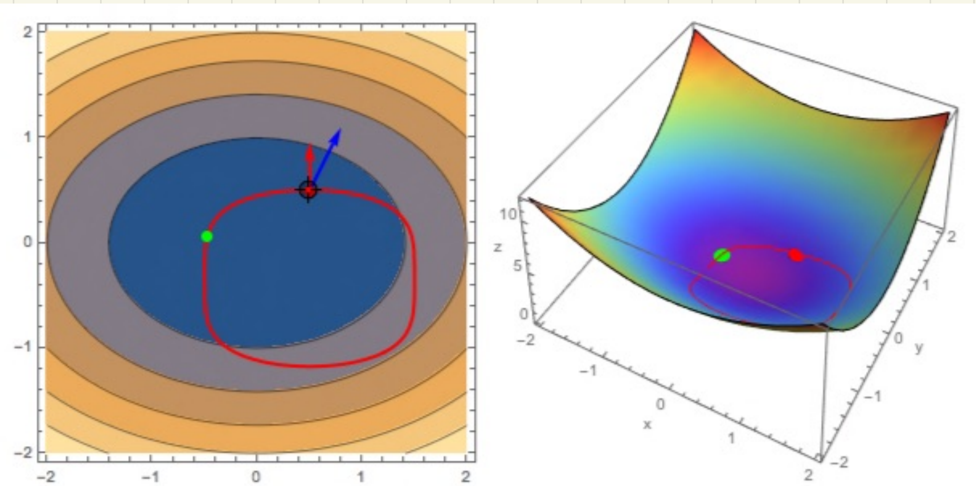
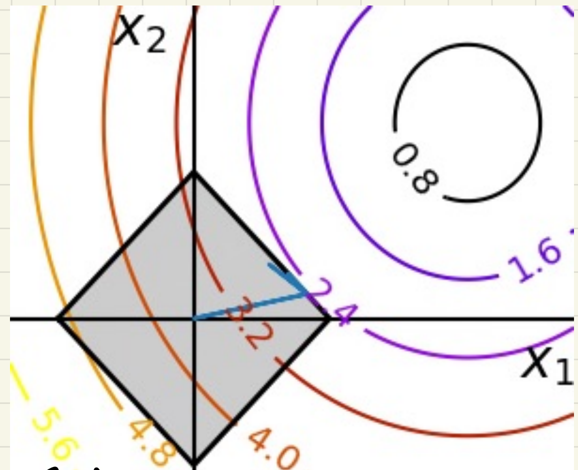
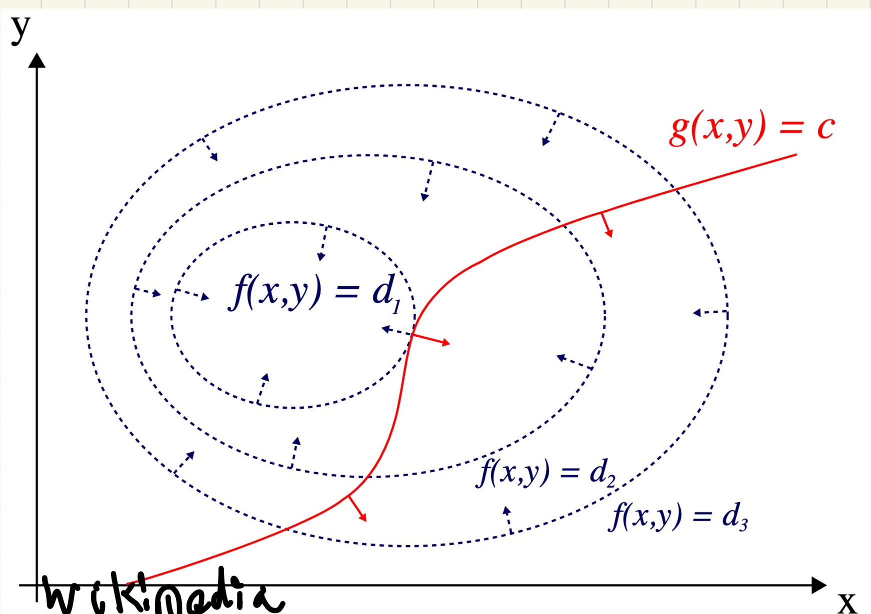
Constrained Optimization



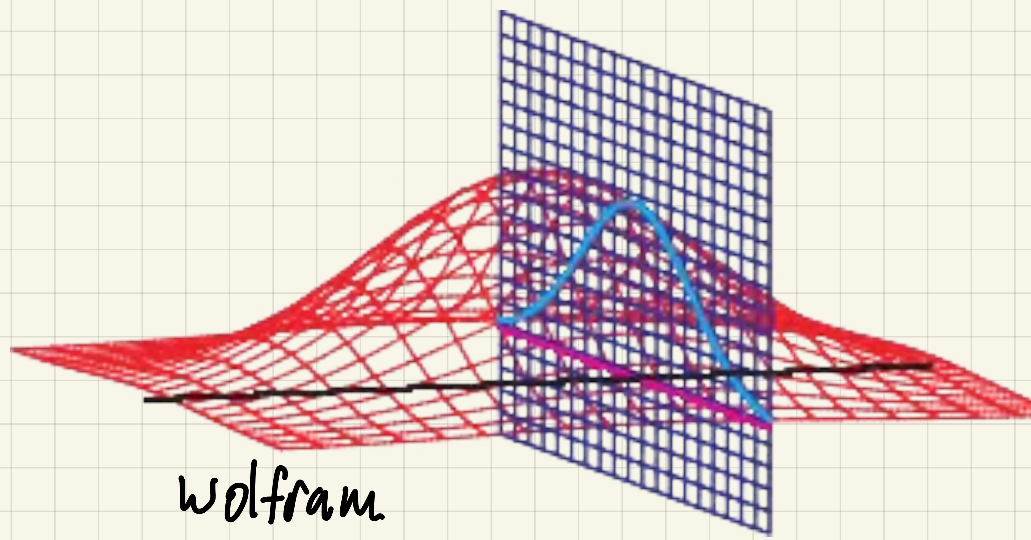
Wolfram



SciPy



wikipedia



wolfram

Objectives today

1. Step from linear least-squares to a constraint

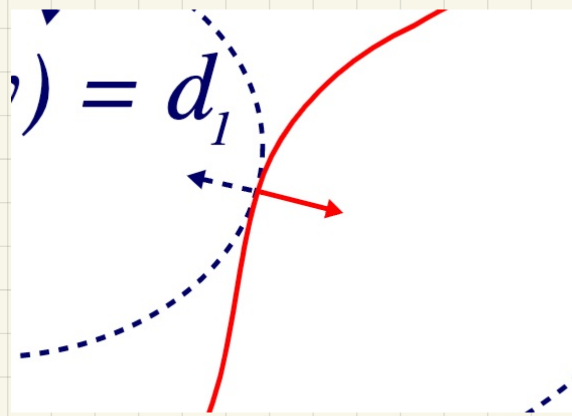
2. Identify the "dual form" of a problem.

3. Introduce Lagrange Multipliers λ

obj. constraint

$\nabla f \sim \nabla g$

λ



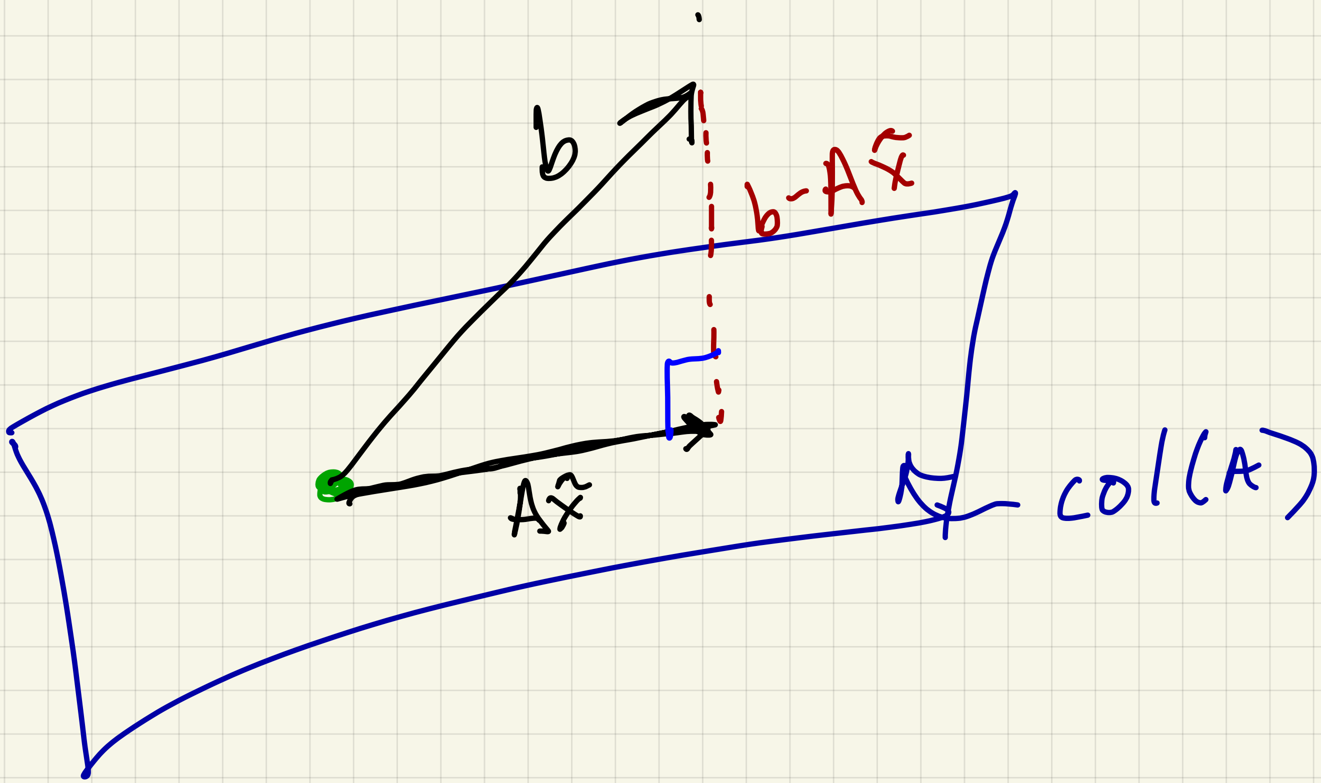
4. Define conditions for a minimum.

$$\|Ax - b\|^2 \longrightarrow \hat{x}$$

$\hookrightarrow A^T A x = A^T b$

$\hookrightarrow A^T (b - Ax) = 0$

$\hookrightarrow A^T r = 0$



$$A^TAx = A^Tb$$

$$\Rightarrow \quad \hat{x} = (A^TA)^{-1}A^Tb$$

$$\rightarrow \quad \text{so} \quad Pb = A(A^TA)^{-1}A^Tb$$

projects $\underline{b}$ onto col(A)

Things $\perp$ to col(A)?

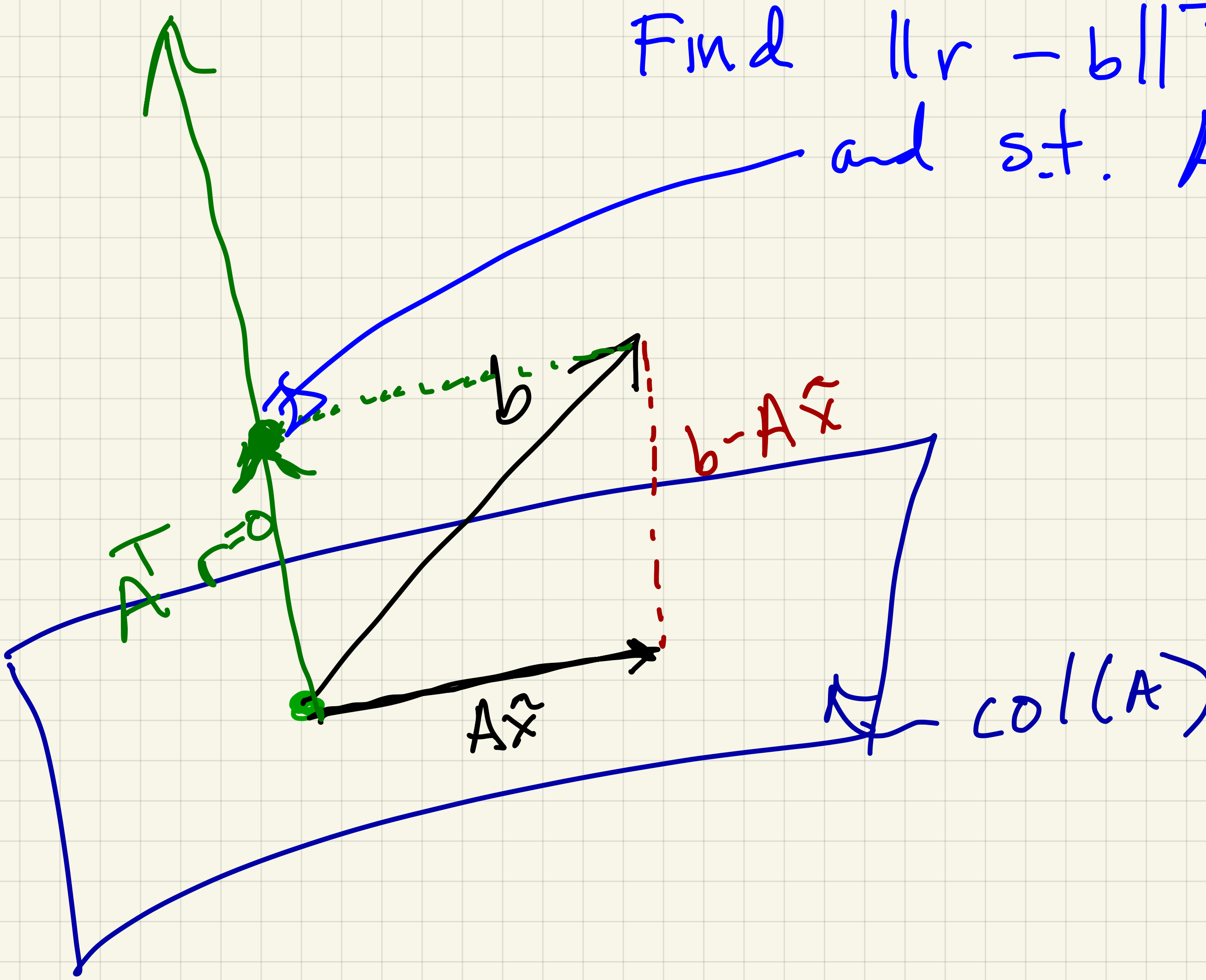
$$\begin{bmatrix} - & a_0 & - \\ - & a_1 & - \\ - & a_2 & - \\ & \vdots & \end{bmatrix} \underline{v} = 0$$

$$\Rightarrow \quad A^T\underline{v} = 0$$

$$\Rightarrow \quad N(A^T)$$

Another way:
project $\underline{b}$ onto $\text{span}(A)^{\perp}$
or $N(A^T)$

Find $\|r - b\|^2$
and s.t. $A^T r = 0$



$$P : b \longrightarrow \mathrm{col}(A)$$

$$P = A(A^TA)^{-1}A^T$$

$$(I-P)^2 = I^2 - 2P + P^2$$

$$= I - 2P + P$$

$$= I - P$$

Want:

$$r = b - A\hat{x}$$

$$A^T r = 0$$

$$\rightarrow \begin{aligned} r + A\hat{x} &= b \\ A^T r &= 0 \end{aligned}$$

$$\rightarrow \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} r \\ \hat{x} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}$$

KKT system

Karush Kuhn Tucker

Example

primal (P) $\frac{1}{2}\|Ax-b\|^2 \Rightarrow \hat{x}$

dual (D) $\frac{1}{2}\|r-b\|^2$ under $A^T r = 0$

$$\boxed{A = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad b = \begin{bmatrix} 3 \\ 4 \end{bmatrix}}$$

(P) $A^T A = [5] \qquad A^T b = [10]$

$$\hat{x} = [2]$$

$$r = \underbrace{\begin{bmatrix} 3 \\ 4 \end{bmatrix}}_{b} - \underbrace{\begin{bmatrix} 4 \\ 2 \end{bmatrix}}_{A\hat{x}} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

$$r \perp A\hat{x}$$

(D) $L = \frac{1}{2}\|r - b\|^2 + \lambda(A^T r)$

$\uparrow$ Lagrange mult.

Set $\nabla_{r_1, r_2, \lambda} L = 0$

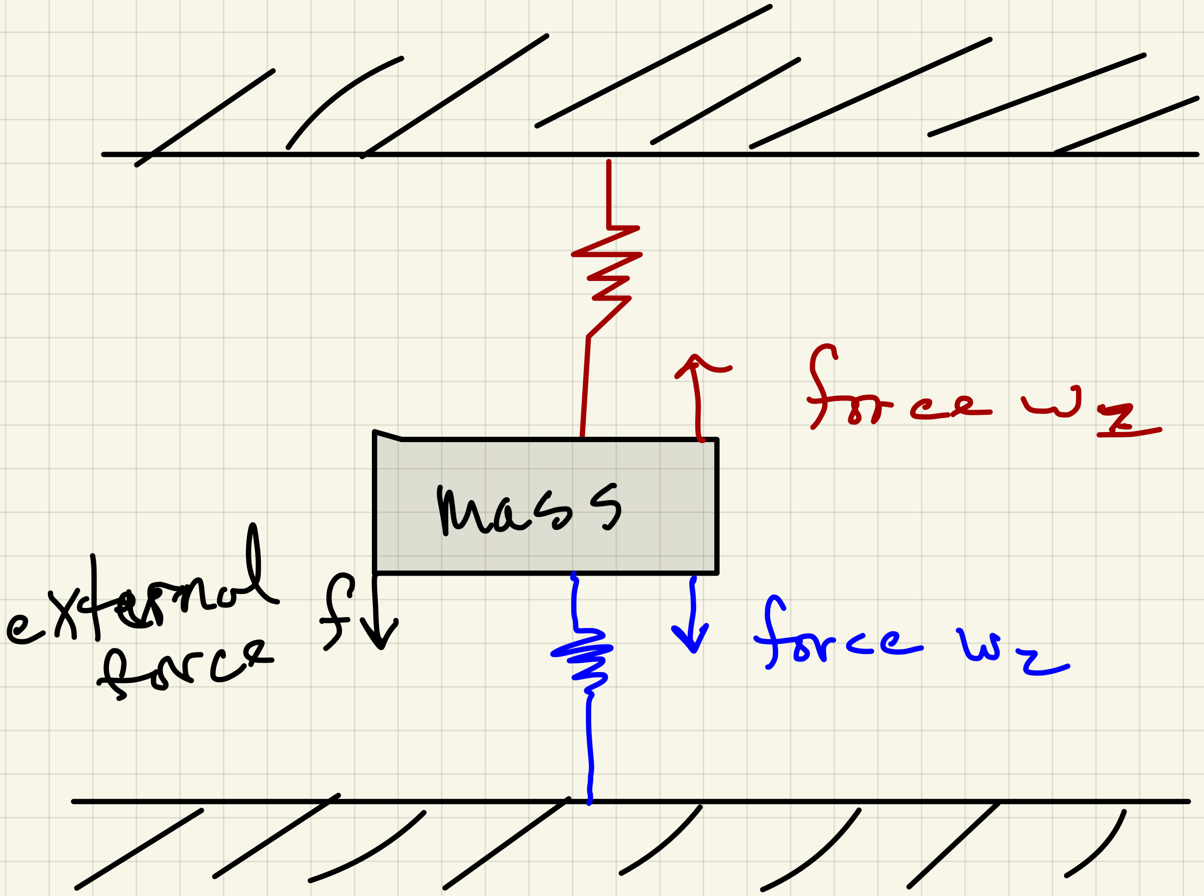
$$L = \frac{1}{2}(r_1 - 3)^2 + \frac{1}{2}(r_2 - 4)^2 + \lambda(2r_1 + r_2)$$

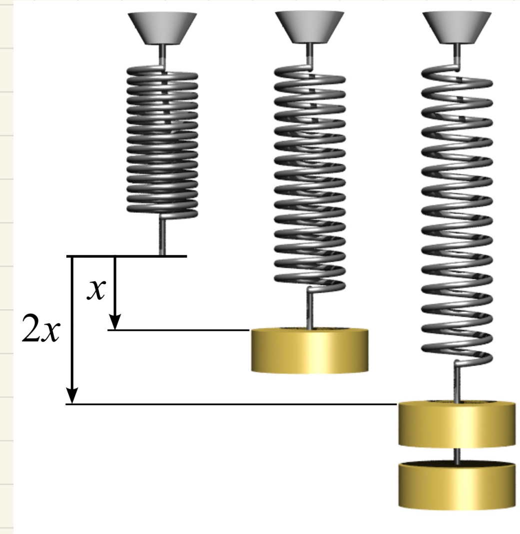
$$\frac{\partial L}{\partial r_1} = r_1 - 3 + 2\lambda$$

$$\frac{\partial L}{\partial r_2} = r_2 - 4 + \lambda \Rightarrow \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$$

$$\frac{\partial L}{\partial \lambda} = 2r_1 + r_2$$

Think about Energy





Hook's law: $w = c \cdot \underset{\uparrow \text{length}}{\ell}$

$$\text{Energy} = \frac{1}{2} c \cdot \ell^2 = \frac{1}{2} \frac{w^2}{c}$$

Goal: minimize

$$E(w) = \frac{1}{2}\frac{w_1^2}{c_1^2} + \frac{1}{2}\frac{w_2^2}{c_2^2}$$

with $w_1 - w_2 = f.$

$$L = \frac{1}{2}\frac{w_1^2}{c_1} + \frac{1}{2}\frac{w_2^2}{c_2} - \lambda(w_1 - w_2 - f)$$

$$\frac{\partial L}{\partial w_1} = \frac{w_1}{c_1} - \lambda$$

$$\frac{\partial L}{\partial w_2} = \frac{w_2}{c_2} - \lambda$$

$$\frac{\partial L}{\partial \lambda} = -w_1 + w_2 + f$$

$c_1 = c_2 = 1$

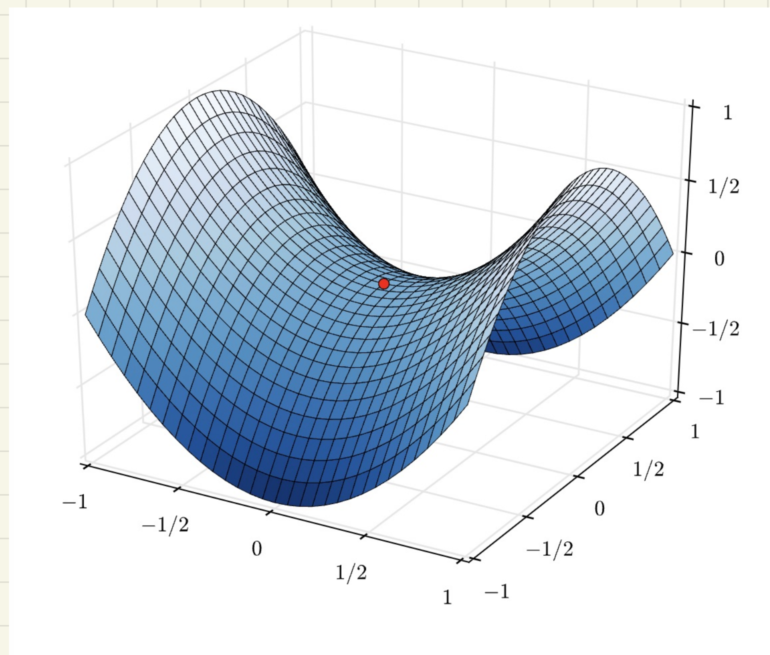
$$\rightarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ u \end{bmatrix}$$

Notize

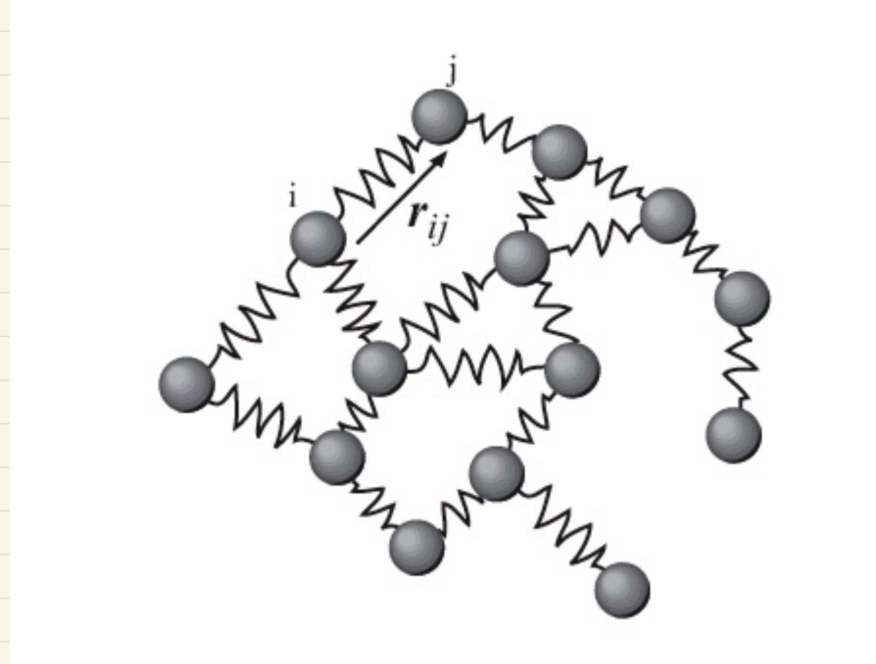
$$\frac{1}{2} w_1^2 + \frac{1}{2} w_2^2 - \lambda w_1 + \lambda w_2$$

$$= \frac{1}{2}\left(\underbrace{(w_1 - \lambda)^2 + (w_2 + \lambda)^2}_{\text{concave up}} \underbrace{- 2\lambda^2}_{\text{down}} \right)$$

$\rightarrow$ saddle



Generalize



n springs

m constraints

$$\underline{w} = (w_1 \; \cdots \; w_n)$$

$$A^T \underline{w} = \underline{f}_{m\times 1}$$

minimize $E = \frac{1}{2} w^T C^{-1} \underline{w}$ $\quad C^{-1} = \begin{bmatrix} \frac{1}{c_1} & \\ & \frac{1}{c_n} \end{bmatrix}$

$$\to L(\underline{w}, \underline{\lambda}) = \frac{1}{2} w^T C^{-1} \underline{w} - \underline{\lambda}^T (A^T \underline{w} - \underline{f})$$

$$\frac{\partial L}{\partial \underline{w}} = C^{-1} \underline{W} - A \underline{\lambda} = 0$$

$$\frac{\partial L}{\partial \underline{\lambda}} = -A^T w + f = 0$$

$$\begin{bmatrix} C^{-1} & -A \\ -A^T & 0 \end{bmatrix}$$

How do we solve

sign trick!

$$\begin{bmatrix} C^{-1} & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} w \\ \lambda \end{bmatrix} = \begin{bmatrix} b \\ f \end{bmatrix}$$

$n+m \times n+m$

Suppose ① C is positive definite.

② A is full rank.

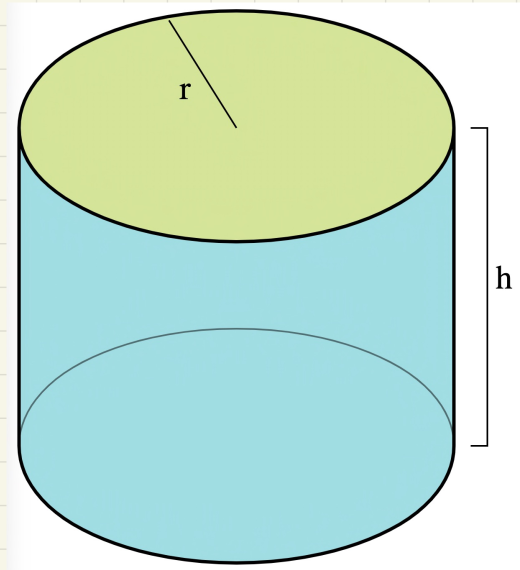
$\hookrightarrow$ $A^T C A$ is invertible

$$= \begin{bmatrix} I & 0 \\ A^T C & I \end{bmatrix} \begin{bmatrix} C^{-1} & 0 \\ 0 & -A^T C A \end{bmatrix} \begin{bmatrix} I & CA \\ 0 & I \end{bmatrix}$$

Example

$n=2 \quad m=1$

Minimize surface area of a cylinder with volume $V=1000$



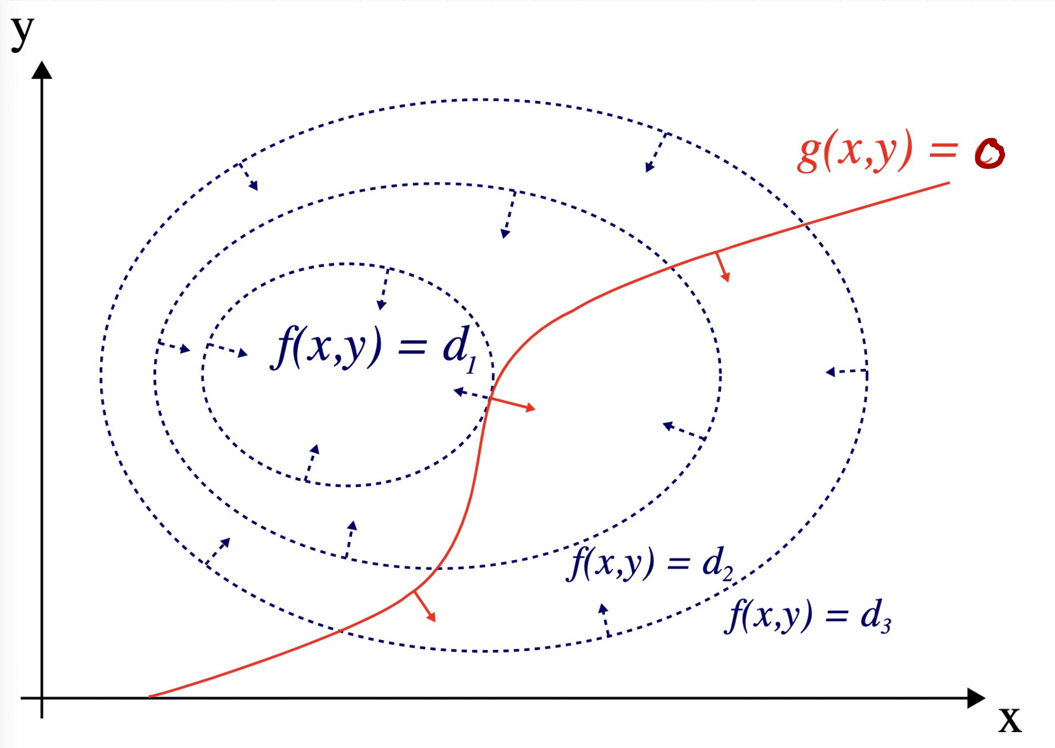
$$f(r,h) = 2\pi rh + 2\pi r^2$$

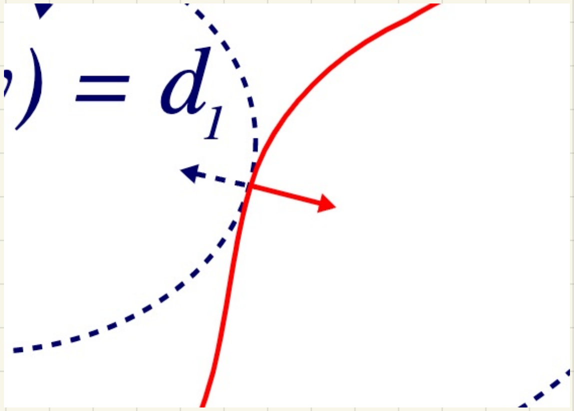
$$g(r,h) = \pi r^2 h - V = 0$$

Intuition

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

View: walk along $g(x,y) = 0$, where is $f(x,y)$ minimized?





$$\nabla f = \lambda \nabla g$$

Ok, so $\nabla f = \lambda \nabla g$ for some λ

and $g = 0$ $\rightarrow$ $\nabla_x (f + \lambda g) = 0$

$$\rightarrow \mathcal{L}(x, y, \lambda) = f(x,y) \pm \lambda g(x,y)$$

Constrained Optimization: Problem Setup

Want $\boldsymbol{x}^*$ so that

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) = 0$$

No inequality constraints just yet. This is *equality-constrained optimization*. Develop a (local) necessary condition for a minimum.

Necessary cond.: "no feasible descent possible". Assume $\boldsymbol{g}(\boldsymbol{x}) = 0$.

Recall unconstrained necessary condition, "no descent possible":

$$\nabla f(\boldsymbol{x}) = 0$$

Look for feasible descent directions from $\boldsymbol{x}$. (Necessary cond.: $\nexists$)

$\boldsymbol{s}$ is a feasible direction at $\boldsymbol{x}$ if

$$\boldsymbol{x} + \alpha \boldsymbol{s} \quad \text{feasible for } \alpha \in [0, r] \quad (\text{for some } r)$$

Constrained Optimization: Necessary Condition

Need: $\nabla f(\boldsymbol{x}) \cdot \boldsymbol{s} \geqslant 0$ ("uphill that way") for any feasible direction $\boldsymbol{s}$.

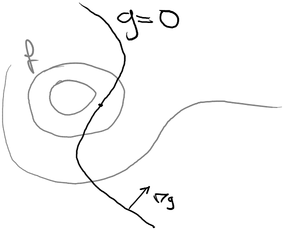
- Not at boundary: $\boldsymbol{s}$ and $-\boldsymbol{s}$ are feasible directions
  $\Rightarrow \nabla f(\boldsymbol{x}) = 0$
  $\Rightarrow$ Only the boundary of the feasible set is different from the unconstrained case (i.e. interesting)
- At boundary: (the common case) $\boldsymbol{g}(x) = 0$. Need:

$$-\nabla f(\boldsymbol{x}) \in \operatorname{rowspan}(J_{\boldsymbol{g}})$$

  a.k.a. "all descent directions would cause a change ($\rightarrow$violation) of the constraints."
  Q: Why 'rowspan'? Think about shape of $J_{\boldsymbol{g}}$.

$$\Leftrightarrow -\nabla f(\boldsymbol{x}) = J_{\boldsymbol{g}}^T \boldsymbol{\lambda} \quad \text{for some } \boldsymbol{\lambda}.$$

Lagrange Multipliers



Seen: Need $-\nabla f(\boldsymbol{x}) = J_{\boldsymbol{g}}^T \boldsymbol{\lambda}$ at the (constrained) optimum.

Idea: Turn constrained optimization problem for $\boldsymbol{x}$ into an *unconstrained* optimization problem for $(\boldsymbol{x}, \boldsymbol{\lambda})$. How?

Need a new function $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ to minimize:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{g}(\boldsymbol{x}).$$

Lagrange Multipliers: Development

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{g}(\boldsymbol{x}).$$

Then: $\nabla \mathcal{L} = 0$ at unconstrained minimum, i.e.

$$0 = \nabla \mathcal{L} = \begin{bmatrix} \nabla_{\boldsymbol{x}} \mathcal{L} \\ \nabla_{\boldsymbol{\lambda}} \mathcal{L} \end{bmatrix} = \begin{bmatrix} \nabla f + J_{\boldsymbol{g}}(\boldsymbol{x})^T \boldsymbol{\lambda} \\ \boldsymbol{g}(\boldsymbol{x}) \end{bmatrix}.$$

Convenient: This matches our necessary condition!

So we could use any unconstrained method to minimized $\mathcal{L}$.
For example: Using Newton to minimize $\mathcal{L}$ is called *Sequential Quadratic Programming*. ('SQP')

Demo: Sequential Quadratic Programming [cleared]

$$\mathcal{L}(x, y, \lambda) = f(\underline{x}) + \lambda g(\underline{x})$$

$$\mathcal{L}(r, h, \lambda) = 2\pi r^2 + 2\pi r h$$

$$= \mathcal{L}(x_1, x_2, \lambda) = 2\pi x_1^2 + 2\pi x_1 x_2 + \lambda\left(\pi x_1^2 x_2 - V\right)$$

$$\nabla \mathcal{L} = \begin{bmatrix} 4\pi x_1 + 2\pi x_2 + 2\pi \lambda x_1 x_2 \\ 2\pi x_1 + \pi \lambda x_1^2 \\ \pi x_1^2 x_2 - V \end{bmatrix}$$

$$H_f = \ldots$$

$$\boxed{\begin{bmatrix} H_f + \sum \lambda_i H_{g_i}(x) & J_g^T \\ J_g & 0 \end{bmatrix}}$$

$$H_f = \begin{bmatrix} 4\pi & 2\pi \\ 2\pi & 0 \end{bmatrix} \quad H_g = \begin{bmatrix} 2\pi x_2 & 2\pi x_1 \\ 2\pi x_1 & 0 \end{bmatrix}$$

$$J_g = \begin{bmatrix} 2\pi x_1 x_2 & \pi x_1^2 \end{bmatrix}$$

$$g(r,h) = \pi r^2 h - V$$

$$\mathcal{L} = \mathcal{L}(\underline{x}, \underline{\lambda})$$

$$= f(\underline{x}) + \underline{\lambda}^T \underline{g}(\underline{x})$$

$$\nabla \mathcal{L} = \begin{bmatrix} \nabla_{\underline{x}} \mathcal{L} \\ \nabla_{\underline{\lambda}} \mathcal{L} \end{bmatrix} = \begin{bmatrix} \nabla_{\underline{x}} f + J_g^T \underline{\lambda} \\ \underline{g}(\underline{x}) \end{bmatrix}$$

$$H_{\mathcal{L}} = \begin{bmatrix} \nabla_{xx} \mathcal{L} & J_g^T(\underline{x}) \\ J_g & 0 \end{bmatrix}$$

$$= \begin{bmatrix} H_f + \sum \lambda_i H_{g_i}(x) & J_g^T \\ J_g & 0 \end{bmatrix}$$

Inequality-Constrained Optimization

Want $\boldsymbol{x}^*$ so that

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) = 0 \quad \text{and} \quad \boldsymbol{h}(\boldsymbol{x}) \leq 0.$$

Develop a necessary condition for a minimum.

Again: Assume we're at a feasible point, on the boundary of the feasible region. Must ensure descent directions are *infeasible*.

Motivation: $\boldsymbol{g} = 0 \Leftrightarrow$ two inequality constraints: $\boldsymbol{g} \leq 0 \wedge \boldsymbol{g} \geq 0$.

Consider the condition $-\nabla f(\boldsymbol{x}) = J_{\boldsymbol{h}}^T \boldsymbol{\lambda}_2$.

- Descent direction must start violating constraint.
  But only one direction is dangerous here!
- $-\nabla f$: descent direction of f, ∇h_i: ascent direction of h_i
- If we assume $\boldsymbol{\lambda}_2 > 0$, going towards $-\nabla f$ would increase $\boldsymbol{h}$ (and start violating $\boldsymbol{h} \leq 0$)

Lagrangian, Active/Inactive

Put together the overall Lagrangian.

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) := f(\boldsymbol{x}) + \boldsymbol{\lambda}_1^T \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{\lambda}_2^T \boldsymbol{h}(\boldsymbol{x})$$

What are active and inactive constraints?

- Active: active $\Leftrightarrow h_i(\boldsymbol{x}^*) = 0 \Leftrightarrow$ at 'boundary' of ineq. constraint
  (Equality constrains are always 'active')
- Inactive: If h_i inactive ($h_i(\boldsymbol{x}^*) < 0$), must force $\lambda_{2,i} = 0$. Otherwise: Behavior of h could change location of minimum of $\mathcal{L}$. Use complementarity condition $h_i(\boldsymbol{x}^*)\lambda_{2,i} = 0$.
  $\Leftrightarrow$ *at least one* of $h_i(\boldsymbol{x}^*)$ and $\lambda_{2,i}$ is zero.

Karush-Kuhn-Tucker (KKT) Conditions

Develop a set of necessary conditions for a minimum.

Assuming $J_{\boldsymbol{g}}$ and $J_{\boldsymbol{h},\text{active}}$ have full rank, this set of conditions is *necessary*:

$$
\begin{aligned}
(*) \quad \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}_1^*, \boldsymbol{\lambda}_2^*) &= 0 \\
(*) \quad \boldsymbol{g}(\boldsymbol{x}^*) &= 0 \\
\boldsymbol{h}(\boldsymbol{x}^*) &\leq 0 \\
\boldsymbol{\lambda}_2 &\geqslant 0 \\
(*) \quad \boldsymbol{h}(\boldsymbol{x}^*) \cdot \boldsymbol{\lambda}_2 &= 0
\end{aligned}
$$

These are called the Karush-Kuhn-Tucker ('KKT') conditions.

Computational approach: Solve $(*)$ equations by Newton.

Karush-Kuhn-Tucker (KKT) Conditions

Develop a set of necessary conditions for a minimum.

Assuming $J_{\boldsymbol{g}}$ and $J_{\boldsymbol{h},\text{active}}$ have full rank, this set of conditions is *necessary*:

$$\begin{aligned}
(*) \quad \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}_1^*, \boldsymbol{\lambda}_2^*) &= 0 \\
(*) \quad \boldsymbol{g}(\boldsymbol{x}^*) &= 0 \\
\boldsymbol{h}(\boldsymbol{x}^*) &\leq 0 \\
\boldsymbol{\lambda}_2 &\geqslant 0 \\
(*) \quad \boldsymbol{h}(\boldsymbol{x}^*) \cdot \boldsymbol{\lambda}_2 &= 0
\end{aligned}$$

These are called the Karush-Kuhn-Tucker (‘KKT’) conditions.

Computational approach: Solve $(*)$ equations by Newton.