بسم الله الرحمن الرحیم

* زیرگرادیان

* محاسبات روی زیرگرادیان

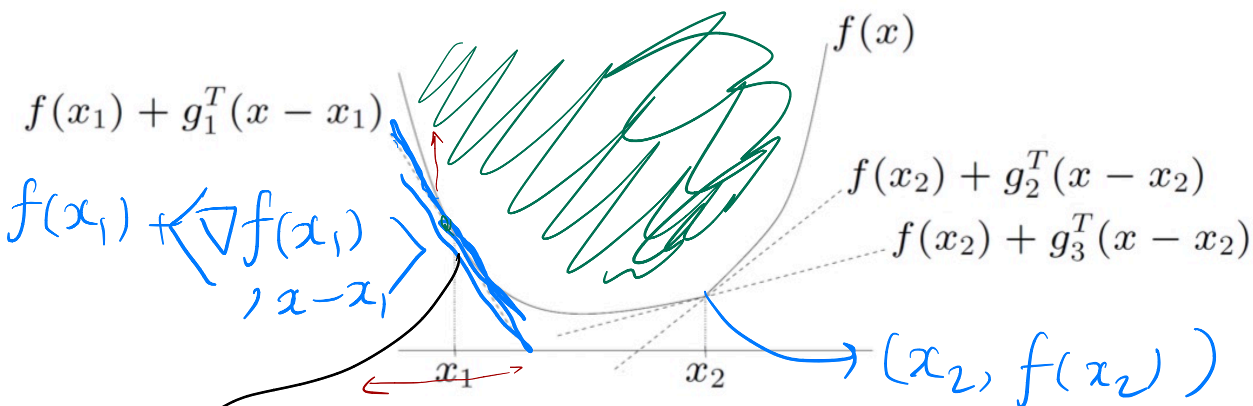
* مثال: مسئله Lasso

$$\left\langle (y, x) - (f(x_1), x_1), \begin{bmatrix} 1 \\ \nabla f(x_1) \end{bmatrix} \right\rangle = 0$$

معادله ابرصفحه (بردار عمود : $\nabla f(x_1)$)

$\rightarrow y =$

$f(x_1) + \langle \nabla f(x_1), x - x_1 \rangle$



$(x_2, f(x_2))$

$(x_1, f(x_1))$

Figure: Subgradients

تابع مشتق پذیر : گرادیان ، ابرصفحه مماس با گرادیان

نقاط مشتق ناپذیری مانند x_2 :

$\{ g : f(x_2) + \langle g, x - x_2 \rangle$ ابرصفحه‌های زیر تابع $\}$

g زیرگرادیان ~~زیرگر~~ در نقطه (x) :

$\forall y \in dom(f)$

$$f(y) - f(x) \geqslant \langle g , y - x \rangle$$

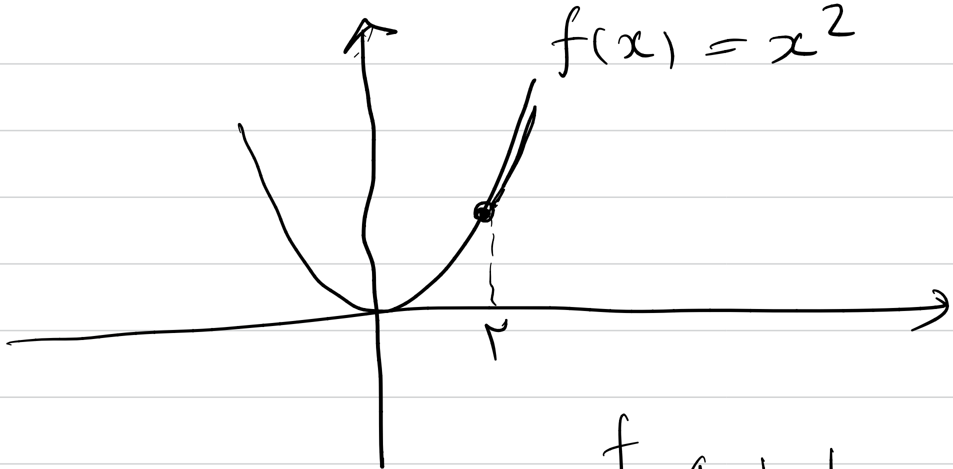
زیردیفرانسیل (زیرمشتق) مجموعه ~~ک~~ همه زیرگرادیان‌های ~~زیرگر~~ در نقطه x .

$$sub\text{-}g^{(f)}(x) = \{ g : \forall y \text{ رابطه برقرار باشد} \}$$

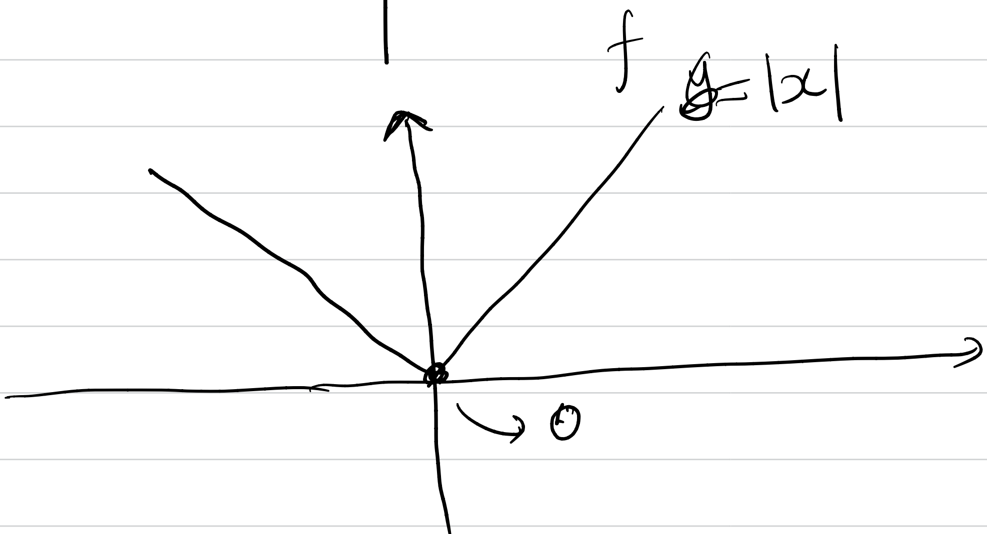
$$= \partial f(x)$$

• $sub\text{-}g(x)$ چه جور مجموعه‌ای ؟ (مجموعه محدب)

مثال ها:



$$(\text{sub-}g)\, f(2) = \{۴\}$$



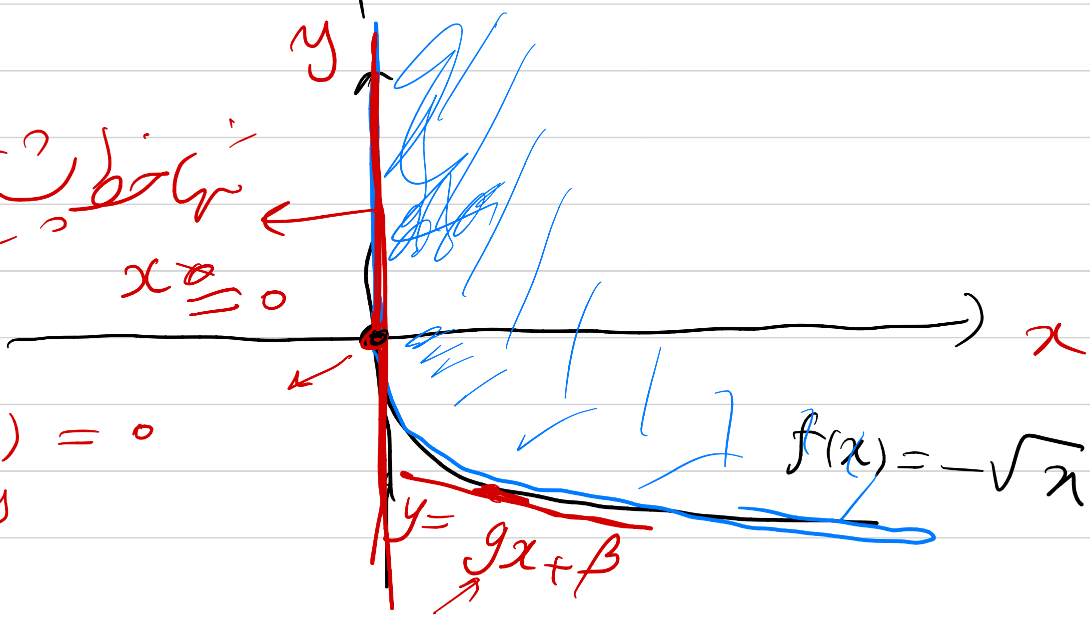
$$\text{sub } f(0) = [-1, 1]$$

$$f(x) - 0 \geq \cancel{g}\; g\, x$$

تنها خط پشتیبان با راستای

$x = 0$

$(g, x) = 0$
y



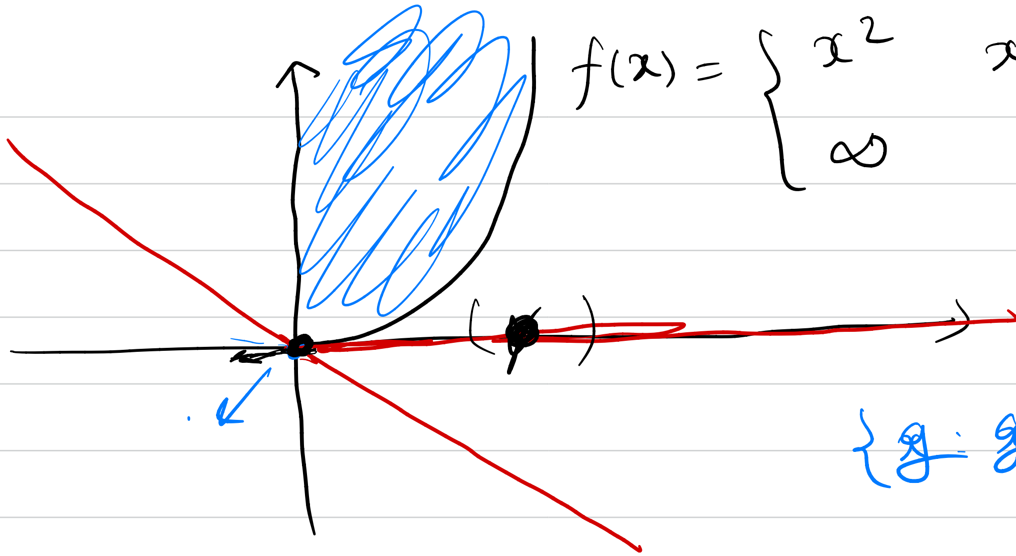
زیرگرادیان تهی است.

$$\begin{bmatrix} g=-1 \\ 0=\beta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \ngeq 0$$

$$-\sqrt{y} + \sqrt{0} \geq \langle g, y - 0 \rangle = g\,y \quad ; \; y > 0$$

$$-\sqrt{y} \geq g\,y \quad ; \; y > 0$$

$$-\frac{1}{\sqrt{y}} \geq g$$



$$f(x) = \begin{cases} x^2 & x \geq 0 \\ \infty & x < 0 \end{cases}$$

زیرگرادیان

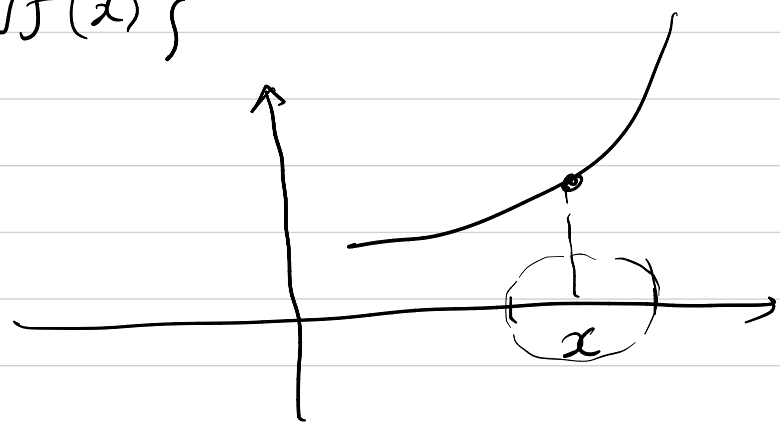
$\{g : g \leq 0\}$ نامثبت همه اعداد

زیرگرادیان توابع مشتق‌پذیر

• اگر $x \in \text{int}(\text{dom}(f))$ ، f مشتق‌پذیر در x

$$\partial f(x) = \text{sub-}g(x) = \{\nabla f(x)\}$$

مجموعه {زیرگرادیان



$g \in \partial f(x)$

$$f(x+\varepsilon \vec{v}) - f(x) \geq \varepsilon \langle g, v \rangle \quad \forall \vec{v}, \varepsilon > 0$$

$$\lim_{\varepsilon \downarrow 0} \frac{f(x+\varepsilon v) - f(x)}{\varepsilon} \geq \langle g, v \rangle$$

$$\langle \nabla f(x), v \rangle \geq \langle g, v \rangle \quad \forall v$$

$$\langle \nabla f - g, v \rangle \geq 0$$

$$\Longrightarrow g = \nabla f$$

خواص زیرگرادیان برای توابع محدب

خاصیت ① : $\partial f(x)$ مجموعه محدب است.

خاصیت ② $x \in int(dom(f))$ ، همیشه مجموعه $\partial f(x)$ ناتهی، محدود و بسته است.

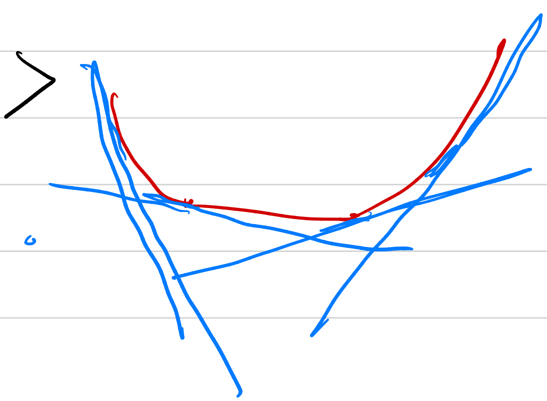
خاصیت ③ : اگر تابع f با دامنه محدب در هر نقطه‌ای شامل زیرگرادیان (مشتق) ناتهی باشد ، f محدب است.

x_0 : $g \in \partial f(x_0)$

x_0 به صورت ترکیب محدب $x_0 = \lambda x + \bar{\lambda} y$.

$$\lambda f(x) + \bar{\lambda} f(y) \geq f(\underbrace{\lambda x + \bar{\lambda} y}_{= x_0})$$

$$\begin{cases} \lambda \times & f(x) \geq f(x_0) + \langle g , x - x_0 \rangle \\ \bar{\lambda} \times & f(y) \geq f(x_0) + \langle g , y - x_0 \rangle \end{cases}$$



• مجموعه زیرگرادیان ∂f ، درصورت تهی نبودن بسته است.

$$\{g_i\} \longrightarrow g$$

$$g_i \in \partial f \longrightarrow g \in \partial f$$

خاصیت (۴): زیرگرادیان ها یکنوا هستند

$$\begin{array}{cc} x_0 \;\;,\;\; x_1 \\ \swarrow \qquad \searrow \\ g \in \partial f(x_0) \qquad h \in \partial f(x_1) \end{array}$$

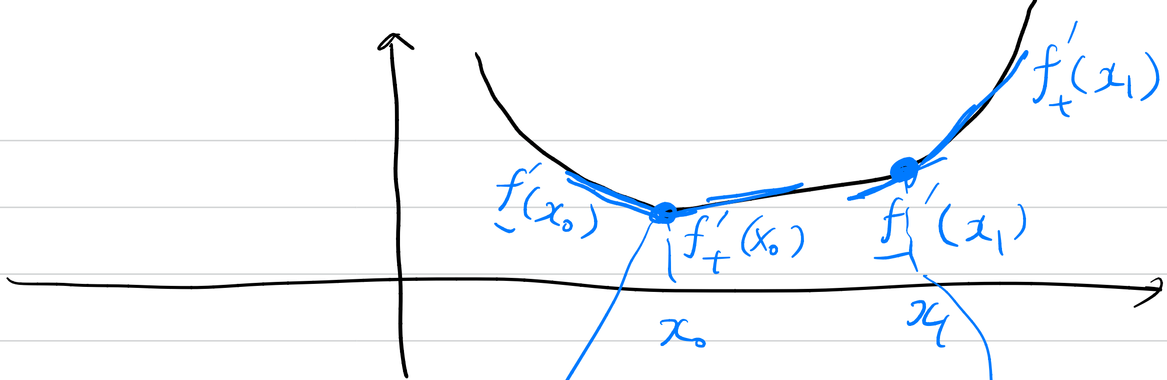
$$\langle g - h \,,\, x_0 - x_1 \rangle \geq 0$$

یادآوری: $\langle \nabla f(x_0) - \nabla f(x_1) \,,\, x_0 - x_1 \rangle \geq 0$

اثبات:

$$\begin{cases} f(x_1) \geq f(x_0) + \langle g \,,\, x_1 - x_0 \rangle \\ f(x_0) \geq f(x_1) + \underbrace{\langle h \,,\, x_0 - x_1 \rangle}_{} \end{cases}$$

$\langle -h \,,\, x_1 - x_0 \rangle$



$$\partial f(x_0) = [\underbrace{f'_-, f'_+}_{x_0}] \quad \partial f(x_1) = [f'_-(x_1), f'_+(x_1)]$$

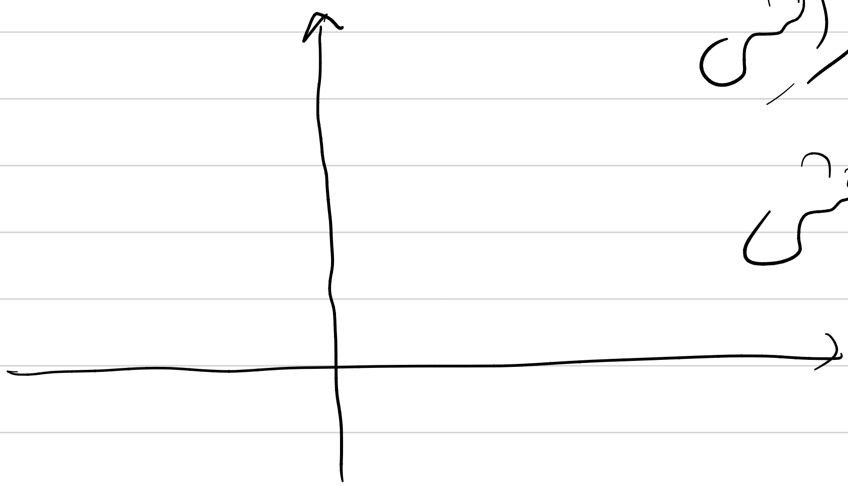
$$f'_+(x_0) \le f'_-(x_1)$$

یادآوری :

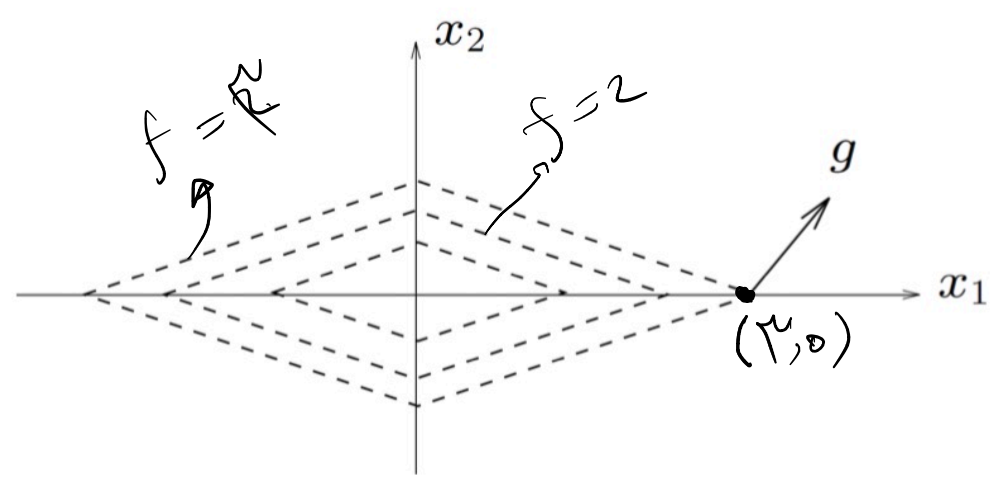
حرکت

جهت گرادیان افزایش

درخلاف جهت گرادیان کاهش



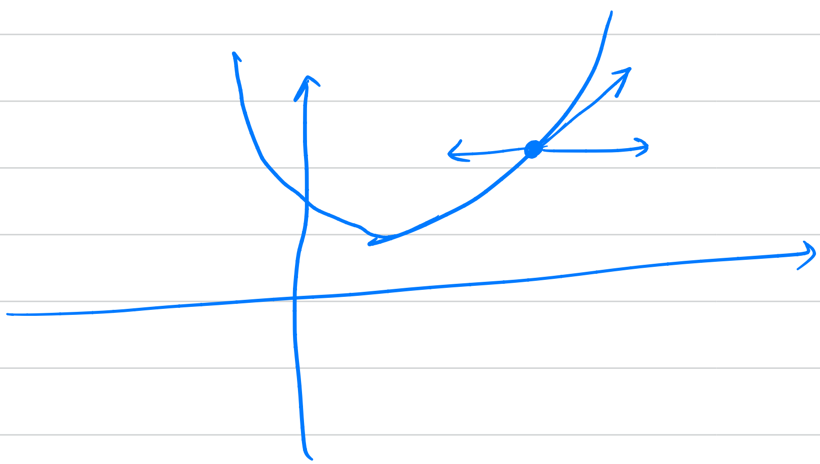
* منفی زیرگرادیان همیشه جهت کاهش نیست!

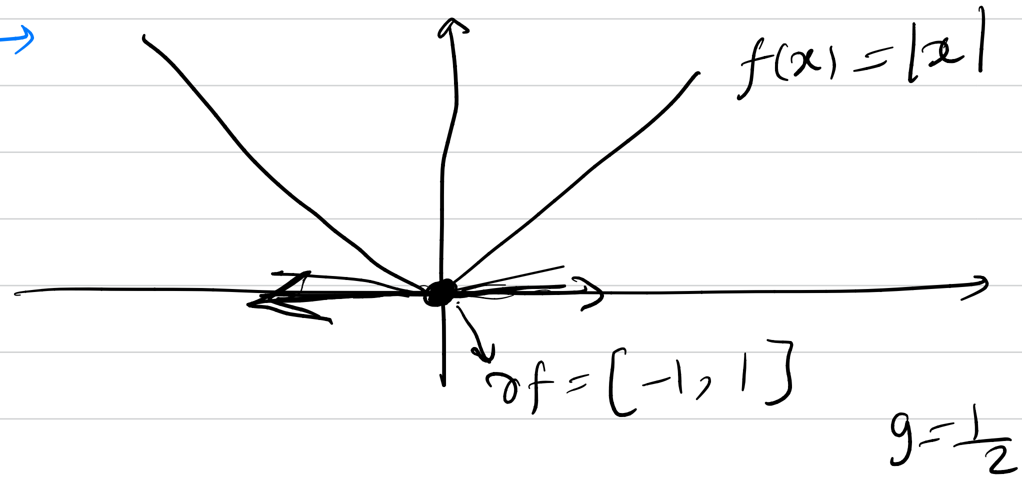


$\to \begin{pmatrix} 1 \\ [-2,2] \end{pmatrix}$

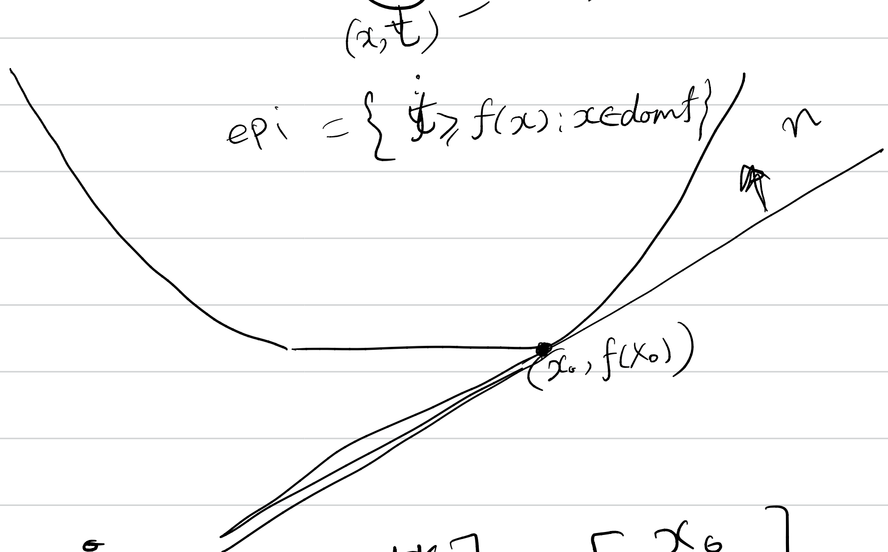
Figure: Contours of function $f(x_1, x_2) = |x_1| + 2|x_2|$

$-g$ جهت کاهش نیست!





ارتباط اپیگراف و زیرگرادیان



معادله خط : $\left\langle \vec{n} , \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_0 \\ f(x_0) \end{bmatrix} \right\rangle = 0$

$(x,t) \in epi$: $\left\langle \vec{n} , \begin{bmatrix} x \\ t \end{bmatrix} - \begin{bmatrix} x_0 \\ f(x_0) \end{bmatrix} \right\rangle \geq 0$

$\vec{n} = \begin{bmatrix} -\alpha \\ \cancel{\beta}\ 1 \end{bmatrix}$; $\alpha \in \mathbb{R}^n$; $\beta \in \mathbb{R}$ $\qquad \forall (x,t)$

$\longrightarrow \left\langle \begin{bmatrix} -\alpha \\ \cancel{\beta}\ 1 \end{bmatrix} , \begin{bmatrix} x - x_0 \\ t - f(x_0) \end{bmatrix} \right\rangle \geq 0$

✓ $\beta > 0$ اگر ; $\beta \geq 0 \longleftarrow t \uparrow \infty$

$$\langle -\alpha , x - x_0 \rangle + t - f(x_0) \geq 0 \quad \forall (x,t)$$

$$t \geq f(x_0) + \langle \alpha , x - x_0 \rangle$$

نکته : اگر $x \in \text{int}(\text{dom}(f))$: آنگاه وقتی بردار نرمال $\begin{bmatrix} -\alpha \\ \beta \end{bmatrix} = \vec{n}$ از ابرصفحه

باشد ؛ $\beta > 0$. در نتیجه $-\dfrac{\alpha}{\beta}$ یک عنصر از زیرگرادیان $\longleftarrow$

زیرگرادیان ناتهی

; $x = x_0 + \epsilon v$ ؛ برای ϵ کوچک ؛ $x \in \text{dom}(f)$

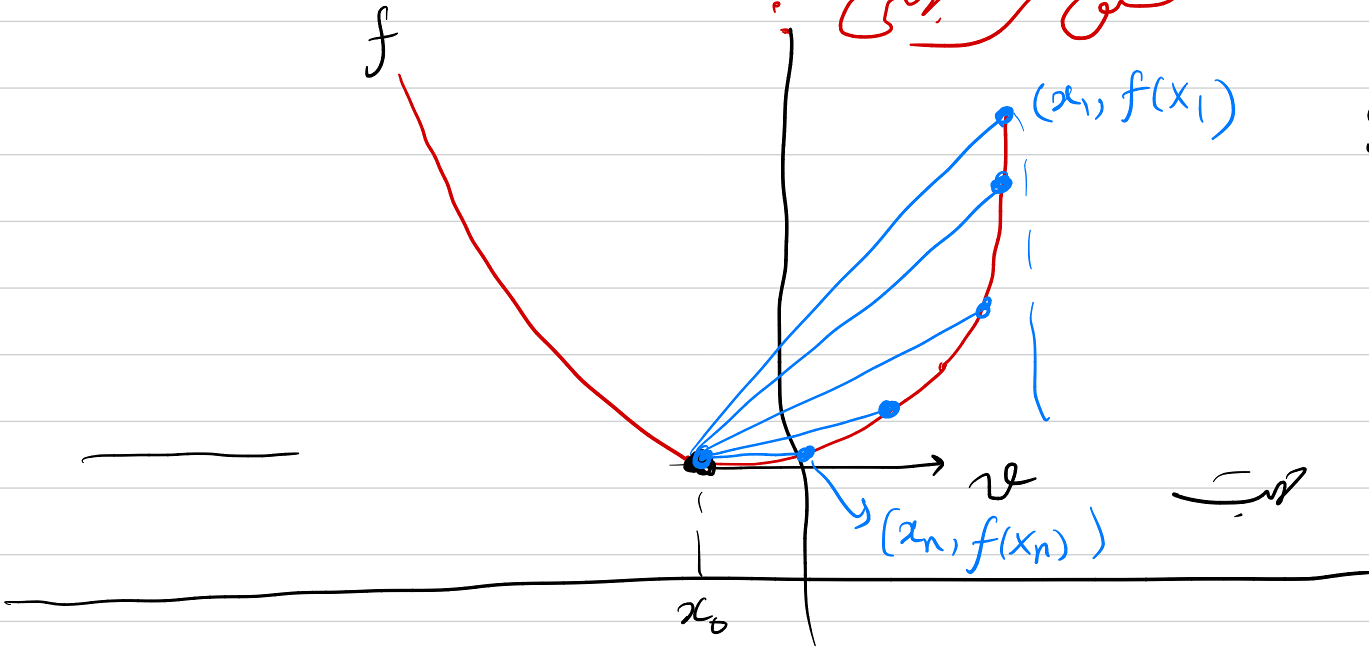
هر v

فرض خلف

$\beta = 0$

$$\longrightarrow \langle -\alpha , x - x_0 \rangle \geq 0 \longrightarrow \alpha = 0 \quad \times$$

مشتق جهتی :



$$g(t) = f(x_0 + tv)$$

$$f'(x_0 ; v) = \lim_{\varepsilon \downarrow 0} \frac{f(x_0 + \varepsilon v) - f(x_0)}{\varepsilon}$$

$$= g'(0^+)$$

$$= \inf_{\varepsilon > 0} \frac{f(x_0 + \varepsilon v) - f(x_0)}{\varepsilon}$$

* زیرگرادیان و مشتق جهتی :

$h \in \partial f(x_0)$

$$f'(x_0; v) = \lim_{\varepsilon \downarrow 0} \frac{f(x_0 + \varepsilon v) - f(x_0)}{\varepsilon} \geq \langle h, \varepsilon v \rangle$$

$$\geq \langle h, v \rangle =$$

$$\longrightarrow \quad f'(x_0; v) \geq \sup_{h \in \partial f(x_0)} \langle h, v \rangle$$

* نکته : $\quad f'(x_0; v) = \sup_{h \in \partial f(x_0)} \langle h, v \rangle$

محاسبات زیرگرادیان:

$$\partial f_1(x) + \partial f_2(x) = \partial(f_1+f_2)(x)$$ ① *

$$A \oplus B = \left\{ a+b : \begin{array}{l} a\in A \\ b\in B \end{array} \right\}$$ جمع مجموعه‌ای +

$\subseteq$ ✓

مثال: $f_1(x) = x$; $f_2(x) = |x|$

$$f(x) = x^2 + |x|$$

$$\partial f(0) = \{1\} \oplus [-1, 1] = [0, 2]$$

اثبات کمی سرراست‌تر:

$$f_1(y) - f_1(x) \geqslant \langle a, y - x \rangle$$

$$f_2(y) - f_2(x) \geqslant \langle b, y - x \rangle$$

$\mathcal{C}$ محدب

$$\min_{x\in\mathcal{C}} f(x) = \min_{x} f(x) + I_{\mathcal{C}}(x)$$

$x\in\mathcal{C}$; زیرگرادیان

مثال: $I_{\mathcal{C}}(x) = \begin{cases} 0 & x\in\mathcal{C} \\ \infty & \text{o.w.} \end{cases}$; $g(x) = f(x) + I_{\mathcal{C}}(x)$

فرض کنید $x\in\mathcal{C}$; $\partial g(x) = \partial f(x) + \partial I_{\mathcal{C}}(x)$

محاسبه $\partial I_{\mathcal{C}}(x)$:

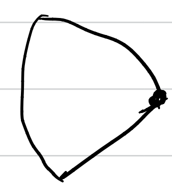
$$I_C(y) - I_C(x) \geq \langle g, y-x\rangle$$

$$g \in \partial I_{\mathcal{C}}(x) \iff \forall y\in\mathcal{C} : \langle g, y\rangle \leq \langle g, x\rangle$$

(یادآوری تعریف مخروط نرمال)

$$\Longrightarrow \partial I_{\mathcal{C}}(x) = N_{\mathcal{C}}(x)$$

$$\boxed{\partial g(x) = \partial f(x) + N_{\mathcal{C}}(x)}$$

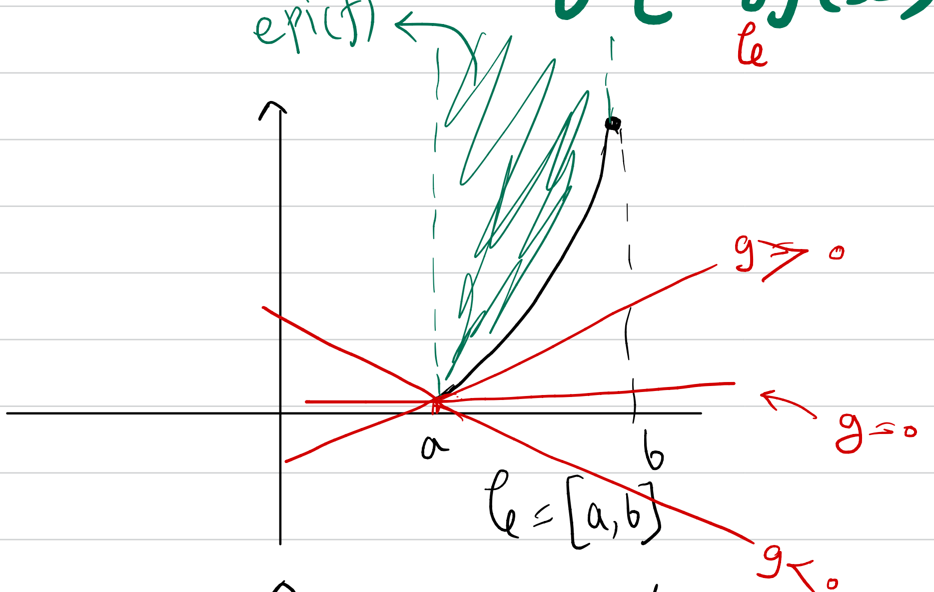


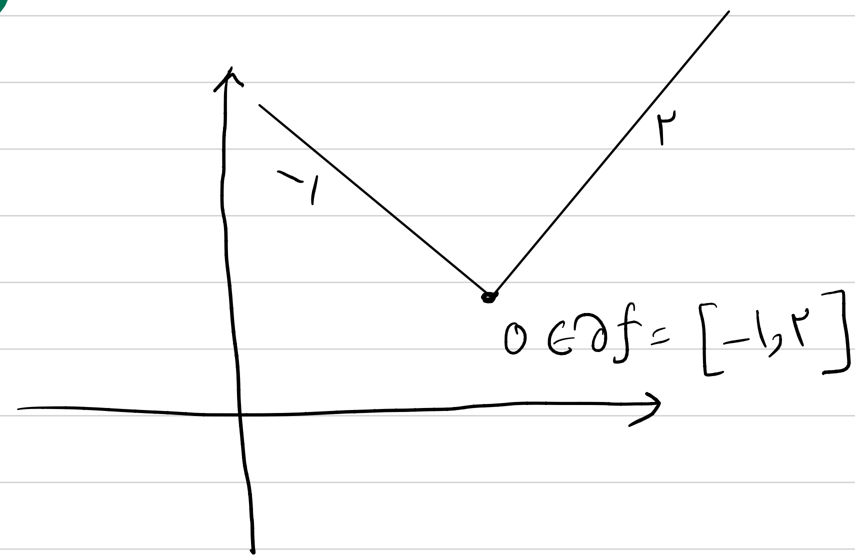
شرط بهینگی.

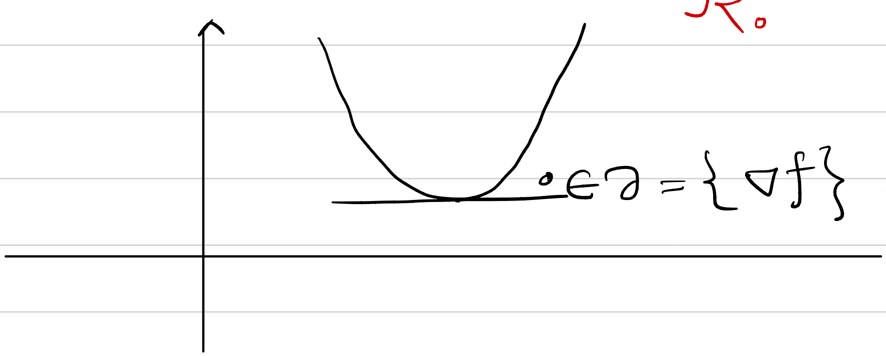
$$\min \; f(x) \qquad ; \operatorname{dom}(f) = C$$

✓ x^* مقدار بهینه است اگر و تنها اگر

$$0 \in \partial f(x^*)$$







$$0 \in \partial = \{\nabla f\} \longrightarrow \nabla f = 0$$

x^* مینیمم کننده باشد

$$f(x) - f(x^*) \geq \langle g, x - x^* \rangle \qquad \forall x$$

$$g = 0 \in \partial f(x)$$

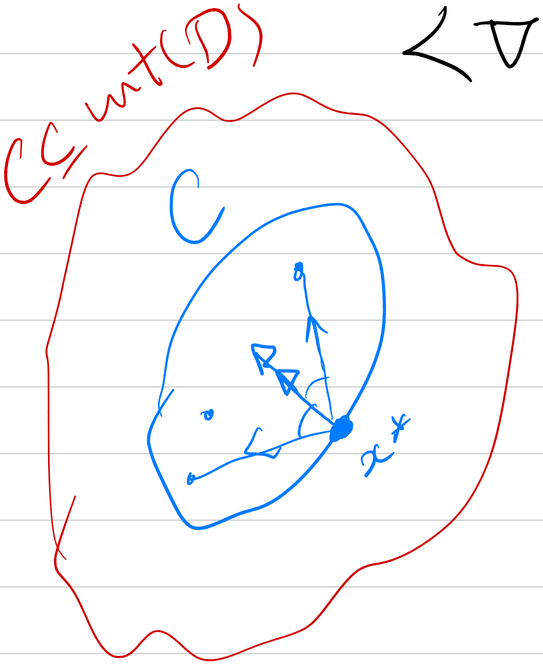
$g = 0$ زیرگرادیان در x^* $\longrightarrow$ $f(x) - f(x^*) \geq 0$

$\longrightarrow$ x^* یک کمینه کننده است.

$$\min_{x\in C} \quad f(x)$$

• درمورد تابع مشتق پذیر :

$$\langle \nabla f(x^*) , y-x^* \rangle \geq 0 \quad \forall y \in C \iff -\nabla f(x^*) \in N_C(x^*)$$



• بررسی با استفاده از زیردیفرانسیل

$$\min_{x} \quad f(x) + I_C(x)$$

$$0 \in \partial \big(f(x^*) + I_C(x^*) \big) \Longrightarrow 0 \in \partial(\cdots) = \underbrace{\partial f(x^*)}_{\nabla f(x^*)} + \underbrace{\partial I_C(x^*)}_{N_C(x^*)}$$

$$-\nabla f(x^*) \in N_C(x^*)$$

$$N_C(x^*) = \{ g : \langle g, y \rangle \leq \langle g, x \rangle \quad \forall y \in C \}$$

$$\langle g, y-x \rangle \leq 0$$

$$A \oplus B = \left\{ a+b : \begin{array}{l} a\in A \\ b \in B \end{array} \right\}$$

محاسبات زیرگرادیان :

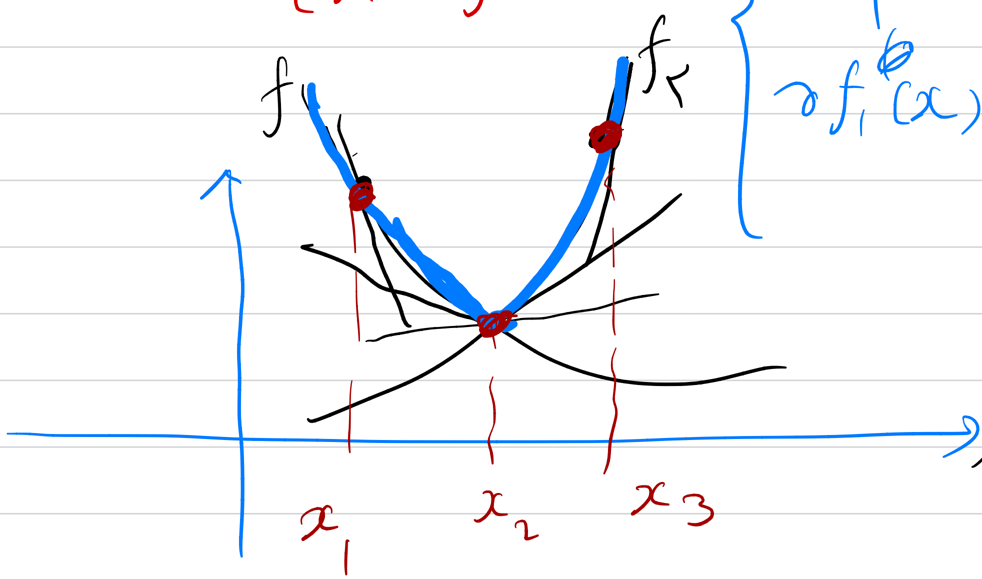
$$\partial \max_i \{f_1,\dots,f_n\} = \text{Conv}\left\{\bigcup \partial f_j(x) : j \in \arg\max_i f_i(x)\right\} \quad (۲) \quad *$$

$$\partial \sup_y f(x,y) = \underbrace{cl}_{\text{بستار}}\left(\text{Conv}\left\{\bigcup \partial f(x,\bar{y}) : f(x,\bar{y}) = \sup_y f(x,y)\right\}\right) \quad (۳) \quad *$$

(۲) : $\supseteq$ (آسان!)

$\subseteq$ ؟؟

$$\partial \max\{f_1,f_2\}(x) = \begin{cases} \text{Conv}\{\partial f_1(x) \cup \partial f_2(x)\} & f_1(x) = f_2(x) \\ \partial f_1(x) & f_1(x) > f_2(x) \end{cases}$$



مثال : $f(x) = |x| = \max(x, -x)$

در نقطه صفر: هر دو تابع $x, -x$ فعال‌اند

$$\partial |x|(0) = \text{Conv}\{1, -1\}$$

$y:$

$f_1(x) \geq f_2(x)$

$$\underbrace{\max\{f_1(y), f_2(y)\} - f_1(\underset{1}{x})}_{\substack{\vee \\ f_1(y) - f_1(x)}} \overset{?}{\geq} \langle g_1, y - \underset{1}{x} \rangle$$

$g_1 \in \partial f_1$

$\bigcup \{ g_i : g_i \in \partial f_i(x)$; $f_i(x)$ برابر با ماکزیمم باشد $\}$ همه زیرگرادیان‌ها است

محاسبات زیرگرادیان :

(۴) *

$$h(x) = f(Ax+b)$$

$$\partial h(x) = A^T \partial f(Ax+b)$$

$$h(y) - h(x) = f(\overbrace{Ay+b}^{z}) - f(\overbrace{Ax+b}^{u})$$

$$g \in \partial f(Ax+b) = \partial f(u) \qquad \geq \langle g , \underbrace{z-u}_{A(y-x)} \rangle$$

$$= \langle A^T g , y-x \rangle$$

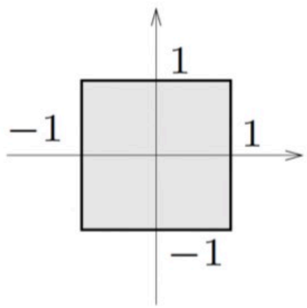
$$\Downarrow$$

$$A^T g \in \partial h(x)$$

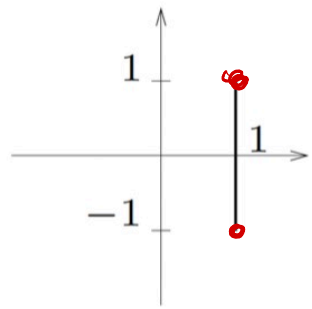
مثال:

$$f(x)=\|x\|_1 = \max_{s\in\{-1,1\}^n} \underbrace{\langle s, x\rangle}_{g_s}$$

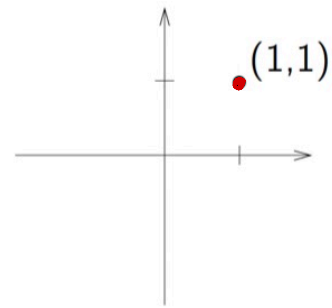
$$= \sum |x_i|$$



$\partial f(x)$ at $x=(0,0)$



at $x=(1,0)$



at $x=(1,1)$

$s=(1,1),(1,-1)$ $\qquad$ $s=(1,1)$

$x_1+x_2 \leftarrow \partial g_{(1,1)} \cup \partial g_{(1,-1)} \rightarrow x_1-x_2$ $\qquad$ $\partial g_{(1,1)}$ $\quad (1,1)$

$\mathrm{Conv}\{\{(1,1)\}\cup\{(1,-1)\}\}$ $\qquad$ x_1+x_2

$$g_{1,-1}(x_1,x_2) = x_1-x_2$$

$$v\in\partial\|x\|_1 = \begin{cases} v_i=1 & x_i>0 \\ v_i\in[-1,1] & x_i=0 \\ v_i=-1 & x_i<0 \end{cases}$$

مثال: زیرگرادیان نرم دوگان

$\| \ \|$ یک نرم داده شده در فضا.

$$\|a\|_* = \sup_{b:\|b\|=1} \langle a, b \rangle$$

به عنوان نمونه: دوگان نرم اقلیدسی، خودش است. دوگان نرم

$$\|a\|_p = \left(\sum |a_i|^p\right)^{1/p}$$

نرم $\|a\|_q$ است که $\frac{1}{p}+\frac{1}{q}=1$. (دوگان نرم l_1، l_∞ است ــ)

$$f(x) = \|x\|_* = \sup_{g:\|g\|=1} \underbrace{\langle x, g \rangle}_{f_g(x)}$$

$$\longrightarrow \partial f(x) = \mathrm{Conv}\{ g : \langle g, x \rangle = \|x\|_*, \|g\|=1 \}$$

حالت خاص:

$$\partial \|x\|_\infty = \mathrm{Conv}\{ g : \|g\|_1 = 1, \langle g, x \rangle = \|x\|_\infty \}$$

$$= \mathrm{Conv}\{ g : x_i \neq \|x_\infty\| \to g_i = 0 \ ; \ \underset{x_i=\|x\|_\infty}{g_i x_i > 0}, \ \sum |g_i| = 1 \}$$

(۱، ۴، ۳، ۴)

۰ ۰

بسم الله

* شکل lasso

* حل تحلیلی شکل بهینه سازی محدب

(روش KKT)

مسئله Lasso انگیزه

$y \in \mathbb{R}^{100}$; ارسال شده : ۵ عنصر ناصفر

نزدیکترین x ممکن به y پیدا کنید که حداکثر ۵ عنصر ناصفر داشته باشد.

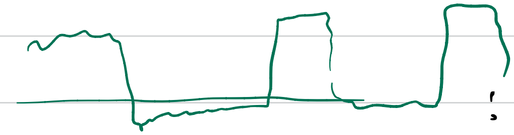
$$\min_{x} \quad \|x-y\|^2$$

$$\text{s.t.} \ \|x\|_0 = \sum 1\{x_i \neq 0\} \leq 5$$

غیر محدب است

Lasso مسئله

حالت خاص:



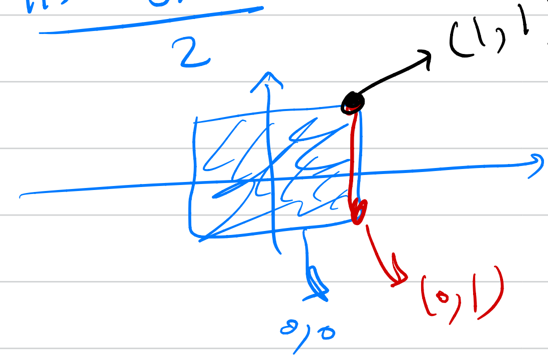
مثال:

$$\min_x \underbrace{\frac{1}{2}\|x-y\|_2^2 + \lambda\|x\|_1}_{f}$$

$$x^* = \arg\min \iff 0 \in \partial f(x^*) = x^* - y + \lambda\,\partial\|x^*\|_1$$

$$\frac{\nabla\|x^*-y\|^2}{2}$$

$$\partial\|x^*\|_1 = \left\{ (g_1,\dots,g_n) : \begin{array}{ll} g_i = 1 & x_i^* > 0 \\ [-1,1] & x_i^* = 0 \\ g_i = -1 & x_i^* < 0 \end{array} \right\}$$



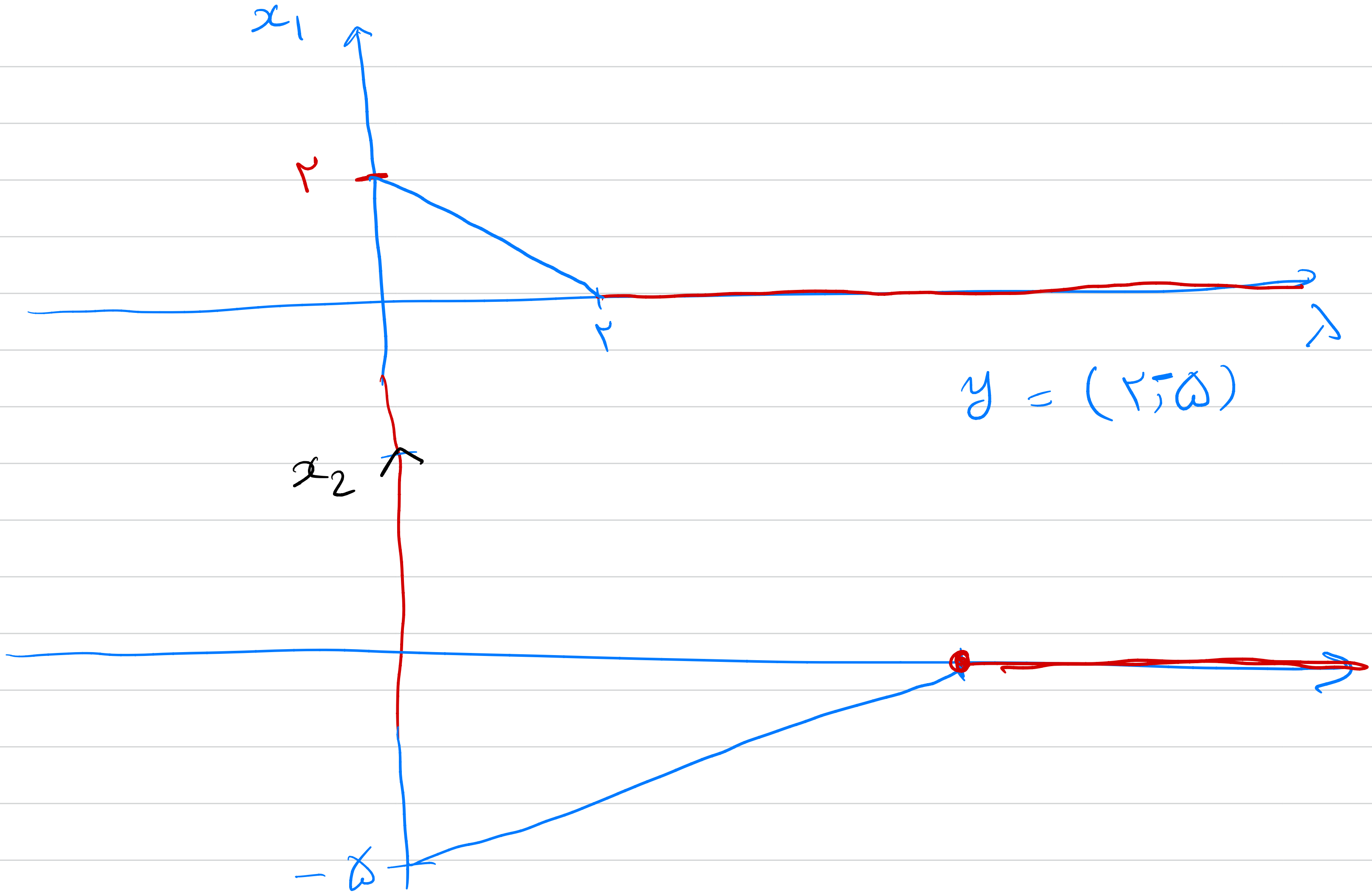
$$\to \exists (g_1,\dots,g_n) : \quad x_i^* + \lambda g_i = y_i$$

$$x_i^* = \begin{cases} y_i - \lambda & g_i = 1 \qquad y_i > \lambda \\ 0 & g_i \overset{?}{=} \frac{y_i}{\lambda} \;\checkmark \qquad |y_i| \le \lambda \\ y_i + \lambda & g_i = -1 \qquad y_i < -\lambda \end{cases}$$

$\partial\|x^*\|$

$$= \sum |x_i|$$

$$= \begin{cases} \mathrm{sign}(x_i)\cdot x_i & x_i \neq 0 \\ ---- & x_i = 0 \end{cases}$$



x_1
۲
۲
λ
$y = (۲,-۵)$
x_2
-۵

Lasso مسئله

مثال:

$$\min_x \underbrace{\frac{1}{2}\|Ax-y\|_2^2 + \lambda\|x\|_1}_{f}$$

$$x^* = \arg\min \iff 0 \in \partial f(x^*) = A^T(Ax^*-y) + \lambda\partial\|x^*\|_1$$

$$\implies A^T(y-Ax^*) = \lambda v$$

برای ; $v \in \partial\|x^*\|$

$$v_i = \begin{cases} 1 & x_i^* > 0 \\ \in[-1,1] & x_i^* = 0 \\ -1 & x_i^* < 0 \end{cases}$$

$$A = \left[A_1 \,|\, \cdots \,|\, A_n \right] \quad ; \quad A^T = \begin{pmatrix} A_1^T \\ \vdots \\ A_n^T \end{pmatrix} \begin{pmatrix} \vdots \end{pmatrix}$$

$$v_i = \frac{1}{\lambda} A_i^T (y - Ax^*) = \langle A_i , y - Ax^* \rangle \cdot \frac{1}{\lambda}$$

$$|\langle A_i , y - Ax^* \rangle| < \lambda \longrightarrow x_i^* = 0$$

شرط کافی

$f(x)=0$; $0 \geq f(y) - f(x) \geq \langle g, y-x \rangle$ $\langle g, y \rangle \leq$

$f(y) \leq 0$

$\longrightarrow$ $g \in -N_C$

$C = \{ y : f(y) \leq 0 \}$ $\longrightarrow$ $\partial f \subseteq -N_C(x)$

~~$\forall$ s.t.~~ $\inf \langle \cancel{g}\, h, y-x \rangle \geq 0$

$h \in -N_C(x)$

محاسبه زیرگرادیان

$I_{\{g(x)\le 0\}}$

- شکل کلی :

$$\min \; f(x)$$
$$g_i(x)\le 0 \quad ; \; i=1,\dots,m$$

- $$\min_x \; f(x) + I_{\{g_1(x)\le 0\}} + I_{\{g_2(x)\le 0\}} + \cdots$$

- شرط بهینگی x^* است

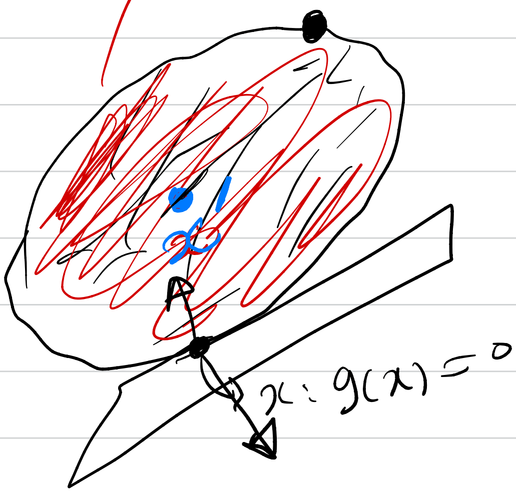
$$0 \in \partial\left(f(x^*) + I_{\{g_1(x^*)\le 0\}} + I_{\{g_2(x^*)\le 0\}} + \cdots \right)$$
$$= \sum \partial\left(I_{\{\le 0\}} \right) + \partial f(x^*)$$

$$I_{\{g_i(x)\le 0\}} = \begin{cases} 0 & g_i(x)\le 0 \\ \infty & g_i(x) > 0 \end{cases}$$

یادآوری :

$$\partial I_C(x) = N_C(x)$$

$C = \{x : g_i(x) \le 0\}$



$N_C(x)$ محاسبه

$N_C(x') = 0$

$x' : \; g(x') < 0 \longrightarrow$ x' داخل مجموعه C

$\longrightarrow N_C(x') = 0$

• x : $g(x) = 0$

• مخروط نرمال : $\{v : \langle v, y-x \rangle \le 0 \}$

$$h \in \partial g(x), \quad y \in C \qquad 0 \ge \underbrace{g(y)}_{\le 0} - \underbrace{g(x)}_{=0} \ge \langle h, y-x \rangle$$

$$\longrightarrow \forall y \in C \quad \langle h, y-x \rangle \le 0 \longrightarrow h \in N_C(x)$$

•] اگر $h' \in N_C(x)$ ، آنگاه $g(y + \varepsilon h') < 0$ برای ε به اندازه کافی کوچک

پس هر ضریبی از داخل N_C در نتیجه $\partial I\{g \le 0\}$ است.