Amath 515, Feb 3

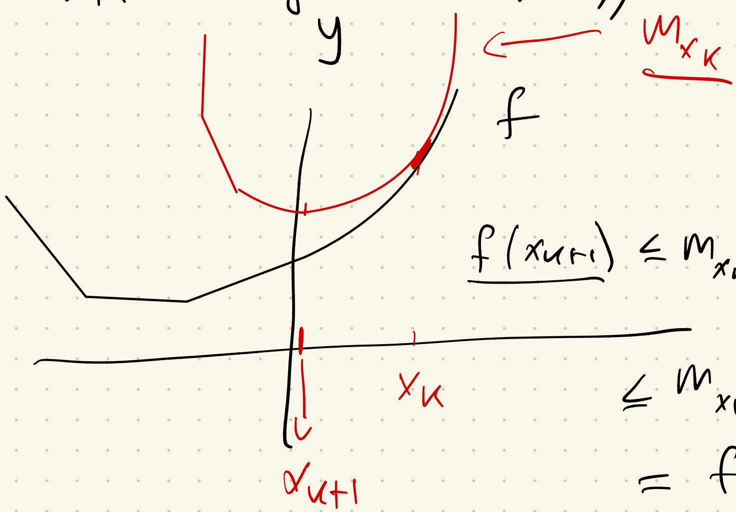
Last time: $\min_x f(x) := \underset{\substack{\downarrow \\ \beta\text{-smooth} \\ \text{cvx}}}{g(x)} + \underset{\text{cvx}}{h(x)}$

Strategy: at x_k, build model for f

$$m_{x_k}(y) = g(x_k) + \langle \nabla g(x_k), y - x_k \rangle + \frac{\beta}{2} \|y - x_k\|^2 + h(y)$$

- $f(y) \le m_{x_k}(y)$ everywhere ✓
- $f(x_k) = m_{x_k}(x_k)$ ✓

Take $x_{k+1} = \underset{y}{\operatorname{argmin}}\ m_{x_k}(y)$



$$\underline{f(x_{k+1})} \le m_{x_k}(x_{k+1})$$

$$\le m_{x_k}(x_k)$$

$$= \underline{f(x_k)}$$

$$m_{x_k}(y) = g(x_k) + \langle \nabla g(x_k), y - x_k \rangle + \frac{\beta}{2}\|y - x_k\|^2 + h(y)$$

$$= g(x_k) + \boxed{\frac{\beta}{2}\left\|y - \left(x_k - \frac{1}{\beta}\nabla g(x_k)\right)\right\|^2 + h(y)} - \frac{1}{2\beta}\|\nabla g(x_k)\|^2$$

$$x_{k+1} = \underset{y}{\operatorname{argmin}} \ \frac{1}{2/\beta}\left\|y - \left(x_k - \frac{1}{\beta}\nabla g(x_k)\right)\right\|^2 + h(y)$$

$$:= \operatorname{prox}_{\frac{1}{\beta}h}\left(x_k - \frac{1}{\beta}\nabla g(x_k)\right)$$

$$\operatorname{prox}_{th}(z) = \underset{y}{\operatorname{argmin}} \ \frac{1}{2t}\|y - z\|^2 + h(y)$$

* will be able to show this works well, just as well as gradient descent for $\underset{x}{\min}\ g(x)$

$$\operatorname{prox}_{th}(z) = \underline{x^+} = \underbrace{\arg\min_y \frac{1}{2t}\|y-z\|^2 + h(y)}$$

$$0 \in \frac{1}{t}(x^+ - z) + \partial h(x^+)$$

$$\boxed{\frac{1}{t}(z - x^+) \in \partial h(x^+)} \; *$$

→ what does it mean if $\underline{x_{k+1} = x_k}$?

$$z = x_k - \frac{1}{\beta}\nabla g(x_k), \quad t = \frac{1}{\beta}$$

$$\cancel{\beta}\left(\cancel{x_k} - \cancel{\frac{1}{\beta}}\nabla g(x_k) - \cancel{x_k}\right) \in \partial h(x_k)$$

$$\boxed{-\nabla g(x_k) \in \partial h(x_k)}$$

$$0 \in \underbrace{\nabla g(x_k) + \partial h(x_k)}_{\partial f(x_k)}$$

$\therefore$ $\underline{x_k \text{ is global minimizer for } f.}$

$$\mathrm{prox}_{th}(z) = \underset{y}{\arg\min} \frac{1}{2t}\|y-z\|^2 + h(y)$$

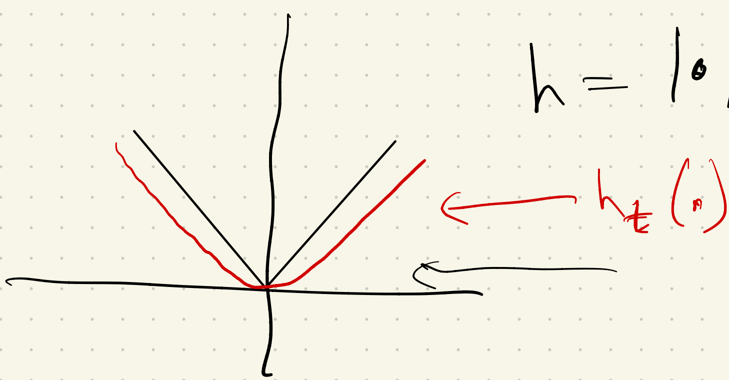
$$h_t(z) = \min_y \frac{1}{2t}\|y-z\|^2 + h(y)$$

Call the prox y_z

$$= \frac{1}{2t}\|y_z - z\|^2 + h(y_z)$$

$h_t(z)$ is called the Moreau envelope of h, and has lots of uses itself.

$h = |\cdot|$



Suppose $h(x) = \delta_C(x)$, C convex closed

$$\text{prox}_{th}(z) = \underset{y}{\operatorname{argmin}} \ \frac{1}{2t}\|z-y\|^2 + \delta_C(y)$$

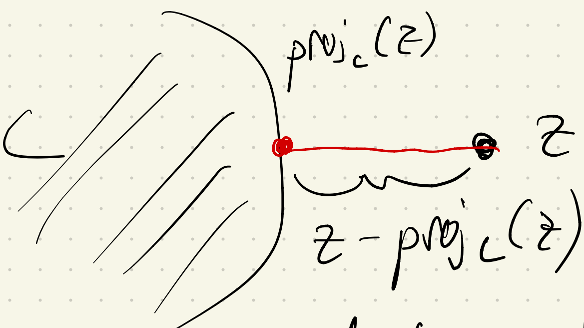
$$\delta_C(y) = \begin{cases} 0 & y \in C \\ \infty & y \notin C \end{cases}$$

$$= \underset{y \in C}{\operatorname{argmin}} \ \frac{1}{2t}\|z-y\|^2$$

$$:= \text{proj}_C(z)$$

What is h_t in this case?

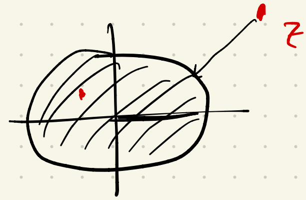
$$h_t(z) = \frac{1}{2t}\|z - \text{proj}_C(z)\|^2$$



$$= \boxed{\frac{1}{2t}\text{dist}^2(z|C)}$$

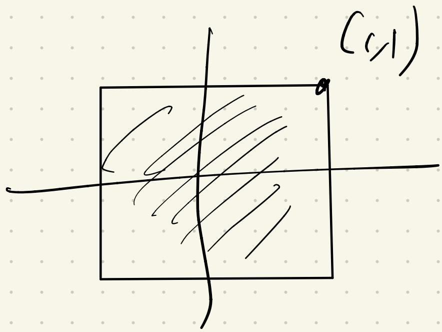
"distance from point z to set C"

Ex 1: $C = \mathbb{B}_2$



$$\text{proj}_{\mathbb{B}_2}(z) = \begin{cases} \frac{z}{\|z\|} & \|z\| > 1 \\ z & \|z\| \le 1 \end{cases}$$

Ex 2: $C = \mathbb{B}_\infty$



$$\text{proj}_{\mathbb{B}_\infty}(z) = \underset{\|y\|_\infty \le 1}{\arg\min} \ \frac{1}{2}\|y - z\|^2$$

$$= \underset{|y_i| \le 1}{\arg\min} \sum_{i=1}^{n} \frac{1}{2}(y_i - z_i)^2$$

every coordinate independent

$$\text{proj}_{[-1,1]}(z_i) = \begin{cases} 1 & z_i > 1 \\ z_i & |z_i| \le 1 \\ -1 & z_i < -1 \end{cases}$$

$$\min_{\|x\|_\infty \le 1} \frac{1}{2}\|Ax-b\|^2$$

$$x_{k+1} = \text{proj}_{B_\infty}\left(x_k - \frac{1}{\beta}A^T(Ax_k - b)\right)$$

$$= \min\left(1, \max\left(-1, x_k - \frac{1}{\beta}A^T(Ax-b)\right)\right)$$

$$h(x) = \lambda\|x\|_1$$

β - smooth

$$\min_x \; g(x) + \lambda\|x\|_1$$

Algorithm

$$x_{k+1} = \text{prox}_{\frac{\lambda}{\beta}\|\cdot\|_1}\left(x_k - \frac{1}{\beta}\nabla g(x_k)\right)$$

$$\text{prox}_{\alpha\|\cdot\|_1} ? \qquad \alpha = \frac{\lambda}{\beta}$$

$$\mathrm{prox}_{\alpha\|\cdot\|_1}(z) = \underset{y}{\arg\min}\ \frac{1}{2\alpha}\|y-z\|^2 + \|y\|_1$$

$$= \underset{y}{\arg\min} \sum_{i=1}^{n} \underbrace{\frac{1}{2\alpha}(y_i - z_i)^2 + |y_i|}_{f_i(y_i)}$$

$$= \underset{y}{\arg\min} \sum_i f_i(y_i)$$

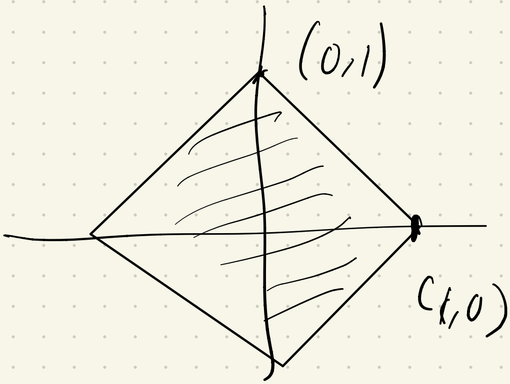
$$\min_y \frac{1}{2\alpha}(y-z)^2 + |y|$$

<u>use cases + calculus</u>

- $\hat{y}$ might be <u>pos</u>
- $\hat{y}$ might be neg
- otherwise $\hat{y} = 0$

Ex: $C = \mathbb{B}_1$

$\{x : \|x\|_1 \leq 1\}$



* No closed form solution

① Deterministic routine with complexity $O(n \log n)$, n dim of C

② Techiques from duality

Analysis of prox-gradient

Problem: $\min_x f(x) = g(x) + h(x)$

Opt: $0 \in \partial f(x) \iff -\nabla g(x) \in \partial h(x)$

Algo: $x_{k+1} = \text{prox}_{\frac{1}{\beta}h}\left(x_k - \frac{1}{\beta}\nabla g(x_k)\right)$

$= \text{prox}_{th}\left(x_k - t\nabla g(x_k)\right), \quad t = \frac{1}{\beta}$

$$x^+ = \text{prox}_{th}\left(x - t\nabla g(x)\right)$$

prox: $\text{prox}_{th}(z) = \underset{y}{\text{argmin}} \frac{1}{2t}\|y - z\|^2 + h(y)$

$$\frac{1}{t}(z - x^+) \in \partial h(x^+)$$

"z" $= \underline{x - t\nabla g(x)}$

$$\frac{1}{t}\left(x - t\nabla g(x) - x^+\right) \in \partial h(x^+)$$

$$\underbrace{\frac{1}{t}(x - x^+)}_{G_t(x)} - \nabla g(x) \in \partial h(x^+)$$

$G_t(x) := \frac{1}{t}(x - x^+)$ "step"

* perfect termination criteria

If $G_t(x) = 0$, that means:

1. $x = x^+$
2. $-\nabla g(x) \in \partial h(x)$, <u>$x$ optimal</u>

G_t acts as a "gradient" for the more complicated $g+h$ problem

Main thm 3.12

$f = g + h$

- g is α-convex ($\alpha = 0$ means cvx)
- g is β-smooth
- h is convex
- $x^+ = \text{prox}_{th}(x - \nabla g(x))$
- $G_t(x) = \frac{1}{t}(x - x^+)$

thm:

$$f(y) \geq f(x^+) + \langle G_t(x), y - x \rangle + t\left(1 - \frac{\beta t}{2}\right)\|G_t(x)\|^2 + \frac{\alpha}{2}\|y - x\|^2$$

- y arbitrary
- x, x^+

Take $\underline{\alpha = 0}$, $t = \frac{1}{\beta}$, $y = x$

$$f(x) \geq f(x^+) + \frac{1}{2\beta}\|G_t(x)\|^2$$

$$f(x^+) \leq f(x) - \frac{1}{2\beta}\|G_t(x)\|^2$$