewis '96 : K convex $\implies$ $\lambda^{-1}(K)$ convex

Prox-regular spectral sets

Let's take the chance to say more on spectral sets

Transfer of structure: if K is invariant by permutation of entries

1. Lewis '96 : K convex $\implies \lambda^{-1}(K)$ convex
2. Daniilidis-Lewis-Malick-Sendov '08 proves

$$K \text{ prox-regular } \implies \lambda^{-1}(K) \text{ prox-regular}$$

Prox-regular spectral sets

Let's take the chance to say more on spectral sets

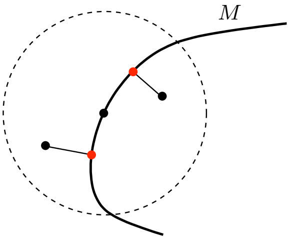
Transfer of structure: if K is invariant by permutation of entries

1. Lewis '96 : K convex $\implies \lambda^{-1}(K)$ convex
2. Daniilidis-Lewis-Malick-Sendov '08 proves

$$K \text{ prox-regular} \implies \lambda^{-1}(K) \text{ prox-regular}$$

General notion of prox-regularity
(eg Rock.-Poliquin-Thibault '00):

P_M is locally unique



⟶ prox-regular spectral sets have locally all the good properties !

(Ex: manifolds...)

Prox-regular spectral sets in practice

$\longrightarrow$ Many spectral sets in alternative nonconvex projections

1. Numerical algebra: nonnegative inverse eigenvalue pb (Orsi '06)
For $\bar{\lambda}$ given, find $X \in M \cap N$

$$\begin{aligned} M &= \{X \in \mathbb{R}^{n\times n} : \ \lambda(X) = \bar{\lambda}\,\} \\ N &= \{X \in \mathbb{R}^{n\times n} : \ X_{ij} \geqslant 0\,\} \end{aligned}$$

Prox-regular spectral sets in practice

$\longrightarrow$ Many spectral sets in alternative nonconvex projections

1. Numerical algebra: nonnegative inverse eigenvalue pb (Orsi '06)
   For $\bar{\lambda}$ given, find $X \in M \cap N$

$$\begin{aligned} M &= \{X \in \mathbb{R}^{n \times n} : \ \lambda(X) = \bar{\lambda} \,\} \\ N &= \{X \in \mathbb{R}^{n \times n} : \ X_{ij} \geqslant 0 \,\} \end{aligned}$$

2. Image processing: design of tight frames (Tropp *et al* '05)
   Find the associated Gram matrix $X \in M \cap N$

$$\begin{aligned} M &= \{X \in \mathbb{C}^{n \times n} : \ \lambda(X) = (n, \ldots, n, 0, \ldots, 0)/d \,\} \\ N &= \{X \in \mathbb{C}^{n \times n} : \ X_{ii} = 1, \ \|X\|_\infty \leqslant \mu \,\} \end{aligned}$$

Prox-regular spectral sets in practice

⟶ Many spectral sets in alternative nonconvex projections

1. Numerical algebra: nonnegative inverse eigenvalue pb (Orsi '06)
   For $\bar\lambda$ given, find $X \in M \cap N$

$$\begin{aligned} M &= \{X \in \mathbb{R}^{n\times n} : \ \lambda(X) = \bar\lambda \,\} \\ N &= \{X \in \mathbb{R}^{n\times n} : \ X_{ij} \geqslant 0 \,\} \end{aligned}$$

2. Image processing: design of tight frames (Tropp *et al* '05)
   Find the associated Gram matrix $X \in M \cap N$

$$\begin{aligned} M &= \{X \in \mathbb{C}^{n\times n} : \ \lambda(X) = (n,\dots,n,0,\dots,0)/d \,\} \\ N &= \{X \in \mathbb{C}^{n\times n} : \ X_{ii} = 1, \ \|X\|_\infty \leqslant \mu \,\} \end{aligned}$$

⟶ Many other types of sets (ex: Phase retrieval)...

Last word on spectral sets

If M is smooth manifold, is $\lambda^{-1}(M)$ smooth as well?

Last word on spectral sets

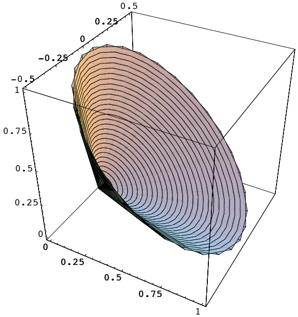
If M is smooth manifold, is $\lambda^{-1}(M)$ smooth as well?

- No, in general

  Ex: $M =]-1,1[\times\{0\} \subset \mathbb{R}^2$,
  and $\lambda^{-1}(M)$ has a kink

- Yes, if M is loccally symmetric !

  a neighbd of M around $x \in M$
  is invariant under permutations
  σ such that $\sigma x = x$



Moreover, we know the dimension of $\lambda^{-1}(M)$

Not straighforward... 43 pages of sheer joy (Daniilidis-Malick-Sendov '09)

Outline

Alternating nonconvex projections

Theorem (local linear convergence)

For closed sets $M, N \subset \mathbb{R}^n$. Assume

- *strong regularity holds at $\bar{x} \in M \cap N$*
- *M is super-regular at $\bar{x}$*
- *initial x_0 near $\bar{x}$*

Then alternating projection method converges R-linearly to $M \cap N$

Alternating nonconvex projections

Theorem (local linear convergence)

For closed sets $M, N \subset \mathbb{R}^n$. Assume

- *strong regularity holds at $\bar{x} \in M \cap N$*
- *M is super-regular at $\bar{x}$*
- *initial x_0 near $\bar{x}$*

Then alternating projection method converges R-linearly to $M \cap N$

Comments

- Super-regular sets: convex sets, smooth manifolds

Alternating nonconvex projections

Theorem (local linear convergence)

For closed sets $M, N \subset \mathbb{R}^n$. Assume

- *strong regularity holds at $\bar{x} \in M \cap N$*
- *M is super-regular at $\bar{x}$*
- *initial x_0 near $\bar{x}$*

Then alternating projection method converges R-linearly to $M \cap N$

Comments

- Super-regular sets: convex sets, smooth manifolds
- The convergence rate is $\cos\theta$ where θ is the minimal angle between $\mathrm{N}_M(\bar{x})$ and $-\mathrm{N}_N(\bar{x})$

Alternating nonconvex projections

Theorem (local linear convergence)

For closed sets $M, N \subset \mathbb{R}^n$. Assume

- *strong regularity holds at $\bar{x} \in M \cap N$*
- *M is super-regular at $\bar{x}$*
- *initial x_0 near $\bar{x}$*

Then alternating projection method converges R-linearly to $M \cap N$

Comments

- Super-regular sets: convex sets, smooth manifolds
- The convergence rate is $\cos\theta$ where θ is the minimal angle between $\mathrm{N}_M(\bar{x})$ and $-\mathrm{N}_N(\bar{x})$
- Rate is $(\cos\theta)^2$ if both M and N are super-regular

Alternating nonconvex projections

Theorem (local linear convergence)

For closed sets $M, N \subset \mathbb{R}^n$. Assume

- *strong regularity holds at $\bar{x} \in M \cap N$*
- *M is super-regular at $\bar{x}$*
- *initial x_0 near $\bar{x}$*

Then alternating projection method converges R-linearly to $M \cap N$

Comments

- Super-regular sets: convex sets, smooth manifolds
- The convergence rate is $\cos\theta$ where θ is the minimal angle between $\mathrm{N}_M(\bar{x})$ and $-\mathrm{N}_N(\bar{x})$
- Rate is $(\cos\theta)^2$ if both M and N are super-regular
- Inexact versions still work

Alternating nonconvex projections

Theorem (local linear convergence)

For closed sets $M, N \subset \mathbb{R}^n$. Assume

- *strong regularity holds at $\bar{x} \in M \cap N$*
- *M is super-regular at $\bar{x}$*
- *initial x_0 near $\bar{x}$*

Then alternating projection method converges R-linearly to $M \cap N$

Comments

- Super-regular sets: convex sets, smooth manifolds
- The convergence rate is $\cos\theta$ where θ is the minimal angle between $\mathrm{N}_M(\bar{x})$ and $-\mathrm{N}_N(\bar{x})$
- Rate is $(\cos\theta)^2$ if both M and N are super-regular
- Inexact versions still work
- More in Lewis-Luke-Malick '08...

Strong regularity

Simple definition:

$$\mathrm{N}_M(\bar{x}) \cap -\mathrm{N}_N(\bar{x}) = \{0\}$$

in other words, the minimal angle between $\mathrm{N}_M(\bar{x})$ and $-\mathrm{N}_N(\bar{x})$ is $\theta > 0$

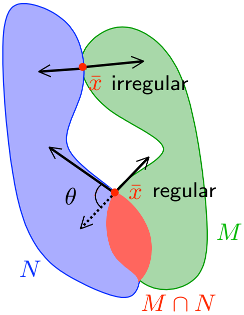
Strong regularity

Simple definition:

$$\mathrm{N}_M(\bar{x}) \cap -\mathrm{N}_N(\bar{x}) = \{0\}$$

in other words, the minimal angle between $\mathrm{N}_M(\bar{x})$ and $-\mathrm{N}_N(\bar{x})$ is $\theta > 0$

Strong regularity is a standard notion of nonsmooth analysis (see eg Kummer '06), useful in theory (ex: normal cone to the intersection)



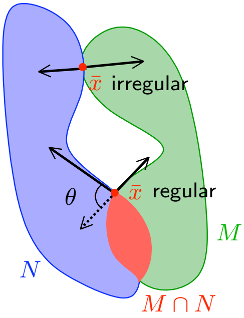
Strong regularity

Simple definition:

$$\mathrm{N}_M(\bar{x}) \cap -\mathrm{N}_N(\bar{x}) = \{0\}$$

in other words, the minimal angle between $\mathrm{N}_M(\bar{x})$ and $-\mathrm{N}_N(\bar{x})$ is $\theta > 0$

Strong regularity is a standard notion of nonsmooth analysis (see eg Kummer '06), useful in theory (ex: normal cone to the intersection)



Examples

1. The intersection of two smooth manifolds is strongly regular
   $\iff$ the manifolds are transverse
2. The intersection of two convex sets is strongly regular
   $\iff$ no separating hyperplane

Super-regularity

Notion (not standard!) introduced in
Lewis-Luke-Malick '08

Examples of super-regular sets

1. convex sets
2. smooth manifolds

Super-regularity

Notion (not standard!) introduced in
Lewis-Luke-Malick '08

Examples of super-regular sets

1. convex sets
2. smooth manifolds
3. prox-regular sets

Super-regularity

Notion (not standard!) introduced in
Lewis-Luke-Malick '08

Examples of super-regular sets

1. convex sets
2. smooth manifolds
3. prox-regular sets
4. constraint sets with Mangasarian-Fromovitz

Super-regularity

Notion (not standard!) introduced in
Lewis-Luke-Malick '08

Examples of super-regular sets

1. convex sets
2. smooth manifolds
3. prox-regular sets
4. constraint sets with Mangasarian-Fromovitz
5. nearly convex sets (Shapiro '93)

Super-regularity

Notion (not standard!) introduced in Lewis-Luke-Malick '08

Examples of super-regular sets

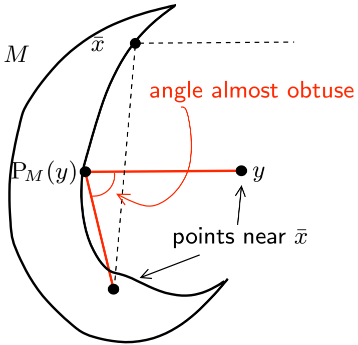
1. convex sets
2. smooth manifolds
3. prox-regular sets
4. constraint sets with Mangasarian-Fromovitz
5. nearly convex sets (Shapiro '93)
6. subsmooth hypomonotone (Aussel-Daniilidis-Thilbault '04)

Super-regularity

Notion (not standard!) introduced in Lewis-Luke-Malick '08

Examples of super-regular sets

1. convex sets
2. smooth manifolds
3. prox-regular sets
4. constraint sets with Mangasarian-Fromovitz
5. nearly convex sets (Shapiro '93)
6. subsmooth hypomonotone (Aussel-Daniilidis-Thilbault '04)

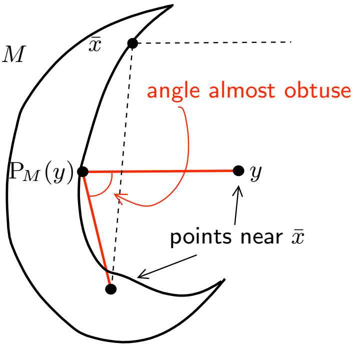


Super-regularity

Notion (not standard!) introduced in
Lewis-Luke-Malick '08

Examples of super-regular sets

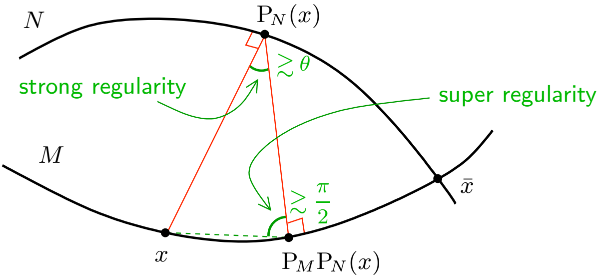
1. convex sets
2. smooth manifolds
3. prox-regular sets
4. constraint sets with Mangasarian-Fromovitz
5. nearly convex sets (Shapiro '93)
6. subsmooth hypomonotone (Aussel-Daniilidis-Thilbault '04)



prox-regular $\subset$ super-regular $\subset$ (Clarke) regular

Sketch proof

The geometry controls the asymptotical improvement:



For $x \in M$ near $\bar{x}$,

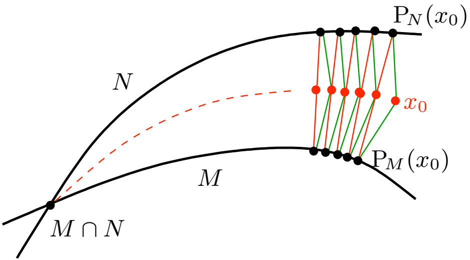
$$\frac{\| P_M P_N(x) - P_N(x)\|}{\| P_N(x) - x\|}$$

is not much larger than $\cos\theta$

Consequence for averaged projections

Method of averaged projections

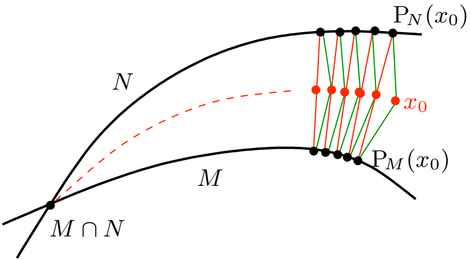
$$
\begin{aligned}
z_M &\in P_M(x) \\
z_N &\in P_N(x) \\
x &\leftarrow \frac{1}{2}(z_N + z_M)
\end{aligned}
$$



Consequence for averaged projections

Method of averaged projections

$$\begin{aligned} z_M &\in P_M(x) \\ z_N &\in P_N(x) \\ x &\leftarrow \frac{1}{2}(z_N + z_M) \end{aligned}$$



Corollary (linear convergence of averaged projections)

For any closed $M, N \subset \mathbb{R}^n$, if strong regularity holds at $\bar{x} \in M \cap N$, then starting with x_0 near $\bar{x}$, averaged projections converges linearly to $M \cap N$

More on averaged projections

1. **Proof:** Following Auslender '69 in the convex case, just consider alternating projections in $\mathbb{R}^n \times \mathbb{R}^n$ between

$$M \times N \qquad \{(x,x) : \ x \in \mathbb{R}^n\} \text{ (super-regular)}$$

More on averaged projections

1. **Proof:** Following Auslender '69 in the convex case, just consider alternating projections in $\mathbb{R}^n \times \mathbb{R}^n$ between

$$M \times N \qquad \{(x,x) :\ x \in \mathbb{R}^n\} \text{ (super-regular)}$$

2. **Interpretation as minimization** (Lewis-Luke-Malick '08)
If M and N both prox-regular, averaged projections is just the steepest descent (with unit step size) applied to

$$f(x) = \frac{1}{4}\left(\mathrm{d}^2_M(x) + \mathrm{d}^2_N(x) \right)$$

$\longrightarrow$ Q-linear convergence: improvements at each iteration

$$\frac{f(x_{k+1})}{f(x_k)} < 1 - \frac{1}{2\kappa^2}$$

Numerical illustration

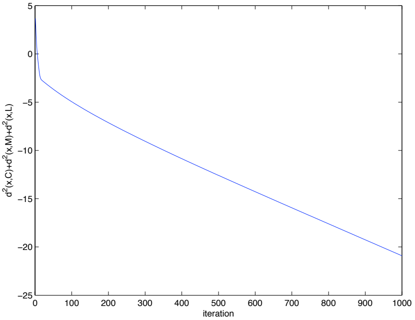
Averaged projections to find d-by-m matrix $U \in L \cap M \cap C$

$$
\begin{array}{rrcl}
\text{(linear)} & L & = & \{U \in \mathbb{R}^{d\times m} :\ U = PW\} \\
\text{(smooth)} & M & = & \{U \in \mathbb{R}^{d\times m} :\ UU^\top = I\} \\
\text{(convex)} & C & = & \{U \in \mathbb{R}^{d\times m} :\ \|U\|_\infty \leqslant \alpha\}
\end{array}
$$

Numerical illustration

Averaged projections to find d-by-m matrix $U \in L \cap M \cap C$

$$\begin{array}{rrcl} \text{(linear)} & L & = & \{U \in \mathbb{R}^{d\times m} : \; U = PW\} \\ \text{(smooth)} & M & = & \{U \in \mathbb{R}^{d\times m} : \; UU^\top = I\} \\ \text{(convex)} & C & = & \{U \in \mathbb{R}^{d\times m} : \; \|U\|_\infty \leqslant \alpha\} \end{array}$$



$$\frac{f(U_{k+1})}{f(U_k)} \leqslant 0.96 < 1$$

where f is the sum of the squared distances

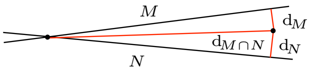
Outline

1. Alternating convex projections
2. Nonconvex projections
3. Alternating nonconvex projections
4. **Regularity and conditionning**

alternating projections + small angle

Weak error bound:

$$\mathrm{d}^2_{M\cap N}(\cdot) \leqslant \rho\big(\,\mathrm{d}^2_M(\cdot) + \mathrm{d}^2_N(\cdot)\big)$$

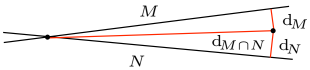


needs ρ large

alternating projections + small angle

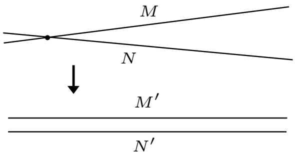
Weak error bound:

$$\mathrm{d}^2_{M\cap N}(\cdot) \leqslant \rho\big(\,\mathrm{d}^2_M(\cdot) + \mathrm{d}^2_N(\cdot)\big)$$



needs ρ large

Small perturbations render the problem ill-posed

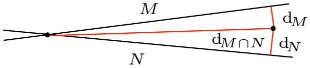


alternating projections + small angle

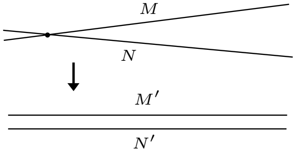
Weak error bound:

$$\mathrm{d}^2_{M\cap N}(\cdot) \leqslant \rho\big(\,\mathrm{d}^2_M(\cdot) + \mathrm{d}^2_N(\cdot)\big)$$

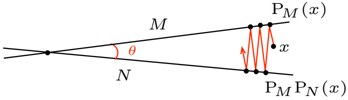
needs ρ large



Small perturbations render the problem ill-posed



Alternating method converges with slow linear rate $\cos\theta$



Conditionning

For linear subspaces for example, it is obvious on picture that instances are not well-conditionning, when the angle θ is small...

... moreover $\cos\theta$ controls so the speed of alternating (and averaged) projection algorithms.

Conditionning

For linear subspaces for example, it is obvious on picture that instances are not well-conditionning, when the angle θ is small...

... moreover $\cos\theta$ controls so the speed of alternating (and averaged) projection algorithms.

$\longrightarrow$ alternating projections turns out be a nice illustration of

- Demmel paradigm (Demmel '87)
- metric regularity (eg Rockafellar-Wets '02)

Conditionning

For linear subspaces for example, it is obvious on picture that instances are not well-conditionning, when the angle θ is small...

... moreover $\cos\theta$ controls so the speed of alternating (and averaged) projection algorithms.

$\longrightarrow$ alternating projections turns out be a nice illustration of

- Demmel paradigm (Demmel '87)
- metric regularity (eg Rockafellar-Wets '02)

For many computational problems, three equivalent properties characterize "hard" instances:

1. a posteriori error bounds are weak
2. the distance to an ill-posed instance is small
3. basic algorithms are slow

Simple example: solving linear system

Consider the positive-definite linear system

$$A\,x = y$$

1. a posteriori error bound is weak

$$\|x - A^{-1}y\| \leqslant \frac{1}{\lambda_{\min}(A)}\|Ax - y\|$$

Simple example: solving linear system

Consider the positive-definite linear system

$$A\,x = y$$

1. a posteriori error bound is weak

$$\|x - A^{-1}y\| \leqslant \frac{1}{\lambda_{\min}(A)} \|Ax - y\|$$

2. the distance from A to singular matrices $(= \lambda_{\min}(A))$ is small

Simple example: solving linear system

Consider the positive-definite linear system

$$A\,x = y$$

1. a posteriori error bound is weak

$$\|x - A^{-1}y\| \leqslant \frac{1}{\lambda_{\min}(A)} \|Ax - y\|$$

2. the distance from A to singular matrices ($= \lambda_{\min}(A)$) is small

3. Basic algorithms converge with slow linear rate. Eg:

$$\left(\frac{\kappa - 1}{\kappa + 1}\right)^2 \quad \text{and} \quad \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^2$$

for steepest descent and conjuguate gradients, where $\kappa = \lambda_{\max}(A)/\lambda_{\min}(A)$

General framework

A very general framework: invertion
Given set-valued $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^p$, suppose

- $F(x)$ is easy to compute
- $F^{-1}(x)$ is hard to compute

Problem: Given a point $\bar{y}$, find some point $\bar{x} \in F^{-1}(\bar{y})$

General framework

A very general framework: invertion
Given set-valued $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^p$, suppose

- $F(x)$ is easy to compute
- $F^{-1}(x)$ is hard to compute

Problem: Given a point $\bar{y}$, find some point $\bar{x} \in F^{-1}(\bar{y})$

Example: Solve $A\,\bar{x} = \bar{y}$

General framework

A very general framework: invertion
Given set-valued $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^p$, suppose

- $F(x)$ is easy to compute
- $F^{-1}(x)$ is hard to compute

Problem: Given a point $\bar{y}$, find some point $\bar{x} \in F^{-1}(\bar{y})$

Example: Solve $A\,\bar{x} = \bar{y}$

Example: (Lewis-Luke-Malick '08)
Given $M, N \subset \mathbb{R}^n$, find $\bar{x} \in M \cap N$

General framework

A very general framework: invertion
Given set-valued $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^p$, suppose

- $F(x)$ is easy to compute
- $F^{-1}(x)$ is hard to compute

Problem: Given a point $\bar{y}$, find some point $\bar{x} \in F^{-1}(\bar{y})$

Example: Solve $A\,\bar{x} = \bar{y}$

Example: (Lewis-Luke-Malick '08)
Given $M, N \subset \mathbb{R}^n$, find $\bar{x} \in M \cap N$
Define $G\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^n$ by

$$G(x) = (M - x) \times (N - x)$$

Finding $\bar{x} \in G^{-1}(0,0) = M \cap N$ is solving the generalized equation

$$(0,0) \in G(x)$$

To quantify local error bounds

Definition (Metric regularity)

Suppose $\bar{y} \in F(\bar{x})$. We say that F is metrically regular at $(\bar{x}, \bar{y})$ if the local error bound

$$\mathrm{d}_{F^{-1}(y)}(x) \leqslant \rho \, \mathrm{d}_{F(x)}(y) \qquad \text{for all } (x, y) \text{ near } (\bar{x}, \bar{y})$$

holds for some ρ. The modulus $\operatorname{reg} F(\bar{x}, \bar{y})$ is the infinum of such ρ

To quantify local error bounds

Definition (Metric regularity)

Suppose $\bar{y} \in F(\bar{x})$. We say that F is metrically regular at $(\bar{x}, \bar{y})$ if the local error bound

$$\mathrm{d}_{F^{-1}(y)}(x) \leqslant \rho \, \mathrm{d}_{F(x)}(y) \qquad \text{for all } (x, y) \text{ near } (\bar{x}, \bar{y})$$

holds for some ρ. The modulus $\operatorname{reg} F(\bar{x}, \bar{y})$ is the infinum of such ρ

For alternating projections (ie: $(0,0) \in G(x)$)

- G is metrically regular at $\big(\bar{x}, (0,0)\big)$
  $\iff$ closed M and N have strongly regular intersection

To quantify local error bounds

Definition (Metric regularity)

Suppose $\bar{y} \in F(\bar{x})$. We say that F is metrically regular at $(\bar{x}, \bar{y})$ if the local error bound

$$\mathrm{d}_{F^{-1}(y)}(x) \leqslant \rho\, \mathrm{d}_{F(x)}(y) \qquad \text{for all } (x, y) \text{ near } (\bar{x}, \bar{y})$$

holds for some ρ. The modulus $\operatorname{reg} F(\bar{x}, \bar{y})$ is the infinum of such ρ

For alternating projections (ie: $(0,0) \in G(x)$)

- G is metrically regular at $\big(\bar{x}, (0,0)\big)$
  $\iff$ closed M and N have strongly regular intersection
- With θ the minimum angle between $\mathrm{N}_M(\bar{x})$ and $-\mathrm{N}_N(\bar{x})$
  $$\operatorname{reg} G(\bar{x}, (0,0)) = \frac{1}{\sqrt{1 - \cos\theta}}$$
  hence the modulus controls local linear convergence rates

Condition number

Modulus of metric regularity quantifies hardness of instances:

1. the a posteriori error bounds (definition of reg)
2. the distance to ill-posedness is $1/\mathrm{reg}$ by a general theorem (Dontchev-Lewis-Rockafellar '03)
3. the rate of convergence of basic algorithms is governed by reg

Condition number

Modulus of metric regularity quantifies hardness of instances:

1. the a posteriori error bounds (definition of reg)
2. the distance to ill-posedness is $1/\operatorname{reg}$ by a general theorem (Dontchev-Lewis-Rockafellar '03)
3. the rate of convergence of basic algorithms is governed by reg

Illustrations of this quantified Demmel's paradigm:

- linear systems
- alternating projections (Lewis-Malick '07, Lewis-Luke-Malick '08)

Condition number

Modulus of metric regularity quantifies hardness of instances:

1. the a posteriori error bounds (definition of reg)
2. the distance to ill-posedness is $1/\operatorname{reg}$ by a general theorem (Dontchev-Lewis-Rockafellar '03)
3. the rate of convergence of basic algorithms is governed by reg

Illustrations of this quantified Demmel's paradigm:

- linear systems
- alternating projections (Lewis-Malick '07, Lewis-Luke-Malick '08)
- proximal point methods (Aragon-Artacho-Dontchev-Geoffroy '05)
- several conceptual algorithms (Klatte-Kummer '07)
- errors bounds and descent methods (Luo-Tseng '93)
- more ?...

Summary

- $\longrightarrow$ Nonconvex projections are tractable in some usual situations
- $\longrightarrow$ Alternating nonconvex projections is a tempting natural heuristic, often converges linearly, and is thus popular !
- $\longrightarrow$ The linear rate reflects the "condition number"

distance ill-posedness $\leftrightarrow$ error bound $\leftrightarrow$ rate

Summary

$\longrightarrow$ Nonconvex projections are tractable in some usual situations

$\longrightarrow$ Alternating nonconvex projections is a tempting natural heuristic, often converges linearly, and is thus popular !

$\longrightarrow$ The linear rate reflects the "condition number"

distance ill-posedness $\leftrightarrow$ error bound $\leftrightarrow$ rate

A. Lewis and J. Malick
Alternating projections on manifolds
Mathematics of Operations Research, 2007

A. Lewis, R. Luke, and J. Malick
Local convergence of nonconvex averaged and alternating projections
Foundations of Computational Mathematics, 2008

Summary

$\longrightarrow$ Nonconvex projections are tractable in some usual situations

$\longrightarrow$ Alternating nonconvex projections is a tempting natural heuristic, often converges linearly, and is thus popular !

$\longrightarrow$ The linear rate reflects the "condition number"

distance ill-posedness $\leftrightarrow$ error bound $\leftrightarrow$ rate

A. Lewis and J. Malick
Alternating projections on manifolds
Mathematics of Operations Research, 2007

A. Lewis, R. Luke, and J. Malick
Local convergence of nonconvex averaged and alternating projections
Foundations of Computational Mathematics, 2008

thanks !