Lec #14

Pointwise max: $f(\underline{x}) = \max_{i=1,\dots,m} \{f_1(\underline{x}), \dots f_m(\underline{x})\}$

$$\partial f(\underline{x}) = \operatorname{conv}\{\partial f_i(\underline{x}) \mid f_i(\underline{x}) = f(\underline{x})\}$$

If each $f_i(\underline{x})$ is differentiable, then

$$\partial f(\underline{x}) = \operatorname{conv}\{\nabla f_i(\underline{x}) \mid f_i(\underline{x}) = f(\underline{x})\}$$

Example:

$$f(\underline{x}) = \|\underline{x}\|_1, \quad \underline{x} \in \mathbb{R}^n$$

$$= |x_1| + \dots + |x_n|$$

$$= \max_{i=1,\dots,m=2^n} \{\underline{s}^T \underline{x} \mid \underbrace{\underline{s} \in \{-1, 1\}^n}_{\substack{\rotatebox{90}{$\equiv$} \\ s_i \in \{-1,1\}}}\}$$

$$f(\underline{x}) = \|\underline{x}\|_1 = |x_1| + \dots + |x_n|$$

$$= \max_{i=1,\dots,m=2^n} \{ \underline{s}^T \underline{x} \mid s_i \in \{-1, 1\} \}$$

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

$$= \mathrm{conv}\{ \underline{s} \mid s_i \in \{-1,1\}^n \}$$

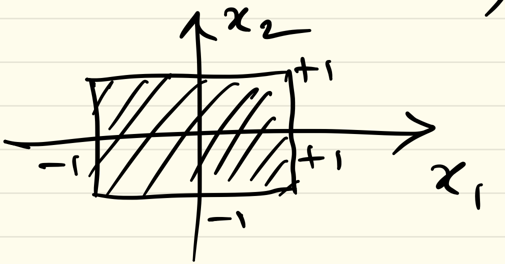
$$= \underbrace{J_1 \times \dots \times J_n}_{\text{product of intervals}}$$

where

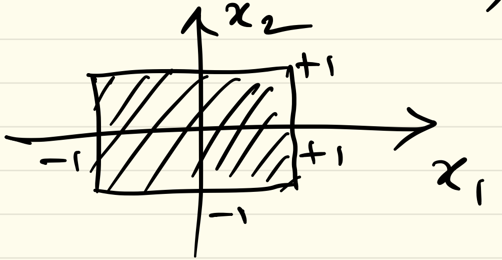
$$J_k = \begin{cases} [-1, 1] & \text{for } x_k = 0 \\ \{1\} & \text{for } x_k > 0 \\ \{-1\} & \text{for } x_k < 0 \end{cases}$$

Say $n = 2$, $(\mathbb{R}^2)$

$\partial f(\underline{x})$ @ $\underline{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

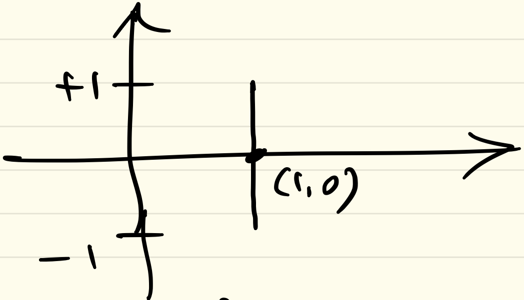


$\partial f(\underline{x})$ @ $\underline{x} = \begin{pmatrix}0\\0\end{pmatrix}$

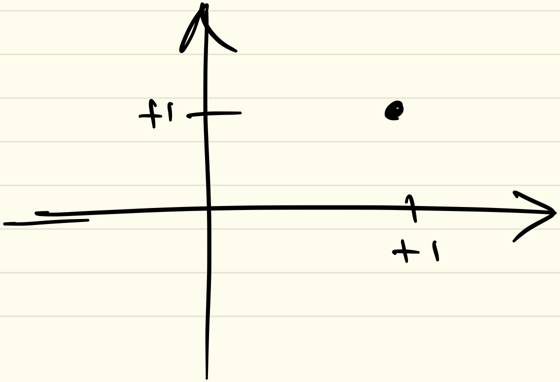


$\partial f(\underline{x})$ @ $\underline{x} = \begin{pmatrix}1\\0\end{pmatrix}$

$\{1\} \times [-1, 1]$



$\partial f(\underline{x})$ @ $\underline{x} = \begin{pmatrix}1\\1\end{pmatrix} = \{(1,1)\}$



Example

$$f(\underline{x}) = \|\underline{x}\|_\infty, \quad \underline{x} \in \mathbb{R}^n$$

$$= \max_{i=1,\dots,n} |x_i|$$

$$f_i(\underline{x}) = |x_i|$$

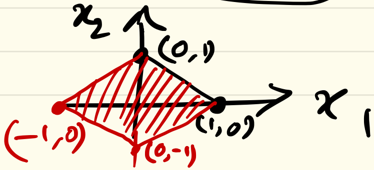
$$= |e_i^T \underline{x}|,$$

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$f(\underline{x}) = \max_{i=1,\dots,n} \{ |e_i^T \underline{x}| \}$$

$$\partial \|\underline{x}\|_\infty = \text{conv}\left\{ \partial |e_i^T x| \;\middle|\; f_i(\underline{x}) = f(\underline{x}) \right\}, \quad \text{say } \underline{x} \in \mathbb{R}^2$$

$\partial \|\underline{x}\|_\infty$ @ $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$



$\text{conv}(\pm e_1, \pm e_2)$

In general:

$$\partial \|\underline{x}\|_\infty \Big|_{\underline{x}=\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}} = \operatorname{conv}\{\pm e_1, \pm e_2, \ldots, \pm e_n\}$$

In general, computing $\partial f(\underline{x})$ is difficult.

Often, we do weak calculus, $\Leftrightarrow$ Find $g \in \partial f(\underline{x})$.

Optimality Condition
(Unconstrained)

If f is cvx, diff. then $f(\underline{x}^*) = \inf\limits_{\underline{x}} f(\underline{x})$

$$\Updownarrow$$

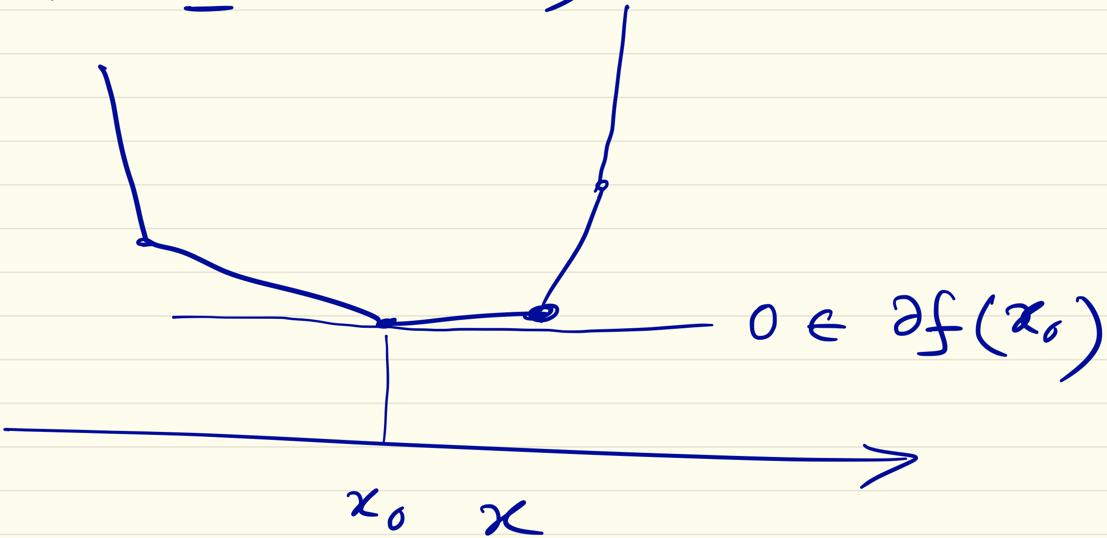
$$\underline{0} = \nabla f(\underline{x}^*)$$

If f is cvx, but NOT diff., then $f(\underline{x}^*) = \inf f(\underline{x}) \Longleftrightarrow \underline{0} \in \partial f(\underline{x}^*)$

Proof : (By defn !!)

$$f(\underline{y}) \geq f(\underline{x}^*) + \underline{0}^T(\underline{y} - \underline{x}^*) \;\forall\, \underline{y}$$

$$\Leftrightarrow \quad \underline{0} \in \partial f(\underline{x}^*)$$



$\blacksquare$

Example : Piecewise linear minimization

$$\min_{\underline{x} \in \mathbb{R}^n} \quad f_0(\underline{x}) = \max_{i=1,\dots,m} \left(\underline{a}_i^T \underline{x} + b_i \right)$$

$\underline{x}^*$ minimizes $f_0(\underline{x}) \Leftrightarrow \underline{0} \in \partial f_0(\underline{x}^*)$

$$= \operatorname{conv}\left\{ a_i \,\middle|\, a_i^T \underline{x}^* + b_i = f_0(\underline{x}^*) \right\}$$

LP
$$\begin{cases} \min\limits_{\underline{x} \in \mathbb{R}^n,\, t \in \mathbb{R}} & t \\ \text{s.t.} & a_i^T \underline{x} + b_i \leq t \quad \forall i = 1, \dots, m \end{cases}$$

Dual problem : also LP

$$\max \quad \underline{b}^T \lambda$$
$$\text{s.t.} \quad \underline{\lambda} \succcurlyeq 0, \quad A^T \underline{\lambda} = 0, \quad \mathbb{1}^T \underline{\lambda} = 1$$

Optimality condition for constrained.

$$\min f_0(\underline{x})$$
$$\text{s.t. } f_i(\underline{x}) \le 0, \quad i = 1, \ldots, m$$

Assume $f_0, f_1, \ldots, f_m \rightarrow$ all cvx but NOT diff.

- Strict primal feasibility (Strong Duality)

Suppose $\underline{x}^*$ is primal optimal
$\underline{\lambda}^*$ is dual optimal

$\Updownarrow$

Generalized KKT condition:

- $\lambda_i^* f_i(\underline{x}^*) = 0$
- $f_i(\underline{x}^*) \le 0, \quad \underline{\lambda}_i^* \succeq 0$ as before
- $\underline{0} \in \partial f_0(\underline{x}^*) + \sum_{i=1}^{m} \lambda_i^* \partial f_i(\underline{x}^*)$

Applications

Least squares: (unconstrained)

$$\min_{\underline{x} \in \mathbb{R}^n} \underbrace{\|A\underline{x} - b\|_2^2}_{f_0(\underline{x})}, \qquad A \text{ has indep. columns}$$

$$f_0(\underline{x}) = x^T A^T A \underline{x} - 2\underline{b}^T A \underline{x} + \underline{b}^T \underline{b}$$

$$\nabla f_0(\underline{x}) = 2A^TA\underline{x} - 2A^T\underline{b} = 0$$

$$\Leftrightarrow \quad \underline{x} = (A^TA)^{-1} A^T \underline{b}$$

ℓ_∞ norm approximation:

$$\min_{\underline{x} \in \mathbb{R}^n} \| A\underline{x} - \underline{b} \|_\infty$$

$$\Updownarrow$$

$$\left.\begin{aligned} &\min_{\underline{x} \in \mathbb{R}^n,\, t \in \mathbb{R}} \quad t \\ &\text{s.t.} \quad -t\mathbb{1} \leq A\underline{x} - \underline{b} \leq +t\mathbb{1} \end{aligned}\right\} \underline{LP}$$

ℓ_1 norm approximation

$$\min_{\underline{x} \in \mathbb{R}^n} \| A\underline{x} - \underline{b} \|_1$$

$$\Updownarrow$$

$$LP \left\{ \begin{aligned} &\min_{\underline{x} \in \mathbb{R}^n,\, \underline{t} \in \mathbb{R}^n} \quad \mathbb{1}^T \underline{t} \\ &\text{s.t.} \quad -\underline{t} \leq A\underline{x} - \underline{b} \leq \underline{t} \end{aligned} \right.$$

Polynomial fitting $\Leftrightarrow$ constrained least sq. approx. (QP)

Smallest vol^m ellipsoid containing finite dataset $\rightarrow$ convex $\rightarrow$ SDP (Dual)

Statistical estimation problem:

maximum likelihood estimators

$$\max_{\underline{x} \in \mathbb{R}^m} \underbrace{l(\underline{x})}_{\text{Log-likelihood}} = \log \underbrace{p}_{\text{PDF}}{}_{\underline{x}}(\underline{y})$$

$\underline{y}$: random vector $\in \mathbb{R}^n$

$\underline{x}$: parameter vector $\in \mathbb{R}^m$

If p as $f^{\underline{n}}$ of $\underline{x}$ (parameter vector) is log-concave then max. likelihood problem is convex optimization problem.

Examples:

linear measurement with i.i.d. noise

$$\boxed{\underbrace{y_i}_{\text{measurements}} = a_i^T \underline{x} + v_i}, \quad i = 1, \ldots, m$$

$\in \mathbb{R}$

$\underline{x} \in \mathbb{R}^n$: parameter vector to be estimated

$v_i \rightarrow$ measurement noise, i.i.d. with density $p(\cdot)$ on $\mathbb{R}$

$$p_{\underline{x}}(\underline{y}) = \prod_{i=1}^{m} p(y_i - a_i^T \underline{x})$$

$$\Rightarrow l(\underline{x}) = \log p_{\underline{x}}(\underline{y}) = \sum_{i=1}^{m} \log p(y_i - a_i^T \underline{x})$$

ML estimate:

$$\max_{\underline{x} \in \mathbb{R}^n} \sum_{i=1}^{m} \log p(y_i - a_i^T \underline{x})$$

Convex optimization if $p(\cdot)$ is log-concave in $\underline{x}$ (param.)

Linear model with
Gaussian noise:

$$v_i \sim \mathcal{N}(0, \sigma^2)$$

$$p(z) = (2\pi\sigma^2)^{-1/2} \exp\left(-z^2/2\sigma^2\right)$$

$$l(\underline{x}) = -\left(\frac{m}{2}\right) \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \|A\underline{x} - \underline{y}\|_2^2$$

where A has rows $\underline{a}_1^T, \ldots, \underline{a}_m^T$

$$\therefore \underline{x}^*_{ML} = \underset{\underline{x} \in \mathbb{R}^n}{\arg\min} \|A\underline{x} - \underline{y}\|_2^2$$

ordinary least sq.