ON LOCAL LINEAR CONVERGENCE OF ELEMENTARY ALGORITHMS WITH SPARSITY CONSTRAINTS

Russell Luke

Universität Göttingen

SIAM IS12, May 20, 2012

Outline

Central Problem: ℓ_0 minimization

Let $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^r$ (not necessarily linear). Solve

$$(\mathcal{P}_0) \qquad \begin{array}{ll} \underset{x\in\mathbb{R}^{m\times n}}{\text{minimize}} & \mathrm{rank}(x) \\ \text{subject to} & \mathcal{A}(x) = b \end{array}$$

where $\mathrm{rank}(x) := \|\sigma(x)\|_0 := \sum_j \mathrm{sign}(\sigma(x)_j)$ with $\mathrm{sign}\,(0) := 0$ for $\sigma(x)$ the vector of singular values of x.

Combinatorial optimization problem $\implies$ NP-complete (Natarajan, 1995).

Central Problem: ℓ_0 minimization

Relaxations

For $0 < p \leq 1$ solve instead

$$(\mathcal{P}_p) \qquad \begin{array}{ll} \underset{x \in \mathbb{R}^{m \times n}}{\text{minimize}} & \|\sigma(x)\|_p \\ \text{subject to} & \mathcal{A}(x) = b \end{array}$$

where $\|x\|_p := \left(\sum_j \sigma(x)_j^p\right)^{1/p}$.
(For $p = 1$ and $m > 1$, $\|x\|_1$ is the *nuclear norm*.)

- $m = 1$, $p = 1$, $\mathcal{A}$ linear and satisfies the Restricted Isometry Property (RIP): optimal solutions of $\mathcal{P}_0$ and $\mathcal{P}_1$ coincide (Candés & Tao, 2006-2007).
- $m = 1$, $0 < p < 1$, $\mathcal{A}$ linear and satisfies the Restricted Isometry Property (RIP): optimal solutions of $\mathcal{P}_0$ and $\mathcal{P}_p$ coincide (LaiWaing, 2010).

Central Problem: ℓ_0 minimization

Rewieghted/greedy algorithms

For $m = 1$, $\mathcal{A}$ linear, $p = 1$ and $\alpha > 0$ fixed, solve

$$(\mathcal{P}_{1,L}) \qquad \begin{array}{ll} \underset{(x,L)\in\mathbb{R}^n\times\mathbb{R}^n}{\text{minimize}} & f(x,L) \\ \text{subject to} & Ax = b \end{array}$$

where

$$f(x,L) := \sum_{j=1}^{n} L_j|x_j| + \iota_{\mathbb{R}_+}(L) + \tfrac{1}{2}\|L-\alpha\|^2$$

- Orthogonal Matching Pursuit (OMP) (Chen, Donoho & Saunders, 1998-99)
- steepest subgradient descent algorithm with exact linesearch and dynamic weights applied to the *dual* (Borwein & L., 2011) (recovers OMP as a limiting case).
- see also Fornasier, Rauhut & Ward (2011), Candés, Wakin, & Boyd (2008), ...

Nonconvex Feasibility

$$\text{find} \quad x \in M_\epsilon \cap \mathcal{C}_s$$

where

$$M_\epsilon = \left\{ x \in \mathbb{R}^{m \times n} \mid d_{\mathbb{R}^r}(\mathcal{A}(x), b) \leq \epsilon \right\}$$

for $d_{\mathbb{R}^r} : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}_+$ a *distance function*, and

$$\mathcal{C}_s := \left\{ x \in \mathbb{R}^{m \times n} \mid \operatorname{rank}(x) \leq s \right\}.$$

- Combettes&Trussell (1990)
- Beck&Teboulle (2011).

Outline

Central Algorithm: Method of Alternating Projections (MAP)

Consider more generally the problem

$$\text{find } x \in A \cap B$$

for two nonempty closed subsets A, B of a Hilbert space $\mathcal{H}$.

For a closed nonempty subset $C \subset \mathcal{H}$ define the *projection* of x onto C by

$$P_C : x \mapsto \text{argmin}_{y \in C}\{\|x - y\|\}$$

Method of Alternating Projections

given $x^0 \in \mathcal{H}$ generate the sequence $\{x^k\}_{k=0}^{\infty}$ in $\mathcal{H}$ by

$$b^k {\color{red}\in} P_B a^k, \qquad a^{2k} {\color{red}\in} P_A b^{k-1}.$$

Method of Alternating Projections (MAP)

- If A and B are subspaces the MAP sequence converges strongly to $P_{A\cap B}x^0$ (von Neumann, 1933).
- If A and B are closed and convex and if one of the sets is compact the MAP sequence converges to a fixed point (Cheney-Goldstein 1959).
- If A and B are closed and convex with nonempty intersection the MAP sequence converges weakly to a point $\overline{x} \in A \cap B$ (Bregman, 1965).

Method of Alternating Projections (MAP)

- If the angle between the subspaces A and B is positive then convergence is linear (Aronszajn, 1950).
- If A and B are closed and convex with $\operatorname{int} A \cap B \neq \emptyset$ the MAP sequence converges linearly to a point $\bar{x} \in A \cap B$ (Gubin-Polyak-Raik, 1967).
- If A and B are manifolds in $\mathbb{E}$ with $T_A(\bar{x}) + T_B(\bar{x}) = \mathbb{E}$ then for initial points x^0 close to $\bar{x}$ the MAP sequence converges linearly (Lewis & Mallick, 2008).
- If $A \subset \mathbb{E}$ is closed, $B \subset \mathbb{E}$ is superregular and $N_A(\bar{x}) \cap -N_B(\bar{x}) = \{0\}$ then for initial points x^0 close to $\bar{x}$ the MAP sequence converges linearly (Lewis, L. & Mallick, 2009).

Method of Alternating Projections (MAP)

Local convergence of MAP requires two things:

- regularity of the intersection $A \cap B \implies$ constraint qualifications
- regularity of the sets A and B near the intersection.

Both of these notions require the normal cone.

Outline

Variational Analysis

Restricted normal cones (Bauschke, L., Phan & Wang, 2012)

Let A and B be nonempty subsets of a Euclidean space $\mathbb{E}$, and let a and u be in $\mathbb{E}$. If $a \in A$, we define the restricted normal cones of A at a as follows:

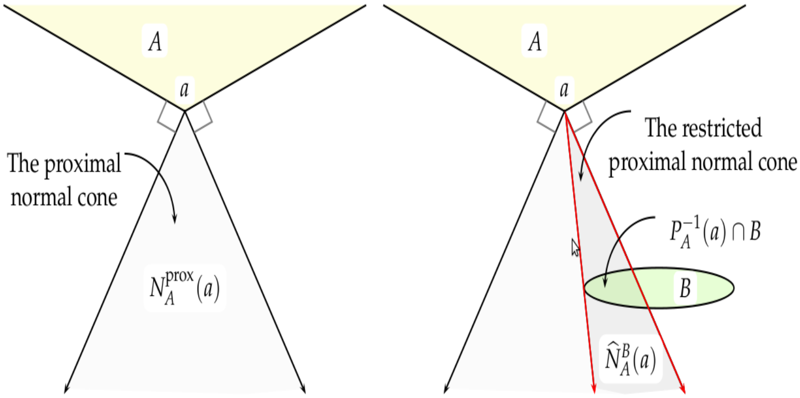
1. The *B-restricted proximal normal cone* of A at a is

$$\widehat{N}_A^B(a) := \operatorname{cone}\Big(\big(B \cap P_A^{-1}a\big) - a\Big) = \operatorname{cone}\Big(\big(B - a\big) \cap \big(P_A^{-1}a - a\big)\Big).$$

2. The *B-restricted normal cone* $N_A^B(a)$ is implicitly defined by $u \in N_A^B(a)$ if and only if there exist sequences $(a_n)_{n\in\mathbb{N}}$ in A and $(u_n)_{n\in\mathbb{N}}$ in $\widehat{N}_A^B(a_n)$ such that $a_n \to a$ and $u_n \to u$.

If $a \notin A$, then all normal cones are defined to be empty.

Variational Analysis



Variational Analysis

Joint constraint qualification number (CQ-number)

Let A, $\widetilde{A}$, B, $\widetilde{B}$, be nonempty subsets of $\mathbb{E}$, let $c \in \mathbb{E}$, and let $\delta \in \mathbb{R}_{++}$. The *CQ-number* at c associated with $(A, \widetilde{A}, B, \widetilde{B})$ and δ is

$$\begin{aligned}\theta_\delta &:= \theta_\delta\big(A, \widetilde{A}, B, \widetilde{B}\big) \\ &:= \sup\left\{\langle u, v\rangle \;\middle|\; \begin{array}{l} u \in \widehat{N}_A^{\widetilde{B}}(a), v \in -\widehat{N}_B^{\widetilde{A}}(b), \|u\| \le 1, \|v\| \le 1, \\ \|a - c\| \le \delta, \|b - c\| \le \delta. \end{array}\right\}.\end{aligned}$$

For nontrivial collections[1] $\mathcal{A} := (A_i)_{i\in I}$, $\widetilde{\mathcal{A}} := (\widetilde{A}_i)_{i\in I}$, $\mathcal{B} := (B_j)_{j\in J}$, $\widetilde{\mathcal{B}} := (\widetilde{B}_j)_{j\in J}$ of nonempty subsets of $\mathbb{E}$, the *joint-CQ-number* at $c \in \mathbb{E}$ associated with $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$ and $\delta > 0$ is

$$\theta_\delta = \theta_\delta\big(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}}\big) := \sup_{(i,j)\in I\times J} \theta_\delta\big(A_i, \widetilde{A}_i, B_j, \widetilde{B}_j\big).$$

[1] The collection $(A_i)_{i\in I}$ is said to be *nontrivial* if $I \neq \varnothing$.

Variational Analysis

The *Friedrichs angle* between subspaces A and B is the number in $[0, \frac{\pi}{2}]$ whose cosine is given by

$$c(A,B) :=$$
$$\sup\left\{ |\langle a,b\rangle| \;\middle|\; a \in A\cap(A\cap B)^{\perp}, b \in B\cap(A\cap B)^{\perp}, \|a\|,\|b\| \leq 1 \right\}$$

CQ-number of two (affine) subspaces and Friedrichs angle (Bauscke, L., Phan & Wang 2012)

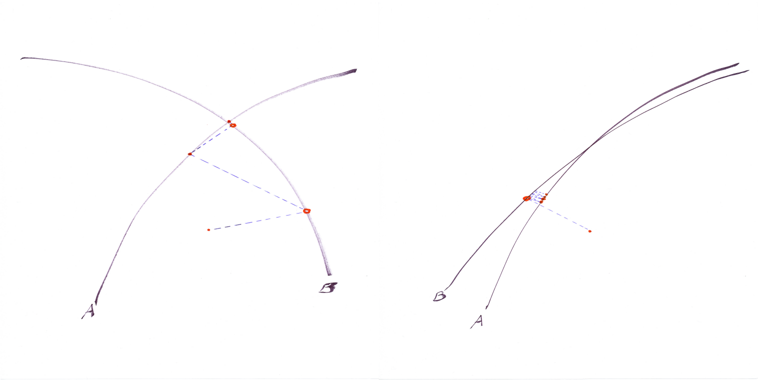
Let A and B be affine subspaces of $\mathbb{E}$, and let $\delta > 0$. Then

$$\theta_\delta(A,A,B,B) = \theta_\delta(A,\mathbb{E},B,B) = \theta_\delta(A,A,B,\mathbb{E}) = c(A,b) < 1. \tag{2}$$

Moreover, if A and B are *affine* subspaces of $\mathbb{E}$ with $\bar{x} \in A\cap B$, and $\delta > 0$, then (2) holds at $\bar{x}$.

Variational Analysis

Angle of intersection



Variational Analysis

special case: two distinct lines through the origin

Suppose that w_a and w_b are two vectors in $\mathbb{E}$ such that $\|w_a\| = \|w_b\| = 1$. Let $A := \mathbb{R}w_a$, $B := \mathbb{R}w_b$, and $\delta > 0$. Assume that $A \cap B = \{0\}$. Then the CQ-number at 0 is

$$\theta_\delta(A, A, B, B) = \theta_\delta(A, \mathbb{E}, B, B) = \theta_\delta(A, A, B, \mathbb{E}) = |\langle w_a, w_b\rangle| < 1.$$

Variational Analysis

Superregularity (Lewis, L. & Mallick, 2009)

We say that C is $(\mathcal{A}, \epsilon, \delta)$*-regular* at $c \in \mathbb{E}$ if $\epsilon \geq 0$, $\delta > 0$, and

$$\left.\begin{array}{c} (y, b) \in B \times B, \\ \|y - c\| \leq \delta, \|b - c\| \leq \delta, \\ u \in \widehat{N}_B^{\mathcal{A}}(b) \end{array}\right\} \quad \Rightarrow \quad \langle u, y - b\rangle \leq \epsilon\|u\| \cdot \|y - b\|. \tag{3}$$

If B is $(\mathbb{E}, \epsilon, \delta)$-regular at c, then we also simply speak of (ϵ, δ)-regularity.

Example: sets for which the projector is single-valued on neighborhoods of the set (prox-regular sets) are (ϵ, δ)-regular. Convex sets are $(0, \infty)$-regular.

Harvest Time

Linear convergence of MAP

Let $\mathcal{A} := (A_i)_{i \in I}$ and $\mathcal{B} := (B_j)_{j \in J}$ be nontrivial collections of nonempty closed subsets of $\mathbb{E}$, and let $\widetilde{\mathcal{A}} := (\widetilde{A}_i)_{i \in I}$ and $\widetilde{\mathcal{B}} := (\widetilde{B}_j)_{j \in J}$ be reasonable collections of subsets of $\mathbb{E}$, $\widetilde{A} := \bigcup_{i \in I} \widetilde{A}_i$ and $\widetilde{B} := \bigcup_{j \in J} \widetilde{B}_j$. Let

$$c \in (\cup_{i \in I}(A_i)) \cap \left(\cup_{j \in J}(B_j)\right)$$

and assume that there exists $\delta > 0$ such that

1. $\mathcal{A}$ is $(\widetilde{B}, 0, 3\delta)$-joint-regular at c;
2. $\mathcal{B}$ is $(\widetilde{A}, 0, 3\delta)$-joint-regular at c; and
3. $\theta < 1$, where $\theta := \theta_{3\delta}$ is the joint-CQ-number at c associated with $(\mathcal{A}, \widetilde{\mathcal{A}}, \mathcal{B}, \widetilde{\mathcal{B}})$

Harvest Time

Linear convergence of MAP

Suppose also that the starting point of the MAP b_{-1} satisfies $\|b_{-1} - c\| \leq \frac{(1-\theta)\delta}{6(2-\theta)}$. Then $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ converge linearly to some point in $\bar{c} \in A \cap B$ with $\|\bar{c} - c\| \leq \delta$ and rate θ^2; in fact,

$$(\forall n \geq 1) \quad \max\left\{\|a_n - \bar{c}\|, \|b_n - \bar{c}\|\right\} \leq \frac{\delta}{2-\theta}\left(\theta^2\right)^{n-1}.$$

Outline

Normal cone to $\mathcal{C}_s$

Recall that $\mathcal{C}_s := \{x \in \mathbb{R}^{m\times n} \,|\, \mathrm{rank}(x) \le s\}$ and define $\mathrm{Supp}\,(x) := \mathrm{range}(x^T)$.

Normal cone to $\mathcal{C}_s$ (L. 2012)

Let $r := \min\{m, n\}$, fix $s \in \{0, 1, \dots, r\}$ and define the set $\mathcal{C}_s := \{x \in \mathbb{R}^{m\times n} \,|\, \mathrm{rank}(x) \le s\}$. At a point $\overline{x} \in \mathcal{C}_s$

$$N_{\mathcal{C}_s}(\overline{x}) = \left\{ v \in \mathbb{R}^{m\times n} \,\middle|\, v \in \mathrm{Supp}\,(\overline{x})^{\perp} \text{ and } \mathrm{rank}(v) \le r - s \right\}.$$

Moreover, $N_{\mathcal{C}_s}(\overline{x}) = \widehat{N}_{\mathcal{C}_s}(\overline{x})$ at every $\overline{x}$ with $\mathrm{rank}(\overline{x}) = s$.

Normal cone to $\mathcal{C}_s$: vector case, $x \in \mathbb{R}^n$

Define

$$\mathcal{J} := 2^{\{1,2,\ldots,n\}} \quad \text{and} \quad \mathcal{J}_s := \mathcal{J}(s) := \{J \in \mathcal{J} \mid |J| = s\}$$

and set

$$(\forall J \in \mathcal{J}) \quad C_J := \operatorname{span}\left\{e_j \mid j \in J\right\}.$$

Define the collections

$$\mathcal{C} := \widetilde{\mathcal{C}} := (C_J)_{J \in \mathcal{J}_s} \quad \text{and} \quad \mathcal{M} := \widetilde{\mathcal{M}} := M := \left\{x \in \mathbb{R}^n \mid Ax = b\right\}.$$

Clearly,

$$\mathcal{C}_s := \left\{x \in \mathbb{R}^n \mid \|x\|_0 \le s\right\} = \widetilde{\mathcal{C}}_s := \bigcup_{J \in \mathcal{J}_s} C_J$$

and

$$M = \left\{x \in \mathbb{R}^n \mid Ax = b\right\}.$$

Normal cone to $\mathcal{C}_s$: vector case, $x \in \mathbb{R}^n$

Then

$$\begin{aligned} N_{\mathcal{C}_s}(\overline{x}) &= \left\{ v \in \mathbb{R}^n \;\middle|\; \|v\|_0 \leq n - s \text{ and } v \in \operatorname{Supp}(\overline{x})^{\perp} \right\} \\ &= \bigcup_{I(\overline{x}) \subseteq J \in \mathcal{J}_s} C_J^{\perp}. \end{aligned}$$

Consequently, if $\|\overline{x}\|_0 = s$, then $N_{\mathcal{C}_s}(\overline{x}) = \operatorname{Supp}(\overline{x})^{\perp} = C_{I(\overline{x})}^{\perp}$.

Application to sparse optimization

MAP for sparse-affine feasibility converges locally linearly

Let $\mathcal{C}_s$, $\mathcal{C}$, $\widetilde{\mathcal{C}}$ M, $\mathcal{M}$ and $\widetilde{\mathcal{M}}$ be defined as above. Suppose that $s \leq n-1$, that $\overline{x} \in \mathcal{C}_s \cap M$, and fix $\delta \in \left(0, \overline{\delta}\right)$ for $\overline{\delta} := \frac{1}{3}\min\left\{d_{C_J}(\overline{x}) \mid \overline{x} \notin C_J, J \in \mathcal{J}_s\right\}$. Then

$$\overline{\delta} = \frac{1}{3}\min\left\{|\overline{x}_j| \mid j \in I(\overline{x})\right\}$$

and

$$\theta_{3\delta}(\mathcal{C}, \widetilde{\mathcal{C}}, M, M) = \max\left\{c(C_J, M) \mid \overline{x} \in C_J, J \in \mathcal{J}_s\right\} < 1,$$

where $\theta_{3\delta}$ denotes the joint-CQ-number at $\overline{x}$ associated with $(\mathcal{C}, \widetilde{\mathcal{C}}, M, M)$.

Application to sparse optimization

MAP for sparse-affine feasibility converges locally linearly

Generate the sequences $\{c_k\}_{k\in\mathbb{N}}$ and $\{m_k\}_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ by the MAP algorithm

$$c_k = P_{\mathcal{C}_s}(m_{k-1}) \quad \text{and} \quad m_k = P_M(c_k)$$

where

$$M = \left\{x \in \mathbb{R}^{m\times n} \mid Ax = b\right\} \text{ and } \mathcal{C}_s := \left\{x \in \mathbb{R}^n \mid \|x\|_0 \leq s\right\}.$$

Suppose the starting point of the MAP $m_{-1} \in M$ satisfies $\|m_{-1} - \overline{x}\| \leq \frac{(1-\overline{\theta})\delta}{6(2-\overline{\theta})}$. Then $(c_k)_{k\in\mathbb{N}}$ and $(m_k)_{k\in\mathbb{N}}$ converge linearly to some point in $\bar{c} \in \mathcal{C}_s \cap M \cap \mathbb{B}(\overline{x}, \delta)$ with rate $\overline{\theta}^2$.

MAP for sparse-affine feasibility converges locally linearly

Remarks

1. Regularity of the intersection is not an assumption of the theorem; it is *automatically* satisfied. This is in contrast to the results of Lewis&Mallick (2008) and Lewis,L.&Mallick (2009) where the required regularity is assumed to hold. Simple examples illustrate that the notions of regularity developed in those works are not satisfied
2. Our analysis is the first (to our knowledge) to yield a quantification of the neighborhood on which local linear convergence is guaranteed.
3. Finding the local neighborhood on which linear convergence is guaranteed may well be tantamount of finding the sparsest solution; however, it does open the door to justify combining the MAP with more aggressive algorithms such as Douglas-Rachford

Thanks for your attention.

Outline

References I

H.H. Bauschke, D.R. Luke, H.M. Phan, and X. Wang, Restricted normal cones and the method of alternating projections, arXiv preprint, May 2012, `http://arxiv.org`

H.H. Bauschke, D.R. Luke, H.M. Phan, and X. Wang, Restricted normal cones and sparsity optimization with affine constraints, arXiv preprint, May 2012, `http://arxiv.org`

A. Beck and M. Teboulle, A Linearly Convergent Algorithm for Solving a Class of Nonconvex/Affine Feasibility Problems, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H.H. Bauschke, R.S. Burachick, P.L. Combettes, V. Elser, D.R. Luke and H. Wolkowicz (editors), Springer, 33–48, 2011.

References II

- J. M. Borwein and D. R. Luke. Entropic Regularization of the ℓ_0 function *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*). *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H.H. Bauschke, R.S. Burachick, P.L. Combettes, V. Elser, D.R. Luke and H. Wolkowicz (editors), Springer, 65–92, 2011.
- E. Candes and T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies?, *IEEE Transactions on Information Theory* 52 (2006), 5406–5425.
- P.L. Combettes and H.J. Trussell, Method of successive projections for finding a common point of sets in metric spaces, *Journal of Optimization Theory and Applications* 67 (1990), 487–507.

References III

- M.J. Lai and J. Wang, An Unconstrained ℓ_q Minimization for Sparse Solution of Underdetermined Linear Systems, *SIAM Journal of Optimization*, 21 (2010), 82–101.
- 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.
- B.K. Natarajan, Sparse approximate solutions to linear systems, *SIAM Journal on Computation*, 24(1995), 227–234.