NONCONVEX NOTIONS OF REGULARITY AND SPARSE AFFINE FEASIBILITY: LOCAL TO GLOBAL CONVERGENCE

Russell Luke
(Heinz Bauschke, Robert Hesse, Patrick Neumann, Hung Phan, and Shawn Wang)

Universität Göttingen

FGMIA, January 15, 2014

Prelude

$$(\mathcal{P}) \qquad \underset{x\in\Omega}{\text{minimize}}\ f_0(x) \qquad \Longleftrightarrow \qquad \underset{x\in\mathbb{E}}{\text{minimize}}\ f(x) := f_0(x) + \iota_\Omega(x)$$

where

$$\iota_\Omega(x) = \begin{cases} 0 & \text{if } x \in \Omega \\ +\infty & \text{else.} \end{cases}$$

Nonsmooth Fermat's rule

If a proper, subdifferentially regular function $f : \mathbb{E} \to (-\infty, +\infty]$ has a local minimum at $\overline{x}$, then

$$0 \in \partial f(\overline{x}) := \{ v \in \mathbb{E} \mid \langle v,\ x - \overline{x}\rangle \le f(x) - f(\overline{x}) + o(|x - \overline{x}|) \}$$

If f is convex, this is also sufficient for global optimality.

Prelude

Optimization with Sparsity

Given a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ full-rank with $0 < m < n$, solve

$$(\mathcal{P}_{\ell_0}) \qquad \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|x\|_0 \qquad \text{subject to} \quad Ax = b$$

where $\|x\|_0 := \sum_j |\text{sign}(x_j)|$ with $\text{sign}\,(0) := 0$.

Every point is a critical point:

$$0 \in \partial \|x\|_0 \ \forall x \in \mathbb{R}^n$$

So first order optimality conditions are not informative.

Prelude

Optimization with Sparsity (Candes-Tao, '05)

If there exists $0 \leq \delta \leq 1$ such that

$$(RIP) \qquad (1-\delta)\left\|x\right\|^2 \leq \left\|Ax\right\|^2 \leq (1+\delta)\left\|x\right\|^2 \quad \forall x \in S$$

then

$$\operatorname{argmin}_{\{Ax=b\}} \|x\|_0 = \operatorname{argmin}_{\{Ax=b\}} \|x\|_1 = \{\overline{x}\},$$

that is

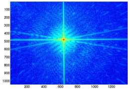
$$\operatorname{argmin}_{\{Ax=b\}} \|x\|_0 = \left\{x \;\middle|\; 0 \in \partial\|x\|_1 + A^* N_{\{b\}}(Ax)\right\} = \{\overline{x}\}$$

Prelude

Phase retrieval

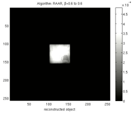
Given

- $|(Ax)_j|^2 = b_j$ for $b_j \in \mathbb{R}_+$ $(j = 1, 2, \ldots, m)$ given by



- some qualitative constraint ($x \in \mathbb{R}^n{}_+$ or supp $x \subset D$ (or $|x_j| = 1$).)

Find



(L. 2012)

Prelude

Conic programming:

$$|(Ax)_j|^2 = \textbf{Tr}\,(|\langle a_{j\cdot},\ x\rangle|^2) = \textbf{Tr}\,(x^* a^*_{\cdot,j} a_{j\cdot} x) = \textbf{Tr}\,(a^*_{\cdot,j} a_{j\cdot} x x^*)$$

so

$$x \in \left\{ x \in \mathbb{R}^{2n} \,\middle|\, |(Ax)_j|^2 = b_j, j = 1, 2, \ldots, m \right\}$$
$$\Longleftrightarrow$$
$$X \in \left\{ X \in \mathbb{R}^{2n \times 2n} \,\middle|\, \mathcal{A}_j X = b_j, j = 1, 2, \ldots, m \right\} \cap \{ X \mid \mathrm{rank}(X) = 1 \}$$

Prelude

Phase-Lift (Candés, Eldar, Strohmer, Voroninski - 2013)

For some convex qualitative constraint $O_1 \subset \mathbb{R}^{2n\times 2n}$ solve

$$\text{argmin}_{\{\mathcal{A}X=B\}\cap O_1} \operatorname{rank}(X) \stackrel{?}{=} \text{argmin}_{\{\mathcal{A}X=B\}\cap O_1} \|X\|_1 \subset \mathbb{R}^{2n\times 2n}$$

Conventional formulation

$$\underset{x\in\mathbb{R}^{2n}}{\text{minimize}}\ f_1(x) + f_2(x)$$

where

$$f_j = \iota_{\Omega_j} \text{ or } \tfrac{1}{2}\text{dist}(x,\Omega_j)^2 \ (j = 1, 2)$$

for $\Omega_1 = \{\text{qualitative constraint}\}$ and
$\Omega_2 := \left\{x \in \mathbb{R}^{2n} \;\middle|\; |(Ax)_j| = b_j, j = 1, \ldots, m\right\}.$

Prelude

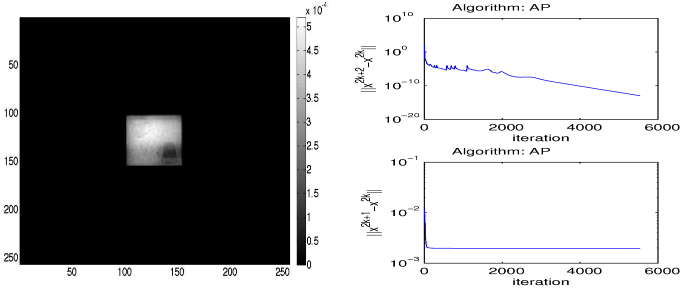
The generalized Fermat's Rule conventional formulation $\Longrightarrow$

$$\text{solve} \quad 0 \in \partial f_1(x) + \partial f_2(x) = \begin{cases} (\mathrm{Id} - P_{\Omega_1})(x) + (\mathrm{Id} - P_{\Omega_2})(x) & \text{if } f_j = \frac{1}{2}\mathrm{dist}(x, \Omega_j)^2 \\ N_{\Omega_1}(x) + N_{\Omega_2}(x) & \text{if } f_j = \iota_{\Omega_j}. \end{cases}$$

Prelude

Solve the phase problem via the *method of alternating projections (AP):*

$$x^{2k+1} = P_{\Omega_1} x^{2k}, \quad x^{2k} = P_{\Omega_2} x^{2k-1}.$$



(L. 2012)

CPU times on the order of seconds. Convergence theory...

Outline

Outline

Goals

$$(\mathcal{P}) \qquad \underset{x\in\mathbb{E}}{\text{minimize}}\ f(x) := f_1(x) + f_2(x)$$

- #1. Nonsmooth nonconvex optimality criteria

If f_1 and f_2 satisfy **?** then $0 \in \partial f(\overline{x})$ is sufficient for global optimality of $\overline{x}$.

- #2. Convergence (with rates and radii) of nonmonotone fixed point iterations

Algorithm $x^{k+1} \in T_\square x^k$ converges to Fix $T_\square$ and there is some mapping Π such that $\Pi(\text{Fix } T_\square) \subset \partial f$.

- #3. Some useful/practical results along the way.

Outline

Basics

Building blocks

- Prox operator: for a regular function $f : X \to \overline{\mathbb{R}}$, define

$$\mathrm{prox}_{\eta f}(x) := \mathrm{argmin}_{\,y}\, f(y) + \frac{1}{2\eta}\|y - x\|^2$$

- Proximal reflector: $R_{\eta f} := 2\,\mathrm{prox}_{\eta f} - \mathrm{Id}$
- Projector: if $f = \iota_\Omega$ for $\Omega \subset X$ closed and nonempty, then $\mathrm{prox}_{\eta f}(\overline{x}) = P_\Omega \overline{x}$ where

$$\begin{aligned} P_\Omega x &:= \{\overline{x} \in \Omega \mid \|x - \overline{x}\| = \mathrm{dist}(x, \Omega)\} \\ \mathrm{dist}(x, \Omega) &:= \inf_{y \in \Omega} \|x - y\|. \end{aligned}$$

- Reflector: if $f = \iota_\Omega$ for some closed, nonempty set $\Omega \subset X$, then $R_\Omega := 2P_\Omega - \mathrm{Id}$

General algorithms

Fixed point iterations

Given $x_0 \in X$ generate the sequence $(x_n)_{n\in\mathbb{N}}$ by

$$x_{n+1} \in (1-\lambda_n)x_n + \lambda_n\left(Tx_n + \epsilon_n\right)$$

Examples: $\lambda_n = 1$ for all $n \in \mathbb{N}$ and define $f(x) := \alpha\varphi(x) + \mathcal{F}_b \circ \mathcal{A}(x)$,

- $T_{AProx} = \text{prox}_{\eta_1 f_1} \circ \text{prox}_{\eta_2 f_2}$
  (Alternating prox)
- $T_{FB} = \text{prox}_{\eta_1 f_1}\left(Id - \eta_2 \partial f_2\right)$
  (proximal gradients/forward-backward)
- $T_{DR\lambda} = \frac{\lambda}{2}\left(R_{\eta_1 f_1} R_{\eta_2 f_2} + Id\right) + (1-\lambda)\,\text{prox}_{\eta_2 f_2}$
  (Relaxed Averaged Alternating Proximal operators)

Global convergence - Optimization with sparsity

Scaled restricted Isometry (Blumensath-Davies, 2009; Beck-Teboulle 2011)

Let $\mathbb{E}$ and $\mathbb{Y}$ be Euclidean spaces with $\dim(\mathbb{E}) = n > m = \dim(\mathbb{Y})$. Define the set $S_s := \{x \in \mathbb{E} \mid \|x\|_0 \leq s\}$ where $\|x\|_0$ is either the rank function or the function that counts the number of nonzero elements in x. The mapping $\mathcal{A} : \mathbb{E} \to \mathbb{Y}$ satisfies the scaled restricted isometry property (SRIP) of order (s, α), if for $\alpha > 1$ there exist $\nu_s, \mu_s > 0$ with $1 \leq \frac{\mu_s}{\nu_s} < \alpha$ such that

$$(SRIP) \qquad \nu_s \|x\|^2 \leq \|\mathcal{A}x\|_2^2 \leq \mu_s \|x\|^2 \quad \forall x \in S_s.$$

Global convergence - Optimization with sparsity

$$\text{find} \quad \overline{x} \in \text{argmin}_{S_s} \ \|\mathcal{A}x - b\|^2 . \qquad (1)$$

Forward-Backward: Sparse Projected Gradient/Iterative Hard Thresholding

Given a set $S \subset \mathbb{E}$, a continuously differentiable function $f : \mathbb{E} \to \mathbb{R}$ and a positive real number τ, we call the mapping

$$T_{PG}(x;\tau) = P_{S_s}\left(x - \frac{1}{\tau}\nabla f(x)\right)$$

the projected gradient operator for the problem (1). We call PG the iteration: given x_0

$$x^{k+1} \in T_{PG}(x^k;\tau_k) = P_{S_s}\left(x^k - \frac{1}{\tau_k}\nabla f(x^k)\right), \ k = 0, 1, 2, \ldots$$

Global convergence - Optimization with sparsity

Global convergence of PG (Blumensath-Davies, 2009; Beck-Teboulle, 2011)

If $\mathcal{A}$ satisfies SRIP of order $(2s, 2)$, then the function values of PG with $f(x) = \frac{1}{2}\|\mathcal{A}x - b\|_2^2$ converge linearly to the globally optimal value of (1) with linear rate constant given by

$$\left\|\mathcal{A}x^{k+1} - b\right\|_2^2 \leq \left(\frac{\tau_k}{\nu_{2s}} - 1\right)\left\|\mathcal{A}x^k - b\right\|_2^2$$

where $\tau_k \in [\mu_{2s}, 2\nu_{2s})$.

Corollary: global convergence of AP (Hesse-L.-Neumann, submitted)

Let the matrix $\mathcal{A}$ satisfy SRIP of order $(2s, 2)$ with $\mu_{2s} = 1$ and $\mathcal{A}\mathcal{A}^\top = \text{Id}$. The function values of the iterates of the AP algorithm converge linearly to zero for every initial point x^0.

Global convergence - Optimization with sparsity

Define $\Omega := \{x \in \mathbb{E} \mid \mathcal{A}x = b\}$.

$$\text{find} \quad \overline{x} \in \operatorname{argmin}_{S_s} \frac{1}{2}\operatorname{dist}(x, \Omega)^2. \tag{2}$$

Relaxed Alternating Projections, *$AP\lambda$*

For two sets $\Omega, S_s \subset \mathbb{E}$ we call the mapping

$$T_{AP\lambda}x := P_{S_s}(x - \lambda(I - P_\Omega)x)$$

the relaxed alternating projections ($AP\lambda$) operator for the problem (2). The $AP\lambda$ algorithm is the iteration: given x^0, for $k = 0, 1, 2, \ldots$ compute $x^{k+1} \in T_{AP_\lambda}x^k$.

Global convergence - Optimization with sparsity

Global convergence of AP$_\lambda$ (Hesse-L.-Neumann, submitted)

Suppose that $S_1 \cap \Omega \neq \emptyset$ and that the matrix $\mathcal{A}$ in the definition of the set Ω satisfies

$$(RIP^\dagger) \qquad (1-\delta_{2s})\|v\|^2 \leq \|\mathcal{A}^\dagger \mathcal{A} v\|^2 \leq (1+\delta_{2s})\|v\|^2$$

with $\delta_{2s} < 0.453975$ for all vectors v of sparsity $2s$. Then there exists a relaxation parameter $\lambda > 0$ such that the sequence $(x^k)_{k\in\mathbb{N}}$ generated by AP_λ for any initial value $x^0 \in \mathbb{E}$ converges to $S_1 \cap \Omega$ with at least linear rate $\rho(\delta_{2s}, \lambda)$ defined by

$$\rho(\delta_{2s}, \lambda) := \left(\frac{1}{(1-\delta_{2s})} + \lambda(\lambda - 2)\right) < 1.$$

Application: phase retrieval

Recall

Phase-Lift

For some convex qualitative constraint $O_1 \subset \mathbb{R}^{2n\times 2n}$ solve

$$\text{argmin}_{\{\mathcal{A}X=B\}\cap O_1} \operatorname{rank}(X) \stackrel{?}{=} \text{argmin}_{\{\mathcal{A}X=B\}\cap O_1} \|X\|_1 \subset \mathbb{R}^{2n\times 2n}$$

Reformulate this on $\mathcal{S}^{2n\times 2n}$ as

$$(P_{PL}) \qquad \text{find} \quad \overline{X} \in \text{argmin}_{O_1\cap S_1} \tfrac{1}{2}\text{dist}(X,\Omega)^2$$

where $\Omega := \{X \mid \mathcal{A}X = B\}$. If O_1 is a subspace then we can absorb this constraint into the affine set Ω and we have

$$(P'_{PL}) \qquad \text{find} \quad \overline{X} \in \text{argmin}_{S_1\subset\mathcal{S}^{2n\times 2n}} \tfrac{1}{2}\text{dist}(X,\Omega)^2.$$

Application: phase retrieval

$$(P'_{PL}) \qquad \text{find} \quad \overline{X} \in \operatorname{argmin}_{S_1} \tfrac{1}{2}\mathrm{dist}(X,\Omega)^2 \subset \mathcal{S}^{2n\times 2n}$$

Local solutions $\overline{X}$ to (P'_{PL}) satisfy $0 \in (\mathrm{Id} - P_\Omega)\overline{X} + N_{S_1}(\overline{X})$.

If, in addition, $S_1 \cap \Omega \neq \emptyset$ and $\mathcal{A}$ in the definition of the set Ω satisfies

$$(RIP^\dagger) \qquad (1-\delta_2)\|v\|^2 \leq \|\mathcal{A}^\dagger\mathcal{A}v\|^2 \leq (1+\delta_2)\|v\|^2$$

with $\delta_2 < 0.453975$ for all vectors v of sparsity 2, then the subdifferential inclusion is sufficient for global optimality.

Moreover,

$$X^k \to S_1 \cap \Omega$$

where $X^{k+1} \in T_{AP\lambda}X^k$ for any X^0 and λ appropriately chosen.

Application: phase retrieval

The global convergence result for either AP or AP_λ applies immediately to (P'_{PL}), as long as $S_1 \cap \Omega \neq \emptyset$ and the $SRIP/RIP^\dagger$ condition can be verified to hold.

No need for the convex relaxation. Is there a need for lifting?

Application: phase retrieval

Let $\Omega' := \left\{ x \in \mathbb{R}^{2n} \;\middle|\; |(Ax)_j|^2 = b_j, j = 1, \ldots, m' \right\}$.

$$\begin{array}{ccc}
\operatorname{argmin}_{S_1 \subset \mathcal{S}^{2n \times 2n}} \frac{1}{2}\operatorname{dist}(X, \Omega)^2 & \leftrightarrow & \operatorname{argmin}_{\mathbb{R}^{2n}} \frac{1}{2}\operatorname{dist}(x, \Omega')^2 \\
\updownarrow & & \uparrow \\
X^{k+1} \in T_{AP\lambda} X^k \subset \mathcal{S}^{2n \times 2n} & \leftrightarrow & x^{k+1} \in T_{\Box} x^k \subset \mathbb{R}^{2n}
\end{array}$$

In other words, there is a nonconvex algorithm on $\mathbb{R}^n$ that is guaranteed to converge globally linearly to a globally optimal solution $\implies$ no need for lifting.

Application: phase retrieval

To do:

- What are weaker conditions to guarantee that solutions to the subdifferential inclusion are global solutions to the nonconvex problem?
- Given your favorite algorithm $T_{\square}$ for phase retrieval in $\mathbb{R}^{2n}$ what is the corresponding algorithm in the lifted space?
- How to guarantee that the lifted analog converges (globally) points from which solutions to the subdifferential inclusion can be computed easily?
- What if the set Ω in (PL') is a cone or polyhedral instead of a subspace?
- If $S_1 \cap \Omega = \emptyset$?

Outline

Local convergence analysis

(C,ϵ)-(firmly-)nonexpansive mappings (Hesse-L. 2013)

Let $C \subset D \subset \mathbb{E}$ be nonempty and let T be a (multi-valued) mapping from D to $\mathbb{E}$.

i) T is called (C,ε)-nonexpansive on D if

$$\begin{aligned} &\|x_+ - \overline{x}_+\| \le \sqrt{1+\varepsilon}\,\|x - \overline{x}\| \\ &\forall x \in D,\ \forall \overline{x} \in C,\ \forall x_+ \in Tx,\ \forall \overline{x}_+ \in T\overline{x}. \end{aligned} \tag{3}$$

If (3) holds with $\epsilon = 0$ then we say that T is C-nonexpansive on D.

Local convergence analysis

(C,ϵ)-(firmly-)nonexpansive mappings (Hesse-L. 2013)

ii) T is called (C,ε)-firmly nonexpansive on D if

$$\|x_+ - \overline{x}_+\|^2 + \|(x - x_+) - (\overline{x} - \overline{x}_+)\|^2 \le (1+\varepsilon)\,\|x - \overline{x}\|^2$$
$$\forall x \in D,\ \forall \overline{x} \in C,\ \forall x_+ \in Tx,\ \forall \overline{x}_+ \in T\overline{x}. \tag{4}$$

If (4) holds with $\epsilon = 0$ then we say that T is C-firmly nonexpansive on D.

□

Special case: $C = \text{Fix } T$, mappings satisfying (3) or (4) are called quasi-(firmly-)nonexpansive.

The classical (firmly) nonexpansive operator on D is $(D,0)$-(firmly) nonexpansive on D.

Local convergence analysis

Noncontractive fixed point iterations (Hesse-L., 2013)

Let $D \subset \mathbb{E}$, $T : D \rightrightarrows \mathbb{E}$, $C \subset \text{Fix } T$ and $U \subset D$. If

(a) T is (C, ε)-firmly nonexpansive on U and

(b) for some $\lambda > 0$, T satisfies the coercivity condition

$$\|x - x_+\| \geq \lambda \text{dist}(x, C) \quad \forall\, x_+ \in Tx, \forall x \in U, \tag{5}$$

then

$$\text{dist}(x_+, C) \leq \sqrt{(1 + \varepsilon - \lambda^2)}\text{dist}(x, C) \quad \forall\, x_+ \in Tx, \forall x \in U. \tag{6}$$

General Notions

Proof. Choose any $x_+ \in Tx$ and $\overline{x}_+ \in T\overline{x}$. Combine inequalities (5) and (4) to get

$$\left\|x_+ - \overline{x}_+\right\|^2 + \left(\lambda \left\|x - \overline{x}\right\|\right)^2 \leq \tag{7}$$

$$\left\|x_+ - \overline{x}_+\right\|^2 + \left\|x - x_+ - (\overline{x} - \overline{x}_+)\right\|^2 \leq (1 + \varepsilon) \left\|x - \overline{x}\right\|^2 \tag{8}$$

for all $x \in U$, which immediately yields (6). □

Convergence of AP/DR: ingredient #1, set regularity

(ε, δ)-(sub)regularity of underlying sets $\Longrightarrow$ regularity of the corresponding fixed-point operator

i) A nonempty set $\Omega \subset \mathbb{E}$ is (ε, δ)-subregular at $\overline{x}$ with respect to $C \subset \mathbb{E}$, if there exists $\varepsilon > 0, \delta > 0$, and

$$\langle v, z - y \rangle \leq \varepsilon \, \|v\| \, \|z - y\| \tag{9}$$

holds for all $y \in \mathbb{B}_\delta(\overline{x}) \cap \Omega$, $z \in C \cap \mathbb{B}_\delta(\overline{x})$, $v \in N_\Omega(y)$. We simply say Ω is (ε, δ)-subregular at $\overline{x}$ if $C = \{\overline{x}\}$.

ii) If $C = \Omega$ in i) then we say that the set Ω is (ε, δ)-regular at $\overline{x}$.

iii) If for all $\epsilon > 0$ there exists a $\delta > 0$ such that (9) holds for all $y, z \in \mathbb{B}_\delta(\overline{x}) \cap \Omega$ and $v \in N_\Omega(y)$, then Ω is said to be super-regular.

Convergence of AP/DR: ingredient #1, set regularity

- (ε, δ)-regularity was introduced in (Bauschke-L.-Phan-Wang, 2012).
- super-regularity was introduced in (Lewis-L.Malick, 2009)

Relations

- Prox-regularity $\implies$ super-regularity $\implies$ Clarke-regularity.
- super-regularity $\implies$ (ε, δ)-regularity
- (ε, δ)-regularity $\not\Rightarrow$ Clarke regularity
- $(0, \delta)$-regularity is local convexity

First Results

Projectors and reflectors onto (ε, δ)-subregular sets

Let $\Omega \subset \mathbb{E}$ be nonempty closed and (ε, δ)-subregular at each point $\overline{x} \in C \subset \Omega$.

(i) The projector is $(C, \tilde{\varepsilon}_1)$-nonexpansive on $\mathbb{B}_\delta(C)$ where $\tilde{\varepsilon}_1 := 2\varepsilon + \varepsilon^2$.

(ii) The projector is $(C, \tilde{\varepsilon}_2)$-firmly nonexpansive on $\mathbb{B}_\delta(C)$, where $\tilde{\varepsilon}_2 := 2\varepsilon + 2\varepsilon^2$.

(iii) The reflector R_Ω is $(C, \tilde{\varepsilon}_3)$-nonexpansive on $\mathbb{B}_\delta(C)$, where $\tilde{\varepsilon}_3 := 4\varepsilon + 4\varepsilon^2$.

Convergence of AP/DR: ingredient #2, regularity of the intersection

Linear regularity

A collection of closed, nonempty sets $\Omega_1, \Omega_2, \ldots, \Omega_m$ is locally linearly regular at $\overline{x} \in \cap_{j=1}^m \Omega_j$ if there exists a $\kappa > 0$ and a $\delta > 0$ such that

$$\text{dist}\left(x, \cap_{j=1}^m \Omega_j\right) \le \kappa \max_{i=1,\ldots,m} \text{dist}\,(x, \Omega_i)\,, \quad \forall x \in \mathbb{B}_\delta(\overline{x}). \tag{10}$$

If (10) holds for any $\delta > 0$ the collection is called linearly regular. The infimum over all κ such that (10) holds is called regularity modulus. □

What we call local linear regularity at $\overline{x}$ has appeared in various forms elsewhere. See for Ioffe (2000), Ngai-Thera (2001), and Kruger (2006). This is a localization of (bounded) linear regularity defined in Bauschke-Borwein (1996).

Outline

Results: AP

Local Linear convergence of AP (Hesse-L., 2013)

Let A, B be closed nonempty subsets of $\mathbb{E}$ that are locally linearly regular with modulus κ at $\overline{x} \in C := A \cap B$. For any $x_0 \in \mathbb{B}_\delta(\overline{x})$, generate the sequence $\{x_n\}_{n\in\mathbb{N}}$ by
$x_{2n+1} \in P_A x_{2n}$ and $x_{2n+2} \in P_B x_{2n+1} \quad (\forall n = 0, 1, 2, \dots)$.

(a) If A and B are $(\varepsilon, \delta)-$subregular w.r.t. C, then

$$\text{dist}\,(x_{2n+2}, C) \leq (1 - \tfrac{1}{\kappa^2} + \varepsilon)\text{dist}\,(x_{2n}, C)\,.$$

(b) If A is $(\varepsilon, \delta)-$subregular w.r.t. C and B is convex, then

$$\text{dist}\,(x_{2n+2}, C) \leq \sqrt{1 - \tfrac{1}{\kappa^2} + \varepsilon}\sqrt{1 - \tfrac{1}{\kappa^2}}\text{dist}\,(x_{2n}, C)\,.$$

(c) If A and B are convex, then

$$\text{dist}\,(x_{2n+2}, C) \leq (1 - \tfrac{1}{\kappa^2})\text{dist}\,(x_{2n}, C)\,.$$

Results: DR

Recall

$$T_{DR\lambda} = \frac{\lambda}{2}(R_{\Omega_1}R_{\Omega_2} + Id) + (1-\lambda)P_{\Omega_2}.$$

Denote $T_{DR} := T_{DR1}$.

$(C, \tilde{\varepsilon})$-firm nonexpansiveness of T_{DR}

Let $A, B \subset \mathbb{E}$ be closed and nonempty. Let A and B be (ε_A, δ)- and (ε_B, δ)-subregular respectively at each $\overline{x} \in C \subset A \cap B$. The DR operator $T_{DR} : \mathbb{E} \rightrightarrows \mathbb{E}$ is $(C, \tilde{\varepsilon})$-firm nonexpansive on $\mathbb{B}_\delta(C)$ where

$$\tilde{\varepsilon} = 2\varepsilon_A(1+\varepsilon_A) + 2\varepsilon_B(1+\varepsilon_B) + 8\varepsilon_A(1+\varepsilon_A)\varepsilon_B(1+\varepsilon_B).$$

Results: Douglas-Rachford

Douglas-Rachford for an affine subspace and a superregular set (Hesse-L., 2013)

Assume $B \subset \mathbb{E}$ is a subspace and that $A \subset \mathbb{E}$ is closed and $(\epsilon, \overline{\delta})$-super-regular at $\overline{x} \in C := A \cap B$. If the collection $\{A, B\}$ is linearly regular on C with modulus κ, then there is a $\delta > 0$ with $\delta \leq \overline{\delta}$ and a $c \in [0, 1)$ such that, $\frac{(1-c)}{\kappa^2} > 2\varepsilon + 2\varepsilon^2$ and hence

$$\operatorname{dist}(x_+, C) \leq \tilde{c} \operatorname{dist}(x, C) \quad \forall\, x_+ \in T_{DR}x, \tag{11}$$

with $\tilde{c} = \sqrt{1 + 2\varepsilon + 2\varepsilon^2 - \frac{(1-c)}{\kappa^2}} < 1$ for all $x \in \left(P_C^{-1}\overline{x}\right) \cap \mathbb{B}_{\frac{\delta}{1+\epsilon}}(\overline{x})$.

Application to sparse affine feasibility

Local linear convergence of AP/DR: affine sparse feasibility (Hesse-L.-Neumann, 2013)

Let $\Omega := \{x \in \mathbb{R}^n \mid Ax = b\}$ and $S_s = \{x \in \mathbb{R}^n \mid \|x\|_0 \le s\}$ with nonempty intersection and let $\overline{x} \in \Omega \cap S_s$. Choose $0 < \delta < \min\left\{|\overline{x}_j| \;\middle|\; j \in I(\overline{x})\right\}$ and $x^0 \in \mathbb{B}_{\delta/2}(\overline{x})$.

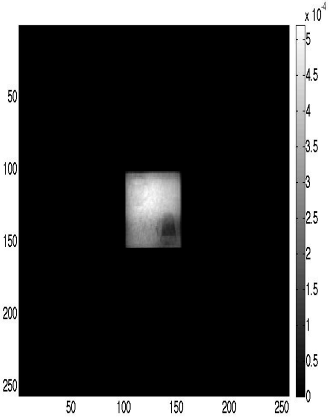
- The AP iterates converge linearly to the intersection $\Omega \cap S_s$ with rate $\left(1 - \frac{1}{\kappa^2}\right)$ where κ is the modulus of regularity of the intersection.
- If $\|\overline{x}\|_0 = s$, then the DR iterates converge linearly to $\mathrm{Fix}\ T_{DR}$. Moreover, for any $\hat{x} \in \mathrm{Fix}\ T_{DR} \cap \mathbb{B}_{\delta/2}(\overline{x})$, we have $P_\Omega \hat{x} \in S_s \cap \Omega$.

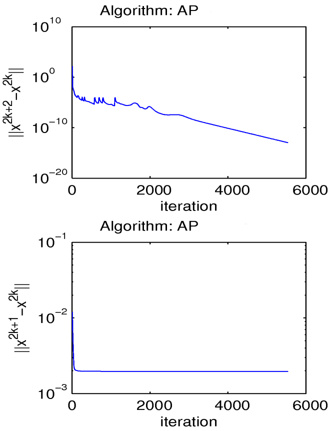
Application to sparse affine feasibility

- *For AP, if there is a solution $x \in \Omega \cap S_s$, then $\|\overline{x}\|_0$ can be smaller than s.*
- *If Ω is a subspace, the point 0 is trivially a solution. The set S_s is not convex on any neighborhood of 0, however the assumptions of the theorem hold, and AP indeed converges locally linearly to 0, regardless of the size of the parameter s.*
- *DR does not appear to be sensitive to the parameter s in practice, but our proof does not establish this.*

Application: phase retrieval

Recall





Linear convergence to fattened sets I

Approximately noncontractive fixed point iterations (L., 2013)

Let $D \subset \mathbb{E}$, $T : D \rightrightarrows \mathbb{E}$ and $S \subset \operatorname{Fix} T \subset \operatorname{int} D$. For $0 < \delta < \overline{\delta}$ fixed, define $S_\delta := \delta\mathbb{B} + S$ where $\overline{\delta}$ is such that $S_{\overline{\delta}} \subset D$. If

(a') T is $(S, \varepsilon_{\overline{\delta}})$-firmly nonexpansive on $S_{\overline{\delta}}$ and

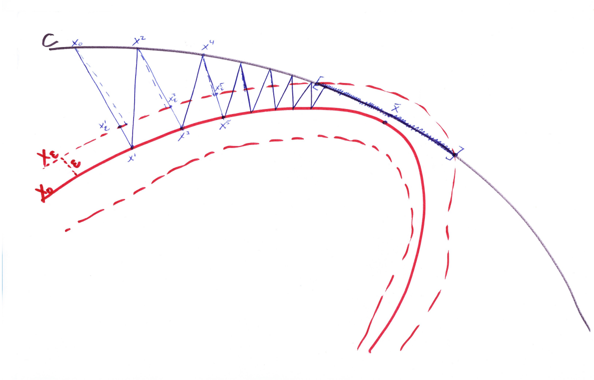
(b') there exists a λ_δ such that T satisfies

$$\|x - x_+\| \geq \lambda_\delta \operatorname{dist}(x, S) \quad \forall\, x_+ \in Tx,\ \forall\, x \in S_{\overline{\delta}} \setminus S_\delta, \tag{12}$$

then T is $(\{\overline{x}\}, \varepsilon_{\overline{\delta}} - \lambda_\delta^2)$-nonexpansive on $S_{\overline{\delta}} \setminus S_\delta$: for all $x_+ \in Tx$, for all $x \in S_{\overline{\delta}} \setminus S_\delta$,

$$\operatorname{dist}(x_+, S) \leq \sqrt{(1 + \varepsilon_{\overline{\delta}} - \lambda_\delta^2)} \operatorname{dist}(x, S). \tag{13}$$

Phat AP



Linear convergence to approximate solutions

L. (2013)

Let $D \subset \mathbb{E}$, $T : D \rightrightarrows \mathbb{E}$ and $S \subset \text{Fix } T$ closed and nonempty. If for all $\overline{\delta} > 0$ small enough there is a $\gamma > 1$ and a triplet

$$(\epsilon, \delta, \lambda) \in \mathbb{R}_+ \times [0, \gamma\overline{\delta}] \times (\sqrt{\epsilon}, \sqrt{1+\epsilon}]$$

such that (a') and (b') are satisfied, then for any $x^{(0)}$ close enough to S the sequence $(x^{(k)})_{k\in\mathbb{N}}$ defined by $x^{(k+1)} \in T(x^k)$ converges to S (and in particular converges in finitely many steps to $S_{\delta'}$ for any fixed $\delta' \in (0, \text{dist}(x^{(0)}, S)]$).

Thanks for your attention.

Outline

References I

F. A. Artacho and J. Borwein.
Global convergence of a nonconvex Douglas-Rachford iteration.
J. of Global Optim., 2012.

H. H. Bauschke and J. M. Borwein.
On projection algorithms for solving convex feasibility problems.
SIAM Rev., 38(3):367–426, 1996.

H.H. Bauschke, D.R. Luke, H.M. Phan, and X. Wang,
Restricted normal cones and the method of alternating projections: theory,
Set-Valued and Variational Analysis (2013)

References II

H.H. Bauschke, D.R. Luke, H.M. Phan, and X. Wang,
Restricted normal cones and the method of alternating projections: applications,
Set-Valued and Variational Analysis (2013)

H.H. Bauschke, D.R. Luke, H.M. Phan, and X. Wang,
Restricted normal cones and sparsity optimization with affine constraints,
Foundations of Computational Mathematics (2013)

J. M. Borwein and B. Sims.
The Douglas-Rachford algorithm in the absence of convexity.
Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer Optimization and its Applications, 49:93–109, 2011.

References III

- F. Deutsch and H. Hundal.
  The rate of convergence for the cyclic projections algorithm III: Regularity of convex sets.
  J. Approx. Theory, 155(1):155–184, 2008.
- R. Hesse and D. R. Luke,
  Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems,
  SIAM J. Optimization, 23(4):2397–2419 (2013).
- R. Hesse, D. R. Luke and P. Neumann
  Projection Methods for Sparse Affine Feasibility: Results and Counterexamples,
  submitted.

References IV

- A. Y. Kruger.
  Weak stationarity: Eliminating the gap between necessary and sufficient conditions.
  Optimization, 53:147–164, 2004.
- A.S. Lewis, D.R. Luke, and J. Malick,
  Local linear convergence for alternating and averaged nonconvex projections,
  Foundations of Computational Mathematics 9 (2009), 485–513.
- A.S. Lewis and J. Malick,
  Alternating projection on manifolds,
  Mathematics of Operations Research 33 (2008), 216–234.

References V

B. Mordukhovich.
Variational Analysis and Generalized Differentiation, I: Basic Theory; II: Applications.
Grundlehren der mathematischen Wissenschaften.
Springer-Verlag, New York, 2006.