LECTURE 13

ALGORITHMS FOR CONVEX NONSMOOTH OPTIMIZATION

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

$f : \mathbb{R}^n \to \mathbb{R}$, convex and nonsmooth (Lipschitz but not continuously differentiable)

Subgradient method

(1) Choose $s_k \in \partial f(x^{(k)})$

(2) Choose a step-length $\alpha_k > 0$ s.t.

$$f\left(x^{(k)} + \alpha_k \left(\frac{-s_k}{\|s_k\|}\right)\right) < f(x^{(k)})$$

if possible

(3) $x^{(k+1)} \leftarrow x^{(k)} + \alpha_k \left(\frac{-s_k}{\|s_k\|}\right)$, $k \leftarrow k+1$

(4) Repeat 1–4

Let x_* be a local (and global) minimizer of f

$$\|x^{(k+1)} - x_*\|_2^2 = \left\| x^{(k)} - x_* - \alpha_k \frac{s_k}{\|s_k\|} \right\|_2^2$$

$$= \|x^{(k)} - x_*\|_2^2 + \alpha_k^2 - \frac{2\alpha_k}{\|s_k\|} s_k^T (x^{(k)} - x_*)$$

$$\leqslant \|x^{(k)} - x_*\|_2^2 + \alpha_k^2 + \frac{2\alpha_k}{\|s_k\|} \underbrace{\left(f(x_*) - f(x^{(k)})\right)}_{\leqslant 0}$$

$$\leqslant \|x^{(k)} - x_*\|_2^2 + \alpha_k^2 .$$

If $\alpha_k > 0$ are chosen such that $\cancel{\alpha_k \to 0}$, $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$

$$\{\|x^{(k)} - x_*\|^2\}$$

(i) is bounded,

(ii) indeed convergent.

<u>THM</u>

Let $f: \mathbb{R}^n \to \mathbb{R}$ have a global minimizer. Furthermore, suppose α_k are chosen such that

$$\sum_{k=1}^{\infty} \alpha_k = \infty \quad \text{and} \quad \cancel{\lim_{k\to\infty} \alpha_k = 0}, \ \sum_{k=1}^{\infty} \alpha_k^2 < \infty .$$

Then the subgradient method generates a sequence $\{x^{(k)}\}$ converging to a global minimizer x_* at a sublinear rate.

Cutting plane method

Based on the piece-wise linear model

$$\phi_k(x) := \max\{\, l_j(x) \mid j = 1, \dots, k \,\}$$

where

$$l_j(x) := f(x^{(j)}) + s_j^T (x - x^{(j)})$$

$$s_j \in \partial f(x^{(j)})$$

(1) Choose $x^{(1)}$

(2) Find the global minimizer $x^{(k+1)}$ of $\phi_k(x)$ on S (compact subset of $\mathbb{R}^n$ containing a global min.)

(3) Define (Choose $s_{k+1} \in \partial f(x^{(k+1)})$)

$$\phi_{k+1}(x) := \max\{\, \phi_k(x), f(x^{(k+1)}) + s_{k+1}^T (x - x^{(k+1)}) \,\}$$

(4) Let

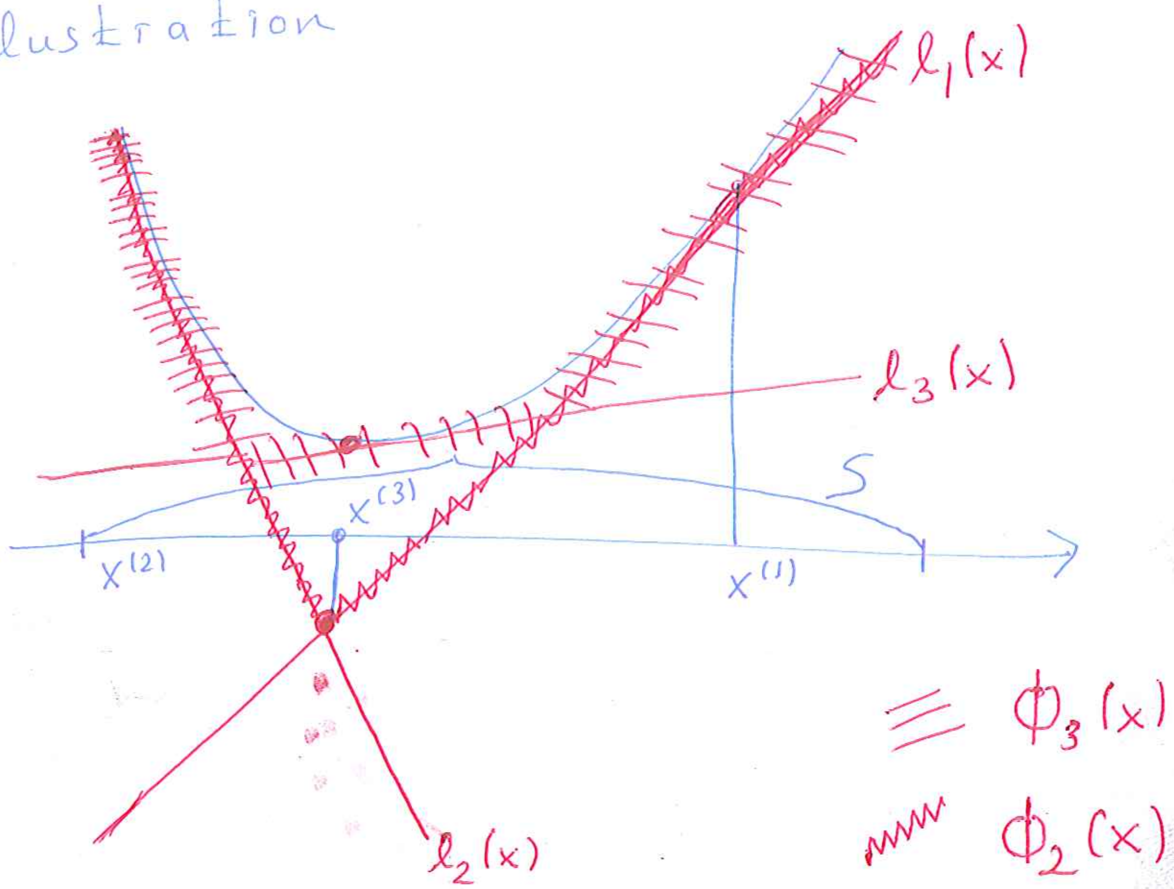
$$l_k := \min_{x \in S} \phi_k(x)$$

$$u_k := \min\{\, f(x^{(j)}) \mid j = 1, \dots, k \,\}$$

and $k \to k+1$

(5) Repeat 2-5.

Illustration



Notice

$$s_j \in \partial f(x_j) \implies$$

$$f(x) \geq \underbrace{f(x^{(j)}) + s_j^T (x - x^{(j)})}_{\ell_j(x)} \quad \forall x \implies$$

$$f(x) \geq \phi_k(x) \quad \forall x \implies$$

$$\min_{x \in S} f(x) \geq \min_{x \in S} \phi_k(x) = \ell_k$$

Moreover

$$\min_{x \in S} f(x) \leq \min \{ f(x^{(j)}) \mid j = 1, \ldots, k \} = u_k$$

Also observe

(i) $\ell_k \leq \ell_{k+1}$ ($\{\ell_k\}$ is non-decreasing)

(ii) $u_k \geq u_{k+1}$ ($\{u_k\}$ is non-increasing)

THM

Let $f : \mathbb{R}^n \to \mathbb{R}$ have a global minimizer x_*.
Cutting plane method generates sequences $\{l_k\}$ and $\{u_k\}$ such that

$$\lim_{k \to \infty} l_k = \lim_{k \to \infty} u_k = f(x_*).$$

(Penalized) Bundle Methods

Based on the model

$$P_k(x) := \phi_k(x) + \frac{\mu_k}{2} \|x - x^{(k)}\|_2^2$$

$\mu_k > 0$ - penalty parameter.

(1) Choose $x^{(0)}$, $\underline{y}^{(0)} \leftarrow x^{(0)}$ ($m \in (0,1)$ fixed parameter)

(2) Find the global minimizer $\underline{y}^{(k+1)}$ of $P_k(x)$

(3) $\delta_{k+1} \leftarrow f(x^{(k)}) - P_k(\underline{y}^{(k+1)})$ $\left(\begin{array}{l}\text{Note:} \\ f(x^{(k)}) = P_k(x^{(k)}) \\ \Rightarrow \delta_{k+1} \geq 0\end{array}\right)$

(4) If $f(x^{(k)}) - f(\underline{y}^{(k+1)}) \geq m \cdot \delta_{k+1}$

$\quad x^{(k+1)} \leftarrow \underline{y}^{(k+1)}$

$\quad K_s \leftarrow K_s \cup \{k+1\}$ (descent step)

else

$\quad x^{(k+1)} \leftarrow x^{(k)}$ (null step)

end

(5) Choose $s_{k+1} \in \partial f(y^{(k+1)})$, also $\mu_{k+1} > 1$ and define

$$\phi_{k+1}(x) := \max\{\phi_k(x), f(y^{(k+1)}) + s_{k+1}^T (x - y^{(k+1)})\}$$

$$P_{k+1}(x) := \phi_{k+1}(x) + \frac{\mu_{k+1}}{2} \|x - x^{(k+1)}\|_2^2$$

and

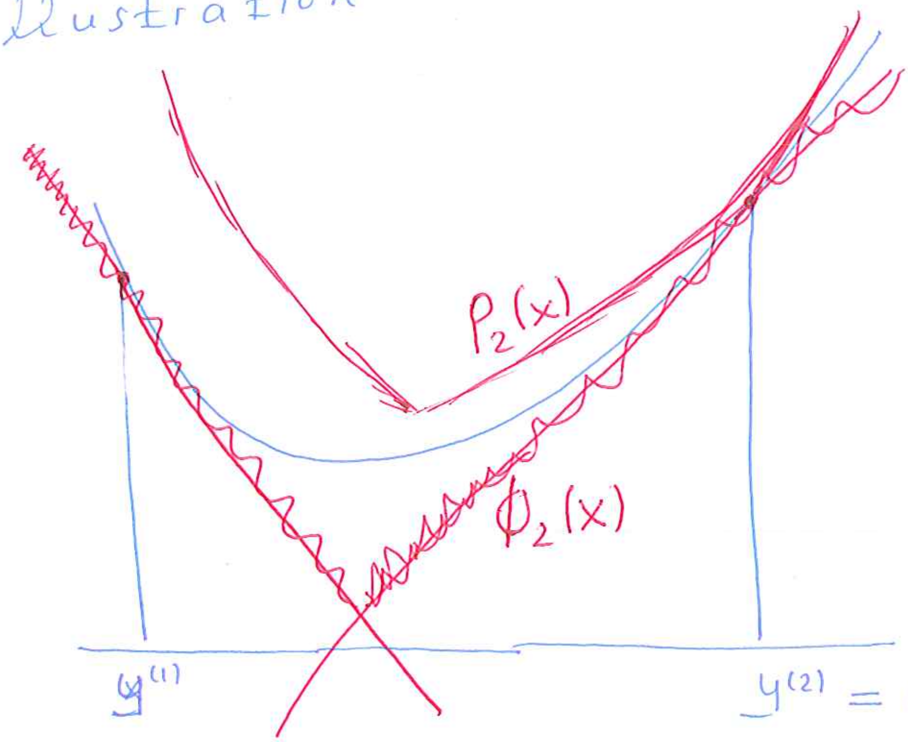
let $k \leftarrow k+1$

(6) Repeat 2-5

$\{y^{(k)}\}$ — sequence of global minimizers of $P_{k-1}(x)$

$\{x^{(k)}\}$ — sequence of descent points

Illustration



THM

Suppose $f: \mathbb{R}^n \to \mathbb{R}$ has a global minimizer, and K_s contains infinitely many integers. Let $f_* := \lim_{\ell \to \infty} f(x^{(k_\ell)})$ where $K_s = \{k_\ell\}$. We have $\sum_{k \in K_s} \delta_k \leq \frac{f(x^{(0)}) - f_*}{m}$.

Suppose $\{y^{(k)}\}$ is convergent. Then $\{x^{(k)}\}$ is also convergent. Thm above implies

$$\delta_{k_\ell} \to 0 \quad \text{as } \ell \to \infty$$

$$\Longrightarrow \lim_{k \to \infty} f(x^{(k_\ell)}) = \lim_{k \to \infty} P_{k_\ell}(y^{(k_\ell+1)})$$

$$= \lim_{k \to \infty} \phi_{k_\ell}(z^{(k_\ell+1)})$$

where $z^{(k_\ell+1)}$ is a global min of $\phi_{k_\ell}(x)$.

Consequently,

$$f(x^{(k_\ell)}) \geq \min_{x \in \mathbb{R}^n} f(x) \geq \phi_k(z^{(k_\ell+1)})$$

$$\Longrightarrow \lim_{\ell \to \infty} f(x^{(k_\ell)}) = \min_{x \in \mathbb{R}^n} f(x) = \lim_{\ell \to \infty} \phi_k(z^{(k_\ell+1)})$$

$$\Longrightarrow \lim_{k \to \infty} f(x^{(k)}) = \min_{x \in \mathbb{R}^n} f(x) = \lim_{k \to \infty} \phi_k(z^{(k)})$$

Proof of thm

For each ℓ

$$f(x^{(k_\ell - 1)}) - f(y^{(k_\ell)}) \geq m \cdot \delta_{k_\ell}$$

$$f(x^{(k_{\ell+1}-1)}) - f(y^{(k_{\ell+1})}) \geq m \cdot \delta_{k_{\ell+1}}$$

But notice that $x^{(k_{\ell+1}-1)} = y^{(k_\ell)}$

~~and $x^{(k_\ell}$~~

$$f(x^{(k_\ell - 1)}) - f(y^{(k_{\ell+1})}) \geq m\,(\delta_{k_\ell} + \delta_{k_{\ell+1}})$$

$$\Longrightarrow f(x^{(k_\ell - 1)}) - f(y^{(k_{\ell+j})}) \geq m\,(\delta_{k_\ell} + \cdots + \delta_{k_{\ell+j}})$$

Taking the limit as $j \to \infty$

$$f(x^{(0)}) - f_* \geq \sum_{k \in K_s} m\,\delta_k \,.$$

$\square$