OPTIMIZATION, GAMES, AND DYNAMICS
Institut Henri Poincaré

November 28-29, 2011

Convergence of descent methods for semi-algebraic and tame problems.

Hedy ATTOUCH
Institut de Mathématiques et Modélisation de Montpellier
UMR CNRS 5149

ANR 2008/2011 OSSDAA

Collaborative papers:

- J. Bolte (GREMAQ, Toulouse I): Math. Programming, Ser. B, 2009;
- J. Bolte, P. Redont (I3M, Montpellier 2), A. Soubeyran (GREQAM, Aix Marseille): Math. of Operations Research, 2010;
- J. Bolte, B. Svaiter (IMPA, Rio de Janeiro, Brésil); Math. Programming, Ser. A, 2011.

Introduction

Goal: design descent algorithms for nonsmooth, nonconvex local optimization.

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, lower semicontinuous, $(f = g + \delta_C)$.

$$\min \{f(x) : x \in \mathbb{R}^n\}.$$

Guideline: Interplay between continuous dynamical systems $(t \to +\infty)$ and algorithms.

Steepest Descent: $(SD) \quad \dot{x}(t) + \partial f(x(t)) \ni 0.$

- Geometrical assumption (Curry, 44; Palis-De Melo, 1982; Absil-Mahony-Andrews, 2005).
- Convexity: Brézis; Baillon; Bruck, JFA. 1975; Quasi-convex: Goudou-Munier, MPB, 2009.
- Analyticity: Łojasiewicz 1984; Tame analysis: Bolte-Daniilidis-Ley-Mazet, TAMS 2010.

Algorithms:

- Forward gradient steps (smooth data), backward proximal steps (nonsmooth data).
- Decomposition methods, high dimension (forward-backward,...): imaging, PDE's...

Presentation of the results

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ lower semicontinuous, proper;

$(x^k)_{k\in\mathbb{N}}$ verifying H1, H2, **KL**, a, b positive constants:

H1. (*Sufficient decrease condition*). For all $k \in \mathbb{N}$,

$$f(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \leq f(x^k);$$

H2. (*Relative error condition*). For all $k \in \mathbb{N}$, there exists $w^{k+1} \in \partial f(x^{k+1})$ such that

$$\|w^{k+1}\| \leq b\|x^{k+1} - x^k\|;$$

KL. (*Kurdyka-Łojasiewicz property*) is satisfied by f (for example f semi-algebraic).

Then,

- $(x^k)_{k\in\mathbb{N}}$ converges to a critical point of f;
- $(x^k)_{k\in\mathbb{N}}$ is of finite length, i.e., $\sum_k \|x^{k+1} - x^k\| < +\infty$;
- x^0 close enough to $\mathrm{Argmin} f \Rightarrow (x^k)_{k\in\mathbb{N}}$ converges to a global minimizer of f.

Plan

1. Łojasiewicz inequality and continuous gradient systems;
2. Kurdyka-Łojasiewicz inequality: nonsmooth case; semi-algebraic functions;
3. Descent algorithms; general convergence results.
4. Gradient methods;
5. Proximal algorithms;
6. Forward-backward algorithms;
7. Application to compressive sensing;
8. Gauss-Seidel methods.
9. Open questions, perspectives.

1. Łojasiewicz inequality and continuous gradient systems

Theorem (Łojasiewicz inequality, 1963) $f : U \subset \mathbb{R}^n \to \mathbb{R}$ real analytic, U open, $\bar{x} \in U$ critical point of f. There exists $\theta \in [\frac{1}{2}, 1)$, $C > 0$, and a neighbourhood W of $\bar{x}$ such that

$$\forall x \in W \quad |f(x) - f(\bar{x})|^{\theta} \leq C \|\nabla f(x)\|.$$

Theorem (Łojasiewicz, 1984) $f : U \subset \mathbb{R}^n \to \mathbb{R}$ real analytic. Any bounded trajectory of the steepest descent dynamical system

$$(SD) \quad \dot{x}(t) + \nabla f(x(t)) = 0$$

has a finite length and hence converges to a critical point of f.

Related results:

- PDE: Simon (1983), semilinear parabolic equations.
- Second order gradient-like system with damping, Haraux-Jendoubi J.Diff.Eq. (1998)

$$\ddot{x}(t) + \lambda \dot{x}(t) + \nabla f(x(t)) = 0.$$

The gradient conjecture of R. Thom

Thom, 1972; Publ. Math IHES, 1989.

Theorem (Kurdyka-Mostowski-Parunsinski, Annals. of Math. 2000)

- $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ real analytic.
- $t \mapsto x(t)$ trajectory of (SD) which converges to a critical point $\bar{x}$ of f.

Then the directional convergence property holds: there exists $d \in S^{n-1}$ such that

$$\lim_{t \to +\infty} \frac{x(t) - \bar{x}}{\|x(t) - \bar{x}\|} = d.$$

Thom's conjecture fails for convex functions, Daniilidis-Ley-Sabourau, JMPA, 2010:

There exists $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ convex, $\mathcal{C}^\infty$, and a trajectory of (SD) which turns infinitely many times around its limit.

Łojasiewicz inequality

f real-analytic , $\nabla f(\bar{x}) = 0$. There exists $\theta \in [\frac{1}{2}, 1)$, $C > 0$, $W \in \mathcal{V}(\bar{x})$ such that

$$\forall x \in W \quad |f(x) - f(\bar{x})|^{\theta} \leq C \|\nabla f(x)\|.$$

Proof $n = 1$, elementary: $\bar{x} = 0$. Analyticity yields $a_k \in \mathbb{R}$, $p_0 \geq 2$, et $a_{p_0} \neq 0$

$$f(x) - f(\bar{x}) = \sum_{k=p_0}^{+\infty} a_k x^k$$

Derivating term by term

$$f'(x) = \sum_{k=p_0}^{+\infty} k a_k x^{k-1}.$$

Taking $\theta \in \mathbb{R}_*^+$ and $x \neq 0$ close to zero,

$$\frac{|f(x) - f(\bar{x})|^{\theta}}{|f'(x)|} \approx \frac{1}{p_0 |a_{p_0}|^{1-\theta}} |x|^{p_0(\theta-1)+1}.$$

By taking $1 > \theta > 1 - \frac{1}{p_0}$ and x sufficiently small, one obtains

$$|f(x) - f(\bar{x})|^{\theta} \leq |f'(x)|.$$

Łojasiewicz inequality and gradient systems

f real-analytic , $\nabla f(\bar{x}) = 0$. There exists $\theta \in [\frac{1}{2}, 1)$, $C > 0$, $W \in \mathcal{V}(\bar{x})$ such that

$$\forall x \in W \quad |f(x) - f(\bar{x})|^{\theta} \leq C\|\nabla f(x)\|.$$

Equivalent formulation: $\varphi(s) = cs^{1-\theta}$ (desingularizing function)

$$\varphi'(f(x) - f(\bar{x}))\|\nabla f(x)\| \geq 1.$$

∞ **Convergence** of (SD): $\dot{x}(t) + \nabla f(x(t)) = 0.$

Lyapunov function: $h(t) = \varphi(f(x(t)) - f(\bar{x}))$, ($\bar{x}$ limit point of the trajectory)

$$\dot{h}(t) = \varphi'(f(x(t)) - f(\bar{x}))\, \langle \nabla f(x(t)), \dot{x}(t) \rangle;$$
$$\dot{h}(t) + \varphi'(f(x(t)) - f(\bar{x}))\, \| \nabla f(x(t)) \|^2 = 0;$$
$$\dot{h}(t) + \| \nabla f(x(t)) \| \leq 0;$$
$$\dot{h}(t) + \| \dot{x}(t) \| \leq 0.$$

Hence $\dot{x} \in L^1(0, +\infty)$.

2. Kurdyka-Łojasiewicz inequality: the nonsmooth case

Tools from variational analysis:

- Fréchet subdifferential of f at $x \in \mathrm{dom} f$:

$$\hat{\partial} f(x) := \left\{ x^* \in \mathbb{R}^n : \quad \liminf_{\substack{y \neq x \\ y \to x}} \frac{1}{\|x - y\|}[f(y) - f(x) - \langle x^*, y - x \rangle] \geq 0 \right\}.$$

- Limiting subdifferential (shortly subdifferential) of f (Mordukhovich):

$$\partial f(x) := \{x^* \in \mathbb{R}^n : \exists x_k \to x,\ f(x_k) \to f(x),\ x_k^* \in \hat{\partial} f(x_k) \to x^*\}.$$

- Closedness property of ∂f: $(x^k, v^k) \in \mathrm{Graph}\partial f \subset \mathbb{R}^n \times \mathbb{R}^n$

$$(x^k, v^k) \to (x, v) \text{ and } f(x^k) \to f(x) \Rightarrow (x, v) \in \mathrm{Graph}\partial f.$$

- Optimality condition: a necessary condition for $x \in \mathbb{R}^n$ to be a (local) minimizer of f is

$$\partial f(x) \ni 0.$$

Such a point is said to be *critical*. The set of critical points of $f = \mathrm{crit} f$.

KL inequality

Definition $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ lsc. has the **KL** property at $\bar{x} \in \operatorname{dom}\partial f$ if there exists $\eta \in (0,+\infty]$, $U \in \mathcal{V}(\bar{x})$, $\varphi:[0,\eta) \to \mathbb{R}_+$ (desingularizing function):

- $\varphi(0)=0$; $\varphi:[0,\eta)\to\mathbb{R}_+$ continuous, $\varphi \in \mathcal{C}^1(0,\eta)$;
- φ increasing: $\varphi'(s)>0$ for all $s\in(0,\eta)$;
- φ concave;

such that for all x in $U \cap [f(\bar{x}) < f < f(\bar{x})+\eta]$, the **KL** inequality holds:

(**KL**) $\varphi'(f(x)-f(\bar{x}))\ \operatorname{dist}(0,\partial f(x)) \geq 1.$

- Łojasiewicz (1963): real analytic functions, $\varphi(s)=s^{1-\theta}$, $\theta\in[\frac{1}{2},1)$.
- Kurdyka (Ann. I. Fourier, 1998): differentiable functions definable in an o-minimal structure (semi-algebraic, subanalytic,...).
- Bolte-Daniilidis-Lewis-Shiota (SIOPT, 2007): Clarke subgradients of nonsmooth functions definable in an o-minimal structure.
- A.-Bolte-Redont-Soubeyran (MOR, 2010): above (**KL**) definition.

Semi-algebraic sets and functions

Definition (a) $S \subset \mathbb{R}^n$ *semi-algebraic* $\Longleftrightarrow$ there exists polynoms $P_{ij}, Q_{ij} : \mathbb{R}^n \to \mathbb{R}$

$$S = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{\, x \in \mathbb{R}^n : P_{ij}(x) = 0,\ Q_{ij}(x) < 0\}.$$

(b) $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *semi-algebraic* $\Longleftrightarrow$ graph$(f) \in \mathbb{R}^{n+1}$ semi-algebraic.

Boolean structure: finite union, intersection, complementary; polynoms: semi-algebraic.

Numerical analysis [50]: cone of positive semidefinite matrices, Stiefel manifold (spheres, orthogonal group [38]), matrices with fixed rank...

Theorem [Tarski-Seidenberg] $A \subset \mathbb{R}^{n+1}$ semi-algebraic. Its canonical projection on $\mathbb{R}^n$

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n : \exists z \in \mathbb{R}, (x_1, \ldots, x_n, z) \in A\}$$

is a semi-algebraic subset of $\mathbb{R}^n$.

Illustration: S and g semi-algebraic $\Rightarrow$ $f(x) = \sup_{y \in S} g(x, y)$ is a semi-algebraic function.

Theorem Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, lower semicontinuous. Then

f semi-algebraic $\Rightarrow$ f satisfies **KL** inequality;

(with $\varphi(s) = cs^{1-\theta}$, $\theta \in [0, 1) \cap \mathbb{Q}$ and $c > 0$).

Further examples of functions satisfying KL

- o-minimal structures (semilinear, semi-algebraic, subanalytic,...): axiomatization of the qualitative properties of semi-algebraic sets, van den Dries (1998).
  Functions definable in a o-minimal structure satisfy **KL**: Kurdyka (1998), BDLS (2007).

- Uniform convexity: for all $x, y \in \mathbb{R}^n$, $x^* \in \partial f(x)$,

$$f(y) \geq f(x) + \langle x^*, y - x \rangle + K\|y - x\|^p, \quad p \geq 1$$

$$\Rightarrow f \in \textbf{KL},\ \phi(s) = cs^{1/p}.$$

  Existence of a smooth convex $f : \mathbb{R}^2 \to \mathbb{R}$ which does not satisfy **KL**;
  Bolte-Daniilidis-Ley-Mazet (2010); Daniilidis-Ley-Sabourau (2010).

- Linearly regular intersection of F_i, transversality, Lewis-Malick (2008):

$$\Rightarrow f(x) := \frac{1}{2}\sum_i \mathrm{dist}(x, F_i)^2 \text{ satisfies } \textbf{KL}.$$

- Metric regularity: $F : \mathbb{R}^n \to \mathbb{R}^m$ metrically regular at $\bar{x} \in \mathbb{R}^n$, if there exists a neighbourhood V of $\bar{x}$ in $\mathbb{R}^n$, a neighbourhood W of $F(\bar{x})$ in $\mathbb{R}^m$ and $k > 0$

$$x \in V,\ y \in W \ \Rightarrow\ \mathrm{dist}\,(x, F^{-1}(y)) \leq k\ \mathrm{dist}\,(y, F(x)).$$

$$\Rightarrow f(x) = \frac{1}{2}\mathrm{dist}^{\,2}(F(x), C) \text{ satisfies } \textbf{KL},\ C \subset \mathbb{R}^m \text{ closed convex},\ \phi(s) = c\sqrt{s},\ ([5]).$$

Sets and functions definable in an o-minimal structure

van den Dries [36] (1998): axiomatization of the qualitative properties of semi-algebraic sets.

Definition $\mathcal{O} = \{\mathcal{O}_n\}_{n\in\mathbb{N}}$, $\mathcal{O}_n$ collection of subsets of $\mathbb{R}^n$. $\mathcal{O}$ is an o-minimal structure if:

(i) Each $\mathcal{O}_n$ is a boolean algebra: $\emptyset \in \mathcal{O}_n$, A, B in $\mathcal{O}_n \Rightarrow A \cup B, A \cap B, \mathbb{R}^n \setminus A \in \mathcal{O}_n$.

(ii) For all A in $\mathcal{O}_n$, $A \times \mathbb{R}$ and $\mathbb{R} \times A$ belong to $\mathcal{O}_{n+1}$.

(iii) For all A in $\mathcal{O}_{n+1}$, $\Pi(A) := \{(x_1, \ldots, x_n) \in \mathbb{R}^n : (x_1, \ldots, x_n, x_{n+1}) \in A\} \in \mathcal{O}_n$.

(iv) For all $i \neq j$ in $\{1, \ldots, n\}$, $\{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_i = x_j\} \in \mathcal{O}_n$.

(v) The set $\{(x_1, x_2) \in \mathbb{R}^2 : \; x_1 < x_2\}$ belongs to $\mathcal{O}_2$.

(vi) The elements of $\mathcal{O}_1$ are exactly finite unions of intervals.

A is definable in $\mathcal{O}$ if A belongs to $\mathcal{O}$.

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is definable if its graph is a definable subset of $\mathbb{R}^n \times \mathbb{R}$.

Theorem (BDLS, SIOPT 2007) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous, definable in an o-minimal structure $\mathcal{O}$. Then, f has the **KL** property at each point of $\operatorname{dom} \partial f$. Moreover, the desingularizing function φ is definable in $\mathcal{O}$.

$\rightarrow$ semilinear, semi-algebraic, subanalytic o-minimal structures.

3. Descent algorithms; general convergence results

$f : \mathrm{I\!R}^n \to \mathrm{I\!R} \cup \{+\infty\}$ proper lower semicontinuous.
a et b fixed positive constants;
We consider sequences $(x^k)_{k\in\mathbb{N}}$ which satisfy **H1**, **H2**, **H3**:

H1. (*Sufficient decrease condition*). For each $k \in \mathbb{N}$,

$$f(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \leq f(x^k);$$

H2. (*Relative error condition*). For each $k \in \mathbb{N}$, there exists $w^{k+1} \in \partial f(x^{k+1})$ such that

$$\|w^{k+1}\| \leq b\|x^{k+1} - x^k\|;$$

H3. (*Continuity condition*). There exists a subsequence $(x^{k_j})_{j\in\mathbb{N}}$ and $\tilde{x}$ such that

$$x^{k_j} \to \tilde{x} \text{ and } f(x^{k_j}) \to f(\tilde{x}) \qquad \text{as } j \to \infty.$$

Remark In most practical algorithms (*e.g.* forward-backward, Gauss-Seidel...) **H3** is satisfied assuming just that f is lower semicontinuous.

Convergence theorems

Theorem 1 (*Convergence to a critical point*) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. Consider a sequence $(x^k)_{k\in\mathbb{N}}$ that satisfies $\mathbf{H1}$, $\mathbf{H2}$, and $\mathbf{H3}$.
If f has the **KL** property, then the sequence $(x^k)_{k\in\mathbb{N}}$ converges, and its limit $\bar{x}$ is a critical point of f. Moreover, the sequence $(x^k)_{k\in\mathbb{N}}$ has a finite length, $i.e.$

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\| < +\infty.$$

Theorem 2 (*Local convergence to a global minima*) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous function which has the **KL** property at x^*, a global minimum point of f. Then for each $r > 0$, there exist $\rho \in (0, r)$, $\mu > 0$ such that the inequalities

$$\|x^0 - x^*\| < \rho, \quad \min f < f(x^0) < \min f + \mu$$

imply that any sequence $(x^k)_{k\in\mathbb{N}}$ that satisfies **H1**, **H2**, and which starts from x^0 satisfies

(i) $x^k \in B(x^*, r)$, $\forall k \in \mathbb{N}$,

(ii) x^k converges to $\bar{x}$ and $\sum_{k=1}^{+\infty} \|x^{k+1} - x^k\| < +\infty$,

(iii) $f(\bar{x}) = \min f$.

Convergence to a local minima

Let x^* be a local minimizer of f and suppose that f satisfies the growth condition:

$$\textbf{H4.} \quad f(y) \geq f(x^*) - \frac{a}{4}\|y - x^*\|^2 \quad \text{for all } y \in \mathbb{R}^n.$$

Theorem 3 (*Local convergence to a local minima*) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function which has the **KL** property at some local minimizer x^*. Assume that **H4** holds at x^*.

Then, for any $r > 0$, there exist $u \in (0, r)$ and $\mu > 0$ such that the inequalities

$$\|x^0 - x^*\| < u, \quad f(x^*) < f(x^0) < f(x^*) + \mu,$$

imply that any sequence $(x^k)_{k\in\mathbb{N}}$ starting from x^0, that satisfies **H1**, **H2** has the finite length property, remains in $B(x^*, r)$ and converges to some $\bar{x} \in B(x^*, r)$ critical point of f with $f(\bar{x}) = f(x^*)$.

4. Gradient methods

$f : \mathbb{R}^n \to \mathbb{R}$ class C^1, ∇f Lipschitz continuous with constant L, $\inf f > -\infty$.

Algorithm 1 Parameters $a > 0, b > 0$, $a > L$. Fix x^0 in $\mathbb{R}^n$. For $k = 0, 1, \ldots$

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{a}{2}\|x^{k+1} - x^k\|^2 \leq 0,$$
$$\|\nabla f(x^k)\| \leq b\|x^{k+1} - x^k\|.$$

Example (steepest descent): $x^{k+1} - x^k = -\lambda_k \nabla f(x_k)$; $\quad 0 < \underline{\lambda} < \lambda_k < \overline{\lambda} < \frac{2}{L}$.

Theorem 4 Suppose that f has the **KL** property. Then each bounded sequence $(x^k)_{k\in\mathbb{N}}$ generated by Algorithm 1 converges to some critical point $\bar{x}$ of f, and has a finite length.

Remark

1. Classical convergence results: $\nabla f(x^k) \to 0$.
   First convergence results for $(x^k)_{k\in\mathbb{N}}$: Absil-Mahony-Andrews, SIOPT, 2005.
2. The conclusion remains unchanged if there exists a closed subset S of $\mathbb{R}^n$ such that
   - $x^k \in S$ for all $k \in \mathbb{N}$; ∇f is L-Lipschitz continuous on $\operatorname{co} S$;
   - f satisfies the **KL** inequality at each point of S,

Average projections for feasibility problems

$F_1, \ldots, F_p$ closed subsets of $\mathbb{R}^n$ such that

$$\bigcap_{i=1}^{p} F_i \neq \emptyset.$$

A classical approach to the problem of finding a common point $\bar{x}$ to the sets $F_1, \ldots, F_p$

$$\bar{x} \in \bigcap_{i=1}^{p} F_i$$

is to find a global minimizer of the function $f : \mathbb{R}^n \to [0, +\infty)$

$$f(x) := \frac{1}{2p} \sum_{i=1}^{p} \operatorname{dist}(x, F_i)^2,$$

where $\operatorname{dist}(\cdot, F_i)$ is the distance function to the set F_i.

- F_i semi-algebraic $\Rightarrow$ dist$(x, F_i)^2$ semi-algebraic $\Rightarrow$ $f \in$ **KL**.
- F_i prox-regular $\Rightarrow$ $\frac{1}{2}$dist$(x, F_i)^2$ locally C^1 function whose gradient is 1-Lipschitz $\Rightarrow$ f idem.

Prox-regular sets

Definition A closed subset F of $\mathbb{R}^n$ is prox-regular if its projection operator P_F is single-valued around each point x in F.

Prominent examples: closed convex sets and C^2 submanifolds of $\mathbb{R}^n$.

Set $g(x) = \frac{1}{2}\text{dist}(x, F)^2$ and suppose that F is prox-regular.

Theorem (Poliquin-Rockafellar-Thibault, Trans. AMS, 2000) Let F be a closed prox-regular set. Then for each $\bar{x}$ in F there exists $r > 0$ such that:

(a) The projection P_F is single-valued on $B(\bar{x}, r)$.

(b) The function g is C^1 on $B(\bar{x}, r)$ and $\nabla g(x) = x - P_F(x)$.

(c) The gradient mapping ∇g is 1-Lipschitz continuous on $B(\bar{x}, r)$.

Inexact averaged projection algorithm

Gradient method for $f(x) := \frac{1}{2p}\sum_{i=1}^{p} \text{dist}(x, F_i)^2$.

Algorithm 2 Take $\theta \in (0,1)$, $\alpha < \frac{1}{2}$, $M > 0$; $x^0 \in \mathbb{R}^n$.

$$x^{k+1} \in (1-\theta)\, x^k + \theta \left(\frac{1}{p} \sum_{i=1}^{p} P_{F_i}(x^k) \right) + \epsilon^k,$$

$(\epsilon^k)_{k\in\mathbb{N}}$ is a sequence of admissible errors which satisfies

$$\begin{aligned} &\langle \epsilon^k, x^{k+1} - x^k \rangle \le \alpha \|x^{k+1} - x^k\|^2 \\ &\|\epsilon^k\| \le M \|x^{k+1} - x^k\| \end{aligned}$$

Theorem 5 Let $F_1, \ldots, F_p$ be semi-algebraic, and prox-regular subsets of $\mathbb{R}^n$, $\bigcap_{i=1}^{p} F_i \neq \emptyset$. If x^0 is sufficiently close to $\bigcap_{i=1}^{p} F_i$, then Algorithm 2 reduces to the gradient method

$$x^{k+1} = x^k - \theta \nabla f(x^k) + \epsilon^k,$$

which therefore defines a unique sequence. Moreover, this sequence has a finite length and converges to a feasible point $\bar{x}$, *i.e.* such that $\bar{x} \in \bigcap_{i=1}^{p} F_i$.

Linear regular intersection, transversality

Lewis-Malick (Math. Oper. Res., 2008), Lewis-Luke-Malick (Found. Comput. Math., 2009):

Similar results hold for sets F_i having a linearly regular intersection at some point $\bar{x}$:

$$\sum_{i=1}^{p} y_i = 0, \text{ with } y_i \in N_{F_i}(\bar{x}) \Longrightarrow y_i = 0, \forall i = 1, \ldots, p$$

.

Example: transverse manifolds.

Key property in LLM: $f(x) := \frac{1}{2}\sum_i \operatorname{dist}(x, F_i)^2$ locally satisfies the inequality

$$\|\nabla f(x)\|^2 \geq c f(x),$$

$=$ Łojasiewicz inequality with a desingularizing function of the form $\varphi(s) = \frac{2}{\sqrt{c}}\sqrt{s}$.

Compare

- The linear regular intersection property provides linear convergence;
- **KL** approach, algebraic structure (common feature), possible tangent sets, desingularizing function (rate of convergence).

5. Proximal algorithms

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ proper lower semicontinuous, $\inf f > -\infty$, $\lambda > 0$.

$$\text{prox}_{\lambda f} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$$

$$\text{prox}_{\lambda f} x := \operatorname{argmin} \left\{ f(y) + \frac{1}{2\lambda}\|y - x\|^2 : y \in \mathbb{R}^n \right\}.$$

Algorithm 3a (Proximal algorithm, exact version)

$$0 < \underline{\lambda} < \lambda_k < \overline{\lambda} < +\infty;$$
$$x_0 \in \mathbb{R}^n;$$
$$x^{k+1} \in \text{prox}_{\lambda_k f}(x^k).$$

Theorem 6 Suppose that f has the **KL** property, and that the restriction of f to its domain is a continuous function. Then each bounded sequence $(x^k)_{k \in \mathbb{N}}$ generated by Algorithm 3 converges to some critical point $\bar{x}$ of f, and has a finite length.

Rate of convergence

- $x^k \to \bar{x}$ convergent sequence generated by the proximal algorithm;
- $f : U \subset \mathbb{R}^n \to \mathbb{R}$ lower semicontinuous, satisfies **KL** at $\bar{x}$:

  There exists $\theta \in [0,1)$, $C > 0$, $W \in \mathcal{V}(\bar{x})$ such that

$$\forall x \in W,\ \forall w \in \partial f(x) \quad |f(x) - f(\bar{x})|^\theta \leq C\|w\|.$$

Theorem 7 (AB, MPB, 2009)

(i) If $\theta = 0$, the sequence $(x^k)_{k\in\mathbb{N}}$ converges in a finite number of steps.

(ii) If $\theta \in (0, \frac{1}{2}]$ then there exist $c > 0$ and $Q \in [0,1)$ such that

$$\|x^k - \bar{x}\| \leq c\, Q^k.$$

(iii) If $\theta \in (\frac{1}{2}, 1)$ then there exists $c > 0$ such that

$$\|x^k - \bar{x}\| \leq c\, k^{-\frac{1-\theta}{2\theta-1}}.$$

Inexact version of the proximal point method

Algorithm 3b: (Proximal algorithm, inexact version)
Take $x_0 \in \mathbb{R}^n$, $\quad 0 < \underline{\lambda} \le \overline{\lambda} < \infty, \quad 0 \le \sigma < 1, \quad 0 < \theta \le 1.$

For $k = 0, 1, \ldots,$ choose $\lambda_k \in [\underline{\lambda}, \overline{\lambda}]$, and find $x^{k+1} \in \mathbb{R}^n$, $w^{k+1} \in \mathbb{R}^n$ such that

$$f(x^{k+1}) + \frac{\theta}{2\lambda_k}\|x^{k+1} - x^k\|^2 \le f(x^k);$$
$$w^{k+1} \in \partial f(x^{k+1});$$
$$\|\lambda_k w^{k+1} + x^{k+1} - x^k\|^2 \le \sigma(\|\lambda_k w^{k+1}\|^2 + \|x^{k+1} - x^k\|^2).$$

The last condition can be replaced by the weaker condition: for some positive $b > 0$

$$\|\lambda_k w^{k+1}\| \le b\|x^{k+1} - x^k\|.$$

Theorem 8 Suppose that f has the **KL** property, and that the restriction of f to its domain is a continuous function. Then each bounded sequence $(x^k)_{k\in\mathbb{N}}$ generated by the inexact proximal algorithm converges to some critical point $\bar{x}$ of f, and has a finite length.

6. Forward-Backward splitting algorithms

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, structured

$$f = g + h$$

- $h : \mathbb{R}^n \to \mathbb{R}$ $\mathcal{C}^1$, ∇h Lipschitz continuous, $L =$ Lipschitz constant of ∇h.
- $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ lower semicontinuous, minorized.
- f satisfies **KL**.

Forward-Backward splitting algorithm (exact form): $0 < \underline{\gamma} < \gamma_k < \overline{\gamma} < \frac{1}{L}$

$$x^{k+1} \in \text{prox}_{\gamma_k g}(x^k - \gamma_k \nabla h(x^k)).$$

Proximal mapping: $\mathbb{R}^n \rightrightarrows \mathbb{R}^n$, $\quad \text{prox}_{\gamma g} x := \operatorname{argmin} \left\{ \gamma g(y) + \frac{1}{2}\|y - x\|^2 : y \in \mathbb{R}^n \right\}.$

Theorem 9 Each bounded sequence $(x^k)_{k\in\mathbb{N}}$ generated by the forward-backward splitting algorithm converges to a critical point of $f = g + h$.
Moreover, $(x^k)_{k\in\mathbb{N}}$ has a finite length i.e. $\sum_k \|x^{k+1} - x^k\| < +\infty$.

Convergence of the forward-backward algorithm with relative error

Algorithm 4: Take $a, b > 0$ with $a > L$. Take $x^0 \in \operatorname{dom} g$.
For $k = 0, 1, \dots,$ find $x^{k+1} \in \mathbb{R}^n$, $v^{k+1} \in \mathbb{R}^n$ such that

$$
\begin{aligned}
&g(x^{k+1}) + \langle x^{k+1} - x^k, \nabla h(x^k)\rangle + \frac{a}{2}\|x^{k+1} - x^k\|^2 \le g(x^k);\\
&v^{k+1} \in \partial g(x^{k+1});\\
&\|v^{k+1} + \nabla h(x^k)\| \le b\|x^{k+1} - x^k\|;
\end{aligned}
$$

Theorem 10 Under the following assumptions

- $f = g + h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous, minorized, satisfying **KL**;
- $h : \mathbb{R}^n \to \mathbb{R}$ $\mathcal{C}^1$, ∇h Lipschitz continuous, $L =$ Lipschitz constant of ∇h;
- the restriction of g to its domain is continuous;

each bounded sequence $(x^k)_{k\in\mathbb{N}}$ generated by Algorithm 3 converges to a critical point of $f = g + h$. Moreover, $(x^k)_{k\in\mathbb{N}}$ has a finite length i.e. $\sum_k \|x^{k+1} - x^k\| < +\infty$.

Remark a) Forward-Backward splitting algorithm (exact form) $=$ particular case.
b) Forward-Backward algorithm, exact form: the continuity assumption concerning g is useless.

c) Application to splitting methods for coupled systems, A.-Briceno-Combettes, SIOPT 2010.

Nonconvex gradient projection algorithms

- $f = i_C + h$ (C closed subset of $\mathbb{R}^n$). For each $\gamma > 0$, $\text{prox}_{\gamma\, i_C} x = P_C(x)$;
- $h : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function whose gradient is L-Lipschitz continuous;
- C a nonempty closed subset of $\mathbb{R}^n$.
- $\epsilon \in (0, \frac{1}{2L})$, a sequence of stepsizes γ_k such that $\epsilon < \gamma_k < \frac{1}{L} - \epsilon$.

$$(NGP) \quad x^{k+1} \in P_C(x^k - \gamma_k \nabla h(x^k)).$$

Theorem 11 Let$(x^k)_{k\in\mathbb{N}}$ be a bounded sequence that complies with (NGP) algorithm. If $h + i_C$ is a **KL** function, then the sequence $(x^k)_{k\in\mathbb{N}}$ converges to a point x^* in C such that

$$\nabla h(x^*) + N_C(x^*) \ni 0.$$

Remark a) The assumption $f = i_C + h \in$**KL** is very general. It is satisfied for example if h ic $\mathcal{C}^1$ semi-algebraic, and C is closed, semi-algebraic.
b) There is no (variational) regularity assumption on C: C is not supposed to be prox-regular, the projection operator may be multivalued in a neighbourhood of C.

Hard-constrained feasibility problems

- $F, F_1, \ldots, F_p$ finite collection of nonempty closed subsets of $\mathbb{R}^n$;
- $F_1, \ldots, F_p$ convex sets; the hard constraint F is not supposed to be convex;

Combettes-Wajs, Multiscale Model. Simul., 2005: $\omega_i > 0$, $\sum_i \omega_i = 1$,

$$\min_{x \in F} \left\{ h(x) := \frac{1}{2} \sum_{i=1}^{p} \omega_i \mathrm{dist}\,(x, F_i)^2 \right\}.$$

Gradient projection algorithm $\rightarrow$ satisfy the hard constraint F, $\neq$ $F_1, \ldots, F_p$ are relaxed.

$L = 1$ Lipschitz constant of ∇h; $0 < \underline{\gamma} \leq \gamma_k \leq \overline{\gamma} < 1$,

$$(NGP) \quad x^{k+1} \in P_F\left((1-\gamma_k)x^k + \gamma_k \sum_{i=1}^{p} \omega_i P_{F_i}(x^k) \right).$$

Theorem 12 $F, F_1, \ldots, F_p$ semi-algebraic.

- Each bounded sequence $(x^k)_{k\in\mathbb{N}}$ generated by the (NGP) algorithm converges to a critical point of $h + i_F$, i.e, $\nabla h(x^*) + N_F(x^*) \ni 0$.
- If x^0 is sufficiently close to the intersection of the $F, F_1, \ldots, F_p$, then $(x^k)_{k\in\mathbb{N}}$ converges to a point which belongs to the intersection of the $F, F_1, \ldots, F_p$.

7. Application to compressive sensing

Optimization methods: Donoho, (2006), Chartrant (2007), Becker-Bobin-Candes (2009). GDR Opt.-Image, http://www.ceremade.dauphine.fr/ peyre/mspc/mspc-moa-11/slides/.

Recover sparse solutions of under-determined linear systems:

$$(P) \quad \min\{\|x\|_0 : Ax = b\}$$

- $\|\cdot\|_0$: counting norm (ℓ^0 norm): the number of nonzero components of $x \in \mathbb{R}^n$.
- $A \neq 0$: $m \times n$ real matrix $(m < n)$, $b \in \mathbb{R}^m$.

$$(P_\lambda) \quad \min\{\lambda\|x\|_0 + \frac{1}{2}\|Ax - b\|^2\}.$$

Forward-Backward algorithm: $f = g + h$, $g(x) = \lambda\|x\|_0$, $h(x) = \frac{1}{2}\|Ax - b\|^2$

- f is lower semicontinuous: $\|\cdot\|_0$ is lower semicontinuous;
- $f = g + h$ semi-algebraic, **KL** function: h polynomial, $\|\cdot\|_0$ piecewise linear graph.

$$x^{k+1} \in \text{prox}_{\gamma_k\lambda\|\cdot\|_0}\left(x^k - \gamma_k(A^T A x^k - A^T b)\right).$$

Iterative hard thresholding algorithms, Blumensath-Davis (2008), (2009).

Computing $\text{prox}_{\gamma\lambda\|\cdot\|_0}$

$n = 1$, counting function $|\cdot|_0$;

$$\text{prox}_{\gamma\lambda|\cdot|_0} u = \begin{cases} u & \text{if } |u| > \sqrt{2\gamma\lambda} \\ \{0, u\} & \text{if } |u| = \sqrt{2\gamma\lambda} \\ 0 & \text{otherwise.} \end{cases}$$

$n \in \mathbb{N}$, $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$,

$$\text{prox}_{\gamma\lambda\|\cdot\|_0} u = \left(\text{prox}_{\gamma\lambda|\cdot|_0} u_1, \ldots, \text{prox}_{\gamma\lambda|\cdot|_0} u_n\right),$$

Theorem 13 Each bounded sequence (x^k) generated by the hard thresholding algorithm

$$x^{k+1} \in \text{prox}_{\gamma_k\lambda\|\cdot\|_0}\left(x^k - \gamma_k(A^T A x^k - A^T b)\right)$$

$0 < \underline{\gamma} < \gamma_k < \overline{\gamma} < |||A^T A|||^{-1}$, converges to a critical point x^* of $\lambda\|x\|_0 + \frac{1}{2}\|Ax - b\|^2$, $i.e.$, i.e. x^* satisfies

$$(A^T A x^*)_i = (A^T b)_i.$$

for all i such that $x_i^* \neq 0$.

Relaxation, approximation of the counting function

$$(P'_\lambda) \quad \min\{\lambda\|x\|_* + \tfrac{1}{2}\|Ax-b\|^2\}.$$

Algorithm: $x^{k+1} \in \mathsf{prox}_{\gamma_k\|\cdot\|_*}(x^k - \gamma_k\lambda(A^TAx^k - A^Tb))$.

1. $\|x\|_* = \|x\|_1$ convex relaxation (soft thresholding, Chen-Donoho-Saunders, 2004).
2. $\|x\|_* = \|x\|_p = \sum_1^n |x_i|^p$, $p \in (0,1)$, Chartrand (2007), Bredies-Lorenz (2009).

   Separable structure of $\|\cdot\|_p \Rightarrow$ computing $\mathsf{prox}_{\gamma\|\cdot\|_p}(u)$ is equivalent to find solve

   $$\min\left\{2\gamma|x|^p + (x-u)^2 : x \in \mathbb{R}\right\}.$$

   $f(x) = \|x\|_p + \frac{\lambda}{2}\|Ax-b\|^2$ satisfies **KL**: There exists a o-minimal structure containing $\{x^\alpha : x > 0,\ \alpha \in \mathbb{R}\}$ and the restricted analytic functions ([37]). $\varphi(s) = cs^\theta$, $\theta \in [0,1)$.
3. Mangasarian (1999), Jokar et Pfetsch (2007) $\|x\|_* = \sum_1^n (1 - e^{-\alpha|x_i|})$.
4. Zhang et al. (2006), $\|x\|_* = \sum_1^n \phi(x_i)$

$$\phi(x_i) = \begin{cases} \lambda|x_i| & \text{if } |x_i| \le \lambda, \\ -(|x_i|^2 - 2a\lambda|x_i| + \lambda^2)/(2(a-1)) & \text{if } \lambda < |x_i| \le a\lambda, \\ (a+1)\frac{\lambda^2}{2} & \text{if } |x_i| > a\lambda \end{cases}$$

8. Regularized Gauss-Seidel methods

Fix an integer $p \geq 2$, and let $n_1, \dots, n_p$ be positive integers. The current vector x belongs to the product space $\mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$, $x = (x_1, \dots, x_p)$, $x_i \in \mathbb{R}^{n_i}$.

$$\min \left\{ Q(x_1, \dots, x_p) + \sum_{i=1}^{p} f_i(x_i);\ x_i \in \mathbb{R}^{n_i}, i = 1, 2, ..., p \right\}$$

- $Q : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p} \to \mathbb{R}$ C^1 coupling function, ∇Q locally Lipschitz continuous;
- $f_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{+\infty\}$ proper lower semicontinuous function, $i = 1, 2, ..., p$.

A proximal modification of the Gauss-Seidel method (Auslender (1992), ABRS (2010))
Alternating proximal minimization of $f(x) = Q(x_1, \dots, x_p) + \sum_{i=1}^{p} f_i(x_i)$.
$(B_i^k)_{k \in \mathbb{N}}$ symmetric positive definite matrices; $x^0 = (x_1^0, \dots, x_p^0)$ in $\mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$;

$$x_1^{k+1} \in \operatorname{argmin} \{ f(\mathbf{u}_1, x_2^k, \dots, x_p^k) + \frac{1}{2} \langle B_1^k(\mathbf{u}_1 - x_1^k), u_1 - x_1^k \rangle : \mathbf{u}_1 \in \mathbb{R}^{n_1} \}.$$

$$x_i^{k+1} \in \operatorname{argmin} \{ f(x_1^{k+1}, \dots, x_{i-1}^{k+1}, \mathbf{u}_i, x_{i+1}^k, \dots) + \frac{1}{2} \langle B_i^k(\mathbf{u}_i - x_i^k), \mathbf{u}_i - x_i^k \rangle : \mathbf{u}_i \in \mathbb{R}^{n_i} \};$$

$$x_p^{k+1} \in \operatorname{argmin} \{ f(x_1^{k+1}, \dots, x_{p-1}^{k+1}, \mathbf{u}_p) + \frac{1}{2} \langle B_p^k(\mathbf{u}_p - x_p^k), \mathbf{u}_p - x_p^k \rangle : \mathbf{u}_p \in \mathbb{R}^{n_p} \}.$$

A proximal version of the Gauss-Seidel method with relative error

Take $0 < \underline{\lambda} < \overline{\lambda} < \infty$.
$(A_i^k)_{k\in\mathbb{N}}$ symmetric positive definite matrices whose eigenvalues lie in $[\underline{\lambda}, \overline{\lambda}]$.
b_i positive parameters ($i = 1, \dots, p$).
$x^0 = (x_1^0, \dots, x_p^0)$ in $\mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$.
For $k = 0, 1, \dots,$ find x^{k+1} and $v^{k+1} \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$ such that

$$f_i(x_i^{k+1}) + Q(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, \dots, x_p^k) + \frac{1}{2}\langle A_i^k(x_i^{k+1} - x_i^k), x_i^{k+1} - x_i^k\rangle$$
$$\leq f_i(x_i^k) + Q(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_p^k); \tag{1}$$
$$v_i^{k+1} \in \partial f_i(x_i^{k+1}); \tag{2}$$
$$\|v_i^{k+1} + \nabla_{x_i} Q(x_1^{k+1}, \dots, x_i^{k+1}, x_{i+1}^k, \dots, x_p^k)\| \leq b_i \|x_i^{k+1} - x_i^k\|, \tag{3}$$

where i ranges over $\{1, \dots, p\}$.

Theorem 14 [Proximal regularization of Gauss-Seidel method] Suppose that

$$f(x) = Q(x_1, \dots, x_p) + \textstyle\sum_{i=1}^p f_i(x_i).$$

is a **KL** function which is bounded from below. Each bounded sequence $(x^k)_{k\in\mathbb{N}}$ generated by the proximal Gauss-Seidel method converges to some critical point $\bar{x}$ of f.
Moreover the sequence $(x^k)_{k\in\mathbb{N}}$ has a finite length, $i.e.$ $\sum_k \|x^{k+1} - x^k\| < +\infty$.

Perspectives

Numerical aspects

- Discrete version of Thom's conjecture.
- Desingularizing functions: rate of convergence, complexity.
- Accelerating gradient methods, Nesterov [58], Beck-Teboulle [13], Becker-Bobin-Candes [14], Wright [69] ($t_1 = 1$):

$$x^k \in \mathrm{prox}_{\gamma_k g}(y^k - \gamma_k \nabla h(y^k))$$
$$y^k = x^{k-1} + \frac{t_{k-1} - 1}{t_k}(x^{k-1} - x^{k-2})$$
$$t_k = \frac{1 + \sqrt{1 + 4t_{k-1}^2}}{2}$$

- Nonautonomous versions, approximation methods

  Coupling descent methods with penalization:

  forward-backward: A.-Czarnecki-Peypouquet, SIOPT 2011
  relaxed Gauss-Seidel methods: A.-Cabot-Frankel-Peypouquet, JNA 2011.

Applications

- Compressive sensing, rank reduction, imaging, signal, statistics.
- Games: Best response dynamics, cost to change, Nash equilibration, Pareto front.
- Infinite dimension problems

a) Decomposition of domains for PDE's: H.A.-Briceno Arias-Combettes [7].

$f_i \in \Gamma_0(\mathcal{H}_i)$, $\varphi_{ij} \in \Gamma_0(L^2(\Upsilon_{ij}))$,

$$\underset{x_1 \in \mathcal{H}_1, \ldots, x_m \in \mathcal{H}_m}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(x_i) + \sum_{i=1}^{m-1} \sum_{j \in J(i+)} \varphi_{ij}(\mathsf{T}_{ij}\, x_i - \mathsf{T}_{ji}\, x_j),$$

b) Optimal control, optimal design of structure:

$$\min \{ f(y) + g(u) : E(y,u) = 0 \}$$

Penalization of the state equation:

$$\min \left\{ f(y) + g(u) + \lambda \| E(y,u) \|^2 \right\}.$$

Optimal design of structure: Allaire [2], alternating minimization, gradient projection.
Quasi-static brittle fracture: Francfort-Marigo, Ambrosio-Tortorelli variational approach, alternating minimization algorithm: Bourdin-Francfort-Marigo [19], Burke-Ortner-Süli.

References

[1] Absil, P.-A., Mahony, R. , Andrews, B., Convergence of the iterates of descent methods for analytic cost functions, SIAM J. Optim., **16**, no. 2, (2005), 531–547.

[2] Allaire, G., Optimal design of structures, Ecole polytechnique, 2011.

[3] Aragon, A., Dontchev, A. , Geoffroy, M., Convergence of the proximal point method for metrically regular mappings, ESAIM Proc., **17**, EDP Sci., (2007).

[4] Attouch, H., Bolte, J., On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, Math. Program., Ser. B, **116** (2009), 5-16.

[5] Attouch, H., Bolte, J., Redont, P., Soubeyran, A. Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Lojasiewicz inequality, Mathematics of Operations Research, **35**, no. 2, (2010), 438-457.

[6] Attouch, H., Briceño-Arias, L.M., Combettes, P.L. A parallel splitting method for coupled monotone inclusions, SIAM J. Control Optim., **48**, no. 5, (2010), 3246-3270.

[7] Attouch, H., Briceño-Arias, L.M., Combettes, P.L. Domain decomposition splitting methods, 2011.

[8] Attouch, H., Cabot, A., Frankel, P., Peypouquet, J. Alternating proximal algorithms for constrained variational inequalities, Applications to domain decomposition for PDE's, submitted to J. Nonlinear Analysis, 2010.

[9] Attouch, H., Czarnecki, M.O., Peypouquet, J. Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities, 2011.

[10] Attouch, H., Czarnecki, M.O., Peypouquet, J. Prox-penalization and splitting methods for constrained variational problems, SIAM J. Optimization, 2010.

[11] Attouch, H., Soubeyran, A. Local search proximal algorithms as decision dynamics with costs to move, Set Valued and Variational Analysis, Online First, 2010.

[12] Auslender, A., Asymptotic properties of the Fenchel dual functional and applications to decomposition problems, J. Optim. Theory Appl., **73** (1992), 427–449.

[13] Beck, A., Teboulle M., Gradient-based algorithms with applications to signal recovery problems, Preprint, Tel-Aviv University, Technion.

[14] Becker, S., Bobin, J., Candès, J., Nesta: A fast accurate first-order method for sparse recovery, Caltech, (2009).

[15] Benedetti, R., Risler, J.-J., *Real Algebraic and Semialgebraic Sets*, Hermann, Éditeur des Sciences et des Arts, (Paris, 1990).

[16] Blumensath T., Davis, M. E., Iterative Thresholding for Sparse Approximations, J. of Fourier Anal. App. **14** (2008), 629–654.

[17] Blumensath T., Davis, M. E., Iterative hard thresholding for compressed sensing, App. Comput. Harmon. Anal., **27** (2009), 265–274.

[18] Bochnak, J., Coste, M., Roy, M.-F., *Real Algebraic Geometry*, (Springer, 1998).

[19] Bourdin B., Francfort, G., Marigo, J.-J. Numerical experiments in revisited brittle fracture, J. Mech. Phys. Solids, **48** (2000), 797–826.

[20] Bolte, J., Combettes, P.L., Pesquet, J.-C., Alternating proximal algorithm for blind image recovery, Proceedings of the IEEE International Conference on Image Processing. Hong-Kong, September 26-29, 2010.

[21] Bolte, J., Daniilidis, A. , Lewis, A., The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, SIAM J. Optim., **17** , no. 4, (2006), 1205–1223.

[22] Bolte, J., Daniilidis, A., Lewis, A., A nonsmooth Morse-Sard theorem for subanalytic functions, J. Math. Anal. Appl., **321**, no. 2, (2006), 729–740.

[23] Bolte, J., Daniilidis, A., Lewis, A., Shiota, M., Clarke subgradients of stratifiable functions, SIAM J. Optim., **18**, no. 2, (2007), 556–572.

[24] Bolte, J., Daniilidis, A., Ley, O., Mazet, L., Characterizations of Łojasiewicz inequalities: Subgradient flows, talweg, convexity, Trans. Amer. Math. Soc., **362**, (2010), 3319-3363.

[25] Bredies, K., Lorenz, D.A., Minimization of nonsmooth, nonconvex functionals by iterative thresholding, preprint http://www.uni-graz.at/ bredies/publications.html.

[26] Chartrand, R. Exact reconstruction of sparce signals via nonconvex minimization, Signal Processing Letters IEEE, 14 (2007), 707–710. **53**, (2003), 1017–1039.

[27] Chill, R., Jendoubi, M.A. Convergence to steady states in asymptotically autonomous semilinear evolution equations, Nonlinear Analysis, **53**, (2003), 1017–1039.

[28] Clarke, F.H., Ledyaev, Yu., Stern, R.I. , Wolenski, P.R., *Nonsmooth analysis and control theory*, Graduate texts in Mathematics **178**, (Springer-Verlag, New-York, 1998).

[29] Combettes, P.L., Quasi-Fejerian analysis of some optimization algorithms, in Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, (D. Butnariu, Y. Censor, and S. Reich, Eds.), New York: Elsevier, 2001, 115-152.

[30] Combettes, P.L., Wajs, V.R., Signal recovery by proximal forward-backward splitting., Multiscale Model. Simul., 4 (2005), 1168–1200.

[31] Coste, M., *An introduction to o-minimal geometry*, RAAG Notes, 81 p., Institut de Recherche Mathématiques de Rennes, November 1999.

[32] Curry, H.B., The method of steepest descent for non-linear minimization problems, Quart. Appl. Math., **2** (1944), 258–261.

[33] Dedieu, J.P., Méthodes d'analyse globale en algèbre linéaire et optimisation, Cours DEA, 126 pages, Université Toulouse Paul Sabatier (en ligne).

[34] Palis, J.,& De Melo, W., *Geometric theory of dynamical systems. An introduction,* (Translated from the Portuguese by A. K. Manning), Springer-Verlag, New York-Berlin, 1982.

[35] Donoho, D. L., Compressed Sensing, IEEE Trans. Inform. Theory **4** (2006), 1289–1306.

[36] van den Dries, L., Tame topology and o-minimal structures. London Mathematical Society Lecture Note Series, **248**, Cambridge University Press, Cambridge, (1998) x+180 pp.

[37] van den Dries, L., & Miller, C., Geometric categories and o-minimal structures, *Duke Math. J.* **84** (1996), 497-540.

[38] Edelman, A., Arias, A., Smith, S.T., The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl. **20** (2) (1999), pp. 303–353.

[39] Grippo, L., Sciandrone, M., Globally convergent block-coordinate techniques for unconstrained optimization, Optimization Methods and Software, **10** (4), (1999), 587–637.

[40] Hare, W., Sagastizábal, C. Computing proximal points of nonconvex functions, Math. Program., **116** (2009), **1-2**, Ser. B, 221–258.

[41] Haraux, A., Jendoubi, M.A. Convergence of solutions of second-order gradient-like systems with analytic nonlinearities, J. Differential Equations, **144** (2), (1999), 313–320.

[42] Huang, S.-Z., Takač, P. Convergence in gradient-like systems which are asymptotically autonomous and analytic, Nonlinear Anal., Ser. A, Theory Methods, **46**, (2001), 675–698.

[43] Ioffe, A.D., An invitation to tame optimization, SIAM Journal on Optimization, **19**, no. 4, (2009), 1080–1917.

[44] Iusem A.N., Pennanen T., Svaiter, B.F. Inexact variants of the proximal point algorithm without monotonicity, SIAM Journal on Optimization, **13**, no. 4 (2003), 1894–1097.

[45] Jokar S., Pfetsch M.E., Exact and approximate sparse solutions of underdetermined linear equations, ZIB-report 07-0 ZIB, March 2007.

[46] Kruger, A.Y., About regularity of collections of sets, Set Valued Analysis, **14**, (2006), 187–206.

[47] Kurdyka, K., On gradients of functions definable in o-minimal structures, Ann. Inst. Fourier, **48**, (1998), 769-783.

[48] Lageman, C., Pointwise convergence of gradient-like systems, Math. Nachr., **280**, (2007), no. 13-14, 1543-1558.

[49] Lewis, A.S., Active sets, nonsmoothness and sensitivity, SIAM Journal on Optimization, **13**, (2003), 702–725.

[50] Lewis, A.S., Malick, J., Alternating projection on manifolds, Mathematics of Operations Research, **33**, no. 1, (2008), 216-234.

[51] Lewis, A.S., Luke, D.R., Malick, J., Local linear convergence for alternating and averaged nonconvex projections., Found. Comput. Math. **9**, (2009), 485–513.

[52] Lewis, A.S., Wright, S.J., A proximal method for composite minimization, 2010.

[53] Łojasiewicz, S., Une propriété topologique des sous-ensembles analytiques réels, in: *Les Équations aux Dérivées Partielles*, pp. 87–89, Éditions du centre National de la Recherche Scientifique, Paris 1963.

[54] Łojasiewicz, S., Sur la géométrie semi- et sous-analytique, Ann. Inst. Fourier **43**, (1993), 1575-1595.

[55] Mangasarian, L., Minimal support solutions of polyhedral concave programs, Optimization **45**, (1999), 149-162.

[56] Mordukhovich, B., Maximum principle in the problem of time optimal response with nonsmooth constraints, J. Appl. Math. Mech., **40** (1976), 960–969 ; [translated from Prikl. Mat. Meh. 40 (1976), 1014–1023].

[57] Mordukhovich, B., *Variational analysis and generalized differentiation. I. Basic theory*, Grundlehren der Mathematischen Wissenschaften, **330**, Springer-Verlag, Berlin, 2006.

[58] Nesterov, Yu., Accelerating the cubic regularization of Newton's method on convex problems, Math. Program., **112** (2008), no. 1, Ser. B, 159–181.

[59] Nesterov, Yu., Nemirovskii, A., *Interior-point polynomial algorithms in convex programming*, SIAM Studies in Applied Mathematics, **13**, Philadelphia, PA, 1994.

[60] Pennanen, T., Local convergence of the proximal point algorithm and multiplier methods without monotonicity, Math. Oper. Res. **27**, (2002), 170–191 .

[61] Peypouquet, J., Sorin, S., Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time, J. Convex Analysis, **17**, (2010), 1113–1163.

[62] Poliquin, R.A., Rockafellar, R.T., Thibault, L., Local differentiability of distance functions, Trans. AMS, **352**, (2000), 5231–5249.

[63] Rockafellar, R.T. , Wets, R., *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften, **317**, Springer, 1998.

[64] Simon, L., Asymptotics for a class of non-linear evolution equations, with applications to geometric problems, Ann. of Math., **118** (1983), 525–571.

[65] Solodov, M.V., Svaiter, B.F., A hybrid projection-proximal point algorithm, Journal of Convex Analysis, **6**, no. 1, (1999), 59–70.

[66] Solodov, M.V., Svaiter, B.F., A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator, Set-Valued Analysis, **7**, (1999), 323–345.

[67] Solodov, M.V., Svaiter, B.F., A unified framework for some inexact proximal point algorithms, Numerical Functional Analysis and Optimization, **22**, (2001), 1013-1035.

[68] Wright, S.J., Identifiable surfaces in constrained optimization. SIAM Journal on Control and Optimization, **31**, (1993), 1063-1079.

[69] Wright, S.J., Accelerated block-coordinate relaxation for regularized optimization, 2010.

[70] Zhang, H.H., Ahn, J., Lin, X., Park, C. Gene selection using support vector machines with non-convex penalty, Bioinformatics, **22**, (2006), 88-95.