POLITECNICO DI MILANO

Math Tools:

Basics on constrained and convex optimization – Part 2

Maria Prandini

Outline

- Constrained convex optimization: non differentiable setting
- Proximal algorithm

Main references:

D. Bertsekas. Nonlinear programming. Athena scientific, 1999

D. Bertsekas. Convex Optimization Theory. Athena Scientific, 2009

Remark: pictures are taken from the reference books

Convex optimization

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \to R$ is a continuously differentiable **convex function** over X

X is a non-empty closed convex subset of R^n

What about non differentiable functions?

Subgradient of a function

g is a subgradient of f (not necessarily convex) at x if

$$f(y) \geq f(x) + g'(y - x), \forall y \in R^n$$

Subgradient of a function

g is a subgradient of f (not necessarily convex) at x if

$$f(y) \geq f(x) + g'(y - x), \forall y \in R^n$$

- subgradient provides an affine global underestimator of f

Subgradient of a function

g is a subgradient of f (not necessarily convex) at x if

$$f(y) \geq f(x) + g'(y - x), \forall y \in R^n$$

- subgradient provides an affine global under-estimator of f
- if f convex, then, there is at least one subgradient at every interior point of X
- If f is convex and differentiable, $\nabla f(x)$ is a subgradient of f at x

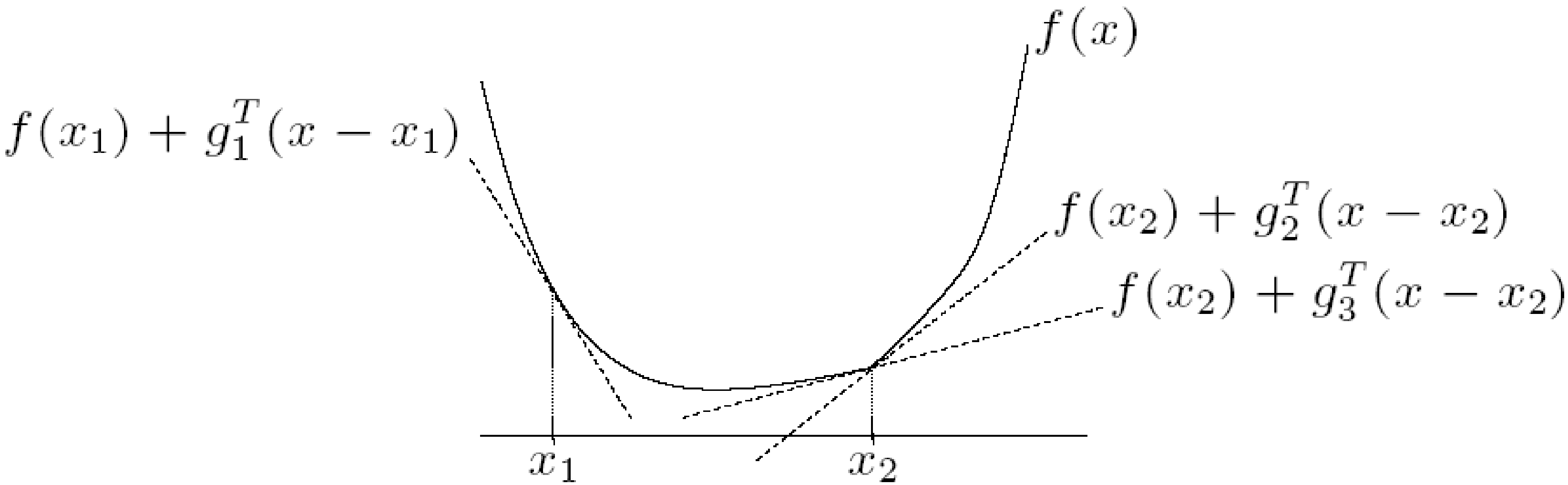
Subgradient of a function

g is a subgradient of f (not necessarily convex) at x if

$$f(y) \geq f(x) + g'(y - x), \forall y \in R^n$$

- subgradient provides an affine global under-estimator of f
- if f convex, then, there is at least one subgradient at every interior point of X
- If f is convex and differentiable, $\nabla f(x)$ is a subgradient of f at x
  This follows from the basic inequality for convex differentiable f
  $$f(y) \geq f(x) + \nabla f(x)'(y - x), \forall y \in R^n, \forall x \in R^n$$

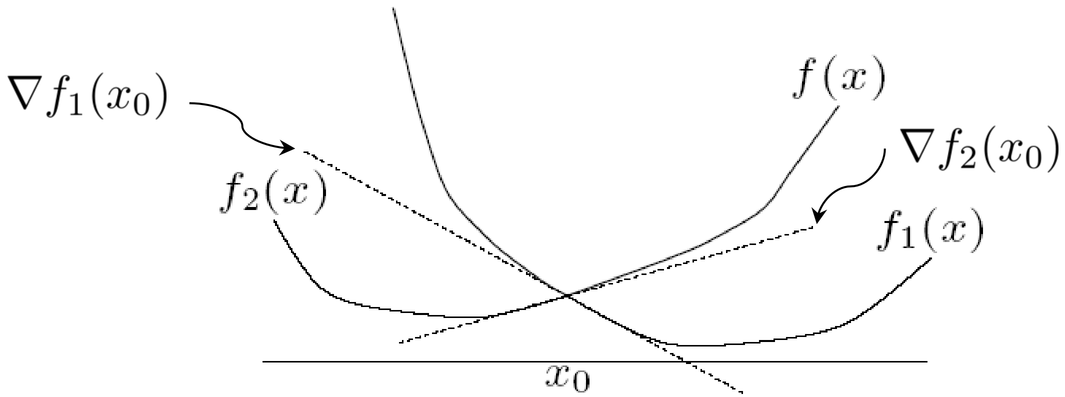
Subgradient: an example



g_2, g_3 are subgradients at x_2; g_1 is a subgradient at x_1

Subgradient: an example

$f = \max\{f_1, f_2\}$, with f_1, f_2 convex and differentiable



- $f_1(x_0) > f_2(x_0)$: unique subgradient $g = \nabla f_1(x_0)$
- $f_2(x_0) > f_1(x_0)$: unique subgradient $g = \nabla f_2(x_0)$
- $f_1(x_0) = f_2(x_0)$: subgradients form a line segment $[\nabla f_1(x_0), \nabla f_2(x_0)]$

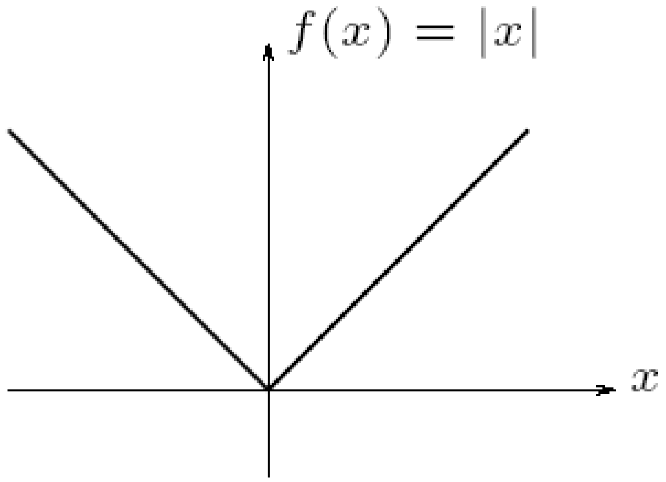
Subdifferential of a function

The set of all subgradient of f at x at is called the subdifferential of f at x. It is denoted as $\partial f(x)$

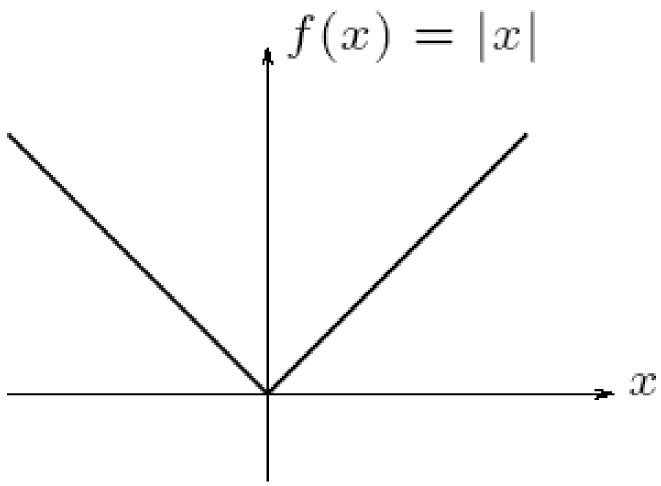
$$f(y) \geq f(x) + g'(y-x), \forall y \in R^n$$

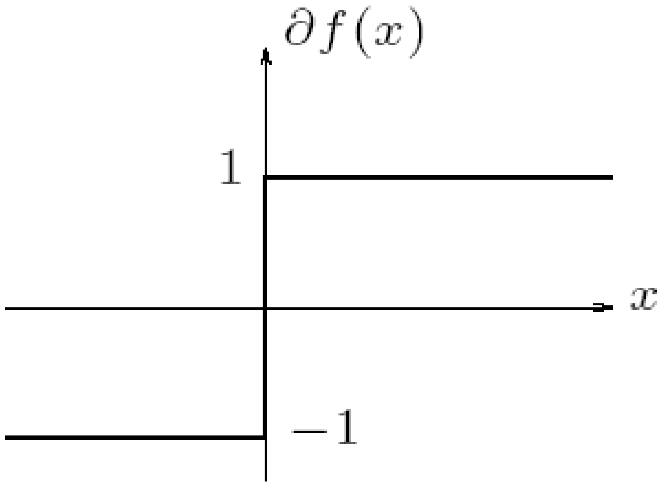
- $\partial f(x)$ is a closed convex set (intersection of closed half spaces, one for each y)
- $\partial f(x)$ nonempty if f convex
- $\partial f(x) = \{\nabla f(x)\}$ if f is differentiable at x
- if $\partial f(x) = \{g\}$, then f is differentiable at x and $g = \nabla f(x)$

Example



Example





Optimality conditions – unconstrained convex

Let $X \subseteq R^n$ be a convex set and $f: X \to R$ convex.

Necessary and sufficient condition for x^* to be a global minimum:

(i) $\nabla f(x^*) = 0$ [if f differentiable]

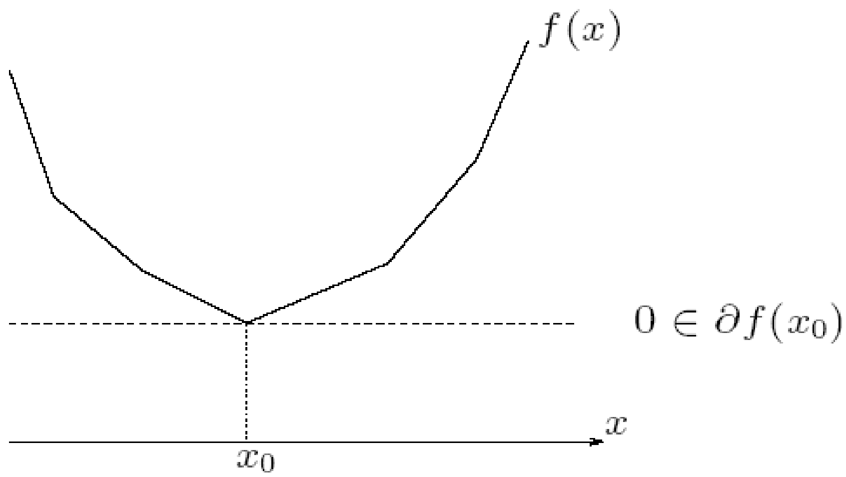
Optimality conditions – unconstrained convex

Let $X \subseteq R^n$ be a convex set and $f: X \rightarrow R$ convex.

Necessary and sufficient condition for x^* to be a global minimum:

(i) $\nabla f(x^*) = 0$ [if f differentiable]

(ii) $0 \in \partial f(x^*)$ [if f non differentiable]



Optimality conditions – unconstrained convex

Let $X \subseteq R^n$ be a convex set and $f: X \to R$ convex.

Necessary and sufficient condition for x^* to be a global minimum:

(i) $\nabla f(x^*) = 0$ [if f differentiable]

(ii) $0 \in \partial f(x^*)$ [if f non differentiable]

Proof [(ii)]

Optimality conditions – unconstrained convex

Let $X \subseteq R^n$ be a convex set and $f: X \to R$ convex.

Necessary and sufficient condition for x^* to be a global minimum:

(i) $\nabla f(x^*) = 0$ [if f differentiable]

(ii) $0 \in \partial f(x^*)$ [if f non differentiable]

Proof [(ii)]

by definition of subgradient g at x^* with $g = 0$, we get

$$f(y) \geq f(x^*) + 0'(y - x^*), \forall y \in X \implies x^* \text{ global minimum}$$

viceversa if x^* is a global minimum then

$$f(y) \geq f(x^*) = f(x^*) + 0'(y - x^*), \forall y \in X \implies 0 \in \partial f(x^*)$$

Optimality conditions – constrained convex

Let $X \subseteq R^n$ be a convex set and $f: X \to R$ convex.

Necessary and sufficient condition for x^* to be a global minimum:

(i) $\nabla f(x^*)'(y - x^*) \geq 0, \forall y \in X$ [if f differentiable]

Optimality conditions – constrained convex

Let $X \subseteq R^n$ be a convex set and $f: X \to R$ convex.

Necessary and sufficient condition for x^* to be a global minimum:

(i) $\nabla f(x^*)'(y - x^*) \geq 0, \forall y \in X$ [if f differentiable]

(ii) there exists a subgradient $g \in \partial f(x^*)$ such that

$$g'(y - x^*) \geq 0, \forall y \in X$$

[if f non differentiable]

Optimality conditions – constrained convex

Let $X \subseteq R^n$ be a convex set and $f: X \to R$ convex.

Necessary and sufficient condition for x^* to be a global minimum:

(i) $\nabla f(x^*)'(y - x^*) \geq 0, \forall y \in X$ [if f differentiable]

(ii) there exists a subgradient $g \in \partial f(x^*)$ such that

$$g'(y - x^*) \geq 0, \forall y \in X$$

[if f non differentiable]

Proof [(ii) only sufficient part]

by definition of subgradient g at x^*, we get

$$f(y) \geq f(x^*) + g'(y - x^*) \geq f(x^*), \forall y \in X$$

→ x^* global minimum

Convex optimization

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \to R$ is a **non differentiable convex function** over X

X is a non-empty closed convex subset of R^n

Projected subgradient – constrained opt

Subgradient method:

$$x_{k+1} = \mathrm{P}_X[x_k - c_k g_k]$$

where $g_k \in \partial f(x_k)$

Proposition [convergence for a diminishing stepsize]

Let $\{x_k\}$ be a sequence generated by the subgradient projection method with c_k asymptotically vanishing satisfying $\sum_{k=0}^{\infty} c_k = \infty$ and $\sum_{k=0}^{\infty} c_k^2 < \infty$.

If function f is Lipschitz continuous, then, $\{x_k\}$ converges to a global minimum.

Note: this results holds also for f differentiable

Projected subgradient – constrained opt

Proof.

Assume for simplicity that the global minimum is unique and denote it by x^*.

$$\|x_{k+1} - x^*\| = \|\mathrm{P}_X[x_k - c_k g_k] - x^*\| = \|\mathrm{P}_X[x_k - c_k g_k] - \mathrm{P}_X[x^*]\|$$
$$\leq \|[x_k - c_k g_k - x^*\|$$

since the projection operator is non-expansive.

Projected subgradient – constrained opt

Proof.

Assume for simplicity that the global minimum is unique and denote it by x^*.

$$\|x_{k+1} - x^*\| = \|\mathrm{P}_X[x_k - c_k g_k] - x^*\| = \|\mathrm{P}_X[x_k - c_k g_k] - \mathrm{P}_X[x^*]\|$$
$$\leq \|[x_k - c_k g_k - x^*\|$$

since the projection operator is non-expansive.

We then have

$$\|x_{k+1} - x^*\|^2 \leq \|[x_k - c_k g_k - x^*\|^2$$
$$= \|x_k - x^*\|^2 - 2c_k g_k'(x_k - x^*) + c_k^2\|g_k\|^2$$

Projected subgradient – constrained opt

Proof.

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2c_k g_k'(x_k - x^*) + c_k^2 \|g_k\|^2$$

Given that f is Lipschitz continuous, there exists L>0 such that

$$|f(x_k) - f(y)| \leq L\|x_k - y\|, \forall y \in R^n$$

By the definition of subgradient, we have that

$$f(y) \geq f(x_k) + g_k{}'(y - x_k), \forall y \in R^n$$

We then get

$$g_k{}'(y - x_k) \leq |f(y) - f(x_k)| \leq L\|x_k - y\|, \forall y \in R^n$$

which leads to $\|g_k\| \leq L$

Projected subgradient – constrained opt

Proof.

$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 - 2c_k g_k'(x_k-x^*) + c_k^2L^2$$

Since

$$g_k{}'(x^*-x_k) \le f(x^*)-f(x_k)$$

then

$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 + 2c_k\big(f(x^*)-f(x_k)\big) + c_k^2L^2$$

From this inequality we get that $\|x_k-x^*\|$ converges since $f(x^*)-f(x_k) < 0$.

Also, by computing $\sum_{k=0}^{\infty}$ of both sides, we get:

$$2\sum_{k=0}^{\infty} c_k\big(f(x_k)-f(x^*)\big) \le \|x_0-x^*\|^2 - \|x_\infty-x^*\|^2 + L^2\sum_{k=0}^{\infty} c_k^2$$

Projected subgradient – constrained opt

Proof.

$2\sum_{k=0}^{\infty} c_k\big(f(x_k) - f(x^*)\big) \le \|x_0 - x^*\|^2 - \|x_\infty - x^*\|^2 + L^2 \sum_{k=0}^{\infty} c_k^2$

entails that $\sum_{k=0}^{\infty} c_k\big(f(x_k) - f(x^*)\big)$ is bounded.

Since $\sum_{k=0}^{\infty} c_k = \infty$, then, $\liminf\limits_{k\to\infty}\big(f(x_k) - f(x^*)\big) = 0$.

This means

$f(x_k) \to f(x^*)$ across a subsequence and, due to the continuity of f, $x_k \to x^*$ along that subsequence.

Since $\|x_k - x^*\|$ converges, $x_k \to x^*$ along every subsequence.

Constrained optimization

$$\text{minimize } f(x)$$
$$\text{subject to } x \in X$$

$f: R^n \to R$ is convex function over X

X is a non-empty closed convex set of R^n

Proximal algorithm

At each k, a (feasible) tentative solution is computed as follows

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

where c_k is a positive scalar parameter weighting the quadratic regularization term that is added to $f(x)$

Proximal algorithm

At each k, a (feasible) tentative solution is computed as follows

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

where c_k is a positive scalar parameter weighting the quadratic regularization term that is added to $f(x)$

The regularization term makes the function to be minimized strictly convex and coercive so that it has a unique global minimum.

Proximal algorithm

At each k, a (feasible) tentative solution is computed as follows

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

where c_k is a positive scalar parameter weighting the quadratic regularization term that is added to $f(x)$

Cost decreases:

$$f(x_{k+1}) + \frac{1}{2c_k} \|x_{k+1} - x_k\|^2 \leq f(x_k) + \frac{1}{2c_k} \|x_k - x_k\|^2 = f(x_k)$$

Proximal algorithm

At each k, the (feasible) tentative solution is computed as follows

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

if $X = R^n$ a necessary and sufficient condition that x_{k+1} has to satisfy is

$\nabla h(x_{k+1}) = 0$ where $h(x) = f(x) + \frac{1}{2c_k} \|x - x_k\|^2$

if f differentiable

Proximal algorithm

At each k, the (feasible) tentative solution is computed as follows

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

if $X = R^n$ a necessary and sufficient condition that x_{k+1} has to satisfy is

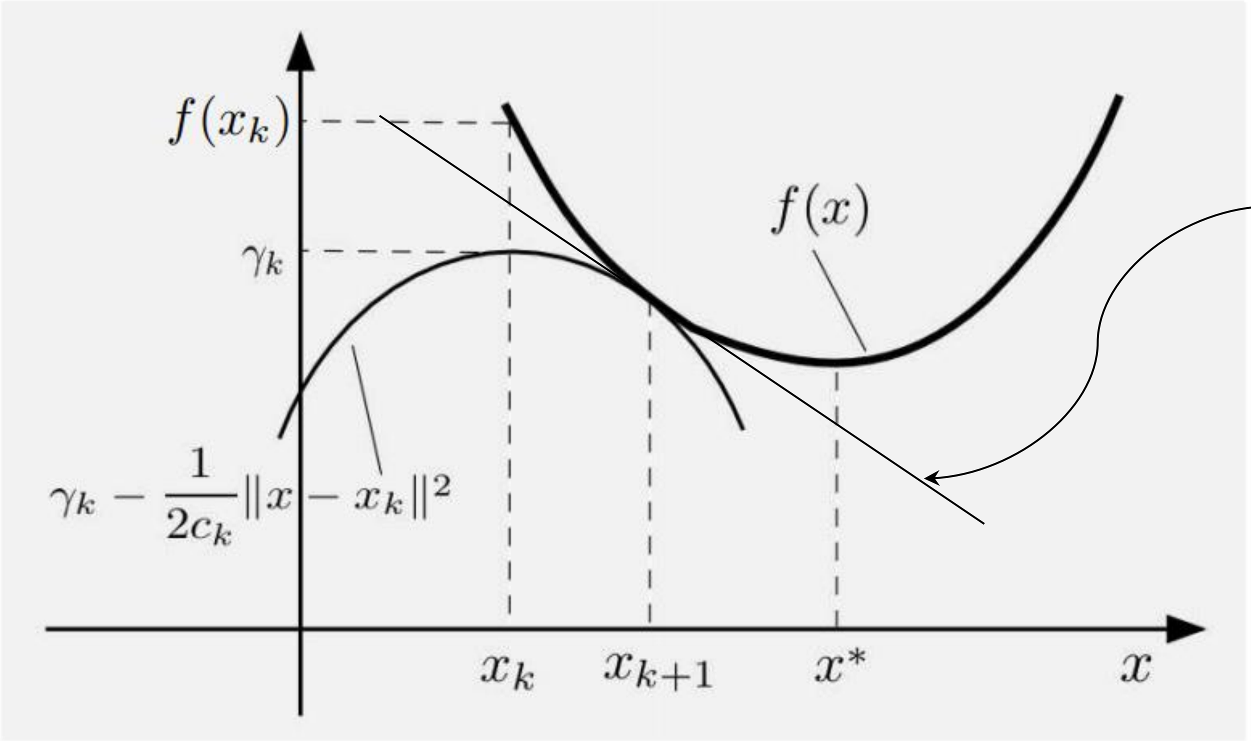
$\nabla h(x_{k+1}) = 0$ where $h(x) = f(x) + \frac{1}{2c_k} \|x - x_k\|^2$

if f differentiable

This entails that

$$\frac{x_k - x_{k+1}}{c_k} = \nabla f(x_{k+1})$$

Proximal algorithm



slope is the common gradient of $f(x)$ and $-\frac{1}{2c_k}\|x - x_k\|^2$ at x_{k+1} that is

$$\frac{x_k - x_{k+1}}{c_k}$$

Proximal algorithm

At each k, the (feasible) tentative solution is computed as follows

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

if $X = R^n$ a necessary and sufficient condition that x_{k+1} has to satisfy is

$\nabla h(x_{k+1}) = 0$ where $h(x) = f(x) + \frac{1}{2c_k} \|x - x_k\|^2$

if f differentiable

This entails that

$$\frac{x_k - x_{k+1}}{c_k} = \nabla f(x_{k+1}) \leftrightarrow x_{k+1} = x_k - c_k \nabla f(x_{k+1})$$

nearly a steepest gradient step

Proximal algorithm

At each k, the (feasible) tentative solution is computed as follows

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

if $X = R^n$ a necessary and sufficient condition that x_{k+1} has to satisfy is

$0 \in \partial h(x_{k+1})$ where $h(x) = f(x) + \frac{1}{2c_k} \|x - x_k\|^2$

if f non differentiable

Proximal algorithm

At each k, the (feasible) tentative solution is computed as follows

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

if $X = R^n$ a necessary and sufficient condition that x_{k+1} has to satisfy is

$0 \in \partial h(x_{k+1})$ where $h(x) = f(x) + \frac{1}{2c_k} \|x - x_k\|^2$

if f non differentiable

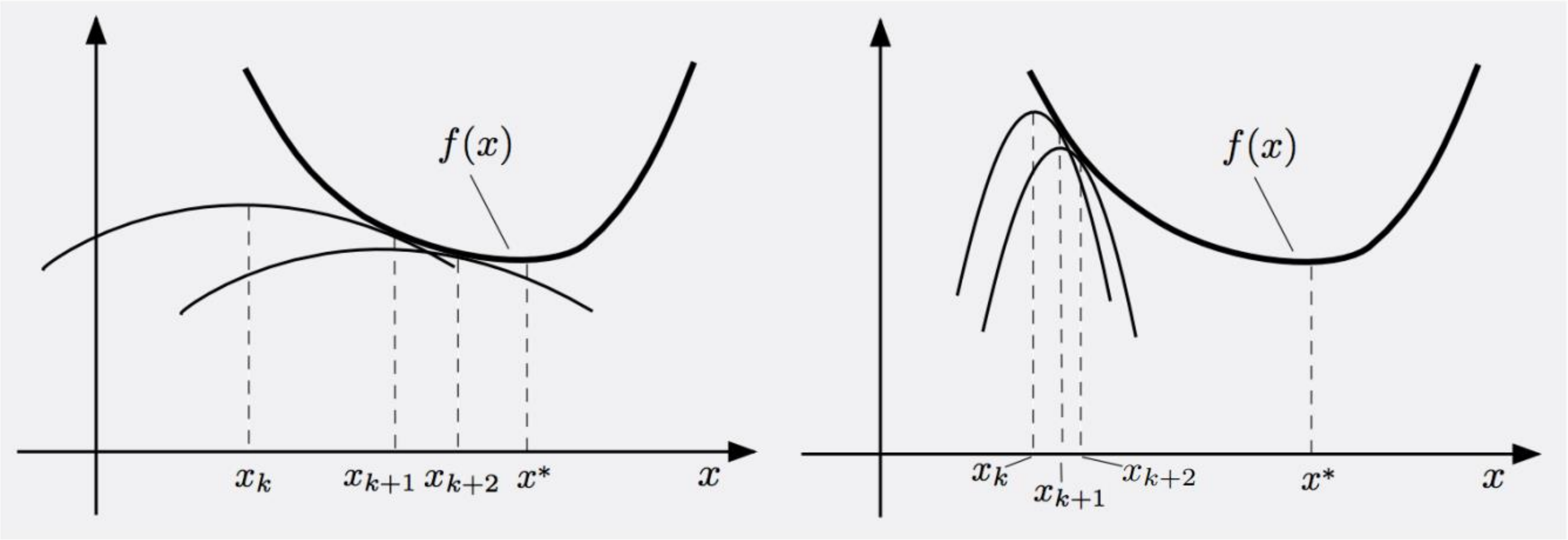
This entails that

$$\frac{x_k - x_{k+1}}{c_k} \in \partial f(x_{k+1})$$

→ a specific subgradient is chosen when performing a sort of subgradient step

Proximal algorithm

Role of the parameter c_k



Large values (on the left) cause a faster progress towards the optimum than smller values (on the right)

Constrained optimization

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } x \in X \end{aligned}$$

$f: R^n \to R$ is convex function over X

X is a non-empty closed convex set of R^n

Proximal minimization

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

Proposition [convergence for a diminishing stepsize]

Let $\{x_k\}$ be a sequence generated by the proximal algorithm with c_k asymptotically vanishing satisfying $\sum_{k=0}^{\infty} c_k = \infty$ and $\sum_{k=0}^{\infty} c_k^2 < \infty$.

Then, $\{x_k\}$ converges to a global minimum.

Note: this results holds for f differentiable or not differentiable

Proximal algorithm

Proof.

Instrumental Lemma

$$y^* = \underset{y \in Y}{\operatorname{argmin}} \{J_1(y) + J_2(y)\}$$

J_1 and J_2 convex, J_2 continuously differentiable. Then,

$$y^* = \underset{y \in Y}{\operatorname{argmin}} \{J_1(y) + \nabla J_2(y^*)' y\}$$

Proof.

Instrumental Lemma

$$y^* = \underset{y \in Y}{\operatorname{argmin}}\{J_1(y) + J_2(y)\}$$

J_1 and J_2 convex, J_2 continuously differentiable. Then,

$$y^* = \underset{y \in Y}{\operatorname{argmin}}\{J_1(y) + \nabla J_2(y^*)' y\}$$

If we apply this lemma to

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}}\left\{f(x) + \frac{1}{2c_k}\|x - x_k\|^2\right\}$$

we get

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}}\left\{f(x) + \frac{1}{c_k}(x_{k+1} - x_k)' x\right\}$$

Proximal algorithm

Proof.

$$x_{k+1} = \underset{x \in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{c_k}(x_{k+1} - x_k)' x \right\}$$

Then,

$$f(x_{k+1}) + \frac{1}{c_k}(x_{k+1} - x_k)' x_{k+1} \le f(x) + \frac{1}{c_k}(x_{k+1} - x_k)' x, x \in X$$

By re-ordering terms and multiplying by 2, we get

$$2(x_{k+1} - x_k)'(x_{k+1} - x) \le 2c_k(f(x) - f(x_{k+1}))$$

Proximal algorithm

Proof.

$$x_{k+1} = \underset{x\in X}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{c_k}(x_{k+1} - x_k)'x \right\}$$

Then,

$$f(x_{k+1}) + \frac{1}{c_k}(x_{k+1} - x_k)'x_{k+1} \leq f(x) + \frac{1}{c_k}(x_{k+1} - x_k)'x, x \in X$$

By re-ordering terms and multiplying by 2, we get

$$2(x_{k+1} - x_k)'(x_{k+1} - x) \leq 2c_k(f(x) - f(x_{k+1})), x \in X$$

Since

$$2(x_{k+1} - x_k)'(x_{k+1} - x) = \|x_{k+1} - x\|^2 + \|x_{k+1} - x_k\|^2 - \|x_k - x\|^2$$

we have

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - \|x_{k+1} - x_k\|^2 + 2c_k(f(x) - f(x_{k+1}))$$

Proximal algorithm

Proof.

Assume for simplicity that the global minimum of f is unique and denote it by x^*.

$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \|x_{k+1} - x_k\|^2 + 2c_k(f(x^*) - f(x_{k+1}))$

Then, $\|x_k - x^*\|$ converges (and is bounded)

By neglecting $-\|x_{k+1} - x_k\|^2$ and computing $\sum_{k=0}^{\infty}$ of both sides, we obtain

$$2\sum_{k=0}^{\infty} c_k(f(x_{k+1}) - f(x^*)) \leq \|x_0 - x^*\|^2 - \|x_\infty - x^*\|^2 < \infty$$

Proximal algorithm

Proof.

Since $\sum_{k=0}^{\infty} c_k = \infty$, then, $\liminf_{k\to\infty}\big(f(x_k) - f(x^*)\big) = 0$.

This means

$f(x_k) \to f(x^*)$ across a subsequence and, due to the continuity of f, $x_k \to x^*$ along that subsequence.

Since $\|x_k - x^*\|$ converges, $x_k \to x^*$ along every subsequence.

POLITECNICO
DI MILANO

Math Tools: Basics on constrained and convex optimization

Maria Prandini