TOPICS:

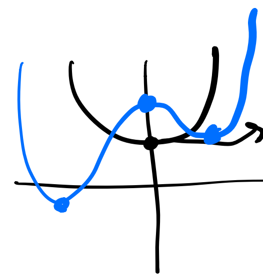
- SYSTEMS OF EQUATIONS
- REVIEW LAGRANGE MULTIPLIERS
- RANK OF MATRICES

LAGRANGE MULTIPLIERS –

BASIC VECTOR OPTIMIZATION

$$\min_{x \in \mathbb{R}^n} f(x) \qquad f: \mathbb{R}^n \to \mathbb{R}$$

LOCAL OPTIMALITY COND: $\boxed{\dfrac{\partial f}{\partial x} = 0}$



$\dfrac{\partial f}{\partial x} = 0$

ex. $f(x) = x^T \boxed{Q} x \quad x \in \mathbb{R}^2$

$Q = Q^T \succ 0$

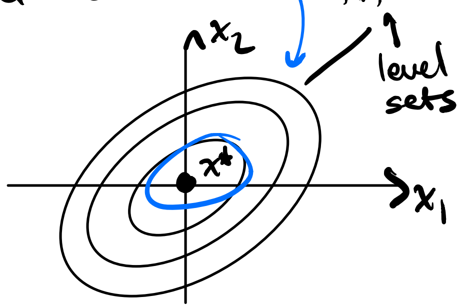
Positive definite:

$\Rightarrow$ bowl curves up

f(x) = constant

level sets

top down

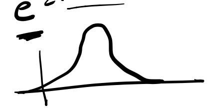


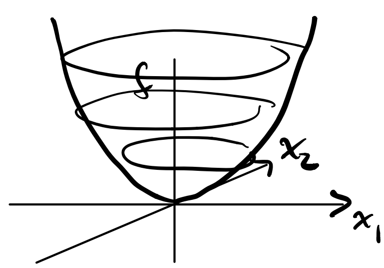
Notes:

- prob. connection

$f(x) \sim e^{-\frac{1}{2}(x^TQx)}$

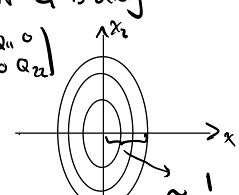
density





- if Q is diagonal

$Q = \begin{bmatrix} Q_{11} & 0 \\ 0 & Q_{22} \end{bmatrix}$



$\sim \dfrac{1}{Q_{11}}$

$\dfrac{\partial f}{\partial x} = x^T(Q + Q^T) = 2x^TQ = 0$

$\Rightarrow x^* = 0$ since Q is PD.

$f(x) = [x_1\, x_2] \begin{bmatrix} Q_{11} & 0 \\ 0 & Q_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= Q_{11} x_1^2 + Q_{22} x_2^2 = \text{constant}$

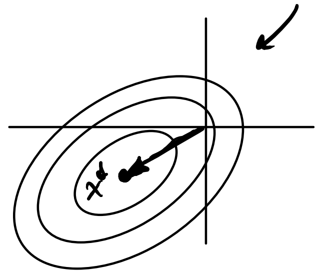
if $x_2 = 0$ $\quad Q_{11} x_1^2 = \text{constant}$

$x_1^2 = \dfrac{\text{constant}}{Q_{11}}$

ex. $f(x) = \frac{1}{2} x^T Q x + c^T x \qquad Q = Q^T \succ 0$

$$\frac{\partial f}{\partial x} = x^T Q + c^T = 0 \Rightarrow \left[\, x^T Q = -c^T \,\right] Q^{-1} \Rightarrow \boxed{x^{*T} = -c^T Q^{-1}}$$

$$\rightarrow x^* = -Q^{-1} c$$



could see this shift by "completing the square"

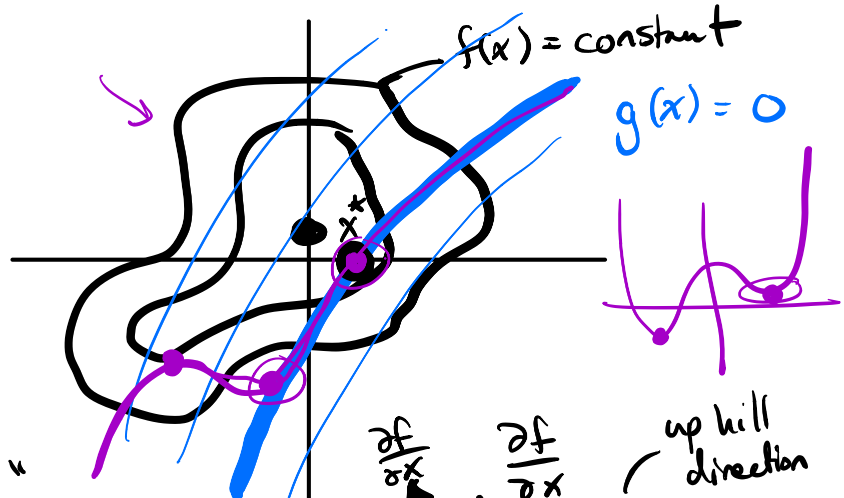
$$f(x) = \frac{1}{2} x^T Q x + c^T x \text{ + const } - \text{ const}$$

$$f(x) = \frac{1}{2}(x - x^*)^T Q (x - x^*) - \text{const}$$

$$\frac{\partial f}{\partial x} = (x - x^*)^T Q = 0$$

Constraints

$$\min_{x \in \mathbb{R}^n} f(x)$$
s.t. $\boxed{g(x) = 0}$



Optimality Conds:

"the gradient of a function is ⊥ to the level sets"

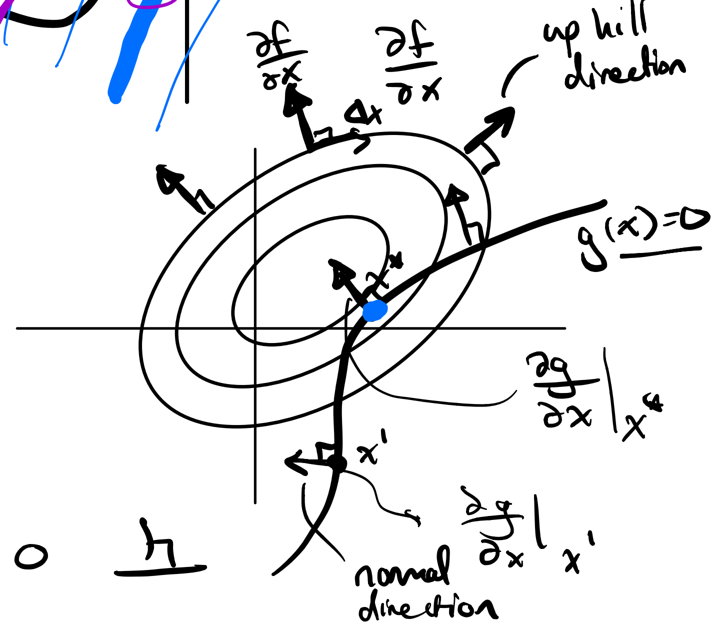
$f(x) = \text{constant}$

Δx along level set...

$\Rightarrow \Delta f = 0$

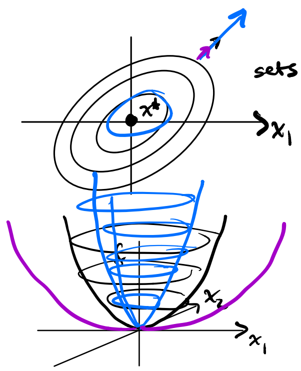
$$\Delta f = \frac{\partial f}{\partial x} \Delta x = 0$$

$$= \left[\, \frac{\partial f}{\partial x} \,\right] \left[\, \Delta x \,\right] = 0 \quad \perp$$



Question:

$\frac{\partial f}{\partial x}, \frac{\partial g}{\partial x} \Rightarrow$ optimality cond.?



"$\left.\frac{\partial f}{\partial x}\right|_{x^*}$ and $\left.\frac{\partial g}{\partial x}\right|_{x^*}$ are parallel."

Notes:

- sign of g doesn't matter

$g(x) = 0 \Leftrightarrow -g(x) = 0$

- mag of g doesn't matter

$g(x) = 0 \Leftrightarrow \text{const} \times g(x) = 0$

want optimality cond to ensure $\frac{\partial f}{\partial x}$ and $\frac{\partial g}{\partial x}$ are parallel, but we also want to be agnostic to their magnitudes...

Lagrange Multipliers:

before: $\frac{\partial f}{\partial x} = 0$

now: $\boxed{\frac{\partial f}{\partial x} = \lambda^T \frac{\partial g}{\partial x}}$

derivative of f is linear comb of derivative of g

$\frac{\partial f}{\partial x}$: a row vector

$g(x) = 0$ can be a scalar equation

or

$g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ vector equation

scalar case: $\frac{\partial g}{\partial x} = [\text{- row vector -}]$

vector case: $\frac{\partial g}{\partial x} = [\text{matrix}]$

scalar $g(x)$: $\quad \underline{\dfrac{\partial f}{\partial x}} = \lambda \underline{\dfrac{\partial g}{\partial x}} \Rightarrow \Big[\underset{\substack{\uparrow \\ \text{row}}}{\dfrac{\partial f}{\partial x}} \Big] = \underset{\substack{\downarrow \\ \text{scalar}}}{\lambda} \Big[\underset{\substack{\uparrow \\ \text{row}}}{\dfrac{\partial g}{\partial x}} \Big]$

vector $g(x)$: $\quad \underset{\nearrow \text{ row}}{\Big[\dfrac{\partial f}{\partial x} \Big]} = \underbrace{[\lambda_1 \cdots \lambda_m]}_{\substack{\lambda^T \\ \text{row vector}}} \underbrace{\begin{bmatrix} - & \frac{\partial g_1}{\partial x} & - \\ & \vdots & \\ - & \frac{\partial g_m}{\partial x} & - \end{bmatrix}}_{\text{matrix}}$

$$\frac{\partial f}{\partial x} = \lambda_1 \frac{\partial g_1}{\partial x} + \cdots + \lambda_m \frac{\partial g_m}{\partial x}$$

Vector Example

$\min_{x \in \mathbb{R}^3} f(x) \qquad f: \mathbb{R}^3 \to \mathbb{R}$

s.t. $\underline{g(x) = 0} \qquad g: \mathbb{R}^3 \to \mathbb{R}^2$

$$\begin{bmatrix} g_1(x) \\ g_2(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Exs $g(x)$ ...

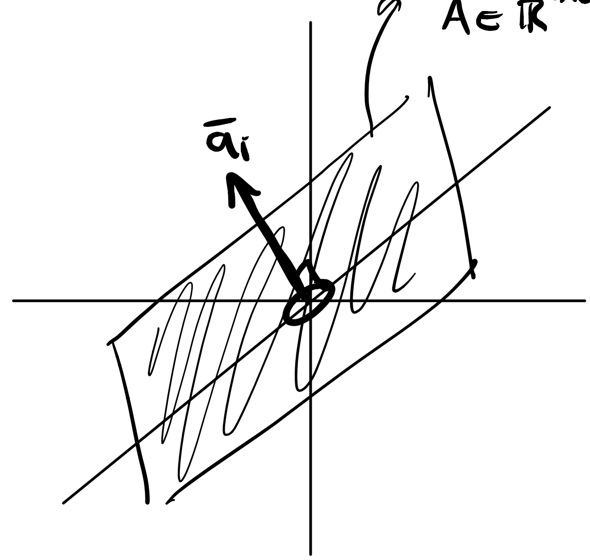
$g(x) = Ax = 0$

$A \in \mathbb{R}^{1 \times 3}$

$$[\underbrace{- \bar{a}_1^T -}_{\text{row}}] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

$\{x \mid Ax = 0,\ x \in \mathbb{R}^3\}$

$A \in \mathbb{R}^{1 \times 3}$

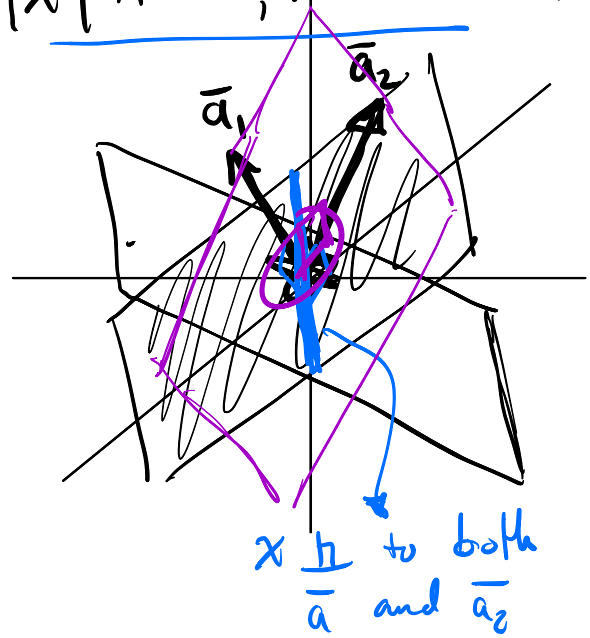


$$g(x) = Ax = 0$$

$$A \in \mathbb{R}^{2\times 3}$$

$$\underbrace{\begin{bmatrix} - \bar{a}_1^T - \\ - \bar{a}_2^T - \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

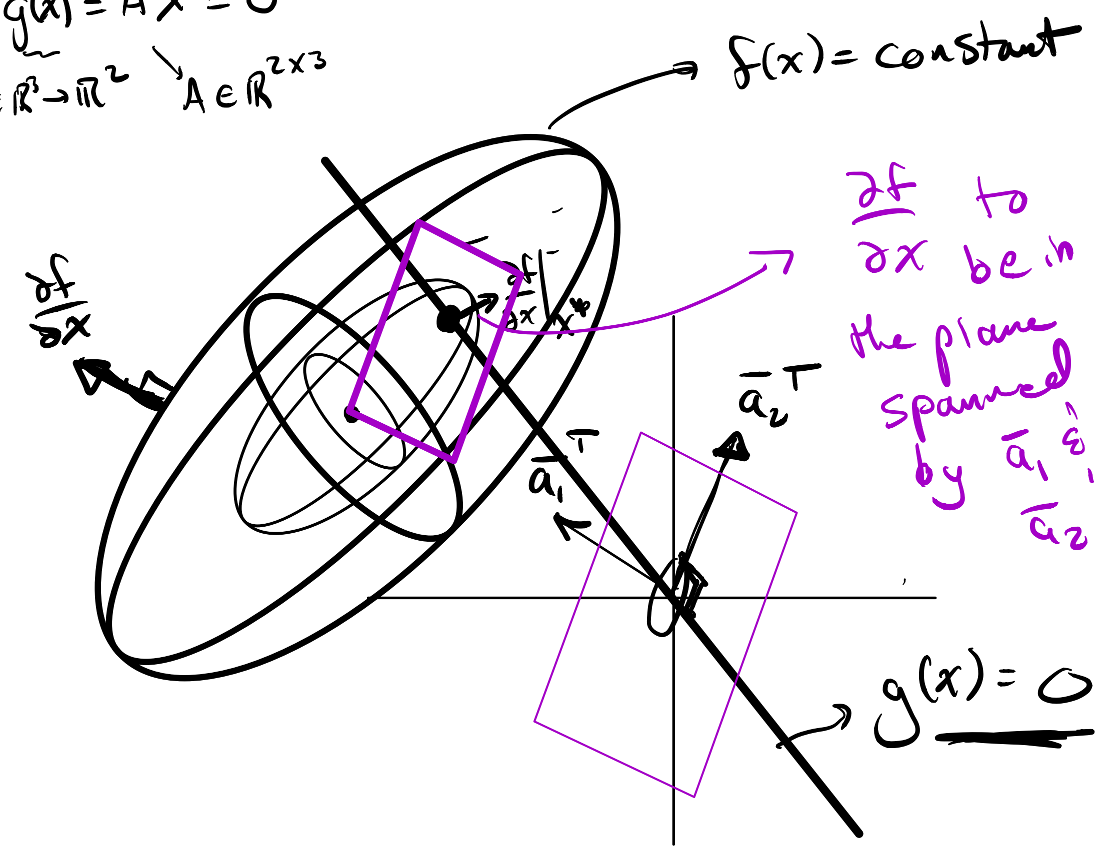
$\{x \mid Ax = 0,\ x \in \mathbb{R}^3\}$ $\quad A \in \mathbb{R}^{2\times 3}$



$$\min_{x \in \mathbb{R}^3} f(x)$$

s.t. $g(x) = Ax = 0$

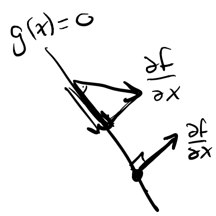
$g: \mathbb{R}^3 \to \mathbb{R}^2$ $\quad A \in \mathbb{R}^{2\times 3}$



$$\frac{\partial f}{\partial x} = \lambda_1 \bar{a}_1^T + \lambda_2 \bar{a}_2^T = [\lambda_1, \lambda_2]\begin{bmatrix} - \bar{a}_1^T - \\ - \bar{a}_2^T - \end{bmatrix}$$

$$\boxed{\frac{\partial f}{\partial x} = \lambda^T \frac{\partial g}{\partial x}} \qquad = \lambda^T A = \lambda^T \frac{\partial g}{\partial x}$$

Constraints push back against the gradient, but they can only push in certain directions given by span of $\frac{\partial g_1}{\partial x}, \ldots \frac{\partial g_m}{\partial x}$. How much they push at optimum and direction they push is given by $\lambda_1 \ldots \lambda_m$



λ: Lagrange multipliers
dual variables

x: primal variables

$$\min_x f(x) \quad \text{s.t. } g(x) = 0$$

Lagrangian:

$$\mathcal{L}(x,\lambda) = f(x) - \lambda^T g(x)$$

before: $\frac{\partial f}{\partial x} = 0$

optimality
"stationarity"

Now:

$$\frac{\partial \mathcal{L}}{\partial x} = 0 : \frac{\partial f}{\partial x} - \lambda^T \frac{\partial g}{\partial x} = 0 \Rightarrow \boxed{\frac{\partial f}{\partial x} = \lambda^T \frac{\partial g}{\partial x}}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 : -\frac{\partial}{\partial \lambda}(\lambda^T g(x)) = \boxed{g(x) = 0}$$

$$-\frac{\partial}{\partial \lambda}(g(x)^T \lambda)$$

constraints are satisfied
"feasibility"

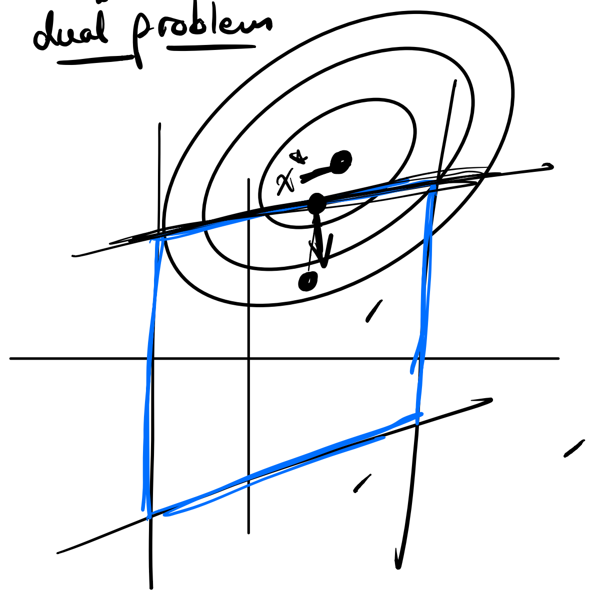
equality constraints

$\frac{\partial \mathcal{L}}{\partial x} = 0$: "stationarity"

$\frac{\partial \mathcal{L}}{\partial \lambda} = 0$: "feasibility"

switch roles of x & λ

dual problem



$\min_x f(x)$
s.t. $g(x) = 0$

$\min_x f(x)$
s.t. $\underbrace{g(x) = 0}_{\text{equality constraints}}$ $\underbrace{h(x) \leq 0}_{\text{inequality constraints}}$

optimality conds

$\frac{\partial \mathcal{L}}{\partial x}$: stationarity

$\frac{\partial \mathcal{L}}{\partial \lambda}$: feasibility

complementary slackness

$\boxed{\lambda^T h(x) = 0}$ ← not linear
encode constraints switching on or off

KKT conditions