Lecture 4   2025-09-04

Last time: smooth unconstrained optimization

Today: smooth optimization with linear equality constraints

$$\min_{x} f(x) \quad f \in \mathcal{C}^2$$

subject to: $Ax = b$ $\quad A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$

(so m-equality constraints)

How to interpret Ax?

$$A = \begin{pmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{pmatrix}$$ where $a_i \in \mathbb{R}^m$ are columns of A

Ax $\begin{pmatrix} | & | & & | \\ a & & & \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \end{pmatrix}$ $\sum^{n}$

$$Ax = \begin{pmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} x_i a_i$$

so Ax is a linear combination of the columns of A

Define range and null space:

$\text{range}(A) = \{ Ax \mid x \in \mathbb{R}^n \} \subseteq \mathbb{R}^m$

$\rightarrow$ When is $\text{range}(A) = \mathbb{R}^m$? $\rightarrow$ when A is full rank

$\text{rank } A = \dim \text{range } A = \min(m,n)$

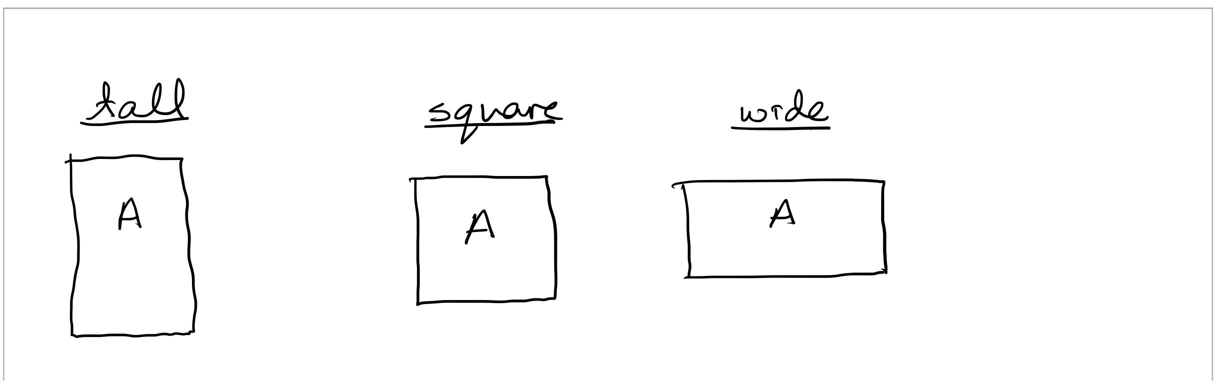
\# of linearly independent columns of A

null space: $\mathcal{N}(A) = \{ x \mid x \in \mathbb{R}^n \text{ and } Ax = 0 \}$

Rank-nullity theorem:

$\text{rank } A + \dim \mathcal{N}(A) = n$

Why does this matter? consider three cases of full rank matrices



$m \geq n$	$m = n$	$m \leq n$
rank $A = n$	rank $A = n$	rank $A = m$
$\dim \mathcal{N}(A) = 0$	$\dim \mathcal{N}(A) = 0$	$\dim \mathcal{N}(A) = n - m$

e.g., suppose A is wide ($m \leq n$) and full rank,

then

$\dim \mathcal{N}(A) = n - m$

so $\exists \; x_2 \in \mathcal{N}(A)$ such that $Ax_2 = 0$

then if we have some x_1 such that $Ax_1 = b$

then note that $A(x_1 + x_2) = Ax_1 + \cancel{Ax_2}^{=0} = Ax_1 = b$

$\rightarrow$ in this case, there are many solutions to $Ax = b$

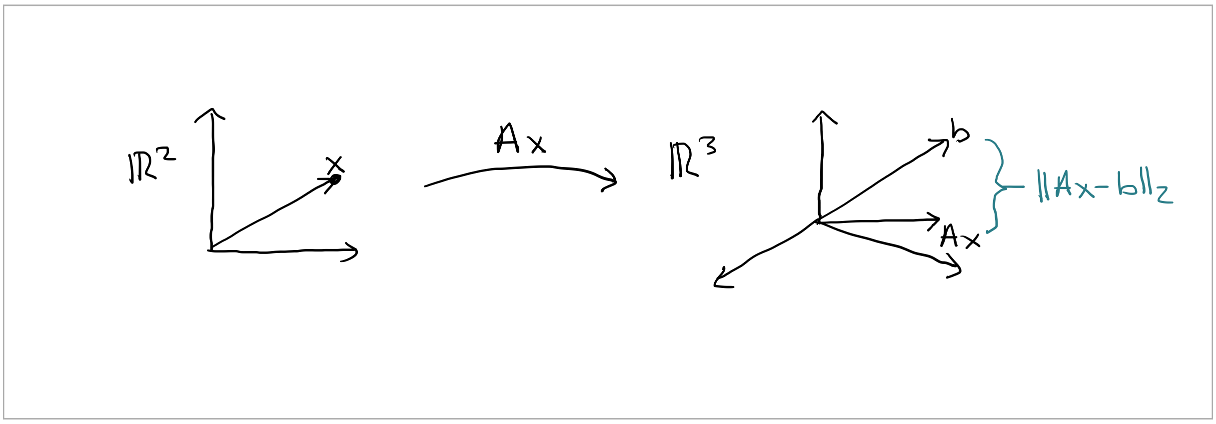
Now consider optimization problem for three cases of $Ax=b$ linear equality constraints:

Case 1: A is square and full rank

$$\min_{x} f(x) \quad \text{subj. to: } Ax=b \quad \rightarrow \quad x^* = A^{-1}b$$

Case 2: A is tall and full rank

e. g. $A \in \mathbb{R}^{3\times 2}$ (i.e. maps from 2D → 3D)



Don't necessarily have an x such that Ax=b

→ minimize squared residual $\|Ax-b\|_2^2$

→ $f(x) = \frac{1}{2}\|Ax-b\|_2^2$

$$= \frac{1}{2}(Ax-b)^T(Ax-b)$$

$$= \frac{1}{2}(x^TA^TAx - b^TAx - x^TAb + b^Tb) \quad b^TAx = x^TAb$$

$$= \frac{1}{2}(x^TA^TAx - 2(A^Tb)^Tx + b^Tb)$$

$$\nabla_x f(x) = \frac{1}{2}(2A^TAx - 2A^Tb)$$

$$= A^TAx - A^Tb := 0$$

$\rightarrow$ $\boxed{x^* = (A^TA)^{-1}A^Tb}$ Linear Least Squares

e.g., applications of LLS include signal reconstruction

e.g., when the residual is of form $f(x) = \frac{1}{2}\|r(x)\|_2^2$ where $r(x)$ is a nonlinear function of x, least squares can iteratively be applied

$\rightarrow$ nonlinear least squares (NLLS) is commonly used in SLAM with the Gauss-Newton method

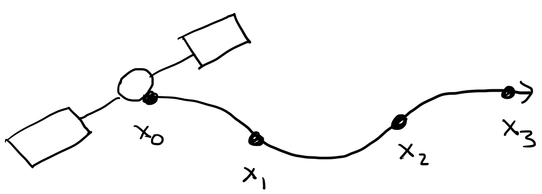
Case 3: A is wide and full rank

$\boxed{\quad A \quad}$

$A \in \mathbb{R}^{m \times n}$ $\qquad$ $\text{rank}(A) = m$

$m \leq n$ $\qquad$ $\dim \mathcal{N}(A) = n - m$

e.g., consider the dynamics constraint matrix for a 2D spacecraft double integrator



$$A = \begin{pmatrix} I & \Delta h \\ 0 & I \end{pmatrix}$$

$$B = \begin{pmatrix} \frac{1}{2}\Delta h^2 \\ \Delta h \end{pmatrix}$$

$x_0 = x_{init}$ $\quad \rightarrow \quad$ $x_0 = x_{init}$

$$x_1 = Ax_0 + Bu_0 \rightarrow Ax_0 + Bu_0 - x_1 = 0$$

$$x_2 = Ax_1 + Bu_1 \rightarrow Ax_1 + Bu_1 - x_2 = 0$$

$$\vdots$$

$$x_N = Ax_{N-1} + Bu_{N-1} \rightarrow Ax_{N-1} + Bu_{N-1} - x_N = 0$$

$$\Rightarrow \underbrace{\begin{pmatrix} I & 0 & 0 & \cdots & \cdots & 0 \\ \hline A & B & -I & 0 & \cdots & 0 \\ 0 & 0 & A & B & -I \cdots & 0 \\ & & \vdots & & & \\ 0 & 0 & \cdots & A & B & -I \end{pmatrix}}_{\bar{A}} \underbrace{\begin{pmatrix} x_0 \\ u_0 \\ x_1 \\ u_0 \\ \vdots \\ x_{N-1} \\ u_{N-1} \\ x_N \end{pmatrix}}_{\bar{x}} = \underbrace{\begin{pmatrix} x_{init} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\bar{b}}$$

$$\rightarrow \bar{A}\bar{x} = \bar{b} \quad \text{where} \quad \begin{aligned} &\bar{x} \in \mathbb{R}^{(N+1)n_x + Nn_u} \\ &\bar{b} \in \mathbb{R}^{(N+1)n_x} \\ &\bar{A} \in \mathbb{R}^{(N+1)n_x \times (N+1)n_x + Nn_u} \end{aligned}$$

OK, A is full rank and wide.

What are necessary conditions of optimality (N.C.O.)

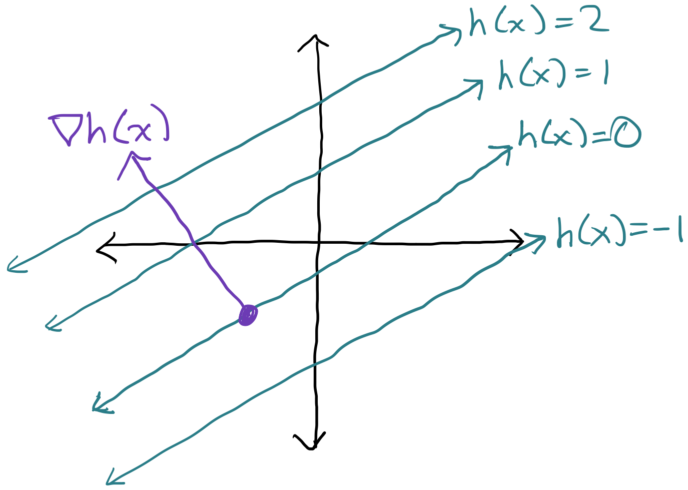
Last time: for unconstrained optimization with $f \in \mathcal{C}^2$,

$\rightarrow$ N.C.O: $\nabla f(x^*) = 0$

suppose we have linear constraints $h(x) = c^T x - b = 0 \in \mathbb{R}$

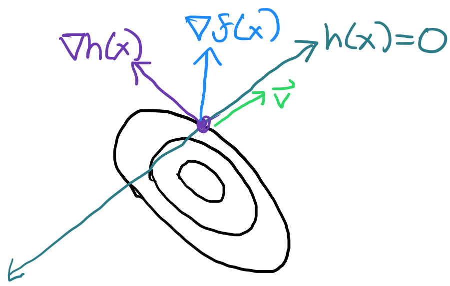
consider level surfaces of $h(x)$:

$h(x) = c^T x - b$



The gradient $\nabla h(x)$ points in the direction of greatest increase

What's the connection between $\nabla f(x)$ and $\nabla h(x)$ at a local optimizer x^*?

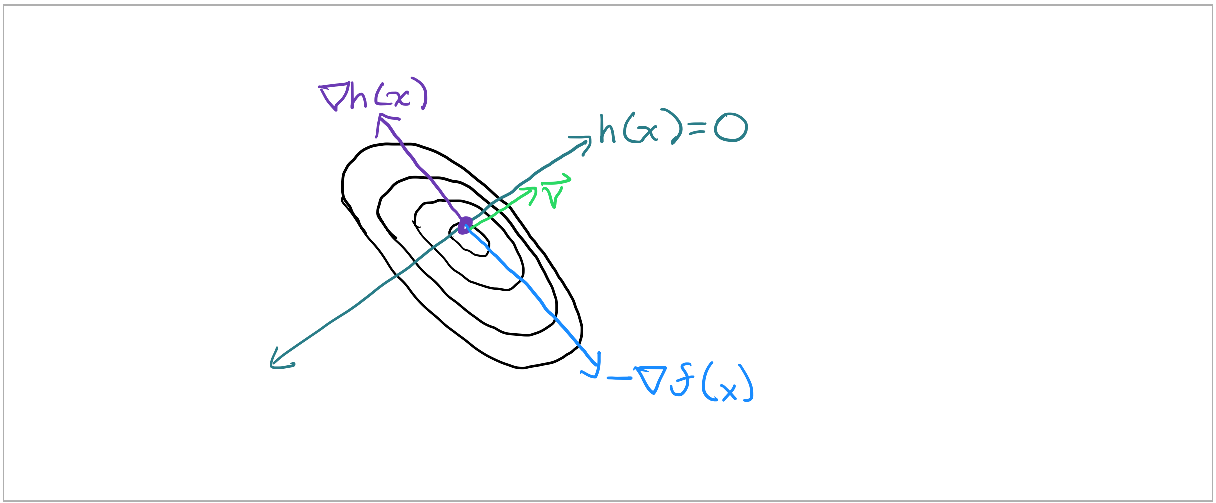


To remain along the constraint set $h(x)=0$, we cannot move along $\nabla h(x)$, which necessarily moves along direction of greatest increase, and can only move along the tangent direction $\vec{v}$ to $h(x)$.

Thus, given some iterate $x^{(k)}$, consider expressing the descent direction $d^{(k)} = -\nabla_x f(x^{(k)})$ along:

$$d^{(k)} = \gamma \vec{v} + \lambda \nabla h(x) \quad \text{for } \gamma \in \mathbb{R} \ \& \ \lambda \in \mathbb{R}$$

at the optimal solution x^*, we cannot possibly have a descent direction along $\vec{v}$



Thus, if x^* is a local optimizer,

$$\nabla f(x^*) = \lambda \nabla h(x^*)$$

$$\rightarrow \quad \boxed{\nabla f(x^*) + \lambda \nabla h(x^*) = 0}$$

where the sign on λ has been swapped for convenience

if there are m equality constraints, then this condition holds as:

$$\boxed{\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h(x^*) = 0}$$

Suppose we integrate this expression w.r.t. x;

define the Lagrangian:

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

$$= f(x) + \lambda^T h(x)$$

we can now succintly rewrite the N.C.O.:

$$\nabla_x \mathcal{L}(x^*) = 0$$