The Bundle Proximal Point Method: An efficient method for solving nonsmooth convex and nonconvex problems

Jean-Jacques Strodiot

University of Namur, Belgium

The 7th International Workshop on Optimization and Its Applications, Hanoi, July 31 - August 2, 2008

OUTLINE

Section 1

Bundle Concept

Nondifferentiable convex minimization problems

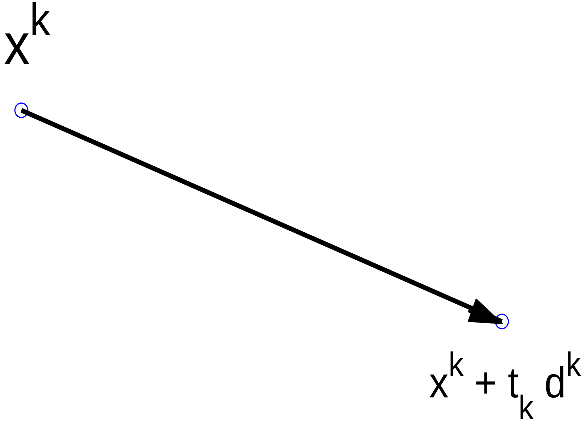
Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a nondifferentiable convex function

- Aim : design an efficient numerical method for finding the minimum of f
- Method : generate a sequence $\{x^k\}$ from a starting point x^0 converging to a minimum of f
- Strategy to pass from x^k to x^{k+1} :
  - define a search direction d^k
  - set $x^{k+1} = x^k + t_k d^k$ where $t_k > 0$ is a well-chosen stepsize
- Usually d^k is chosen in order to reduce the value of f

Descent direction

$d^k \in \boldsymbol{R}^n$ is a descent direction at x^k for f if

$$\exists\, \delta > 0 \text{ such that } \forall t \in (0, \delta] \quad f(x^k + td^k) < f(x^k)$$



When f is convex and differentiable

- $d \in \mathbb{R}^n$ is a descent direction at x for $f \Leftrightarrow \nabla f(x)^T d < 0$
  $d = -\nabla f(x) \neq 0 \Rightarrow d$ is a descent direction at x for f
- Gradient method : $x^{k+1} = x^k + t_k d^k$ with $d^k = -\nabla f(x^k)$
  and $t_k > 0$ well chosen
- x^* is a minimum of $f \Leftrightarrow \nabla f(x^*) = 0$

When f is convex and nondifferentiable

- The subdifferential of f at x :

$$\partial f(x) = \{s \in \mathbb{R}^n \,|\, \forall y \in \mathbb{R}^n \; f(y) \geq f(x) + \langle s, y - x\rangle\}$$

- Elements of $\partial f(x)$ are called subgradients of f at x
- **Geometric interpretation** The inequality

$$\forall y \in \boldsymbol{R}^n \quad f(y) \geq f(x) + \langle s, y - x\rangle$$

means that s is the slope of an affine function

  - which is below f
  - which passes through the point $(x, f(x))$

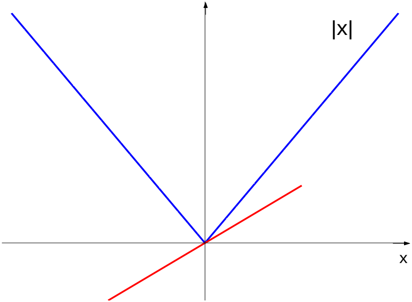
Examples

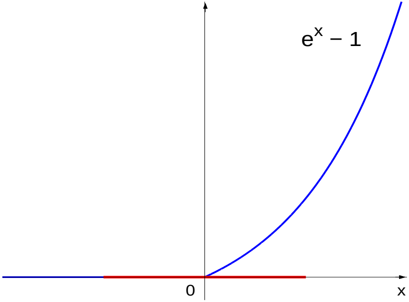
- $f(x) = |x|$

$$\partial f(0) = [-1, 1], \quad \partial f(x) = \{1\} \text{ if } x > 0, \quad \partial f(x) = -1 \text{ if } x < 0$$

- $f(x) = e^x - 1$ if $x \geq 0$ and 0 if $x < 0$

$$\partial f(0) = [0, 1], \quad \partial f(x) = \{e^x\} \text{ if } x > 0, \quad \partial f(x) = 0 \text{ if } x < 0$$





Descent directions and Optimality

The following properties are equivalent :

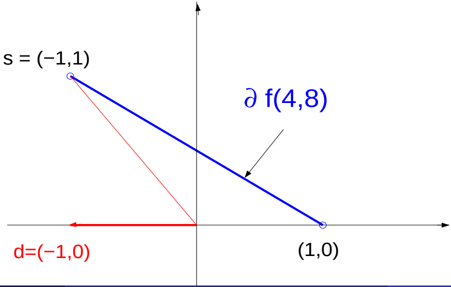
- d is a descent direction at x for f
- $f'(x; d) < 0$
- $\langle s, d \rangle < 0$ for all $s \in \partial f(x)$.

There exists a descent direction at x for f if and only if $0 \notin \partial f(x)$

Optimality : x^* is a minimum of $f \Leftrightarrow 0 \in \partial f(x^*)$

Opposite of a subgradient

- We know : when f is differentiable at x, $d = -\nabla f(x)$ is a descent direction at x if $\nabla f(x) \neq 0$
- When f is not differentiable, the opposite of a subgradient at x is not necessarily a descent direction at x
- Example : $f(x) = \max\{-x_1 - x_2, -x_1 + x_2, x_1\}$
  - the subdifferential $\partial f(4,8)$ is the convex hull of $(-1,1)$ and $(1,0)$
  - the vector $(1,0)$ belongs to $\partial f(4,8)$ but $d = -(1,0)$ is not a descent direction. Indeed, for $s = (-1,1) \in \partial f(4,8)$, we have $\langle s, d\rangle = 1 > 0$



Steepest descent direction

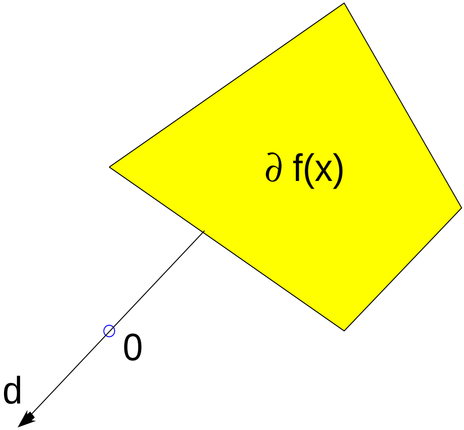
Let $x \in \mathbb{R}^n$ such that $0 \notin \partial f(x)$. To get the steepest descent direction at x for f,

$$\text{replace} \quad \min_{\|d\| \leq 1} \langle \nabla f(x), d \rangle \quad \text{by} \quad \min_{\|d\| \leq 1} \max_{s \in \partial f(x)} \langle s, d \rangle$$

Let $x \in \mathbb{R}^n$ such that $0 \notin \partial f(x)$. Then

- the steepest descent direction at x for f is the vector $-\frac{m}{\|m\|}$ where m is the vector of minimum norm in $\partial f(x)$
- the vector of minimum norm in $\partial f(x)$ exists and is unique. It is the orthogonal projection of 0 onto $\partial f(x)$

Steepest descent direction. Illustration



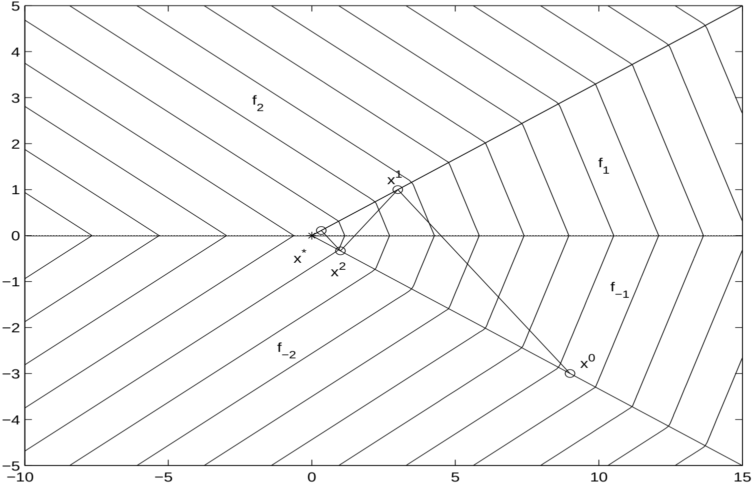
Steepest descent method

0. Choose a starting point x^0 and set $k = 0$,
1. Compute m, the vector of minimum norm in $\partial f(x^k)$
2. If $m = 0$, Stop, x^k is a minimum of f
3. Set $d^k = -m$ and find t_k solution of the problem

$$\min_{t>0} f(x^k + td^k)$$

4. Set $x^{k+1} = x^k + t_k d^k$
5. Set $k := k + 1$ and go to Step 1.

Nonconvergence of the steepest descent method



$\{x^k\}_k$ converges (very slowly) to the origin $x^* = 0$ which is not optimal

Evaluating the whole subdifferential is too expensive

- For computing the vector of minimum norm of $\partial f(x)$, it is supposed that the whole subdifferential is known.
  Very often it is too expensive
- Example :
  Let $\lambda_{\max}(M) =$ the largest eigenvalue of a symmetric matrix M.
  It is easy to see that

$$\partial \lambda_{\max}(M) = \text{ conv } \{ qq^T \mid q^T q = 1, \; Mq = \lambda_{\max}(M) q \}$$

  - To compute this set, all the normalized eigenvectors associated with $\lambda_{\max}$ must be found. This is too expensive
  - However, computing one subgradient is much cheaper because it amounts to only determine one eigenvector

Conclusion

It would be important to design an algorithm which is convergent and where, at each iteration,

- only the value of f and
- one subgradient of f

are used.

The procedure that gives $f(x)$ and one subgradient of f at x is called an oracle

Strategy : Use the subgradients given by the oracle at points near x to build a descent direction at x, i.e., to approximate $\partial f(x)$

We need to introduce the approximate subdifferential of f at x^k

Approximate subdifferential

Let $f : \boldsymbol{R}^n \to \boldsymbol{R}$ be convex and let $\varepsilon \geq 0$.

- The ε–subdifferential of f at $x \in \boldsymbol{R}^n$ is the set

$$\partial_\varepsilon f(x) = \{s \in \boldsymbol{R}^n \,|\, f(y) \geq f(x) + \langle s, y - x\rangle - \varepsilon \;\forall y \in \boldsymbol{R}^n\}$$

  Each element $s \in \partial_\varepsilon f(x)$ is called an ε–subgradient of f at x

- **Geometric interpretation** The inequality

$$\forall y \in \boldsymbol{R}^n \quad f(y) \geq f(x) + \langle s, y - x\rangle - \varepsilon$$

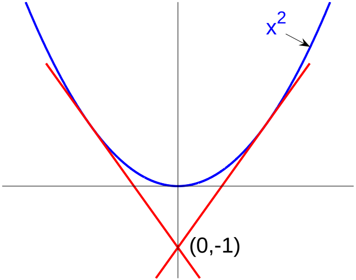
  means that s is the slope of an affine function
  - which is below f
  - which passes through the point $(x, f(x) - \varepsilon)$

Example $f(x) = x^2$

The ε–subdifferential of f at $x = 0$ is

$$\partial_\varepsilon f(0) = [-2\sqrt{\varepsilon}, 2\sqrt{\varepsilon}]$$

This set is reduced to the gradient of f at 0 when $\varepsilon = 0$

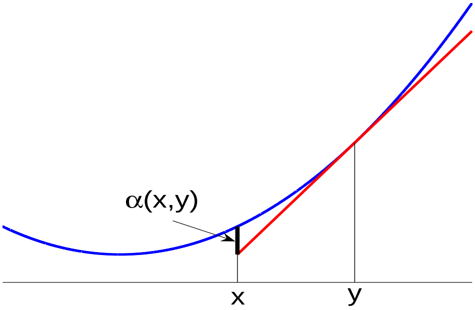


$$\partial_1 f(0) = [-2, 2]$$

Transportation formula

Let $x, y \in \boldsymbol{R}^n$ and let $s(y) \in \partial f(y)$. Then $s(y) \in \partial_{\alpha(x,y)} f(x)$ where $\alpha(x, y)$ is the linearization error

$$\alpha(x, y) \equiv f(x) - f(y) - s(y)^T(x - y)$$



Let $\varepsilon > 0$. Then $s(y) \in \partial_\varepsilon f(x) \iff \alpha(x, y) \leq \varepsilon$

The direction-finding problem

Basic Assumption. At every point $y \in \boldsymbol{R}^n$, only the value $f(y)$ and a subgradient $s(y) \in \partial f(y)$ are available (by means of an oracle)

Using the approximate subdifferential, we replace the direction-finding problem

$$\left\{ \begin{array}{ll} \min & \|s\| \\ \text{s.t.} & s \in \partial f(x) \end{array} \right. \quad \text{by} \quad \left\{ \begin{array}{ll} \min & \|s\| \\ \text{s.t.} & s \in \partial_\varepsilon f(x) \end{array} \right.$$

Our aim is

- to construct an approximation of $\partial_\varepsilon f(x)$ thanks to the oracle mentioned in the Basic Assumption.
- This will be done by using subgradients computed at points in a neighborhood of x
- to get a descent direction when the approximation is "good"

Dual approach of Bundle Methods

- Suppose x^k is the current iteration point and that $y^1, \ldots, y^p$ are points in a neighborhood of x^k. For simplicity, suppose $y^p = x^k$.
- Let $s^j \in \partial f(y^j)$, $j = 1, \ldots, p$. We have

$$s^j \in \partial_{\alpha_j^k} f(x^k), \quad j = 1, \ldots, p$$

  where $\alpha_j^k = \alpha(x^k, y^j) = f(x^k) - f(y^j) - s^{jT}(x^k - y^j)$ is the linearization error. (Here $\alpha_p^k = 0$)
- The set $\{(s^j, \alpha_j^k)\}_{1 \leq j \leq p}$ is called a bundle. It represents a collection of approximate subgradient information available around the point x^k.
- Assume $\alpha_j^k \leq \varepsilon$, $j = 1, \ldots, p$. Then $s^j \in \partial_\varepsilon f(x^k)$ for $j = 1, \ldots, p$

Inner approximation of the ε-subdifferential

The bundle $\{(s^j, \alpha_j^k)\}_{1 \le j \le p}$ allows us to build the following inner approximation of $\partial_\varepsilon f(x^k)$:

$G(x^k, \varepsilon) =$

$$\left\{ \sum_{j=1}^{p} \lambda_j s^j \mid \lambda_j \ge 0, j = 1, \dots, p, \sum_{j=1}^{p} \lambda_j = 1, \sum_{j=1}^{p} \lambda_j \alpha_j^k \le \varepsilon \right\}$$

- $G(x^k, \varepsilon)$ is a convex subset of $\partial_\varepsilon f(x^k)$
- Replace $\partial_\varepsilon f(x^k)$ by $G(x^k, \varepsilon)$ to compute the search direction

Search of a descent direction

Strategy : Replace $\partial_\varepsilon f(x^k)$ by its approximation $G(x^k, \varepsilon)$

$\Rightarrow$ direction $d^k \equiv$ the opposite of the vector of minimum norm in $G(x^k, \varepsilon)$. This can be done as follows :

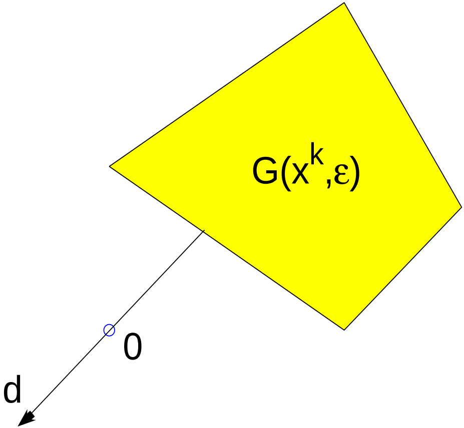
<u>Step 1.</u> Solve the convex quadratic problem

$$QD(x^k, \varepsilon) \left\{ \begin{array}{ll} \min & \frac{1}{2} \| \sum_{j=1}^{p} \lambda_j s^j \|^2 \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j = 1, \ \lambda_j \geq 0, j = 1, \ldots, p \\ & \sum_{j=1}^{p} \lambda_j \alpha_j^k \leq \varepsilon \end{array} \right.$$

to obtain the solution $\lambda_j^k, j = 1, \ldots, p$

<u>Step 2.</u> Set $d^k = -\sum_{j=1}^{p} \lambda_j^k s^j$

Search of a descent direction. Illustration



Serious step versus null step

- $G(x^k, \varepsilon) \approx \partial_\varepsilon f(x^k) \Rightarrow d^k$ may not be a descent direction at x^k
- The linesearch must have two exits corresponding to :
  - (**a serious step**) there exists $t_k > 0$ not too small such that the reduction $f(x^k) - f(x^k + t_k d^k)$ is sufficiently large, i.e., satisfies an Armijo-type condition. In that case : $x^{k+1} = x^k + t_k d^k$
  - (**a null step**) no such t_k exists. In that case
    - $x^{k+1} = x^k$ and the approximation $G(x^k, \varepsilon)$ must be improved.
    - Practically the step $t_k > 0$ is reduced along d^k until the subgradient $s(t_k) \in \partial f(x^k + t_k d^k)$ given by the oracle belongs to $\partial_{m_3 \varepsilon} f(x^k)$ where $0 < m_3 < 1$.
    - Add $(s(t_k), \alpha(x^k, x^k + t_k d^k))$ to the bundle

Linearly constrained problems

Consider the problem (P) : $\min f(x)$ s.t. $Ax \leq b$
where $f : \mathbb{R}^n \to \mathbb{R}$ is convex, A is an $m \times n$ matrix of rank m and $b \in \mathbb{R}^m$.
We have

- x^* optimal solution to $(P) \Leftrightarrow 0 \in \partial(f + \psi_S)(x^*)$ where S is the feasible set and ψ_S denotes the indicator function of S.
- $\partial_\varepsilon (f + \psi_S)(x) =$

$$\cup_{0 \leq \varepsilon_0 \leq \varepsilon} \{\partial_{\varepsilon_0} f(x) + \{A^T v \mid v \geq 0, v^T(b - Ax) \leq \varepsilon - \varepsilon_0\}\}$$

The bundle $\{(s^j, \alpha_j^k)\}_{1 \leq j \leq p}$ allows us to build the following inner approximation of $\partial_\varepsilon(f + \psi_S)(x^k)$:

$G(x^k, \varepsilon) =$

$$\left\{ \sum_{j=1}^p \lambda_j s^j + A^T v \;\middle|\; \begin{array}{l} \lambda_j \geq 0, j = 1, \ldots, p, \sum_{j=1}^p \lambda_j = 1, v \geq 0 \\ \sum_{j=1}^p \lambda_j \alpha_j^k + v^T(b - Ax^k) \leq \varepsilon \end{array} \right\}$$

Linearly constrained problems

Strategy : Replace $\partial_\varepsilon(f + \psi_S)(x^k)$ by its approximation $G(x^k, \varepsilon)$

$\Rightarrow$ direction $d^k \equiv$ the opposite of the vector of minimum norm in $G(x^k, \varepsilon)$. This can be done as follows :

Step 1. Solve the convex quadratic problem

$$QD(x^k, \varepsilon) \begin{cases} \min & \frac{1}{2} \| \sum_{j=1}^{p} \lambda_j s^j + A^T v \|^2 \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j = 1,\ \lambda_j \geq 0, j = 1, \ldots, p, v \geq 0 \\ & \sum_{j=1}^{p} \lambda_j \alpha_j^k + v^T (b - Ax^k) \leq \varepsilon \end{cases}$$

to obtain the solution $\lambda_j^k, j = 1, \ldots, p$

Step 2. Set $d^k = -\sum_{j=1}^{p} \lambda_j^k s^j - A^T v$

References

- Strodiot, J.J., Nguyen, V.H., and Heukemes, N., *Epsilon-optimal solutions in nondifferentiable convex programming and some related questions*. Mathematical Programming, 1983, Vol.25, pp.307 – 328.
- Nguyen, V.H. and Strodiot, J.J., *A linearly constrained algorithm not requiring derivative continuity*. Engineering Structures, 1984, Vol.6, pp.7 – 11.

Primal Approach : Cutting Plane Model

The bundle $\{(s^j, \alpha_j^k)\}_{1 \leq j \leq p}$ allows us to build the following inner approximation of $\partial_\varepsilon f(x^k)$:

$$G(x^k, \varepsilon) = \left\{ \sum_{j=1}^{p} \lambda_j s^j \mid \lambda_j \geq 0, \sum_{j=1}^{p} \lambda_j = 1, \sum_{j=1}^{p} \lambda_j \alpha_j^k \leq \varepsilon \right\}$$

The bundle $\{(s^j, \alpha_j^k)\}_{1 \leq j \leq p}$ also allows us to build the following piecewise linear convex approximation of f :

$$f^p(x) = \max_{1 \leq j \leq p} \{ f(y^j) + \langle s^j, x - y^j \rangle \} \leq f(x)$$

We have : $\partial_\varepsilon f^p(x^k) = G(x^k, \varepsilon)$

We will consider other approximations of f in the Proximal Point Method

Nonconvex unconstrained problems

Assume $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz, i.e., f satisfies the property : for each $x \in \mathbb{R}^n$, there exist $\varepsilon_x > 0$ and $L_x \geq 0$ s.t.

$$|f(y) - f(z)| \leq L_x \|y - z\| \quad \text{for all } y, z \in x + \varepsilon_x B$$

where B denotes the open unit ball in $\mathbb{R}^n$.
The generalized gradient of f at x is defined as

$$\partial f(x) = \{s \in \mathbb{R}^n \,|\, f^\circ(x; v) \geq \langle s, v \rangle \text{ for all } v \in \mathbb{R}^n\}$$

where $f^\circ(x; v)$ denotes the generalized directional derivative of f at x :

$$f^\circ(x; v) = \limsup_{y \to x, \lambda \downarrow 0} \frac{f(y + \lambda v) - f(y)}{\lambda}$$

Difficulties for nonconvex problems

- When f is nonconvex, we do not have the subgradient inequality :

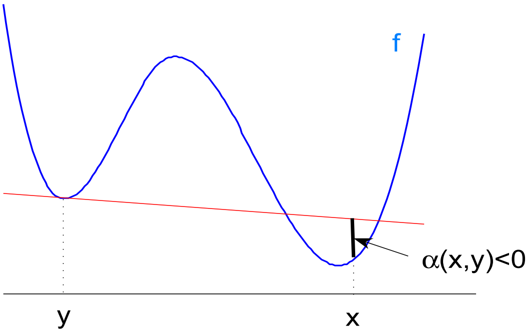
$$s(x) \in \partial f(x) \quad \Leftrightarrow \quad f(y) \geq f(x) + \langle s(x), y - x \rangle \quad \text{for all } y \in R^n$$

- As a consequence the linearization error at x :
  $\alpha(x, y) = f(x) - f(y) - s(y)^T(x - y)$ may become negative and a subgradient computed very far from x can be considered as an approximating subgradient at x.
  Furthermore the cutting plane model is no longer an approximation of f from below.
- To cope with this difficulty, we replace $\alpha(x, y)$ by

$$\beta(x, y) = \max\{\alpha(x, y),\ c\|y - x\|^2\}$$

where $c > 0$ (c can be set to 0 when f is convex)

Linearization Error



The direction-finding problem

The bundle $\{(s^j, \beta_j^k)\}_{1 \le j \le p}$ allows us to build the following inner approximation of the generalized gradient $\partial f(x^k)$:

$G(x^k, \varepsilon) =$

$$\left\{ \sum_{j=1}^{p} \lambda_j s^j \mid \lambda_j \ge 0, j = 1, \ldots, p, \sum_{j=1}^{p} \lambda_j = 1, \sum_{j=1}^{p} \lambda_j \beta_j^k \le \varepsilon \right\}$$

Step 1. Solve the convex quadratic problem

$$QD(x^k, \varepsilon) \begin{cases} \min & \frac{1}{2} \| \sum_{j=1}^{p} \lambda_j s^j \|^2 \\ \text{s.t.} & \sum_{j=1}^{p} \lambda_j = 1, \ \lambda_j \ge 0, j = 1, \ldots, p \\ & \sum_{j=1}^{p} \lambda_j \beta_j^k \le \varepsilon \end{cases}$$

to obtain the solution $\lambda_j^k, j = 1, \ldots, p$

Step 2. Set $d^k = -\sum_{j=1}^{p} \lambda_j^k s^j$

Convergence and references

- Convergence of $\{x^k\}$ to a stationary point x^* ($0 \in \partial f(x^*)$) is obtained when f is weakly semi-smooth, i.e., when, for any x and v, $f'(x; v)$ exists and

$$t_j \downarrow 0, \quad s_j \in \partial f(x + t_j v) \quad \text{imply that } \langle s_j, v\rangle \to f'(x; v)$$

- References (Generalization to the linearly constraint case)
  - Strodiot, J.J. and Nguyen, V.H. *On the numerical treatment of the inclusion* $0 \in \partial f(x)$. In : Topics in Nonsmooth Mechanics (ed. by Moreau J.J., Panagiotopoulos, P.D., and Strang, G.) Birkhauser Verlag Basel, 1988, pp.267 − 294.
  - Bihain, A., Nguyen, V.H. and Strodiot, J.J., *A reduced subgradient algorithm*. Mathematical Programming Study, 1987, Vol.30, pp.127 − 149.

Section 2

Bundle Proximal Point Methods for Minimization Problems

Moreau-Yosida Regularization

- Strategy : Construct a differentiable convex function F approximating the nondifferentiable convex function f in such a way that the minima of f and F coincide
- Classical methods as gradient methods or BFGS methods can be used for minimizing F.
  However these methods are often non implementable for minimizing F
- In this section, other approximations than polyhedral models can be considered

Moreau-Yosida Regularization. Definition

- Let $f : \boldsymbol{R}^n \to \boldsymbol{R}$ convex and $c > 0$. The function $F : \boldsymbol{R}^n \to \boldsymbol{R}$ defined by
$$F(x) = \min_{y \in \boldsymbol{R}^n} \{f(y) + \frac{c}{2} \parallel y - x \parallel^2\}$$
is called the Moreau–Yosida regularization of f
- The unique minimum denoted by $p_f(x)$ is called the proximal point of x associated with f
- When $f = \psi_C$ is the indicator function associated with a convex subset C :
$$F(x) = \min_{y \in \mathbb{R}^n} \{\psi_C(y) + \frac{c}{2}\|y - x\|^2\} = \min_{y \in C} \frac{c}{2}\|y - x\|^2$$
In that case, $p_f(x)$ is the orthogonal projection of x on C (hence the name proximal point of x)

Moreau-Yosida Regularization. Properties

- The Moreau–Yosida regularization F is finite everywhere, convex and differentiable
- Its gradient is

$$\nabla F(x) = s_f(x) = c\ [x - p_f(x)]\ \in\ \partial f(p_f(x))$$

- Its conjugate is $F^* : \boldsymbol{R}^n \to \boldsymbol{R} \quad F^*(s) = f^*(s) + \frac{1}{2c}\|s\|^2$
- Moreover, for all x and x' in $\boldsymbol{R}^n$,

$$\|\nabla F(x) - \nabla F(x')\|^2 \leq c\langle \nabla F(x) - \nabla F(x'), x - x' \rangle$$

and

$$\|\nabla F(x) - \nabla F(x')\| \leq c\|x - x'\|$$

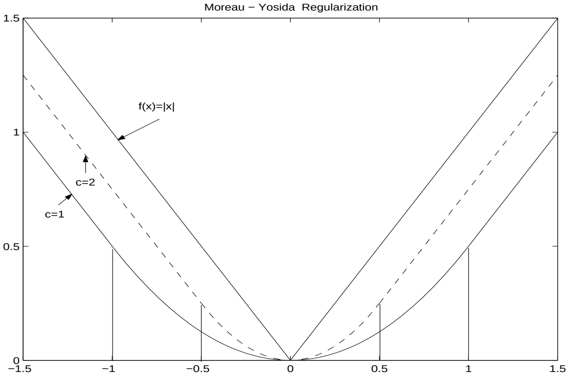
i.e., ∇F is Lipschitz continuous on $\boldsymbol{R}^n$ with constant c

Moreau-Yosida Regularization. Example

Let $f(x) = |x|$. The Moreau-Yosida regularization of f is

$$F(x) = \begin{cases} \frac{c}{2}x^2 & \text{if} \quad |x| \leq \frac{1}{c} \\ |x| - \frac{1}{2c} & \text{if} \quad |x| > \frac{1}{c} \end{cases}$$

The minima of f and F are the same



Main Result

- $\inf_{x\in \boldsymbol{R}^n} F(x) = \inf_{x\in \boldsymbol{R}^n} f(x)$ (equality in $\boldsymbol{R} \cup \{+\infty\}$)
- The following statements are equivalent
    - x minimizes f
    - $p_f(x) = x$
    - x minimizes F
    - $f(p_f(x)) = f(x)$
    - $F(x) = f(x)$

Proximal Point Algorithm

- Minimizing f is equivalent to finding a fixed point of p_f.
  Hence the fixed point iteration : $x^{k+1} = p_f(x^k)$, i.e.,

$$x^{k+1} = \arg\min_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{c}{2} \|y - x^k\|^2 \right\}$$

  This algorithm is called the Proximal Point Algorithm
- Since the gradient of the Moreau-Yosida regularization F at x^k is

$$\nabla F(x^k) = c(x^k - p_f(x^k))$$

  we have

$$x^{k+1} = p_f(x^k) \Leftrightarrow x^{k+1} = x^k - \frac{1}{c} \nabla F(x^k)$$

  So the proximal point algorithm is nothing else that the gradient method with fixed stepsize applied to the Moreau-Yosida regularization

Proximal Point Algorithm

Step 1. Choose $x^0 \in \boldsymbol{R}^n$ and $t_0 > 0$. Set $k = 0$.

Step 2. Compute $x^{k+1} = p_f(x^k)$ by solving the problem

$$\min_{y \in \boldsymbol{R}^n} \{f(y) + \frac{1}{2t_k}\|y - x^k\|^2\}$$

Step 3. If $x^{k+1} = x^k$ STOP, x^{k+1} is a minimum of f

Step 4. Choose $t_{k+1} > 0$. Replace k by $k+1$ and go to Step 2.

Interpretation : Since $x^{k+1} = \text{argmin}_y \{f(y) + \frac{1}{2t_k}\|y - x^k\|^2\}$:

$$\gamma^k \equiv \frac{1}{t_k}(x^k - x^{k+1}) \in \partial f(x^{k+1})$$

So the prox–iteration : $x^{k+1} = x^k - t_k\gamma^k$ with $\gamma^k \in \partial f(x_{k+1})$

Convergence

Let $\{x^k\}_{k \in \boldsymbol{N}}$ be the sequence generated by the proximal point algorithm

If $\sum_{k=0}^{+\infty} t_k = +\infty$, then

- $\lim_{k \to \infty} f(x^k) = f^* = \inf_{x \in \boldsymbol{R}^n} f(x)$
- the sequence $\{x^k\}$ converges to some minimum of f (if there is one).

In particular, if $t_k = 1/c$ for all k with $c > 0$, the sequence $\{x^k\}$ generated by the proximal point algorithm converges to some minimum of f (if there exists one)

Approximate Proximal Point Method

- Very often the problem of finding $p_f(x^k)$ i.e., of solving

$$\min_{y \in \boldsymbol{R}^n} \{ f(y) + \frac{1}{2t_k} \|y - x^k\|^2 \}$$

  is as difficult as solving the initial problem

- Strategy : replace f by a simpler convex function φ^k such that the subproblems

$$\min_{y \in \boldsymbol{R}^n} \{ \varphi^k(y) + \frac{1}{2t_k} \|y - x^k\|^2 \}$$

  are easier to solve and the convergence is preserved

- The function φ^k must be built under the assumption :
  At every point $y \in \boldsymbol{R}^n$, only the value $f(y)$ and a subgradient $s(y) \in \partial f(y)$ are available

Example where the subproblems are easy to solve

If φ^k is chosen as a piecewise linear function :

$$\varphi^k(x) = \max_{1 \le j \le m} \{a_j^T x + b_j\}$$

then the subproblem

$$\min_{y \in \mathbb{R}^n} \{\varphi^k(y) + \frac{1}{2t_k} \|y - x^k\|^2\}$$

can be rewritten as

$$\begin{cases} \min & v + \frac{1}{2t_k} \|y - x^k\|^2 \\ \text{s.t.} & a_j^T y + b_j \le v, \; j = 1, \ldots, m. \end{cases}$$

This problem is a convex quadratic problem. Very efficient methods exist for solving it

σ-approximation of f. A General Algorithm

- Let $\sigma \in (0,1)$ and $x^k \in \boldsymbol{R}^n$.
- A convex function φ^k is said to be a σ-approximation of f at x^k if $\varphi^k \leq f$ and

$$f(x^k) - f(x^{k+1}) \geq \sigma \; [\; f(x^k) - \varphi^k(x^{k+1}) \;]$$

where $x^{k+1} = \arg\min\{\varphi^k(y) + \frac{1}{2t_k}\|y - x^k\|^2\}$

A General Algorithm

Let $\sigma \in (0,1)$ and $\{t_k\}_{k \in \boldsymbol{N}_0}$ be a sequence of positive numbers. Choose a starting point x^0 and set $k = 0$.

- Find φ^k a σ-approximation of f at x^k and denote x^{k+1} the unique solution of the subproblem
- Increase k by 1 and start again.

Convergence of the General Algorithm

Let $\{x^k\}$ be the sequence generated by the General Algorithm.

- If $\sum_{k=1}^{+\infty} t_k = +\infty$, then $f(x^k) \searrow \bar{f} = \inf_x f(x)$
- If, in addition, $t_k \leq \bar{t}$ for all k, then $x^k \to x^*$ where x^* is a minimum of f (provided that some minimum exists).

How to construct σ-approximations of f ?

An Example. The Cutting Plane Model

- Let x^k be the current point. Set $y_0^k = x^k$
- First Model $\varphi_1^k(y) = f(y_0^k) + \langle s_0^k, y - y_0^k \rangle$ where $s_0^k \in \partial f(y_0^k)$
- Solve

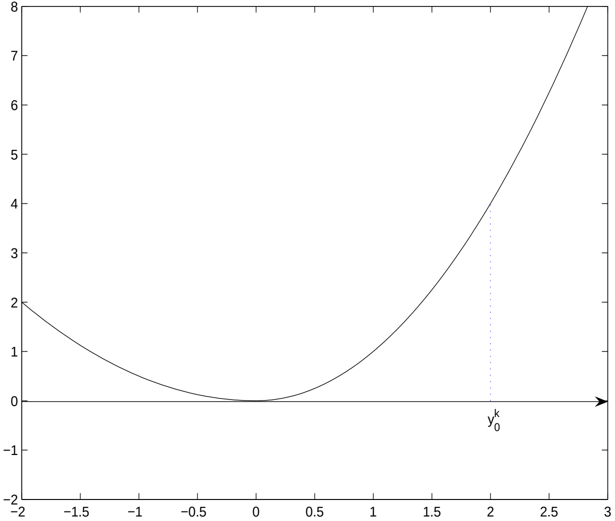
$$(P_1^k) \quad \min_y \{\varphi_1^k(y) + \frac{1}{2t_k}\|y - x^k\|^2\} \quad \text{to get } y_1^k$$

- If $f(x^k) - f(y_1^k) \geq \sigma[f(x^k) - \varphi_1^k(y_1^k)]$, then φ_1^k is a σ-approximation of f at x^k. Set $x^{k+1} = y_1^k$

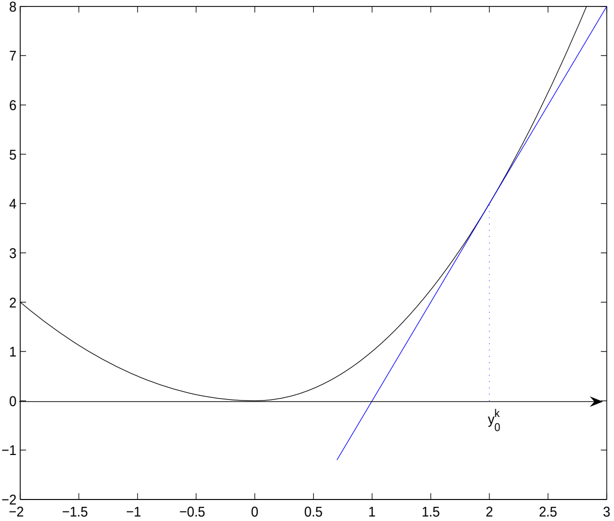
- Otherwise improve the model as follows :

$$\varphi_2^k(y) = \max_{j=0,1}\{f(y_j^k) + \langle s_j^k, y - y_j^k \rangle\}$$

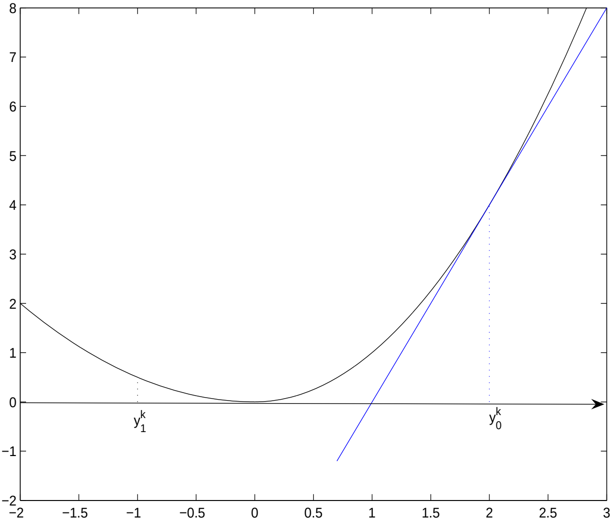
Building σ-approximations



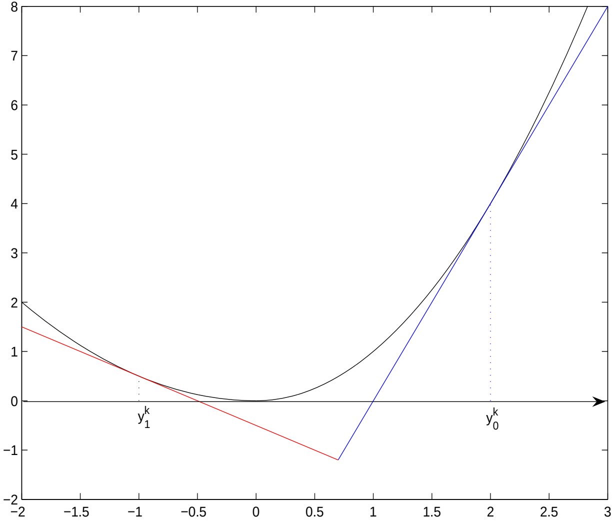
Building σ-approximations



Building σ-approximations



Building σ-approximations



Building σ-approximations at x^k

Serious Step Algorithm.

Let $x^k \in \boldsymbol{R}^n$ and $\sigma \in (0,1)$. Set $i = 0$ and $y_0^k = x^k$

Step 1. Consider the model

$$\varphi_{i+1}^k(y) = \max_{0 \le j \le i} \{ f(y_j^k) + \langle s_j^k, y - y_j^k \rangle \}$$

and solve the problem

$$(P_{i+1}^k) \quad \min_y \{ \varphi_{i+1}^k(y) + \frac{1}{2t_k} \|y - x^k\|^2 \} \text{ to get } \; y_{i+1}^k$$

Step 2. If $f(x^k) - f(y_{i+1}^k) \geq \sigma[f(x^k) - \varphi_{i+1}^k(y_{i+1}^k)]$, then set

$$x^{k+1} = y_{i+1}^k \text{ and STOP}\,; \; x^{k+1} \text{ is a serious step}$$

Step 3. Increase i by 1 and go to Step 1.

Three properties of the model functions φ_i^k

By construction, for each $y \in \mathbb{R}^n$, we have

$$\varphi_{i+1}^k(y) = \max_{0 \le j \le i} \{ f(y_j^k) + \langle s_j^k, y - y_j^k \rangle \} \quad \text{for } i = 0, 1, \ldots$$

- So we get : (**C1**) $\varphi_i^k \le f$ and (**C2**) $\varphi_{i+1}^k \ge f(y_i^k) + \langle s_i^k, \cdot - y^i \rangle$ for $i = 1, 2, \ldots$
- $y_i^k = \arg\min_y \{ \varphi_i^k(y) + \frac{1}{2t_k} \|y - x^k\|^2 \}$

$$\Rightarrow \quad \gamma_i^k := \frac{1}{t_k}(x^k - y_i^k) \in \partial \varphi_i^k(y_i^k)$$

$$\Rightarrow \quad \varphi_i^k(y) \ge \varphi_i^k(y_i^k) + \frac{1}{t_k} \langle \gamma_i^k, y - y_i^k \rangle := l_i^k(y) \quad \text{for each } y \in \mathbb{R}^n$$

Hence (**C3**) $\varphi_{i+1}^k \ge \varphi_i^k \ge l_i^k$ for $i = 1, 2, \ldots$

Properties that must be satisfied by the model functions

- In order to allow other examples of model functions φ_i^k, we will only impose on them the three properties satisfied by the previous example (see previous slide)
- Let us recall them :

(C1) $\varphi_i^k \leq f$ on $\boldsymbol{R}^n$ for $i = 1, 2, \ldots$
(C2) $\varphi_{i+1}^k \geq f(y_i^k) + \langle s(y_i^k), \cdot - y_i^k \rangle$ on $\boldsymbol{R}^n$ for $i = 1, 2, \ldots$
(C3) $\varphi_{i+1}^k \geq l_i^k$ on $\boldsymbol{R}^n$ for $i = 1, 2, \ldots,$

where

- $s(y_i^k)$ denotes the subgradient of f available at y_i^k
- $l_i^k(y) = \varphi_i^k(y_i^k) + \langle \gamma_i^k, y - y_i^k \rangle$ and $\gamma_i^k = \dfrac{1}{t_k}(x^k - y_i^k)$

Another model for the functions φ_i^k

- Another example : for $i = 1, 2, \dots$

$$\varphi_{i+1}^k(y) = \max\left\{ l_i^k(y), f(y_i^k) + \langle s(y_i^k), y - y_i^k \rangle \right\} \quad \forall y \in \boldsymbol{R}^n$$

  where $l_i^k(y) = \varphi_i^k(y_i^k) + \frac{1}{t_k} \langle \gamma_i^k, y - y_i^k \rangle$

- l_i^k plays the same role as the i linear functions

$$f_k(y_j^k) + \langle s(y_j^k), y - y_j^k \rangle, \; j = 0, \dots, i-1$$

  It is the reason why this function l_i^k is called the aggregate affine function

- The advantage of this example is that it limits the size of the bundle to two elements (and thus the number of constraints in the subproblem)
- Many other examples between these two extreme cases can be considered

Serious Step Algorithm

Let $x^k \in \boldsymbol{R}^n$, $t_k > 0$ and $\sigma \in (0,1)$. Set $i = 1$ and $y_0^k = x^k$

Step 1. Choose a convex model φ_i^k satisfying conditions

$(C1)$–$(C3)$ and solve the problem

$$(P_i^k) \quad \min_y \{\varphi_i^k(y) + \frac{1}{2t_k}\|y - x^k\|^2\} \text{ to get } y_i^k$$

Step 2. If $f(x^k) - f(y_i^k) \geq \sigma[f(x^k) - \varphi_i^k(y_i^k)]$, then set

$x^{k+1} = y_i^k$ and STOP ; x^{k+1} is a serious step

Step 3. Increase i by 1 and go to Step 1.

Convergence

Assume that $\sum t_k = +\infty$ and $t_k \leq \bar{t}$ for all k

- If the sequence $\{x^k\}$ generated by the algorithm is infinite, then $\{x^k\}$ converges to some minimum of f

- If after some k has been reached, the criterion

$$f(x^k) - f(y_i^k) \geq \sigma[f(x^k) - \varphi_i^k(y_i^k)]$$

  is never satisfied, then x^k is a minimum of f

Stopping Criterion

- $\bar{x}$ is an ε–stationary point if there exists

$$s \in \partial_\varepsilon f(\bar{x}) \quad \text{with} \quad \|s\| \le \varepsilon$$

- Since, by optimality of y_i^k, $\gamma_i^k \in \partial \varphi_i^k(y_i^k)$, it is easy to prove that

$$\gamma_i^k \in \partial_{\varepsilon_i^k} f(y_i^k)$$

  where $\varepsilon_i^k = f(y_i^k) - \varphi_i^k(y_i^k)$

- Stopping criterion :

$$\left.\begin{array}{l} f(y_i^k) - \varphi_i^k(y_i^k) \le \varepsilon \\ \|\gamma_i^k\| \le \varepsilon \end{array}\right\} \quad \Rightarrow \quad y_i^k \quad \text{is an } \varepsilon\text{-stationary point}$$

Stopping Criterion. Justification

Assume $0 < \underline{t} \leq t_k \leq \bar{t}$ for all k.

- If the sequence $\{x^k\}$ generated by the previous algorithm is infinite, then $f(y_{i_k}^k) - \varphi_{i_k}^k(y_{i_k}^k) \to 0$ and $\|\gamma_{i_k}^k\| \to 0$ when $k \to +\infty$
- If the sequence $\{x^k\}$ is finite with k the latest index, then $f(y_i^k) - \varphi_i^k(y_i^k) \to 0$ and $\|\gamma_i^k\| \to 0$ when $i \to +\infty$

Bundle Proximal Point Algorithm

Let an initial point $x^0 \in C$, together with a tolerance $\sigma \in (0,1)$, $\varepsilon > 0$, and a positive sequence $\{t_k\}_{k \in \boldsymbol{N}}$. Set $y_0^0 = x^0$ and $k = 0, i = 1$.

Step 1. Choose a piecewise linear convex function φ_i^k satisfying $(C1) - (C3)$ and solve

$$(P_i^k) \quad \min_y \{\varphi_i^k(y) + \frac{1}{2t_k}\|y - x^k\|^2\}$$

to obtain the unique optimal solution y_i^k.

Compute $\gamma_i^k = (x^k - y_i^k)/t_k$

If $\|\gamma_i^k\| \leq \varepsilon$ and $f(y_i^k) - \varphi_i^k(y_i^k) \leq \varepsilon$, then STOP, y_i^k is an ε-stationary point

Step 2. If $f(x^k) - f(y_i^k) \geq \sigma[f(x^k) - \varphi_i^k(y_i^k)]$ then set $x^{k+1} = y_i^k$, $y_0^{k+1} = x^{k+1}$, increase k by 1 and set $i = 0$.

Step 3. Increase i by 1 and go to Step 1.

Convergence

Assume $0 < \underline{t} \leq t_k \leq \bar{t}$ for all k.

- The Bundle Proximal Point Algorithm exits after finitely many iterations with an ε–stationary point
- In other words, there exist k and i such that

$$\|\gamma_i^k\| \leq \varepsilon \text{ and } f(y_i^k) - \varphi_i^k(y_i^k) \leq \varepsilon$$

Numerical Results

The function f is the maximum of five quadratic functions :

$$f_j(x) = x^T C^j x - d^{jT} x, \; j = 1, \ldots, 5$$

where C^j is a $n \times n$ symmetric matrix defined by

$$C^j_{ik} = \exp(\frac{i}{k}) \cos(ik) \sin j, \; i < k \qquad C^j_{ii} = \frac{i}{n} \mid \sin j \mid + \sum_{i \neq k} \mid C^j_{ik} \mid$$

and d^j is a vector in $\boldsymbol{R}^n$ whose components are

$$d^j_i = \exp(i/k) \sin(ij)$$

Choice of the Parameters

- the parameter σ is initialized at 0.4
- the starting point is $x_0 = (1, ..., 1)$
- the stopping criterion for the outer loop is

$$\|x^{k+1} - x^k\| \leq \eta \text{ where } \eta = 10^{-3}$$

- the bundle is emptied after each serious step
- the maximal model has been chosen
- the number of variables is $n = 10$.
- the parameter t_k is constant equal to t

Results and Comments

In the next table, k denotes the number of serious steps, μ the average number of null steps by outer iteration and $c = 1/t$

c	k	μ	Optimal value
1	15	55.8	-0.8414065
25	19	9.47	-0.8413951
50	29	7.27	-0.8412801
75	42	7.14	-0.8411583

Large value of $c \Rightarrow$ more serious steps and less null steps

Small value of $c \Rightarrow$ less serious steps and more null steps

Section 3

Bundle Proximal Point Methods for Equilibrium Problems

Equilibrium Problems

Consider

- $C \subset \mathbb{R}^n$ a nonempty closed convex subset and
- $f : C \times C \to \mathbb{R}$ an equilibrium function, i.e., $f(x, x) = 0 \quad \forall x \in C$.

Problem EP : Find $x^* \in C$ such that $f(x^*, y) \geq 0$ for all $y \in C$.

In this talk we assume that

- $f(x, \cdot) : C \to \mathbb{R}$ is convex and lower semicontinuous for all $x \in C$
- $f(\cdot, y) : C \to \mathbb{R}$ is upper semicontinuous for all $y \in C$

Examples of Equilibrium Problems

- Convex Minimization Problems
  Let $C \subset \mathbb{R}^n$ be closed and convex and let $f(x, y) = h(y) - h(x)$ where $h : \mathbb{R}^n \to \mathbb{R}$ is l.s.c. and convex. Then

$$\text{(EP)} \;\Leftrightarrow\; \text{Find } x^* \in C \text{ s.t. } h(x^*) \leq h(y) \text{ for all } y \in C$$

- Variational Inequality Problems
  Let $C \subset \mathbb{R}^n$ be closed and convex and let $f(x, y) = \langle F(x), y - x\rangle$ where $F : C \to \mathbb{R}^n$ is continuous. Then

$$\text{(EP)} \;\Leftrightarrow\; \text{(VIP) Find } x^* \in C \text{ s.t. } \langle F(x^*), y - x^*\rangle \geq 0 \text{ for all } y \in C$$

  When $C = \mathbb{R}^n_+$, then

$$\text{(EP)} \;\Leftrightarrow\; \text{(NCP) Find } x^* \in \mathbb{R}^n_+ \text{ s.t. } F(x^*) \in \mathbb{R}^n_+ \text{ and } \langle F(x^*), x^*\rangle = 0$$

Nash Equilibrium Problem

- N players, each player controls the decision variables $x_\nu \in \mathbb{R}^{n_\nu}$
- $x = (x_1, \ldots, x_N)$; $x_{-\nu} = (x_1, \ldots, \cancel{x_\nu}, \ldots, x_N)$; $n = \sum_{\nu=1}^{N} n_\nu$
- Each player has an objective function $\theta_\nu : \mathbb{R}^n \to \mathbb{R}$ depending on x_ν and $x_{-\nu}$
- Each player's strategy belongs to a set $C_\nu \subset \mathbb{R}^{n_\nu}$
- Aim of player ν : given the other players' strategy $x_{-\nu}$

$$\text{find} \quad x_\nu = \arg\min \{\theta_\nu(x_\nu, x_{-\nu}) \mid x_\nu \in C_\nu\}$$

- Nash equilibrium problem : find $x^* \in C := C_1 \times \cdots \times C_N$ such that

$$\theta_\nu(x^*_\nu, x^*_{-\nu}) \leq \theta_\nu(y_\nu, x^*_{-\nu}) \text{ for all } \nu \text{ and all } y \in C$$

  No player can decrease his objective function by changing x^*_ν

- Here $f(x, y) = \sum_{\nu=1}^{N} \{\theta_\nu(y_\nu, x^*_{-\nu}) - \theta_\nu(x^*_\nu, x^*_{-\nu})\}$

Proximal Point Method for EP

The proximal point algorithm for EP is defined as follows : Given $x^k \in C$

$$\text{Find } x^{k+1} \in C \text{ s.t. } f(x^{k+1}, y) + \frac{1}{c}\langle x^{k+1} - x^k, y - x^{k+1}\rangle \geq 0 \;\forall y \in C.$$

If $C = \mathbb{R}^n$ and $f(x, y) = h(y) - h(x)$ with $h : \mathbb{R}^n \to \mathbb{R}$ l.s.c. and convex, then by definition of $\partial h(x^{k+1})$, we have

$$\frac{1}{c}(x^k - x^{k+1}) \in \partial h(x^{k+1}),$$

which is the optimality condition of the subproblem :

$$x^{k+1} = \arg\min_{y \in C} \{h(y) + \frac{1}{2c}\|y - x^k\|^2\}$$

Convergence

The function f is said to be

- monotone if $\forall x, y \in C \quad f(x,y) + f(y,x) \leq 0$
- strongly monotone if $\forall x, y \in C \quad f(x,y) + f(y,x) \leq -\gamma \|x - y\|^2$

Convergence

- f monotone $\Rightarrow x^k \to x^*$ solution to EP
- f strongly monotone $\Rightarrow x^k \to x^*$ the unique solution to EP

When f is monotone,

- the function $(x,y) \mapsto f(x,y) + \frac{1}{c}\langle x - x^k, y - x\rangle$ is strongly monotone
- So the subproblems are strongly monotone equilibrium problems
- There is a need of an efficient algorithm for solving such problems

Another Generalization of the Proximal Point Method

It is easy to see that $x^* \in C$ is a solution to problem EP if and only if

$$x^* \in \arg\min_{y \in C} \left\{ f(x^*, y) + \frac{1}{2c} \|y - x^*\|^2 \right\} \quad (c > 0)$$

The corresponding algorithm : Auxiliary Problem Principle Algorithm

Data : Let $x^0 \in C$ and $c > 0$. Set $k = 0$.

Step 1 Compute $x^{k+1} = \arg\min_{y \in C} \left\{ f(x^k, y) + \frac{1}{2c} \|y - x^k\|^2 \right\}$.

Step 2 If $x^{k+1} = x^k$, then STOP : x^k is a solution to EP.

Replace k by $k+1$, and go to Step 1.

When $C = \mathbb{R}^n$ and $f(x, y) = h(y) - h(x)$, Step 1 becomes :

$$x^{k+1} = \arg\min_{y \in C} \{ h(y) + \frac{1}{2c} \|y - x^k\|^2 \}$$

Convergence of the Auxiliary Problem Principle Algorithm

Theorem (Mastroeni)

Assume

(a) $f(\cdot, y) : C \to \mathbb{R}$ is continuous for all $y \in C$

(b) f is strongly monotone (with modulus $\gamma > 0$)

(c) There exists $d_1 > 0$ and $d_2 > 0$ such that, for all $x, y, z \in C$,

$$f(x,y) + f(y,z) \geq f(x,z) - d_1 \|y - x\|^2 - d_2 \|z - y\|^2$$

Then $x^k \to x^*$ the unique solution to EP provided that

$$c \leq d_1 \text{ and } d_2 < \gamma$$

This algorithm can be used for solving the subproblems of the proximal point algorithm.

Comments on Assumption (c)

There exists $d_1 > 0$ and $d_2 > 0$ such that, for all $x, y, z \in C$,

$$f(x,y) + f(y,z) \geq f(x,z) - d_1 \|y - x\|^2 - d_2 \|z - y\|^2$$

When $f(x,y) = \langle F(x), y - x \rangle$ with $F : \mathbb{R}^n \to \mathbb{R}^n$, problem EP becomes the variational inequality problem :

$$\text{Find } x^* \in C \text{ s.t. } \langle F(x^*), y - x^* \rangle \geq 0 \text{ for all } y \in C$$

In that case $f(x,y) + f(y,z) - f(x,z) = \langle F(x) - F(y), y - z \rangle$ for all $x, y, z \in C$ and it is easy to see that if F is Lipschitz continuous (with constant L), then for all $x, y, z \in C$,

$$|\langle F(x) - F(y), y - z \rangle| \leq L\|x - y\| \, \|y - z\| \leq \frac{L}{2}\|x - y\|^2 + \frac{L}{2}\|y - z\|^2$$

and thus f satisfies condition (c).

Convergence under weaker assumptions

Recently (*) convergence has been obtained under weaker assumptions than (b) and (c) :

There exist $\gamma, d_1, d_2 > 0$ and a nonnegative function $g : C \times C \to \mathbb{R}$ such that

(i) $f(x, y) \geq 0 \Rightarrow f(y, x) \leq -\gamma g(x, y)$

(ii) $f(x, z) - f(y, z) - f(x, y) \leq d_1 g(x, y) + d_2 \|z - y\|^2$

(*) Nguyen Thi Thu Van, J.J. Strodiot, and V.H. Nguyen, *A Bundle Method for Solving Equilibrium Problems*, Mathematical Programming, 2009, Vol.116, pp.529 – 552.

Approximate Auxiliary Problem Principle

Let $x^k \in C$ and let $f^k := f(x^k, \cdot)$.
Strategy : Approximate f^k in the subproblem

$$(P^k) \qquad x^{k+1} = \arg\min_{y \in C} \left\{ f^k(y) + \frac{1}{2c} \|y - x^k\|^2 \right\}$$

by a simpler function φ^k in such a way that the convergence is preserved.
Definition Let $\sigma \in (0, 1]$. A convex function $\varphi^k : C \to \mathbb{R}$ is a σ-approximation of f^k at x^k if

$$\varphi^k \le f^k \quad \text{and} \quad f^k(y^k) \le \sigma \varphi^k(y^k),$$

where y^k is the unique solution to problem (AP^k) :

$$(AP^k) \qquad \min_{y \in C} \left\{ \varphi^k(y) + \frac{1}{2c_k} \|y - x^k\|^2 \right\}$$

Approximate Auxiliary Problem Principle Algorithm

Since $\varphi^k(x^k) \le f^k(x^k) = 0$, the inequality $f^k(y^k) \le \sigma\varphi^k(y^k)$ implies :

$$f^k(x^k) - f^k(y^k) \ge \sigma\left(\varphi^k(x^k) - \varphi^k(y^k)\right)$$

The reduction on f^k is greater than a fraction of the reduction on φ^k.

Data : Let $x^0 \in C$ and $\sigma \in (0,1]$. Set $k = 0$

Step 1. Find φ^k a σ-approximation of f^k at x^k and solve

$$(AP^k) \qquad x^{k+1} = \arg\min_{y \in C} \left\{ \varphi^k(y) + \frac{1}{2c_k}\|y - x^k\|^2 \right\}$$

to get x^{k+1}.

Step 2. Replace k by $k+1$ and go to Step 1.

Convergence

Assume $c_k \geq \underline{c} > 0$. Then

$$\left.\begin{array}{l} \{x^k\} \text{ bounded} \\ \|x^{k+1} - x^k\| \to 0 \end{array}\right\} \Rightarrow \text{any limit point of } \{x^k\} \text{ is solution to EP}$$

Suppose that there exist $\gamma, d_1, d_2 > 0$ and a nonnegative function $g : C \times C \to \mathbb{R}$ such that

(i) $f(x,y) \geq 0 \Rightarrow f(y,x) \leq -\gamma g(x,y)$

(ii) $f(x,z) - f(y,z) - f(x,y) \leq d_1\, g(x,y) + d_2 \,\|z-y\|^2$

If $\{c_k\}$ is nonincreasing and $c_k < \frac{\sigma}{2d_2}$ and if $\frac{d_1}{\gamma} \leq \sigma \leq 1$,

then $\{x^k\}$ is bounded and $\|x^{k+1} - x^k\| \to 0$

Properties that must be satisfied by the model functions

- As previously, to get φ^k a σ-approximation of f^k, we construct successively model functions φ_i^k, $i = 1, 2, \ldots$ satisfying the conditions

  (C1) $\varphi_i^k \leq f^k$ on $\boldsymbol{R}^n$ for $i = 1, 2, \ldots$

  (C2) $\varphi_{i+1}^k \geq f^k(y_i^k) + \langle s(y_i^k), \cdot - y_i^k \rangle$ on $\boldsymbol{R}^n$ for $i = 1, 2, \ldots$

  (C3) $\varphi_{i+1}^k \geq l_i^k$ on $\boldsymbol{R}^n$ for $i = 1, 2, \ldots,$

  where

  - $s(y_i^k)$ denotes the subgradient of f available at y_i^k
  - $l_i^k(y) = \varphi_i^k(y_i^k) + \langle \gamma_i^k, y - y_i^k \rangle$ and $\gamma_i^k = \dfrac{1}{t_k}(x^k - y_i^k)$

- We stop when for some i_k, the function $\varphi_{i_k}^k$ is a σ-approximation of f^k. In that case we set $\varphi^k = \varphi_{i_k}^k$.

Serious Step Algorithm

Let $x^k \in \boldsymbol{R}^n$ and $\sigma \in (0,1]$. Set $i = 1$ and $y_0^k = x^k$

Step 1. Choose a convex model φ_i^k satisfying conditions

$(C1)-(C3)$ and solve the problem

$$(P_i^k) \quad \min_y \left\{ \varphi_i^k(y) + \frac{1}{2c_k}\|y - x^k\|^2 \right\} \text{ to get } \; y_i^k$$

Step 2. If $f^k(y_i^k) \leq \sigma\varphi_i^k(y_i^k)$, then set $x^{k+1} = y_i^k$ and STOP ;

x^{k+1} is a serious step

Step 3. Increase i by 1 and go to Step 1.

x^k not a solution $\Rightarrow$ after finitely many iterations φ_i^k is a σ-approximation

Convergence

Suppose that there exist $\gamma, d_1, d_2 > 0$ and a nonnegative function $g : C \times C \to \mathbb{R}$ such that

(i) $f(x,y) \geq 0 \Rightarrow f(y,x) \leq -\gamma g(x,y)$

(ii) $f(x,z) - f(y,z) - f(x,y) \leq d_1 g(x,y) + d_2\|z-y\|^2$

If $\{c_k\}$ is nonincreasing and $0 < \underline{c} \leq c_k < \frac{\sigma}{2d_2}$ and if $\frac{d_1}{\gamma} \leq \sigma \leq 1$,

then $\{x^k\}$ converges to some solution to problem EP.

Nguyen Thi Thu Van, Strodiot, J.J., and Nguyen, V.H. *A Bundle Method for Solving Equilibrium Problems*, Mathematical Programming, 2009, Vol.116, pp.529 – 552.

Application to Mixed Variational Inequality Problems

- $(MVIP)$: Find $x^* \in C$ such that for all $y \in C$

$$\langle F(x^*), y - x^* \rangle + h(y) - h(x^*) \geq 0,$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and $h : \mathbb{R}^n \to \mathbb{R}$ is convex.

- Here $f(x, y) = \langle F(x), y - x \rangle + h(y) - h(x)$

- At $x^k \in C$, the function $f^k(y) := f(x^k, y)$ is approximated by

$$\varphi_i^k(y) = \langle F(x^k), y - x^k \rangle + h_i^k(y) - h(x^k),$$

where h_i^k is an approximation of the convex function h at x^k

Application to Mixed Variational Inequality Problems

As previously, the model functions h_i^k, $i = 1, 2, \dots$ satisfy the conditions :

(C1) $h_i^k \leq h$ on $\boldsymbol{R}^n$ for $i = 1, 2, \dots$

(C2) $h_{i+1}^k \geq h(y_i^k) + \langle s(y_i^k), \cdot - y_i^k \rangle$ on $\boldsymbol{R}^n$ for $i = 1, 2, \dots$

(C3) $h_{i+1}^k \geq l_i^k$ on $\boldsymbol{R}^n$ for $i = 1, 2, \dots,$

where

- $s(y_i^k)$ denotes the subgradient of h available at y_i^k
- $l_i^k(y) = h_i^k(y_i^k) + \langle \gamma_i^k, y - y_i^k \rangle$ and $\gamma_i^k = \dfrac{1}{c_k}(x^k - y_i^k) - F(x^k)$

Application to Mixed Variational Inequality Problems

- Let $\sigma \in (0,1)$. The condition of σ-approximation : $f^k(y_i^k) \leq \sigma \varphi_i^k(y_i^k)$ becomes :

$$h(x^k) - h(y_i^k) \geq \sigma \left(h(x^k) - h_i^k(y_i^k) \right) + (1-\sigma) \langle F(x^k), y_i^k - x^k \rangle$$

- Assumption : F is h-co-coercive (with modulus $\gamma > 0$), i.e., for all $x, y \in C$,

$$\langle F(x), y - x \rangle + h(y) - h(x) \geq 0$$
$$\Rightarrow \quad \langle F(y), y - x \rangle + h(y) - h(x) \geq \gamma \| F(y) - F(x) \|^2$$

Assumption and Convergence

- It is easy to see that if F is *h-co-coercive*, then the two following conditions (used for the convergence of the Bundle Proximal Point Algorithm for EP) are satisfied :

  (i) $f(x,y) \geq 0 \quad \Rightarrow \quad f(y,x) \leq -\gamma g(x,y)$

  (ii) $f(x,z) - f(y,z) - f(x,y) \leq \frac{1}{2}\, g(x,y) + \frac{1}{2}\, \|z-y\|^2$

  where $g(x,y) = \|y-x\|^2$.

- So, if F is *h-co-coercive*, $\{c_k\}$ is nonincreasing, $0 < \underline{c} \leq c_k < \sigma$ and $2\sigma\gamma \geq 1$, then the sequence $\{x^k\}$ (if infinite) converges to a solution of $(MVIP)$

Application to Multivalued Variational Inequality Problems

- $(GVIP)$: Find $x^* \in C$ and $\xi^* \in F(x^*)$ such that for all $y \in C$

$$\langle \xi^*, y - x^* \rangle \geq 0,$$

where $F : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is continuous.

- Here $f(x, y) = \sup_{\xi \in F(x)} \langle \xi, y - x \rangle$

- At $x^k \in C$, the function $f^k(y) := f(x^k, y)$ is approximated by

$$\varphi^k(y) = \langle \xi^k, y - x^k \rangle,$$

where ξ^k is any element in $F(x^k)$.

Question : When is φ^k a σ-approximation of f^k ?

σ-Approximation

Assumption : F is co-coercive on C, i.e., there exists $\gamma > 0$ such that for all $x, y \in C$ and for all $\xi_x \in F(x)$ and $\xi_y \in F(y)$, one has :

$$\langle \xi_x - \xi_y, x - y \rangle \geq \gamma\, g(x, y),$$

where $g(x, y) = \sup_{\xi_1 \in F(x)} \inf_{\xi_2 \in F(y)} \|\xi_1 - \xi_2\|^2$

Suppose F is co-coercive on C with constant $\gamma > 0$.

Let $\sigma \in (0, 1)$ and $x^k \in C$. Then

$c_k \leq 4\gamma\,(1 - \sigma) \Rightarrow \varphi^k$ is a σ-approximation of f^k

Algo : Given $x^k \in C$ and $c_k > 0$, choose $\xi^k \in F(x^k)$ and compute :

$$x^{k+1} = \arg\min_{y \in C} \left\{ \langle \xi^k, y - x^k \rangle + \frac{1}{2c_k} \|y - x^k\|^2 \right\}$$

Convergence

- It is easy to see that if F is co-coercive on C, then the two following conditions (used for the convergence of the Bundle Proximal Point Algorithm for EP) are satisfied :

  (i) $f(x,y) \geq 0 \quad \Rightarrow \quad f(y,x) \leq -\gamma g(x,y)$

  (ii) $f(x,z) - f(y,z) - f(x,y) \leq \frac{1}{2}\, g(x,y) + \frac{1}{2}\, \|z-y\|^2$

  where $g(x,y) = \sup_{\xi_1 \in F(x)} \inf_{\xi_2 \in F(y)} \|\xi_1 - \xi_2\|^2$.

- So if F is co-coercive with constant $\gamma > 0$, $\{c_k\}$ is nonincreasing, $0 < \underline{c} \leq c_k < \sigma$ and $c_k < 4(2-\sqrt{3})\gamma$ for all k, then the sequence $\{x^k\}$ converges to a solution of $(GVIP)$

References

- Nguyen Thi Thu Van, Strodiot, J.J., and Nguyen, V.H. *A Bundle Method for Solving Equilibrium Problems*, Mathematical Programming, 2009, Vol.116, pp.529 – 552.
- Salmon, G., Strodiot, J.J., and Nguyen, V.H. *A Bundle Method for Solving Variational Inequalities*, SIAM J. Optimization, 2004, Vol.14, pp.869 – 893.
- Tran Thi Hue, Strodiot, J.J., and Nguyen, V.H. *Convergence of the Approximate Auxiliary Problem Method for Solving Generalized Variational Inequalities*, Journal of Optimization Theory and Applications, 2004, Vol.121, pp.119 – 145.

Extragradient Methods

Our Aim : We do not want to assume hypothesis (i) below (because too strong) to obtain the convergence of the Proximal Point Method.

(i) $f(x,y) \geq 0 \quad \Rightarrow \quad f(y,x) \leq -\gamma \|y-x\|^2$

(ii) $f(x,z) - f(y,z) - f(x,y) \leq d_1 \|y-x\|^2 + d_2 \|z-y\|^2$

Strategy : Add an extra step to obtain the convergence under the sole assumption (ii), i.e., under a Lipschitz-type condition.

Proximal Extragradient Method for VIP

(VIP) : Find $x^* \in C$ such that for all $y \in C$

$$\langle F(x^*), y - x^* \rangle \geq 0$$

Data : Let $x^0 \in C$ and $c > 0$. Set $k = 0$.

Step 1. Compute $y^k = P_C(x^k - c\,F(x^k))$

If $y^k = x^k$, then STOP : x^k is a solution to VIP.

Step 2. Compute $x^{k+1} = P_C(x^k - c\,F(y^k))$.

Replace k by $k+1$ and go to Step 1.

Extragradient Method for VIP. Convergence

Definition : F is said to be pseudomonotone on C if for all $x, y \in C$,

$$\langle F(x), y - x\rangle \geq 0 \quad \Rightarrow \quad \langle F(y), x - y\rangle \leq 0$$

Assume F is pseudomonotone and Lipschitz continuous on C with constant $L > 0$. Then

$0 < c < \frac{1}{L} \quad \Rightarrow \quad \{x^k\}$ converges to a solution of VIP

Extragradient Method for EP

(EP) : Find $x^* \in C$ such that for all $y \in C$, $\quad f(x^*, y) \geq 0$

To get the extragradient method for EP :

- replace $y^k = P_C(x^k - c\,F(x^k))$ by

$$y^k = \arg\min_{y \in C} \{f(x^k, y) + \frac{1}{2c}\|y - x^k\|^2\}$$

- and $x^{k+1} = P_C(x^k - c\,F(y^k))$ by

$$x^{k+1} = \arg\min_{y \in C} \{f(y^k, y) + \frac{1}{2c}\|y - x^k\|^2\}$$

Reminder : for VIP, we have $f(x, y) = \langle F(x), y - x\rangle$

Extragradient Method for EP

Data : Let $x^0 \in C$ and $c > 0$. Set $k = 0$.

Step 1. Find

$$y^k = \arg\min_{y \in C} \{f(x^k, y) + \frac{1}{2c} \|y - x^k\|^2\}$$

If $y^k = x^k$, then STOP : x^k is solution to EP.

Step 2. Find

$$x^{k+1} = \arg\min_{y \in C} \{f(y^k, y) + \frac{1}{2c} \|y - x^k\|^2\}$$

Replace k by $k+1$ and go to Step 1.

Extragradient Method for EP. Convergence

Definition : f is pseudomonotone on $C \times C$ if for all $x, y \in C$,

$$f(x, y) \geq 0 \quad \Rightarrow \quad f(y, x) \leq 0$$

Assume f is pseudomonotone and l.s.c. on $C \times C$. If there exist $d_1, d_2 > 0$ such that

$$f(x, z) - f(y, z) - f(x, y) \leq d_1 \|y - x\|^2 + d_2 \|z - y\|^2,$$

then $\{x^k\}$ converges to a solution of EP

Reference :

Tran Dinh Quoc, Le Dung Muu, and Nguyen Van Hien, *Extragradient Algorithms Extended to Equilibrium Problems*, Optimization, *Online First*.

Approximate Extragradient Method for EP.

- For Step 1, we have $\arg\min_{x\in C}\left\{f(x^k,y)+\frac{1}{2c}\|y-x^k\|^2\right\}$ and we consider a σ-approximation of $f(x^k,y)$.
- For Step 2, we write

$$\arg\min_{x\in C}\left\{f(y^k,y)+\frac{1}{2c}\|y-x^k\|^2\right\}=$$
$$\arg\min_{x\in C}\left\{f(y^k,y)+\frac{1}{c}\langle y-y^k,y^k-x^k\rangle+\frac{1}{2c}\|y-y^k\|^2\right\}$$

  and we consider a σ-approximation of $f(y^k,y)+\frac{1}{c}\langle y-y^k,y^k-x^k\rangle$
- The Bundle Method can be used for building these two σ-approximations.
- Convergence is obtained under the same assumptions as in the exact case.

Extragradient Method for VIP without Lipschitz Continuity

Strategy : At $x^k \in C$

- First compute $y^k = P_C(x^k - c\,F(x^k))$
- Then use an Armijo-type linesearch to get $z^k \in [x^k, y^k]$ such that the hyperplane $H^k = \{x \in \mathbb{R}^n \mid \langle F(z^k), x - z^k \rangle = 0\}$ strictly separates x^k from the solution set
- Compute $w^k = P_{H^k}(x^k)$ and $x^{k+1} = P_C(w^k)$

Armijo Condition : $\langle F(z^k), x^k - y^k \rangle \geq \frac{\alpha}{c} \|y^k - x^k\|^2$

Projection : $w^k = x^k - \frac{\langle F(z^k), x^k - z^k \rangle}{\|F(z^k)\|^2} F(z^k)$

Convergence : If F is continuous and pseudomonotone,

then $\{x^k\}$ converges to a solution of VIP

Extragradient Method for EP without Lipschitz Condition

(EP) : Find $x^* \in C$ such that for all $y \in C$, $\quad f(x^*, y) \geq 0$

Since $f(x,y) = \langle F(x), y - x \rangle$ for VIP,

- the Armijo condition for VIP : $\langle F(z^k), x^k - y^k \rangle \geq \frac{\alpha}{c} \|y^k - x^k\|^2$ becomes
$$f(z^k, x^k) - f(z^k, y^k) \geq \frac{\alpha}{c} \|y^k - x^k\|^2$$
- $F(z^k)$ is replaced by $g^k \in \partial f(z^k, \cdot)(x^k)$
- $w^k = x^k - \frac{\langle F(z^k), x^k - z^k \rangle}{\|F(z^k)\|^2} F(z^k)$ (for VIP) becomes
$$w^k = x^k - \frac{f(z^k, x^k)}{\|g^k\|^2} g^k$$

Convergence : If f is continuous on $C \times C$ and pseudomonotone,

then $\{x^k\}$ converges to a solution of EP

Interior Proximal Algorithms for EP

- Consider the simplest case : $C = \{x \in \mathbb{R}^n \mid x \geq 0\}$
- Use a barrier method for treating the constraint set C :

  The subproblem $\min_{x \in C} \{c_k f(x^k, y) + \frac{1}{2}\|y - x^k\|^2\}$ is replaced by the unconstrained problem :

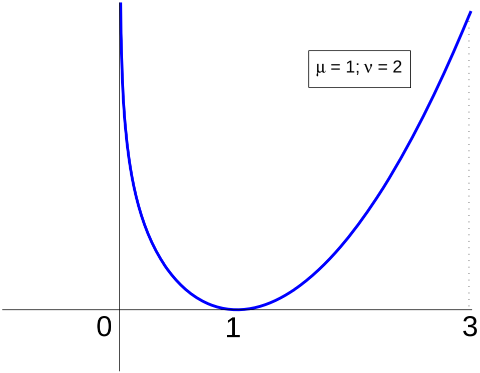
$$\min_{x \in \mathbb{R}^n_{++}} \left\{ c_k f(y^k, y) + \frac{\nu}{2}\|y - x^k\|^2 + \mu \sum_{j=1}^{n} x_j^{k\,2} h\left(\frac{y_j}{x_j^k}\right) \right\}$$

  where $\nu > \mu > 0$ and $h : \mathbb{R}_{++} \to \mathbb{R}$ is defined by $h(t) = t - \log t - 1$

- Notation : $\varphi(t) = \mu h(t) + \frac{\nu}{2}(t-1)^2$ (log-quad function) and

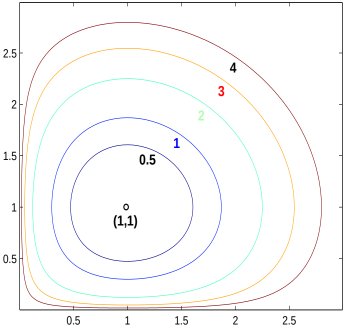
$$D_\varphi(y, x^k) := \sum_{j=1}^{n} x_j^{k\,2} \varphi\left(\frac{y_j}{x_j^k}\right) = \frac{\nu}{2}\|y - x^k\|^2 + \mu \sum_{j=1}^{n} x_j^{k\,2} h\left(\frac{y_j}{x_j^k}\right)$$

Log-quad function

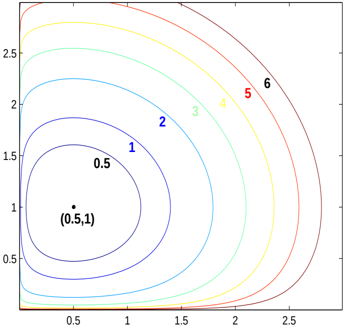


$$\varphi(t) = \mu(t - \log t - 1) + \frac{\nu}{2}(t-1)^2$$

$x \mapsto D_\varphi(x, y)$

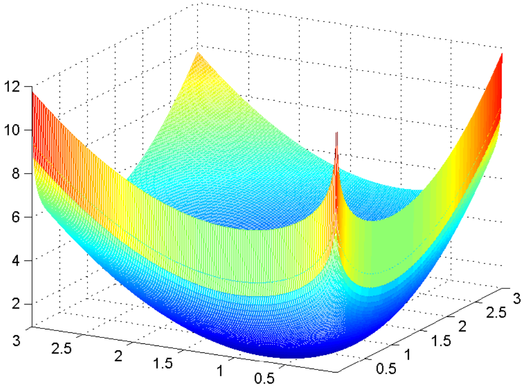


$y = (1, 1)$



$y = (0.5, 1)$

$x \mapsto D_\varphi(x, y)$ with $y = (1, 1)$



Interior Proximal Extragradient Method for EP. Algo IPE

Data : Let $x^0 \in C$ and $c > 0$. Set $k = 0$.

Step 1. Find

$$y^k = \arg \min_{y \in \mathbb{R}^n_{++}} \{c_k f(x^k, y) + D_\varphi(y, x^k)\}$$

If $y^k = x^k$, then STOP : x^k is solution to EP.

Step 2. Find

$$x^{k+1} = \arg \min_{y \in \mathbb{R}^n_{++}} \{c_k f(y^k, y) + D_\varphi(y, x^k)\}$$

Replace k by $k+1$ and go to Step 1.

Convergence

Assume that f is pseudomonotone on $C \times C$ and that there exist $d_1, d_2 > 0$ such that

$$f(x,z) - f(y,z) - f(x,y) \leq d_1 \|y - x\|^2 + d_2 \|z - y\|^2$$

If $0 < c < c_k < \min\left\{\frac{\nu - 5\mu}{2d_1}, \frac{\nu - 3\mu}{2d_2}\right\}$, then $\{x^k\}$ converges to a solution of EP

Interior Proximal Extragradient Method without Lipschitz Continuity. Algo IPLE

At $x^k \in \mathbb{R}^n_{++}$

- First compute $y^k = \arg\min_{y \in \mathbb{R}^n_{++}} \{c_k f(x^k, y) + D_\varphi(y, x^k)\}$
- Then use an Armijo-type linesearch to get $z^k \in [x^k, y^k]$ such that

$$f(z^k, x^k) - f(z^k, y^k) \geq \frac{\alpha}{c_k} D_\varphi(y^k, x^k)$$

- Take $g^k \in \partial f(z^k, \cdot)(x^k)$
- Compute $w^k = x^k - \frac{f(z^k, x^k)}{\|g^k\|^2} g^k$
- Set $x^{k+1} = (1 - \tau) x^k + \tau P_C(w^k)$ where $\tau \in (0, 1)$
  So $x^{k+1} \in \mathbb{R}^n_{++}$

Algo IPLE. Convergence

- If $0 < c \leq c_k \leq \bar{c}$ for all k, then every limit point of $\{x^k\}$ is a solution to problem EP
- If, in addition, f is pseudomonotone, then the whole sequence $\{x^k\}$ converges to a solution of problem EP

Reference :

Nguyen Thi Thu Van, Strodiot J.J., and Nguyen Van Hien, *The Logarithmic-Quadratic Extragradient Method for Solving Equilibrium Problems*, Journal of Global Optimization, *Online First*.

Difficulties

- This time, the subproblems

$$y^k = \arg \min_{y \in \mathbb{R}^n_{++}} \{c_k f(x^k, y) + D_\varphi(y, x^k)\}$$

  are no more quadratic and defined on an open set $\mathbb{R}^n_{++}$.
  So, in general, they are difficult to solve.
- When the conjugate of the convex function $f(x^k, \cdot)$ is finite on $\mathbb{R}^n$ and easily computable, then the strategy is
  - first solve the Fenchel dual

$$\min_{u \in \mathbb{R}^n} \{f(x^k, \cdot)^*(u) + D_\varphi(\cdot, x^k)^*(-u)\} \quad \text{to obtain } u^*$$

    because $\varphi^*(t)$ and $(\varphi^*)'(t)$ can be explicitly computed.
  - then recover the solution y^k by using the formula :

$$(y^k)_j = x^k_j (\varphi^*)' \left(-\frac{u^*_j}{x^k_j} \right) \quad \text{for all } j = 1, \ldots, n$$

Example where Fenchel duality is useful

- Let $f(x, y) = \langle Px + Qy + q,\, y - x\rangle$ for $x, y \in C := \mathbb{R}^n_+$
- The corresponding EP is related to the Nash Cournot equilibrium model. Reference :

  Le Dung Muu, Nguyen Van Hien, and Nguyen Van Quy,
  On Nash-Cournot Oligopolistic Market Equilibrium Models with Concave Cost Functions,
  Journal of Global Optimization, Vol.41, pp.351 − 364, 2008.

- Assumptions : Q symmetric positive definite and $Q - P$ negative semidefinite.

  $\Rightarrow$ f is continuous, monotone and Lipschitz (in the sense of (ii))

  Convergence assumptions are satisfied

Example where Fenchel duality is useful

- The subproblems can be written

$$\min_{y \in \mathbb{R}^n_{++}} \left\{ g(y) + D_\varphi(y, x^k) \right\}$$

where $g(y) = c_k y^T Q y + c_k b^T y$ and $b = (P - Q)x + q$

- $g^*(u) = \frac{1}{4c_k} \langle u - c_k b,\ Q^{-1}(u - c_k b) \rangle$ for $u \in \mathbb{R}^n$

- The Fenchel dual

$$\min_{u \in \mathbb{R}^n} \left\{ g^*(u) + D_\varphi(\cdot, x^k)^*(-u) \right\}$$

can be solved using a unconstrained optimization method

Numerical Results

	Example 1		Example 2		Example 3	
Algorithm	IPE	IPLE	IPE	IPLE	IPE	IPLE
it	19	1305	20	1342	40	228
cpu (sec.)	1.078	26.89	1.296	27.64	10.875	13.25
optimality	-0.00000	-0.00257	-0.00000	-0.00237	-0.00006	-0.00152

- Three examples randomly generated where $n = 5$ and $C = \mathbb{R}^n_+$
- it := number of iterations ; cpu := cpu time (in seconds)
- optimality at $x \Leftrightarrow \min_{y \in \mathbb{R}^n_+} f(x, y) = 0$
- IPE by far better than IPLE

Section 4

Other Applications of the Bundle Proximal Point Method

1. Generalized Fractional Programming Problems
2. Bilevel Problems
3. D.C. Programming Problems

Generalized Fractional Programming Problems

Consider the nonlinear program

$$(P) \qquad \lambda^* = \inf_{x \in X} \left\{ \max_{1 \le i \le m} \left\{ \frac{f_i(x)}{g_i(x)} \right\} \right\}$$

where

- $X \subseteq \mathbb{R}^n$ nonempty closed
- $f_i(x), g_i(x)$ continuous for all $1 \le i \le m$
- $g_i > 0$ on X for all $1 \le i \le m$

When $m = 1$, the problem is called a fractional problem

Question : find λ^* and a solution x^* of (P)

Auxiliary Parametric Problems

For each $\lambda \in \mathbb{R}$, we introduce a parametric problem with a simpler structure :

$$(P_\lambda) \qquad F(\lambda) = \inf_{x \in X} \left\{ \max_{1 \le i \le m} \left\{ f_i(x) - \lambda g_i(x) \right\} \right\}$$

- If $F(\lambda^*) = 0$, then problems (P) and (P_{λ^*}) have the same set of optimal solutions (which may be empty)
  $\Rightarrow$ two steps : first find λ^* a zero of F and then solve (P_{λ^*})
- F is nonincreasing and $F(\lambda) < 0$ if and only if $\lambda > \lambda^*$

Strategy : Let $\lambda_k > \lambda^*$. Then

- solve (P_{λ_k}) to get x^k
- approximate $F(\lambda)$ by $\overline{F}(\lambda, x^k) = \max_{1 \le i \le m} \{f_i(x^k) - \lambda g_i(x^k)\}$
- find λ_{k+1} a zero of $\overline{F}(\lambda, x^k)$

Local Approximation of $F(\lambda)$

Consider again : $F(\lambda) = \inf_{x \in X} \{\max_{1 \le i \le m} \{f_i(x) - \lambda g_i(x)\}\}$
and define

$$\overline{F}(\lambda, x) = \max_{1 \le i \le m} \{f_i(x) - \lambda g_i(x)\} \quad \text{for all } \lambda \in \mathbb{R}, \text{ and } x \in X$$

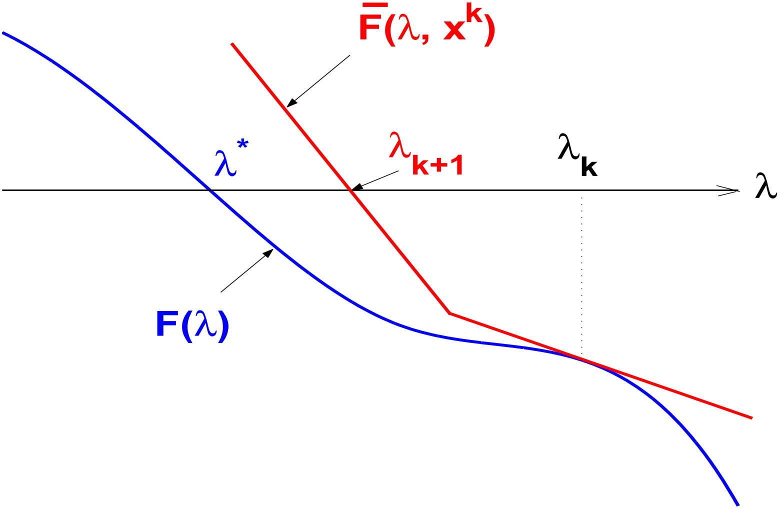
- The function $\lambda \to \overline{F}(\lambda, x^k)$ is decreasing, piecewise linear and convex
- Let $\lambda_k > \lambda^*$. Then

$$x^k \text{ is solution to } (P_{\lambda_k}) \Leftrightarrow x^k \text{ is the minimum over } X \text{ of } \overline{F}(\lambda_k, x)$$

- $\overline{F}(\lambda_k, x^k) = F(\lambda_k) < 0$ and $F(\lambda) \le \overline{F}(\lambda, x^k), \ \forall \lambda$
- Finding λ_{k+1} a zero of $\overline{F}(\lambda, x^k)$ amounts to compute

$$\lambda_{k+1} = \max_{1 \le i \le m} \{f_i(x^k)/g_i(x^k)\}.$$

Geometric Interpretation



Dinkelbach-type algorithm (DTA)

Step 0 Let $x^0 \in X$, $\lambda_1 = \max_{1 \le i \le m}\{f_i(x^0)/g_i(x^0)\}$, and $k = 1$

Step 1 Determine an optimal solution x^k of

$$(P_{\lambda_k}) \quad F(\lambda_k) = \inf_{x \in X} \{\max_{1 \le i \le m} \{f_i(x) - \lambda_k g_i(x)\}\}$$

Step 2 If $F(\lambda_k) = 0$, x^k is an optimal solution of (P) and

λ_k is the optimal value, and STOP

Step 3 Let $\lambda_{k+1} = \max_{1 \le i \le m}\{f_i(x^k)/g_i(x^k)\}$.

Replace k by $k+1$ and repeat Step 1.

The Auxiliary Problems

The performances of the DTA algorithm heavily depend on the effective solution of the auxiliary problems :

$$(P_{\lambda_k}) \quad F(\lambda_k) = \inf_{x \in X} \left\{ \max_{1 \leq i \leq m} \{ f_i(x) - \lambda_k g_i(x) \} \right\}$$

Let us denote $\overline{F}(x, \lambda_k) = \max_{1 \leq i \leq m} \{ f_i(x) - \lambda_k g_i(x) \}$

Difficulties :

- $\overline{F}(x, \lambda_k)$ is in general nonsmooth
- Problems (P_{λ_k}) may have several solutions

Strategy : add a prox-regularization term to $\overline{F}(x, \lambda_k)$ to obtain a strongly convex function.

Here in this talk, we assume that the functions $\overline{F}(x, \lambda_k)$ are convex.

Inexact Proximal Point Method

Given (x^{k-1}, λ_k), the prox-regularization method replaces $\min_{x \in X} \overline{F}(x, \lambda_k)$ by

$$(P_{\lambda_k}) \qquad \min_{x \in X} \{\overline{F}(x, \lambda_k) + \frac{1}{2c_k} \|x - x^{k-1}\|^2\}$$

Strategy : approximate $\overline{F}(\cdot, \lambda_k)$ by a convex function $\varphi^k(\cdot, \lambda_k)$ such that

- the convergence is preserved. As previously, we choose for $\varphi^k(\cdot, \lambda_k)$ a σ-approximation of $\overline{F}_k(\cdot, \lambda_k)$
- the problem

$$(AP_{\lambda_k}) \qquad \min_{x \in X} \{\varphi^k(x, \lambda_k) + \frac{1}{2c_k} \|x - x^{k-1}\|^2\}$$

  is easy to solve exactly. As previously, we choose for $\varphi^k(\cdot, \lambda_k)$ a piecewise linear function

Inexact proximal point algorithm

Step 0 Choose $x^0 \in X$, $c_1 > 0$, $\sigma > 0$, and set $\lambda_1 = \max_i \frac{f_i(x^0)}{g_i(x^0)}$, and $k = 1$

Step 1 Construct a σ-approximation $\varphi^k(\cdot, \lambda_k)$ of $\overline{F}(\cdot, \lambda_k)$ and find $x^k \in X$ the unique solution of problem

$$(AP_{\lambda_k}) \qquad \min_{x \in X} \{\varphi^k(x, \lambda_k) + \frac{1}{2c_k} \|x - x^{k-1}\|^2\}$$

Step 2 Set $\lambda_{k+1} = \max_i \frac{f_i(x^k)}{g_i(x^k)}$, choose $c_{k+1} > 0$

Step 3 Replace k by $k + 1$ and repeat Step 1.

Convergence

Let $\sigma \in (0,1)$. Assume $0 < \nu \leq g_i(x^k) \leq \gamma$ for all k and $1 \leq i \leq p$. Assume also that $\sum_{k \geq 0} c_k = +\infty$ and that either $c_k \leq \bar{c}$ for all k or $c_k \leq c_{k+1}$ for all k

Then

- the sequence $\{\lambda_k\}$ generated by the inexact proximal point algorithm converges to λ^*, the optimal value of problem (P).
- if $c_k \leq \bar{c}$ for all k and the solution set of problem (P) is nonempty, then the sequence $\{x^k\}$ converges to some solution of (P).

Reference

Strodiot, J.J., Crouzeix, J. P., Ferland, J.A., and Nguyen, V.H.

Inexact Proximal Point Method for Solving Generalized Fractional Programs

Journal of Global Optimization, Vol. 42, No 1, pp. 121 – 138, 2008.

Bilevel Problems

Consider the bilevel problem

$$\begin{cases} \min & f_1(x) \\ \text{s.t.} & x \in S_2 := \arg\min\{f_2(x) \,|\, x \in \mathbb{R}^n\}, \end{cases}$$

where $f_1,\ f_2 : \mathbb{R}^n \to \mathbb{R}$ are nondifferentiable convex functions.

The classical convex problem

$$\begin{cases} \min & f_1(x) \\ \text{s.t.} & g_i(x) \leq 0,\ i = 1, \dots, m, \end{cases}$$

where $g_i : \mathbb{R}^n \to \mathbb{R}^n$, $i = 1, \dots, m$ are nondifferentiable convex functions,

is an example of Bilevel Problem : take $f_2(x) = \sum_{i=1}^m \max\{0,\ g_i(x)\}$

Bilevel Problems

For each value of $\tau > 0$, we introduce the penalty function

$$F_\tau(x) = \tau f_1(x) + f_2(x).$$

Given (x^k, τ_k), the prox-regularization method replaces $\min_{x\in\mathbb{R}^n} F_{\tau_k}(x)$ by

$$(P_{k,\tau_k}) \qquad \min_{x\in\mathbb{R}^n} \{F_{\tau_k}(x) + \frac{1}{2c_k}\|x - x^k\|^2\}$$

Strategy : replace F_{τ_k} by a σ-approximation φ^k by using the bundle concept.

Let $x^k = \arg\min_{x\in\mathbb{R}^n} \left\{\varphi^k(x) + \frac{1}{2c_k}\|x - x^k\|^2\right\}$

Convergence

Let f_1 and f_2 be convex functions such that f_1 is bounded below and the solution set of the bilevel problem is nonempty and bounded.
Suppose that $0 < \underline{c} \leq c_k \leq \bar{c}$.
If the sequence $\{x^k\}$ is infinite and if $\tau_k \to 0$ and $\sum_{k=1}^{\infty} \tau_k = +\infty$, then each limit point of $\{x^k\}$ is a solution to the bilevel problem.

Advantage of the method :

- no need of regularity assumptions on constraints, such as the Slater condition.
- So we can consider complementarity constraints which do not satisfy constraint qualifications.

$$-Qx - q \leq 0, \quad -x \leq 0, \quad \langle Qx + q, x \rangle \leq 0,$$

where Q is a positive semidefinite matrix.

Reference

Reference :

M. Solodov, *A bundle method for a class of bilevel nonsmooth convex minimization problems*, SIAM J. Optimization, Vol. 18, pp. 242 – 259, 2007.

D.C. Programming

Consider the D.C. programming problem

$$\begin{cases} \min & f(x) \\ \text{s.t.} & x \in \mathbb{R}^n, \end{cases}$$

where $f = g - h$ with g and h convex from $\mathbb{R}^n$ to $\mathbb{R}$.

Necessary condition :

- x^* optimal solution $\Rightarrow \partial h(x^*) \subset \partial g(x^*) \Rightarrow \partial g(x^*) \cap \partial h(x^*) \neq \emptyset$
- The first necessary condition is hard to obtain.

We try to find a critical point x^* of f, i.e., a point x^* such that

$$\partial g(x^*) \cap \partial h(x^*) \neq \emptyset$$

Two Lemmas

Let $x \in \mathbb{R}^n$ and $c > 0$. Then

$$\forall w \in \partial h(x),\ w \neq 0 \qquad h(x + cw) > h(x)$$

Let $x \in \mathbb{R}^n$, $w \in \partial h(x)$ and $c > 0$. Then

x is a critical point of f if and only if

$$x = \arg\min_{y \in \mathbb{R}^n} \left\{ g(y) + \frac{1}{2c} \|y - (x + cw)\|^2 \right\}$$

Proximal Point Algorithm

Data : Let $x^0 \in \mathbb{R}^n$ and $c_0 > c > 0$. Set $k = 0$.

Step 1. Calculate $w^k \in \partial h(x^k)$ and set $z^k = x^k + c_k w^k$

Step 2. Find

$$x^{k+1} = \arg \min_{y \in \mathbb{R}^n} \left\{ g(y) + \frac{1}{2c_k} \|y - z^k\|^2 \right\}$$

Step 3. If $x^{k+1} = x^k$, then STOP : x^k is a critical point of f

Otherwise replace k by $k+1$, choose $c_k > c$ and go to Step 1.

Inexact Proximal Point Algorithm

Data : Let $x^0 \in \mathbb{R}^n$ and $c_0 > c > 0$. Choose $\alpha \in (0,1)$. Set $k = 0$.

Step 1. Calculate $w^k \in \partial h(x^k)$ and set $z^k = x^k + c_k w^k$

Step 2. Using the bundle concept, choose $\hat{g}^k$ an approximation of g at z^k such that

$$\hat{g}^k \leq g \quad \text{and} \quad g(x^{k+1}) - \hat{g}^k(x^{k+1}) \leq \frac{\alpha}{c_k} \|x^{k+1} - x^k\|^2$$

where

$$x^{k+1} = \arg\min_{y \in \mathbb{R}^n} \left\{ \hat{g}^k(y) + \frac{1}{2c_k} \|y - z^k\|^2 \right\}$$

Step 3. Replace k by $k+1$, choose $c_k > c$ and go to Step 1.

Convergence

- Assume $f = g - h$ is bounded below and $c_k > c > 0$ for all k.

  Then $\{f(x^k)\}$ is convergent and $\lim_{k\to\infty} c_k^{-1} \|x^{k+1} - x^k\| = 0$

- Moreover, if $\{x^k\}$ and $\{w^k\}$ are bounded, then the limit points x^∞ and w^∞ of $\{x^k\}$ and $\{w^k\}$ are critical points of $f = g - h$ and $h^* - g^*$, respectively

Reference :

Wen-yu Sun, R.J.B. Sampaio, and M.A.B. Candido, *Proximal point algorithm for minimization of DC function*, Journal of Computational Mathematics, Vol. 21, pp. 451 – 462, 2003.