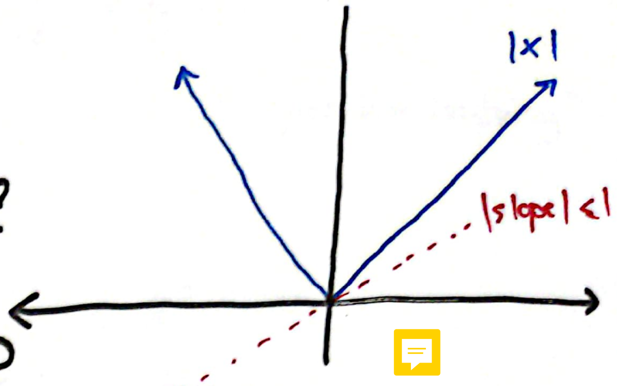
Starting 9/30

Example $f(x) = |x|$



What is $\partial f(x)$ and $\mathrm{prox}_{\alpha f}(x)$?

Claim: $$\partial f(x) = \begin{cases} \{1\} & \text{if } x > 0 \\ \{-1\} & \text{if } x < 0 \\ [-1,1] & \text{if } x = 0. \end{cases}$$

$\partial f(x) \approx$ convex hull nearby gradients

Proof. At $x \neq 0$, the gradient is the unique vector w/

$$\lim_{y \to x} \frac{f(y) - f(x) + g^T(y-x)}{\|y-x\|} = 0.$$

$\Rightarrow \partial f(x) = \{\nabla f(x)\}$.

At $x=0$, check! Proof by picture.

Claim: $$\mathrm{prox}_{\alpha f}(\bar{x}) = \begin{cases} 0 & \text{if } \bar{x} \in (-\alpha, \alpha) \\ \bar{x} - \alpha & \text{if } \bar{x} \geq \alpha \\ \bar{x} + \alpha & \text{if } \bar{x} \leq -\alpha. \end{cases}$$

Proof. Let $\bar{x}^* = \mathrm{prox}_{\alpha f}(\bar{x})$, then $\frac{1}{\alpha}(\bar{x} - \bar{x}^*) \in \partial f(\bar{x}^*)$.
(by HW2, Q2)

If $\bar{x}^* > 0$, $\frac{1}{\alpha}(\bar{x} - \bar{x}^*) = 1 \Leftrightarrow \underline{\bar{x}^* = \bar{x} - \alpha}$
(Further, $\bar{x} > \alpha$).

If $\bar{x}^* < 0$, $\frac{1}{\alpha}(\bar{x} - \bar{x}^*) = -1 \Leftrightarrow \underline{\bar{x}^* = \bar{x} + \alpha}$
(Further $\bar{x} < -\alpha$).

If $\underline{\bar{x}^* = 0}$, $\frac{1}{\alpha}(\bar{x} - \bar{x}^*) \in [-1,1]$, $\Leftrightarrow$ $\bar{x} \in \bar{x}^* + [-\alpha, \alpha]$
$\underline{= [-\alpha, \alpha]}$..

(Recall LASSO, $\min \|Ax-b\|_2^2 + \gamma \underbrace{\|x\|_1}_{=\sum |x_i|}$)

"spersity inducing"

<u>Lemma (Seperable Prox Functions)</u>

$f: \mathbb{R}^d \to \mathbb{R}$

Suppose $f(x) = \sum_{i=1}^{d} f_i(x_i)$, then

$$\text{prox}_{\alpha f}(\bar{x}) = \text{prox}_{\alpha f_1}(\bar{x}_1) \times \ldots \times \text{prox}_{\alpha f_d}(\bar{x}_d).$$

Proof.

$$\min_x \left\{ f(x) + \frac{1}{2\alpha}\|x - \bar{x}\|_2^2 \right\}$$

$$= \min_x \left\{ \sum_{i=1}^{d} f_i(x_i) + \frac{1}{2\alpha} \sum_{i=1}^{d} (x_i - \bar{x}_i)^2 \right\}$$

$$= \min \left\{ \sum \left(f_i(x_i) + \frac{1}{2\alpha}(x_i - \bar{x}_i)^2 \right) \right\}$$

$$= \sum \min_{x_i} \left\{ f_i(x_i) + \frac{1}{2\alpha}(x_i - \bar{x}_i)^2 \right\}.$$

$\Rightarrow$ Minimize seperately over each x_i. $\square$

$$\Rightarrow \left[\text{prox}_{\alpha\|\cdot\|_1}(x)\right]_i = \begin{cases} 0 & \text{if } x_i \in [-\alpha, \alpha] \\ x_i - \alpha & \text{if } x_i > \alpha \\ x_i + \alpha & \text{if } x_i < -\alpha \end{cases}.$$

"Soft Thresholding".

Iterating the prox operation, you get
"the Proximal Point Method"

$$x_{k+1} = \mathrm{prox}_{\alpha_k f}(x_k)$$

By HW2, Q2, $\frac{1}{\alpha_k}(x_k - x_{k+1}) \in \partial f(x_{k+1})$

$$\Leftrightarrow \quad x_k - x_{k+1} \in \alpha_k \cdot \partial f(x_{k+1})$$

$$\Leftrightarrow \quad x_{k+1} \in x_k - \alpha_k \partial f(x_{k+1})$$

Looks like gradient descent
(i) Gradient → Subgradient
(ii) using 1^{st}-order object at $k+1$ instead of k.

(for ODE background, Backward Euler Step)

Lemma Every strongly convex function f has a unique minimizer.

Proof. HW 3

Then for convex f, $f(x) + \frac{1}{2\alpha}\|x - \bar{x}\|_2^2$ is $0 + \frac{1}{\alpha}$-strongly

$\overset{\text{by Lemma}}{\Rightarrow}$ $\mathrm{prox}_{\alpha f}$ is well-defined and a singleton.

Abuse notation, $x_{k+1} = \mathrm{prox}_{\alpha f}(x_k)$.

2. Constraints via Proximal Operator

Suppose we want to minimize $f(x)$ over $x \in S$.

We will use the extended reals $\mathbb{R} \cup \{\pm\infty\}$ to write this as a single objective function.

Define indicator function

$$\delta_S(x) = \begin{cases} +\infty & \text{if } x \notin S \\ 0 & \text{if } x \in S. \end{cases}$$

Then we consider

$$\min_{x \in \mathbb{R}^d} f(x) + \delta_S(x) = \min_{x \in S} f(x).$$

We can handle f via gradients if its smooth, need prox to handle δ_S.

<u>Lemma</u> $\operatorname{prox}_{\alpha\delta_S}(\bar{x}) = \operatorname{proj}_S(\bar{x})$

$$:= \operatorname{argmin}\{\|x-\bar{x}\|_2 \mid x \in S\}$$

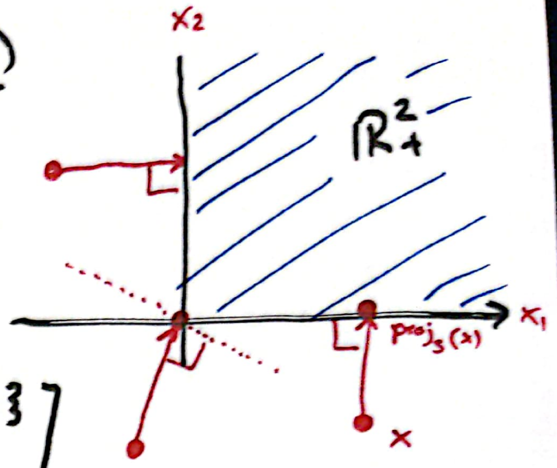
"orthogonal projection" moving to nearest feasible point.

Proof. $\operatorname{prox}_{\alpha\delta_S}(\bar{x}) = \operatorname{argmin}\{\delta_S(x) + \frac{1}{2\alpha}\|x-\bar{x}\|_2^2\}$

$$= \operatorname{argmin}\{\|x-\bar{x}\|_2^2 \mid x \in S\}. \quad \square$$

Example (Nonnegativity Constraints)

$S = \{ x \text{ s.t. } x_i \geq 0 \} = \mathbb{R}^d_+$



Claim: $\text{proj}_S(x) = [x]_+ := \begin{bmatrix} \max\{x_1, 0\} \\ \vdots \\ \max\{x_d, 0\} \end{bmatrix}$

Proof. $\text{proj}_S(x) = \text{prox}_{\delta_S}(x)$

$= \text{prox}_{\delta_{\mathbb{R}^d_+}}(x)$

$= \text{prox}_{\delta_{\mathbb{R}_+}}(x_1) \times \dots \times \text{prox}_{\delta_{\mathbb{R}_+}}(x_d)$

$\delta_{\mathbb{R}^d_+}(x) = \sum_{i=1}^{d} \delta_{\mathbb{R}_+}(x_i) = \begin{cases} +\infty & \text{if any } x_i < 0 \\ 0 & \text{otherwise} \end{cases}$

Check $\text{proj}_{\mathbb{R}_+}(x) = \max\{x, 0\}$. $\square$

Example (Grading Polyhedral)

$S = \{ (H, M, F) \mid Ax \leq b \}$

↖ 7 inequalities in syllabus.

$$\text{proj}_S(\bar{x}) = \underset{\text{s.t. } x \in S.}{\text{argmin}} \; \|x - \bar{x}\|_2^2$$

Quadratic Programming