Lec15-new

Wednesday, February 04, 2009 10:31 AM

Gradient projection algorithm

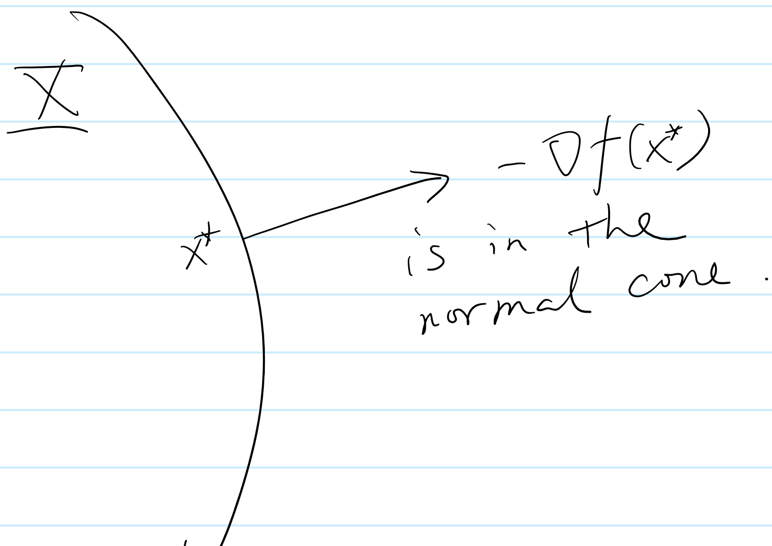
Wednesday, March 04, 2009 3:27 AM

The gradient projection algorithm

$$x(t+1) = [x(t) - \gamma \nabla f(x(t))]^+$$

Lemma: Assume that f is convex & differentiable. x^* minimizes $f(x)$ over X if and only if

$$x^* = [x^* - \gamma \nabla f(x^*)]^+$$



Proof: "$\Leftarrow$" Assume that

$$x^* = [x^* - \gamma \nabla f(x^*)]^+$$

Then $\forall y \in X$

$$(y - x^*)(x^* - \gamma \nabla f(x^*) - x^*) \le 0$$

$$\Leftrightarrow \quad (y - x^*) \cdot \nabla f(x^*) \ge 0$$

$\Leftrightarrow$ x^* if optimal

"$\Rightarrow$" can be shown analogously.

Theorem: If f is convex, ∇f is Lipschitz with parameter L, and there exists at least one point x^* such that

$$x^* = [x^* - \gamma \nabla f(x^*)]^+,$$

then the sequence of points

$$x(t+1) = [x(t) - \gamma \nabla f(x(t))]^+$$

converges if $0 < \gamma < \frac{2}{L}$, and the limit x_0 minimizes $f(x)$ over X.

Proof: Let $y(t+1) = x(t) - \gamma \nabla f(x(t))$
Let x^* be one of the optimal point
Let $y^* = x^* - \gamma \nabla f(x_0)$.

Using the techniques in the proof of the standard gradient algorithm, we have

$$\|y(t+1) - y^*\|^2 \le \|x(t) - x^*\|^2 - \left(\frac{2\gamma}{L} - \gamma^2\right) \|\nabla f(x(t)) - \nabla f(x^*)\|^2$$

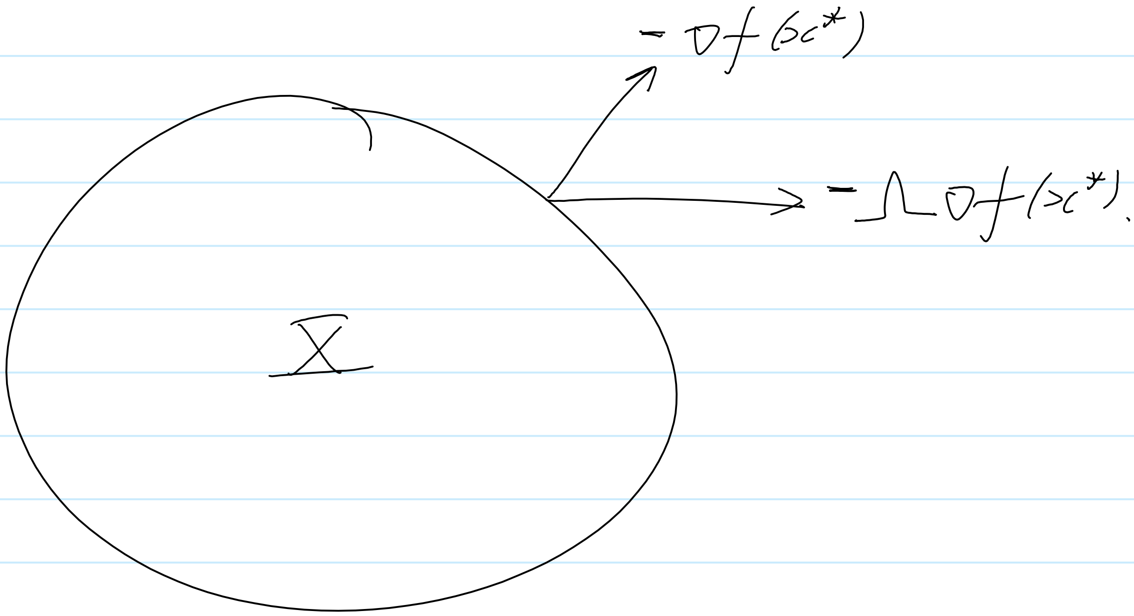
Using the projection theorem, we have

$$\|x(t+1) - x^*\|^2 \le \|y(t+1) - y^*\|^2.$$

The rest of the proof follows the case without projection.

Scaled projection

Wednesday, March 04, 2009 3:37 AM



- We may be tempted to do

$$x(t+1) = [x(t) - \gamma \Omega \nabla f(x(t))]^+$$

where Ω is a positive-definite matrix.

- However, in general the scaled version will not converge to the optimal point.

- We will have to redefine $[x]^+$ according to Ω, i.e.,

$$[x]^+_\Omega = \arg\min \; (z-x)^T \Omega^{-1} (z-x)$$
$$\text{sub to } \; z \in X$$

- In that case,

$$x(t+1) = \left[x(t) - \gamma \Omega \nabla f(x(t))\right]_{\Omega}^{+}$$

will converge to x^*

- Doesn't matter if $\underline{X}$ is a quadrant set & Ω is diagonal.
- Bertsekas p214–215

- For general constraint set $\underline{X}$, however, both the standard projection and the scaled projection are not easy to compute

$\Rightarrow$ The situation will be a lot better for the dual!

Subgradients

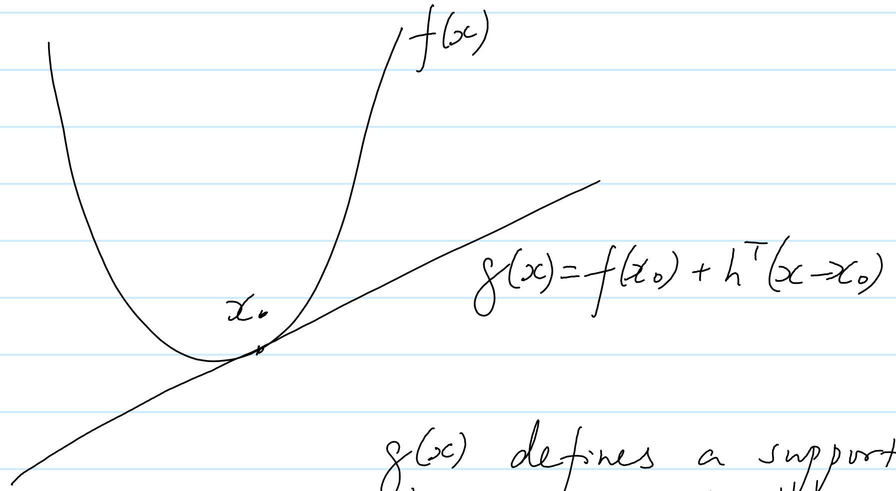
Monday, January 19, 2009 4:57 PM

- What if ∇f is not Lipschitz, or if the convex function f is not even differentiable
  - we will need to use the notion of sub-gradients

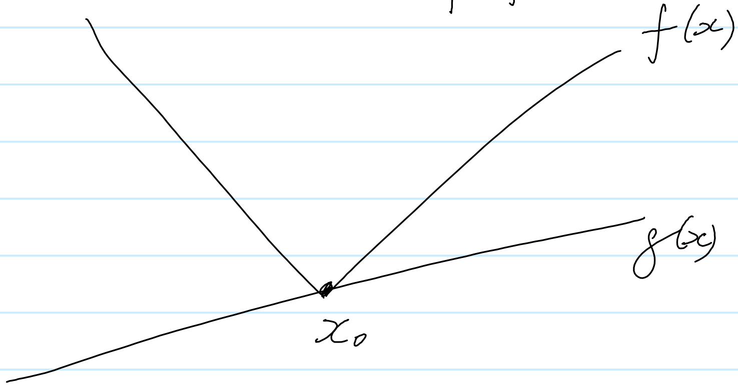
Define the subgradient of f at x_0 as a vector h such that

$$f(x) \geq f(x_0) + h^T(x - x_0)$$

for all $x \in C$.



$g(x)$ defines a supporting hyperplane of the epigraph of f.



The set of all subgradients of f at x_0 is called the sub-differential of f at x_0.
- denoted by $\partial f(x_0)$

Fact: A convex function f has non-empty sub-differentials at any $x \in \text{int}(\text{dom} f)$.

- Will revisit later when we study separation theorem
- If f is differentiable, then

$$\partial f(x_0) = \{\nabla f(x_0)\}$$

First-order necessary and sufficient conditions for unconstrained problems.

$$0 \in \partial f(\bar{x})$$

$$\Leftrightarrow \quad f(x) \geq f(\bar{x}) + 0\,(x - \bar{x}).$$

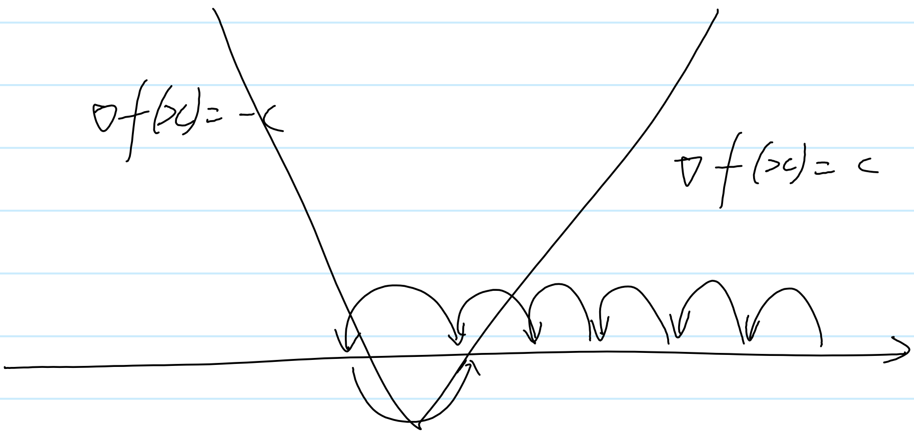
Subgradient descent

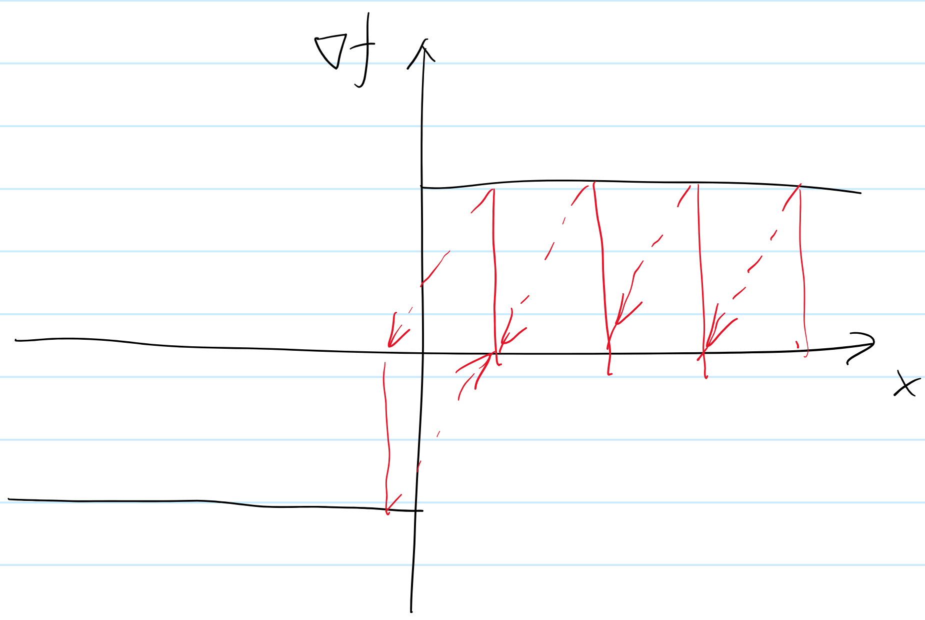
Wednesday, March 04, 2009 3:42 AM

- Subgradient-descent algo

$$x(t+1) = x(t) - \gamma \underbrace{\nabla f(x(t))}_{\text{a subgradient}}$$

- However, it may not converge to a single point.





- For any positive stepsize, $x(t)$ may oscillate around x^*.

- As the stepsize goes to zero, the deviation will go to zero.

- Let Ω be the set of points x such that $f(x) = f(x^*)$.

- Let Ω_ε be the set of points such that

$$f(x) \leq f(x^*) + \varepsilon.$$

- Assume that $\|\nabla f(x(t))\|$ is bounded by M.

- Recall that

$$\|x(t+1) - x^*\|^2 \leq \|x(t) - x^*\|^2 + \gamma^2 \|\nabla f(x(t))\|^2 - 2\gamma \cdot \nabla f(x(t))^T \cdot (x(t) - x^*)$$

- Since f is convex,

$$f(x^*) \geq f(x(t)) + \nabla f(x(t))^T (x^* - x(t)).$$

Hence, if $x(t) \notin \Omega_\varepsilon$, then

$$\nabla f(x(t))'(x^* - x(t)) \le f(x^*) - f(x(t)) \le -\varepsilon$$

– Let $\gamma < \frac{\varepsilon}{M^2} \Rightarrow \gamma\varepsilon \ge \gamma^2 M^2$

then

$$\|x(t+1) - x^*\|^2 \le \|x(t) - x^*\|^2 - \gamma\varepsilon$$

$\Rightarrow$ $x(t)$ must enter Ω_ε eventually.

– Once $x(t) \in \Omega_\varepsilon$, then if $\gamma < \frac{\varepsilon}{M}$,

$$\Rightarrow \|x(t+1) - x(t)\| \le \|\gamma \nabla f(x)\| \le \varepsilon$$

$\Rightarrow$ $x(t+1)$ cannot be far from Ω_ε.

We have proven the following.

Theorem: Assume that f is convex and its subgradients are bounded. Consider the subgradient descent algorithm

$$x(t+1) = x(t) - \gamma \nabla f(x(t))$$

where $\nabla f(x(t))$ is a sub-gradient of f at $x(t)$. Then for any $\varepsilon > 0$, there exists a stepsize γ such that for any initial condition $x(0)$, there exists a time T such that

$$d(x(t), \Omega) < \varepsilon \quad \forall t \ge T.$$

(60)

A similar result can be stated for constrained problems.

Diminishing stepsize

Wednesday, March 04, 2009 11:18 AM

- Alternatively, if we can let the stepsize goes to zero in an appropriate fashion, then we can still claim convergence

<u>Theorem</u>: Assume that f is convex & continuous, and its subgradient is bounded. Consider the subgradient-descent algorithm

$$x(t+1) = x(t) - \gamma_t \nabla f(x(t))$$

where $\nabla f(x(t))$ is a subgradient of f at $x(t)$. Further, assume that

$$\sum_{t=1}^{+\infty} \gamma_t \to +\infty$$

$$\sum_{t=1}^{+\infty} \gamma_t^2 < +\infty$$

then as $t \to +\infty$

$x(t)$ converges to x_0

and $\nabla f(x_0) = 0$, provided that there exists at least one point x^* with $\nabla f(x^*) = 0$.

Proof: $$\|x(t+1) - x^*\|^2 \le \|x(t) - x^*\|^2 + \gamma_t^2 \|\nabla f(x(t))\|^2 - 2\gamma_t \left(\nabla f(x(t))\right)^T (x(t) - x^*)$$

If $\nabla f(x(t))$ is bounded by M, then

$$\|x(t+1) - x^*\|^2 \le \|x(0) - x^*\|^2 + \sum_{k=1}^{t} \gamma_k^2 \cdot M^2$$

$$- 2 \sum_{k=1}^{t} \gamma_k \nabla f(x(k))^T (x(k) - x^*)$$

Note that

$$\sum_{k=1}^{\infty} \gamma_k^2 M^2 < +\infty$$

Hence

$$\sum_{k=1}^{\infty} \gamma_k \nabla f(x(t))^T (x(t) - x^*) < +\infty$$

skip

Further, note that

$$\nabla f(x(t))^T (x(t) - x^*) \geq f(x(t)) - f(x^*) \geq 0$$

Hence, using $\sum_{k=1}^{\infty} \gamma_k = +\infty$, there must exist a sub-sequence t_n such that

$$\lim_{n \to +\infty} \nabla f(x(t_n)) \cdot (x(t_n) - x^*) = 0$$

We can verify that $x(t_n)$ is a bounded sequence, Hence, there exists a limit point x_0.

First, we want to show that $f(x_0)$ is optimal

- Note that

$$f(x^*) \geq f(x(t_n)) + \nabla f(x(t_n))(x^* - x(t_n))$$

$$\Rightarrow \nabla f(x(t_n))(x(t_n) - x^*) \geq f(x(t_n)) - f(x^*)$$

- Hence

$$0 \leq \lim_{n \to +\infty} f(x(t_n)) - f(x^*) \leq \lim_{n \to +\infty} \nabla f(x(t_n))(x(t_n) - x^*)$$

$$= 0$$

$$\Rightarrow f(x_0) = \lim_{n \to +\infty} f(x(t_n)) = f(x^*)$$

Second, we want to show that $x(t) \to x_0$.

- Replacing x^* by x_0, we have

$$\|x(t+1) - x_0\|^2 \le \|x(t) - x_0\|^2 + \gamma_t^2 M^2$$

Summing over $t > t_n$

$$\|x(t+1) - x_0\|^2 \le \|x(t_n) - x_0\|^2 + \underbrace{\sum_{k=t_n+1}^{t} \gamma_k^2 M^2}_{\to 0 \text{ as } t_n \to +\infty}.$$

$$\therefore \quad \lim \|x(t) - x_0\|^2 \le \lim \|x(t_n) - x_0\|^2 = 0$$

#

(75)