AMATH 840, Lecture 08, Jan 31, 2022

RIP — $m \gtrsim O(s \ln(\frac{n}{s}))$

Today

① Examples of matrices that satisfy the RIP cond.

② Sparse Optimization Methods

Coherence $m \simeq O(s^2)$

(i) Proximal Methods : nondiff $\|x\|_1$

(ii) Lagrangian Methods : constrained → unconstrained

Image processing → $\|\nabla x\|_1$ → ADMM

References

① A Mathematical Intro. to Compressive Sensing by Foucart and Rauhut, Chapters 9, 12, 14.

② Convex Optimization, by Boyd & Vandenberghe

③ Slides from | EE 236C, by Vandenberghe, UCLA
CMSC 764, by T. Goldstein, Uni of Maryland

3 ℓ_1 - Optimization Methods

1. $$\min_{z\in\mathbb{C}^n} \|z\|_1 \quad \text{s.t} \quad \|y-Az\|_2 \le \eta, \text{ for } \eta \ge 0$$

   Lagrangian methods

2. $$\min_{z} \|y-Az\|_2^2 \quad \text{s.t} \quad \|z\|_1 \le \tau$$

   APMM

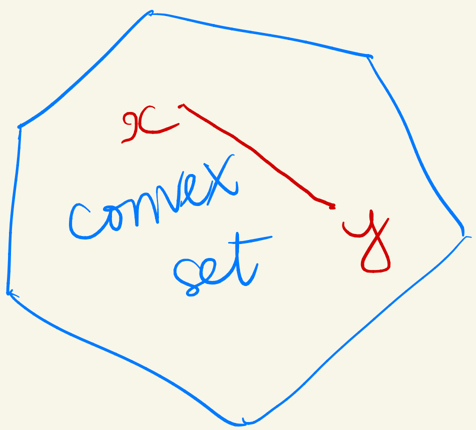
3. $$\min_{z} \underbrace{\frac{1}{2}\|y-Az\|_2^2}_{\text{convex, diff}} + \underbrace{\lambda\|z\|_1}_{\text{convex, nondiff at } z=0} \rightarrow \text{proximal methods}$$

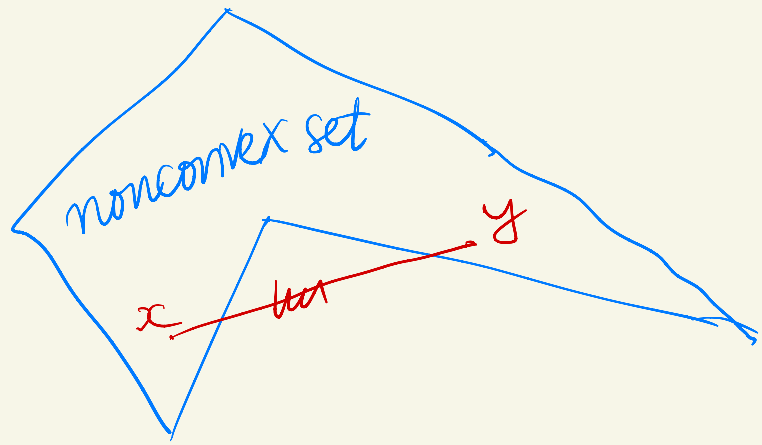
Keywords: Convex, non-differentiable functions, subgradient, constrained / unconstrained optimization prob.

3.1 Recall some definitions from convex optimization

① **Convex set**: $\Omega \subset \mathbb{R}^n$

$$\forall x, y \in \Omega, \quad \forall \theta \in [0,1], \quad \theta x + (1-\theta) y \in \Omega$$





More examples:

Hyperplane: $\{x : a^T x = b\}$

Half-space: $\{x : a^T x \geq b\}$

$\|\cdot\|$ is a norm.

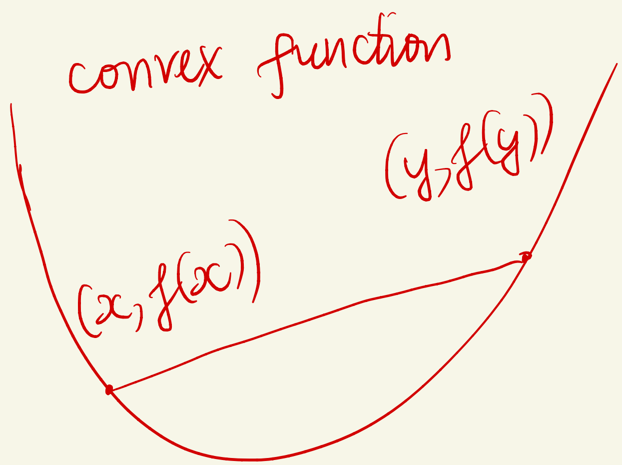
Sphere: $\{x : \|x - x_0\| = b\}$

Ball: $\{x : \|x - x_0\| \leq b\}$

② **Convex function** Let $\Omega \subset \mathbb{R}^n$ be a convex set.

$f : \Omega \to \mathbb{R}$ is convex if $\forall \theta \in [0,1]$, $\forall x, y \in \Omega$, we have

$$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$$



Examples ① affine: $f(x) = a^T x + b$ $a \in \mathbb{R}^n$, $b \in \mathbb{R}$

② $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^{ax}$, $a \in \mathbb{R}$

③ $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^{\alpha}$, $\alpha \geq 1$ or $\alpha \leq 0$

④ $f: \mathbb{R} \to \mathbb{R}$, $f(x) = |x|^p$, $p \geq 1$

⑤ $f: \mathbb{R}_{x>0} \to \mathbb{R}$, $f(x) = x \log x$

⑥ $f: \mathbb{R}^n \longrightarrow \mathbb{R}$, $f(x) = \|x\|_p$, $p \in [1, \infty]$

⑦ $f: \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}$

7.1 $f(X) = tr(A^T X)$

7.2 $f(X) = \|X\|_2 = \sigma_{max}(X)$

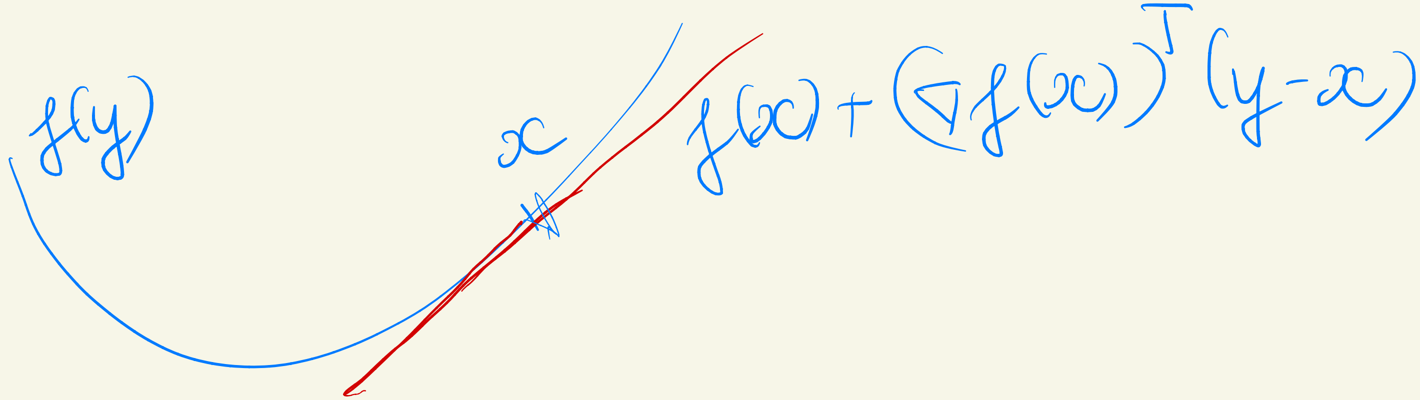
⑧ $f(x) = \frac{1}{2} \|Ax - b\|_2^2$

Important properties:

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_d}\right]^T$$

① If f is differentiable then f is convex iff

$$f(y) \geq f(x) + (\nabla f(x))^T (y-x), \quad \forall x, y \in \operatorname{dom}(f)$$



② If $f \in C^2$ with convex domain then f is convex iff

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in \operatorname{dom} f$$

where $\nabla^2 f(x)_{i,j} := \dfrac{\partial^2 f(x)}{\partial x_i \partial x_j}$, $\quad x = (x_1, \dots, x_n)^T$

③ For a convex function, any minimizer is global and the set of minimizers is convex.

$\Rightarrow$ We can find global minimizer.

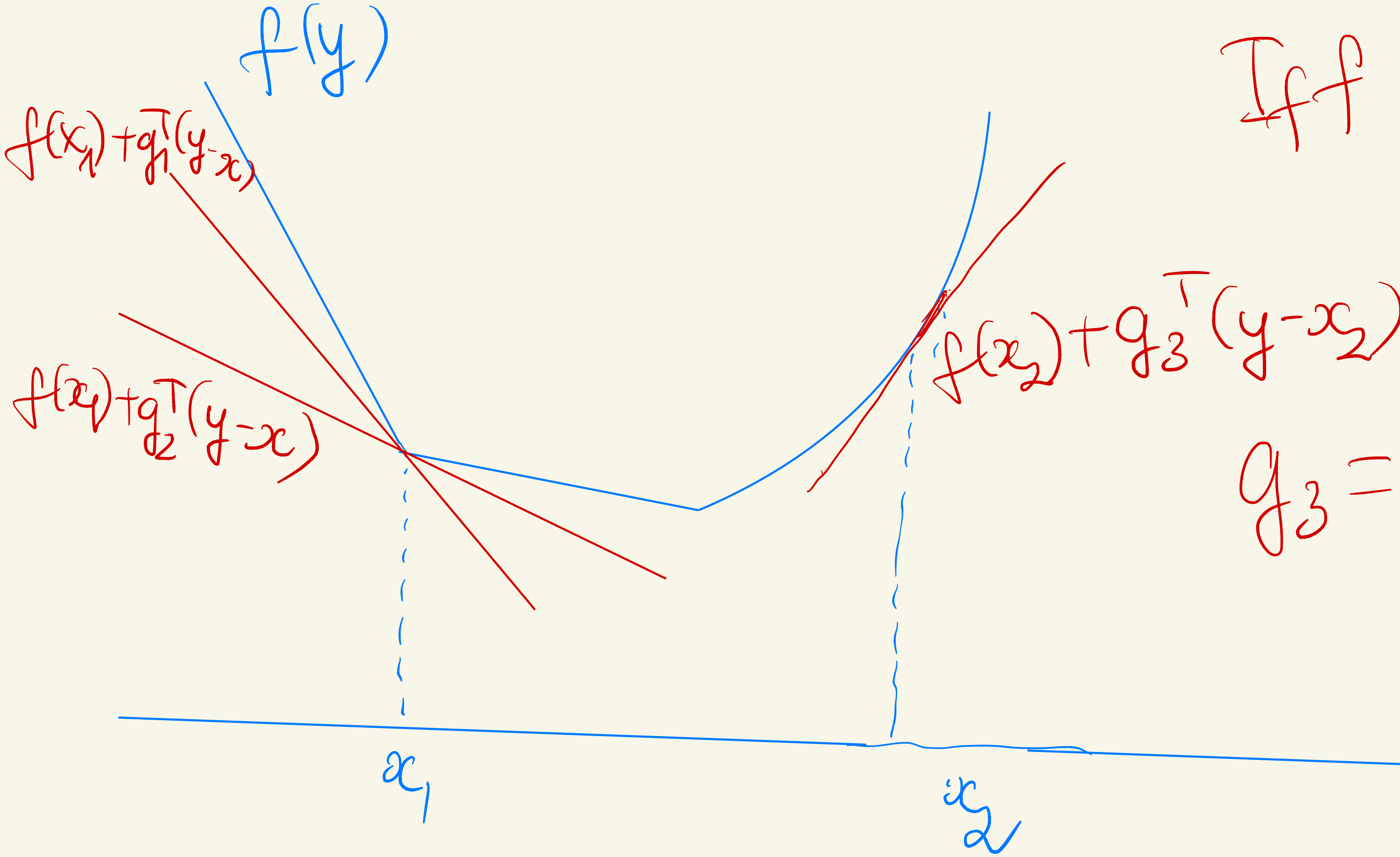
Example $f(x) = \|Ax - b\|_2^2$ is convex since

$$\nabla^2 f(x) = 2A^T A \succeq 0$$

③ Subgradient: Consider $f: \Omega \subset \mathbb{R}^n \to \mathbb{R}$

"$(\nabla f(x))^T$"

$$\underbrace{\partial f(x)}_{\text{subdifferential of } f \text{ at } x} := \{ \underbrace{g}_{\text{a subgradient of } f \text{ at } x} \in \mathbb{R}^n : \ f(y) \geq \underbrace{f(x) + g^T (y-x)}, \ \forall y \in \Omega \}$$



If f is diff at x

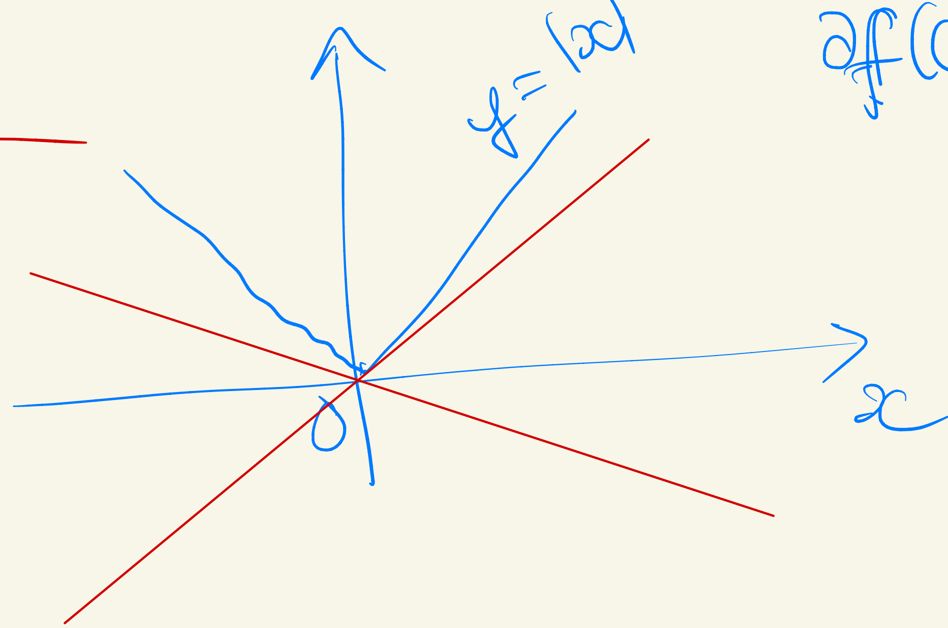
$$\partial f(x) = \{\nabla f(x)\}$$

$$g_3 = \nabla f(x_2)$$

More examples

Example1

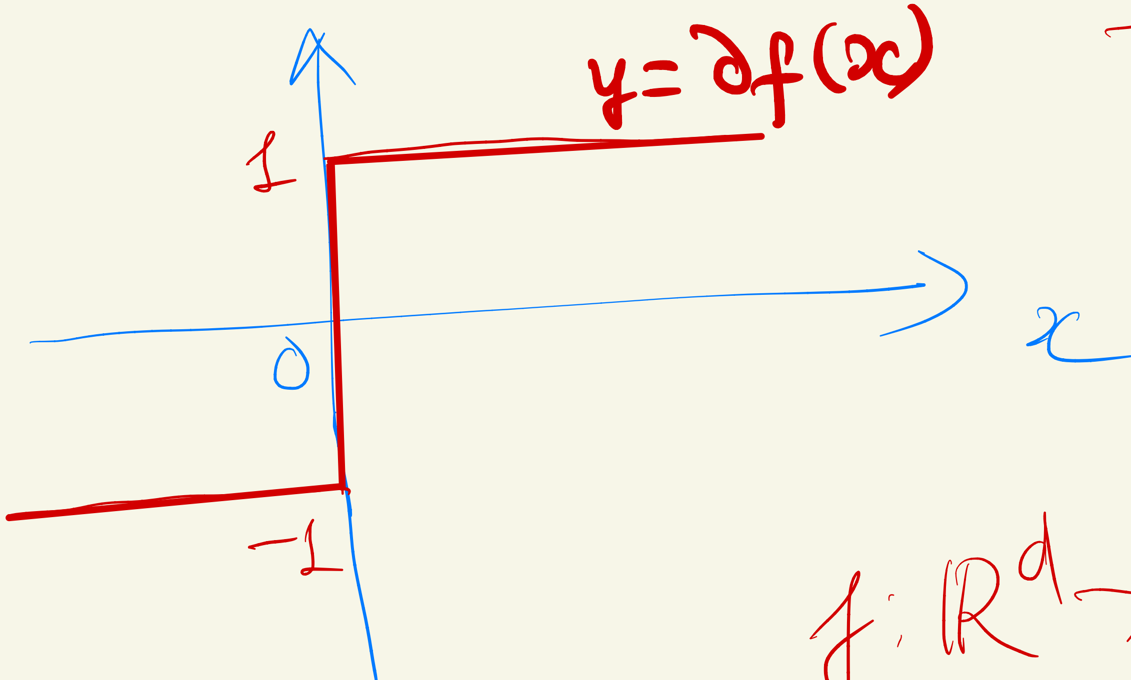
$$f(x) = |x|$$



$$\partial f(0) = [-1, 1]$$

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

$(I + \partial f)^{-1}$ unique



$f = |x|$

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

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

Example 2 $f(x) = \|x\|_2$

$$\partial f(x) = \begin{cases} \left\{ \dfrac{x}{\|x\|_2} \right\} & \text{if } x \neq 0 \\ \{ g : \|g\|_2 \leq 1 \} & \text{if } x = 0. \end{cases}$$

Example 3: $\partial f(x) = \phi$

(i) $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$, $f(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x > 0 \end{cases}$

(ii) $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$, $f(x) = -\sqrt{x}$

Important Properties of subgradients

1. $\partial f(x)$ is a closed, convex set (could be empty)

2. If $x \in \text{int}(\text{dom} f)$, $\partial f(x)$ is nonempty & bounded

3. $\partial f(x)$ is a monotone operator: maximal monotone operator

   $f : \Omega \to \mathbb{R}$

   $x, y \in \Omega$

   $$(u-v)^T (x-y) \geq 0, \quad \forall u \in \partial f(x), \; v \in \partial f(y)$$

4. If $f(x) = \sigma(Ax+b)$, then $\partial f(x) = A^T \partial \sigma(Ax+b)$

⑤ Pointwise maximum: $f(x) = \max\{f_1(x), \dots, f_m(x)\}$

$$\partial f(x) = \text{convex hull} \bigcup_{k:\, f_k(x)=f(x)} \partial f_k(x)$$

$$f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$$

For example $f(x) = \|x\|_1 = \sum_{k=1}^{n} \underbrace{|x_k|_1}$, $x \in \mathbb{R}^n$

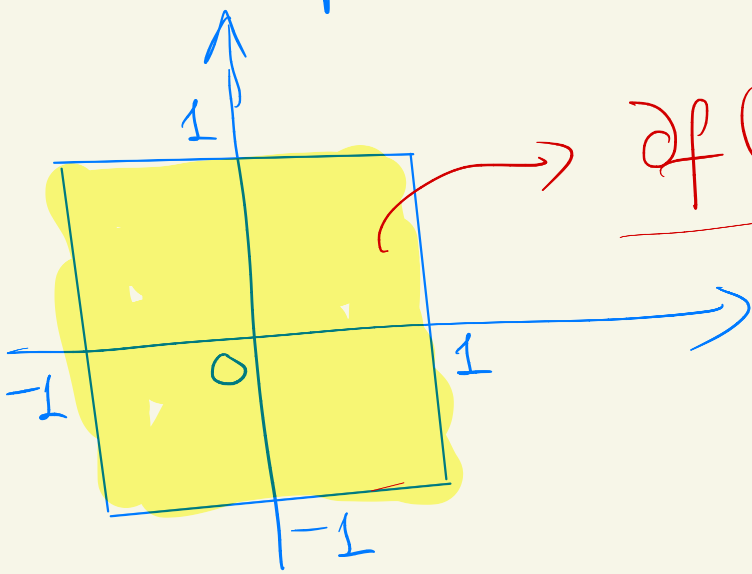
$$= \max_{s \in \{-1,1\}^n} s^T x$$

$$\underline{\partial f(x)} = J_1 \times \underbrace{J_2 \times \dots \times J_n} \quad \text{where } J_k = \begin{cases} [-1,1], & x_k = 0 \\ 1, & x_k > 0 \\ -1, & x_k < 0 \end{cases}$$

For example in $\mathbb{R}^2$

$f(x) = |x_1| + |x_2|$

$x = (x_1, x_2)^T$



$\partial f(0,0)$

④ Proximal operator.

Def Let f be a closed convex function $f: \overset{\text{convex}}{\Omega} \to \mathbb{R}$

$$\operatorname{prox}_f(x;\tau) = \underset{u}{\operatorname{argmin}} \Big(\underbrace{f(u)}_{\text{objective}} + \underbrace{\frac{1}{2\tau}}_{\text{stepsize}} \underbrace{\|u-x\|_2^2}_{\substack{\text{proximal penalty} \\ \text{find } u \text{ close to } x}} \Big) \qquad (*)$$

$\operatorname{prox}_{\tau f}(x)$

Remark ① minimizer exists and unique.

② If u^* is a solution of $(*)$, then

$$0 \in \partial f(u^*) + \frac{1}{\tau}(u^* - x) \iff \frac{1}{\tau}(u^* - x) \in -\partial f(u^*)$$

Example

$f(x) = \|x\|_1$

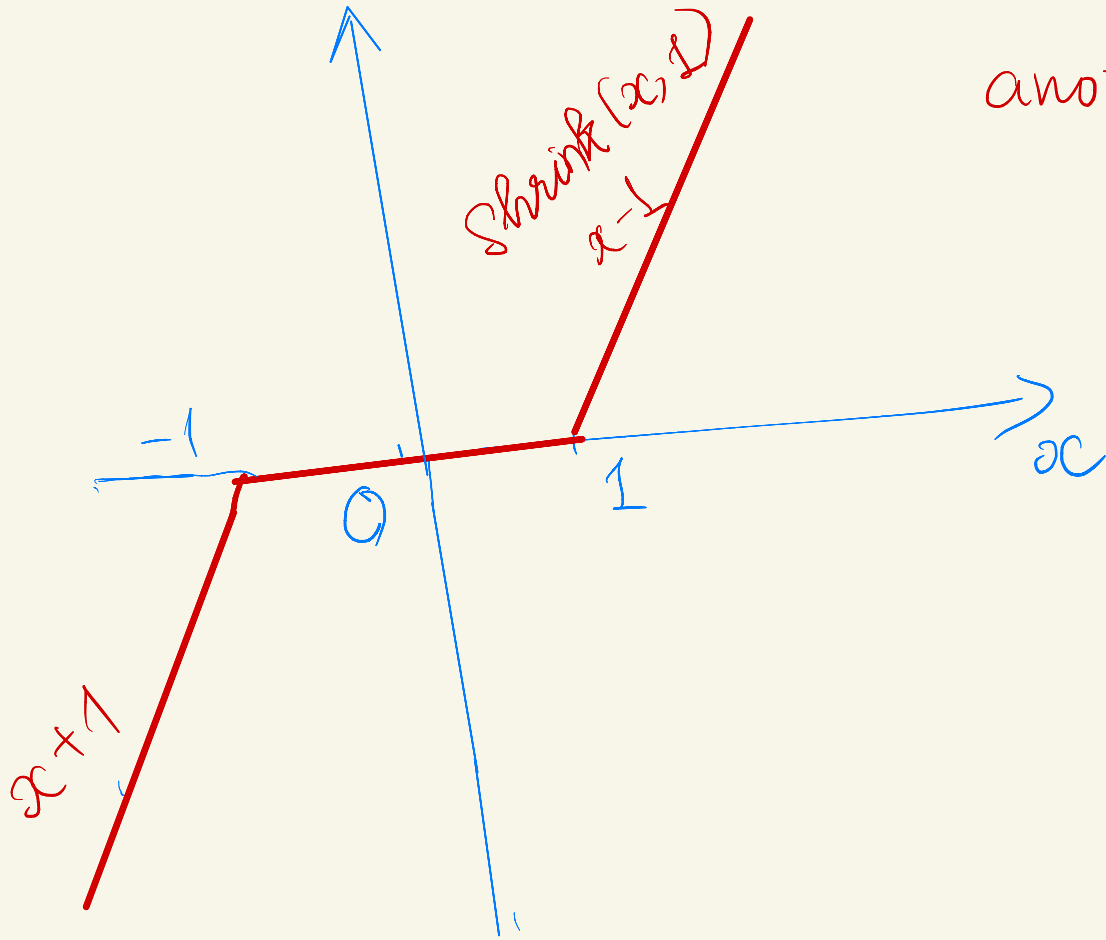
$u^* = \text{prox}_f(x, \tau) = \arg\min \|u\|_1 + \frac{1}{2\tau}\|u - x\|_2^2$

$$\frac{1}{\tau}(u^* - x) \in -\left(\partial f(u^*) = \begin{cases} 1 & \text{if } u^* > 0 \\ -1 & \text{if } u^* < 0 \\ [-1, 1], & \text{if } u^* = 0 \end{cases}\right)$$

Define $\text{Shrink}(x, \tau) := \begin{cases} u^* = x - \tau, & \text{if } x > \tau \\ u^* = x + \tau, & \text{if } x < -\tau \\ 0, & \text{otherwise} \end{cases}$

$u^* > 0$

$\frac{1}{\tau}(u^* - x) = -1$



another name:

soft-thresholding

⑤ Proximal-Gradient Method

$$\min_{x} f(x) = \underset{\uparrow}{g(x)} + h(x)$$

$g(x)$: convex, differentiable

$h(x)$: convex with inexpensive prox-operator

For example, $\min \underbrace{\frac{1}{2}\|y - \overset{m\times n}{A}x\|_2^2}_{g(x)} + \underbrace{\lambda_m \|x\|_1}_{h(x)}$

Proximal gradient algorithm

$$x_{k+1} = \mathrm{prox}_{t_k h}\left(x_k - t_k \nabla g(x_k)\right)$$

t_k: step size

Interpretation ① Denote $x^+ = \mathrm{prox}_{th}(x - t\nabla g(x))$

$$x^+ = \underset{u}{\mathrm{argmin}}\; th(u) + \frac{1}{2}\|u - x + t\nabla g(x)\|_2^2$$

$$= \underset{u}{\mathrm{argmin}}\; h(u) + g(x) + \nabla g(x)^T(u-x) + \frac{1}{2t}\|u-x\|_2^2$$

② $x_{k+1} = \mathrm{prox}_{t_k h}(x_k - t_k \nabla g(x_k))$

Forward step: $\hat{x} := x_k - t_k \nabla g(x_k)$

Backward step: $x_{k+1} = \mathrm{prox}_{t_k h}(\hat{x})$

$$x_{k+1} = x_k - t_k \nabla g(x_k) - t_k \partial h(x_{k+1})$$

Fixed-point property:

$$x_* = x_* - t_k \nabla g(x_*) - t_k \partial h(x_*)$$

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

Example $\min \frac{1}{2}\|Ax-b\|_2^2 + \lambda \|x\|_1$.

① $$\boxed{\begin{aligned} \hat{x} &= x_k - \tau A^T(Ax_k - b) \\ x_{k+1} &= \operatorname{prox}_{\tau\|\cdot\|_1}(\hat{x}) = \operatorname{shrink}(\hat{x}, \lambda\tau) \end{aligned}}$$

② FISTA :

$$x_{k+1} = \operatorname{prox}_h\left(y_k - \tau\nabla g(y_k), \tau\right)$$

$$\alpha_{k+1} = \frac{1}{2}\left(1 + \sqrt{4\alpha_k^2 + 1}\right)$$

$$y_{k+1} = x_{k+1} + \frac{\alpha_k - 1}{\alpha_{k+1}}\left(x_{k+1} - x_k\right)$$