Constrained Optimization with Equality Constraints

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

Look for x_0 s.t. $\nabla f(x_0) \parallel \nabla g(x_0)$ and $g(x_0) = 0$

$\exists \lambda$ s.t. $\nabla f(x_0) = -\lambda \nabla g(x_0)$

λ is called Lagrange multiplier

(Note: $\nabla f(x_0) \parallel \nabla g(x_0) \Rightarrow$

$\exists\, a, b$, not both zero s.t. $a \nabla f(x_0) + b \nabla g(x_0) = 0$

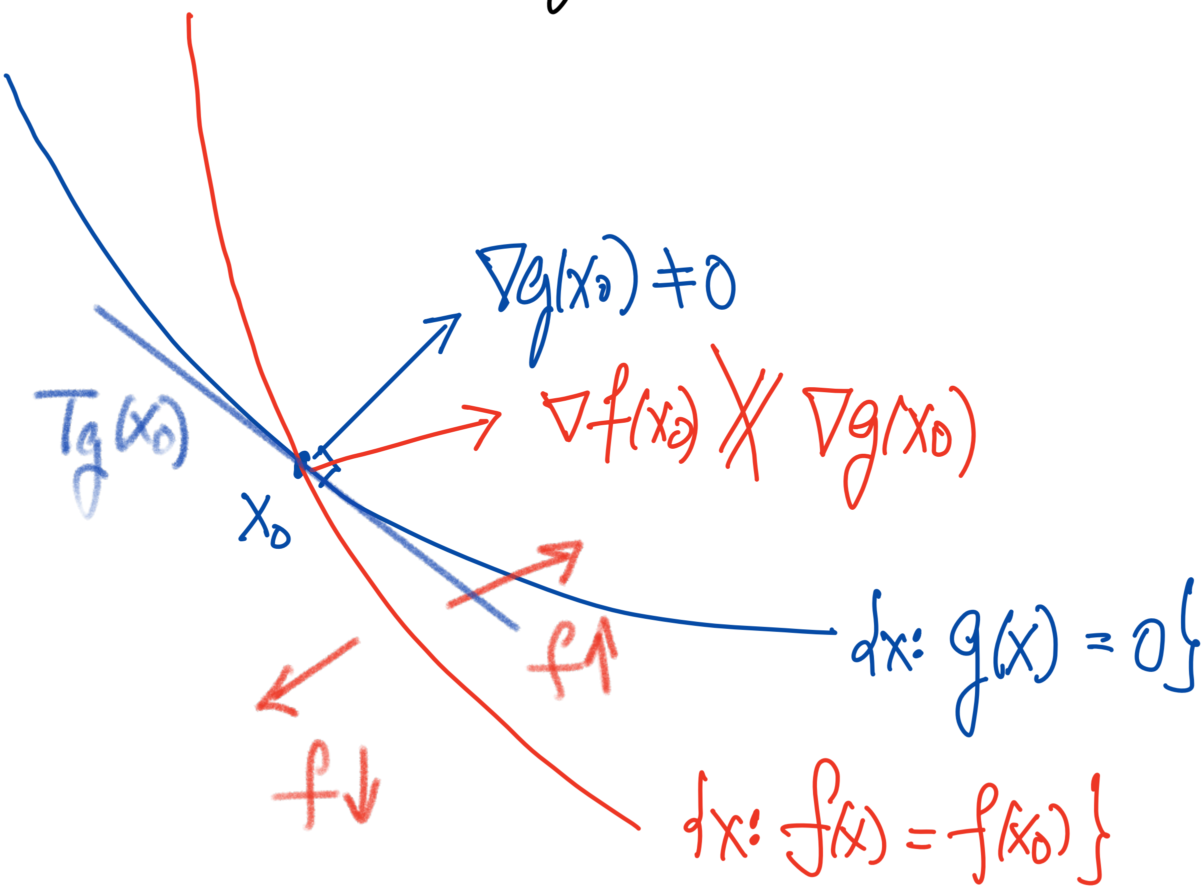
Assume x_0 is regular in the sense that $\nabla g(x_0) \neq 0$.

Then $a \neq 0 \Longrightarrow$

$$\nabla f(x_0) + \overset{\lambda}{\frac{b}{a}} \nabla g(x_0) = 0.$$

)

Intuition If the level set $\{x: f(x) = f(x_0)\}$ cross $\{x: g(x) = 0\}$



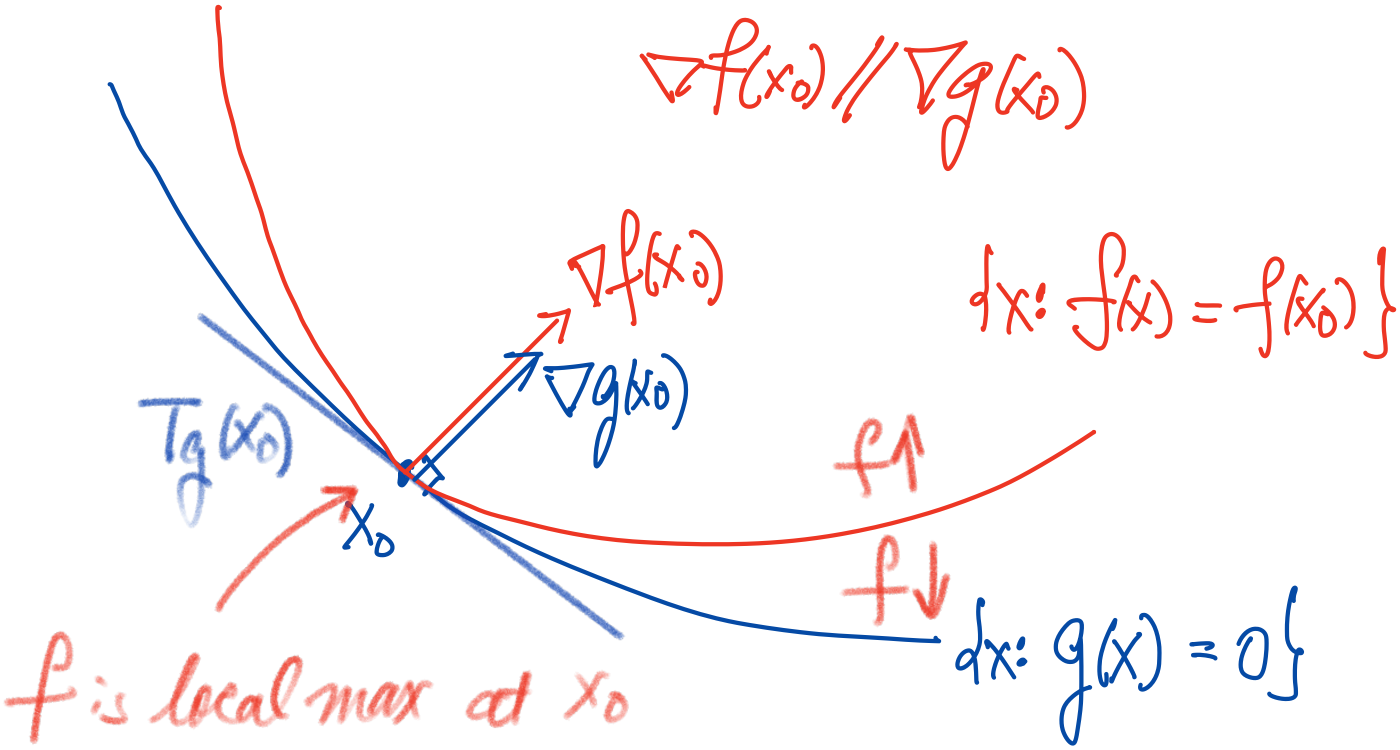
Then x_0 cannot be a critical/min/max of $f(x)$ on $\{x: g(x) = 0\}$

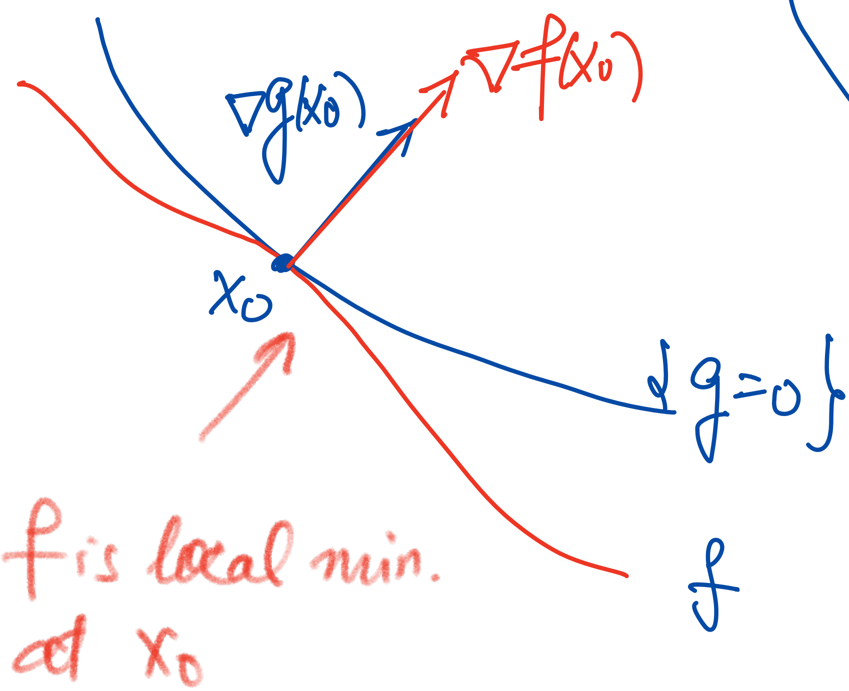
(1) $\nabla f(x_0)$ has non-zero projection onto the tangent plane $T_g(x_0)$ of $\{x: g(x) = 0\}$ at x_0

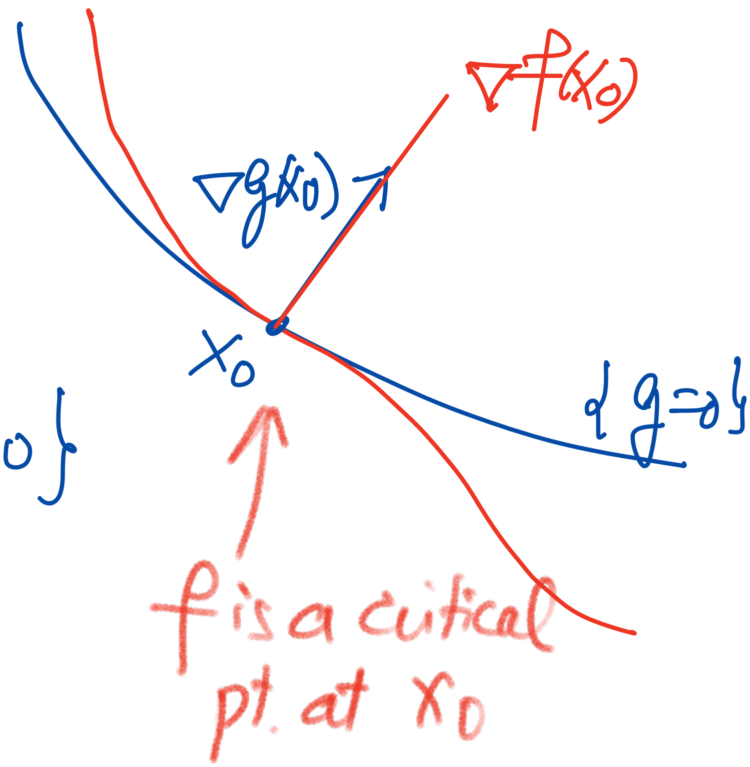
(2) Move x along $\{x: g(x) = 0\}$

$\Rightarrow$ $f(x)$ can increase and decrease.

Hence $\{x : f(x) = f(x_0)\}$ must "touch" $\{x : g(x) = 0\}$:







With multiple equality constraints

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_i(x) = 0, \quad i = 1, \cdots, m$$

Find x_0, $\lambda_1, \lambda_2, \cdots, \lambda_m$ s.t.

$$\boxed{\begin{aligned} &\nabla f(x_0) + \sum_i \lambda_i \nabla g_i(x_0) = 0 \\ &g_i(x_0) = 0, \quad i = 1, \cdots, m \end{aligned}}$$

(1) Assume x_0 is regular in the sense that $\{\nabla g_i(x_0), i = 1, \cdots, m\}$ is linearly independent

(2) $\lambda_1, \cdots, \lambda_m$ — Lagrange multipliers

(3) $\nabla f(x_0) = \underbrace{-\sum_i \lambda_i \nabla g_i(x_0)}_{\in \operatorname{Span}\{\nabla g_i(x_0),\ i = 1, \cdots, m\}}$

A bit of notation

$$\nabla f(x_0) + \sum_i \lambda_i \nabla g_i(x_0) = 0$$

(recall $Df = (\nabla f)^T$)

$$\Updownarrow$$

$$Df(x_0) + \sum_i \lambda_i Dg_i(x_0) = 0$$

row vector

$$Df(x_0) + \lambda^T DG(x_0) = 0$$

where $\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$, $\lambda^T = (\lambda_1, \lambda_2, \cdots \lambda_m)$

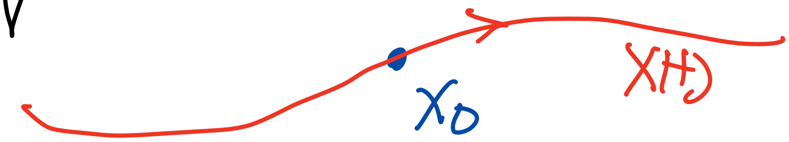
$$G(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}^{m\times 1},$$

$$DG(x) = \begin{pmatrix} Dg_1 \\ \vdots \\ Dg_m \end{pmatrix} = \begin{pmatrix} (\nabla g_1)^T \\ \vdots \\ (\nabla g_m)^T \end{pmatrix}$$

$$= \begin{pmatrix} \partial_{x_1} g_1 & \partial_{x_2} g_2 & \cdots & \partial_{x_n} g_1 \\ \cdots & & & \\ \partial_{x_1} g_m & \partial_{x_2} g_m & \cdots & \partial_{x_n} g_m \end{pmatrix}^{m\times n}$$

(Intuition: let $x(t)$ be any curve s.t.

$$x(0) = x_0, \qquad g_i(x(t)) = 0, \quad i=1,\dots,m.$$



Then computing directional derivative of f along $x(t)$:

$$\frac{d}{dt} f(x(t))\Big|_{t=0} = \langle \nabla f(x_0), \dot{x}(0)\rangle = 0$$

But
$$\frac{d}{dt} g_i(x(t))\Big|_{t=0} = \frac{d}{dt} 0 = 0$$
$$\frac{d}{dt} g_i(x(t))\Big|_{t=0} = \langle \nabla g_i(x_0), \dot{x}(0)\rangle$$

i.e. $\nabla f(x_0) \perp$ any vector V s.t.

$$\langle \nabla g_i(x_0), V\rangle = 0 \quad \text{for } i=1,\dots,m.$$

$$\Longleftrightarrow \quad \nabla f(x_0) \in \operatorname{Span}\{\nabla g_i(x_0),\ i=1,\dots,m\}$$

Related linear algebra fact: $\mathrm{Null}(A)^{\perp} = \mathrm{Row}(A)$)

where $A = DG$.

Examples

(1) Maximizing entropy for probability measures:

$$\max \quad f(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i \log p_i$$

$$\text{s.t.} \quad \sum_{i=1}^{n} p_i = 1 \qquad g(p) = \left(\sum_i p_i - 1\right) = 0$$

$$\nabla f + \lambda \nabla g = 0 \qquad \left(\vec{p} = (p_1, \dots p_n)\right)$$

$$\partial_{p_i} f(p) + \lambda \partial_{p_i} g(p) = 0$$

$$-\log p_i - 1 + \lambda = 0$$

$$\log p_i = \lambda - 1$$

$$p_i = e^{\lambda - 1} \leftarrow \text{a constant, does not depend on } i$$

$$\boxed{p_1 = p_2 = \cdots = p_n = \frac{1}{n} \qquad \left(\sum_i p_i = 1\right)}$$

(2) Max. entropy with energy constraint:

$$\max f(p) = -\sum_i p_i \log p_i$$

s.t. $\sum p_i = 1$, and $\sum_i E_i p_i = E$

energy level   average energy

$$g_1(p) = \sum_i p_i - 1 = 0$$

$$g_2(p) = \sum_i E_i p_i - E = 0$$

Find α, β, $\vec{p}$ s.t.

$$\nabla f(p) + \alpha \nabla g_1(p) + \beta \nabla g_2(p) = 0$$

$$\sum p_i = 1 \quad \text{and} \quad \sum E_i p_i = E$$

$$-\log p_i - 1 + \alpha + \beta E_i = 0$$

$$\log p_i = \alpha + \beta E_i - 1$$

Gibbs distribution

$$p_i = e^{\alpha + \beta E_i - 1} = C e^{\beta E_i} \quad (C = e^{\alpha - 1})$$

Find C, β s.t.

$$\sum_i C e^{\beta E_i} = 1$$

$$\sum_i C E_i e^{\beta E_i} = E$$

$C = ?$

$\beta = ?$

$$C = \frac{1}{\sum_i e^{\beta E_i}}$$

$$\underbrace{\sum_i \frac{E_i e^{\beta E_i}}{\sum_j e^{\beta E_j}}}_{①} = E$$

(Note such a β exists:

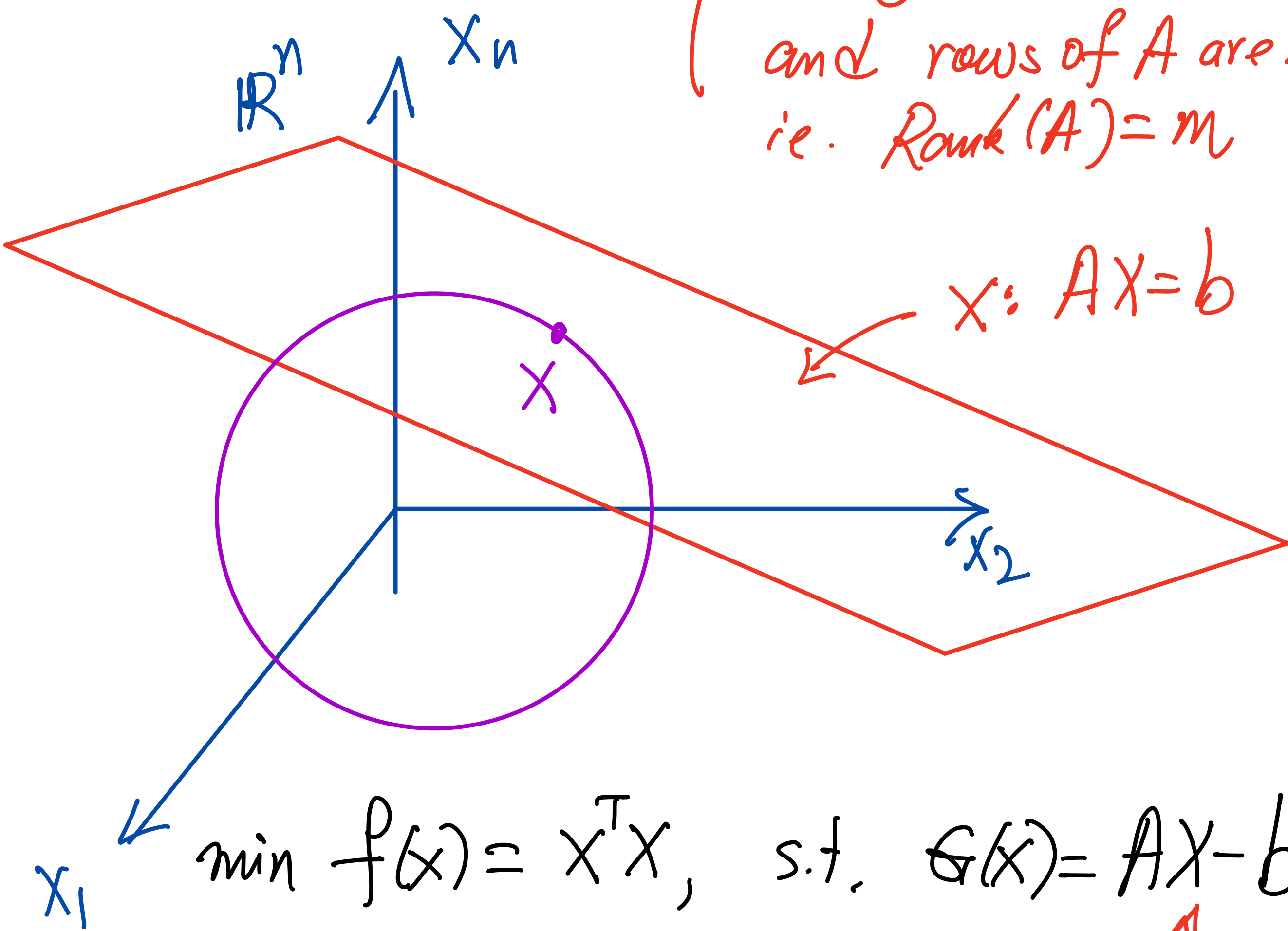
$$① \xrightarrow{\beta \to +\infty} \max E_i \geq E \geq \min E_i \xleftarrow{\beta \to -\infty} ①$$

)

(3) Solution of $AX=b$ with min norm/length

$$\min_{X\in\mathbb{R}^n} \|X\|^2 \quad \text{s.t.} \quad A^{m\times n} X = b$$

(assume $m \le n$
and rows of A are lin ind
i.e. $\text{Rank}(A) = m$)



$$\min f(X) = X^T X, \quad \text{s.t.} \quad G(X) = AX - b = 0$$

$$\underbrace{Df(X)}_{X^T} + \lambda^T \underbrace{DG(X)}_{A} = 0$$

Hence, $\quad X^T + \lambda^T A = 0 \qquad$ i.e. $X = A^T \lambda$

$$AA^T\lambda = b \implies \lambda = (AA^T)^{-1} b$$

$$\implies \boxed{X = A^T (AA^T)^{-1} b}$$

(4) $\min\limits_{x \in \mathbb{R}^n} x^T Q x$ s.t. $\|x\|^2 = 1$

$Q = Q^T$, $n\times n$ sym. matrix

$f(x) = x^T Q x$, $Df(x) = 2x^T Q$

$g(x) = x^T x - 1$, $Dg(x) = 2x^T$

Hence $x^T Q + \lambda x^T = 0$

$$Q^T x + \lambda x = 0$$

$$(Q + \lambda I) x = 0$$

$$Qx = -\lambda x \leftarrow (-\lambda), \text{ eigenvalue of } Q$$

$$f(x) = x^T Q x = x^T(-\lambda x) = -\lambda x^T x = (-\lambda)$$

Try to find min eigenvalue of Q

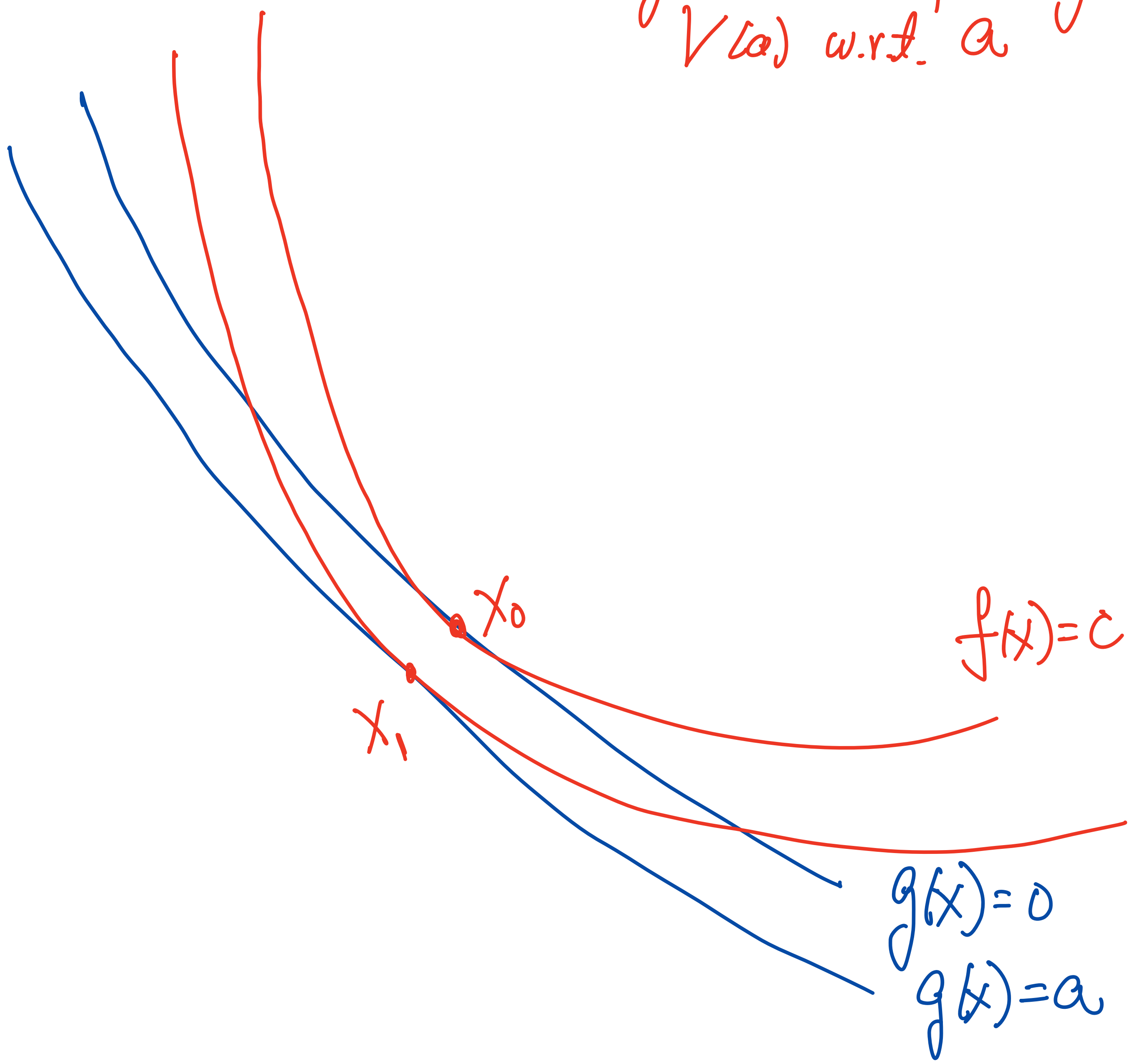
Interpretation of Lagrange Multiplier in terms of Value function (Evans, p. 15)

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

Let the min value be $V(a)$

Then λ_0 (at $a=0$) $= -V'(0)$

negative rate of change $V(a)$ w.r.t. a



Pf: Let $\min f(x)$ s.t. $g(x)=0$ has a solution at x_0 and

$$\nabla f(x_0)+\lambda_0 \nabla g(x_0)=0$$

Consider: $V(a)=\min\{f(y):\ g(y)=a\}$

Let $x\in\mathbb{R}^n$. Then

$$V(g(x))=\min\{f(y):\ g(y)=g(x)\}$$
$$\leq f(x) \quad (\text{let } y=x)$$

But

$$V(g(x_0))=V(0)=f(x_0)$$

Hence the function $h(x)=f(x)-V(g(x))$ is minimized at $x=x_0$

Thus $\nabla h(x_0)=0$

$$\Rightarrow \nabla f(x_0)-\underbrace{V'(g(x_0))}_{\lambda_0 \text{ Lagrange multiplier}}\nabla g(x_0)=0$$

Back to LP

$$\max \quad \zeta(x) = c^T x$$
$$\text{s.t.} \quad Ax \leq b$$
$$x \geq 0$$

At opt: $\zeta^* = \zeta(x_*) = c_B^T B^{-1} b$

If b is changed by Δb,

then ζ^* (opt. value) is changed by

$$\Delta \zeta^* = \underbrace{(c_B^T B^{-1})}_{y^* \text{ (opt. solution for dual problem)}} \Delta b$$

can be viewed as a Lagrange multiplier.