

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

(so m -equality constraints)

How to interpret Ax ?

$$A = \begin{pmatrix} | & | & | \\ a_1 & a_2 & \dots & a_n \\ | & | & | \end{pmatrix} \quad \text{where } a_i \in \mathbb{R}^m \text{ are columns of } A$$
$$Ax = \begin{pmatrix} | & | & | \\ a_1 & a_2 & \dots & a_n \\ | & | & | \end{pmatrix} \begin{pmatrix} | \\ | \\ | \end{pmatrix} x_i \quad \leq n$$

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

so Ax is a linear combination of the columns of A

Define range and null space:

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

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

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

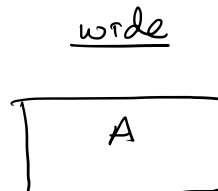
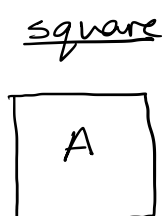
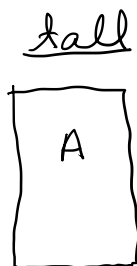
of linearly independent columns of A

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

Rank-nullity theorem:

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

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



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

e. g. suppose A wide 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 $Ax = b$

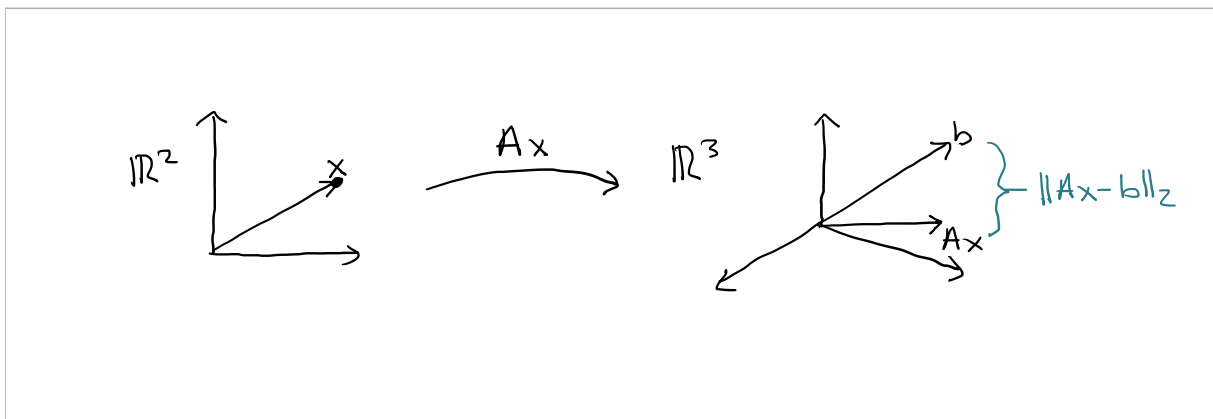
Now consider optimization problem for three cases
of $Ax = b$ linear equality constraints:

Case 1: A is square and full rank

$$\begin{array}{ll} \min_x f(x) & \rightarrow x^* = A^{-1}b \\ \text{subj. to: } Ax=b \end{array}$$

Case 2: A is tall and full rank

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



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

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

$$\rightarrow f(x) = \frac{1}{2} \|Ax - b\|_2^2$$

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

$$\begin{aligned}
&= \frac{1}{2} (x^T A^T A x - b^T A x - x^T A b + b^T b) \quad b^T A x = x^T A b \\
&= \frac{1}{2} (x^T A^T A x - 2(A^T b)^T x + b^T b)
\end{aligned}$$

$$\begin{aligned}
\nabla_x f(x) &= \frac{1}{2} (2 A^T A x - 2 A^T b) \\
&= A^T A x - A^T b := 0
\end{aligned}$$

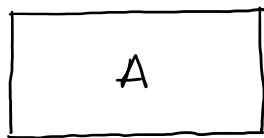
$$\rightarrow \boxed{x^* = (A^T A)^{-1} A^T b} \quad \text{Linear Least Squares}$$

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



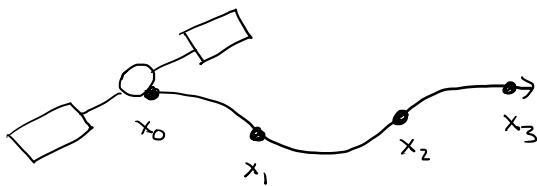
$$A \in \mathbb{R}^{m \times n}$$

$$m \leq n$$

$$\text{rank}(A) = m$$

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

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



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

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

$$x_0 = x_{\text{inst}} \rightarrow x_0 = x_{\text{inst}}$$

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

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

⋮

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

$$\Rightarrow \underbrace{\begin{pmatrix} I & 0 & 0 & \dots & 0 \\ A & B & -I & 0 & \dots & 0 \\ 0 & 0 & A & B & -I & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & A & B & -I \end{pmatrix}}_{\bar{A}} \underbrace{\begin{pmatrix} x_0 \\ u_0 \\ x_1 \\ u_1 \\ \vdots \\ x_{N-1} \\ u_{N-1} \\ x_N \end{pmatrix}}_{\bar{x}} = \underbrace{\begin{pmatrix} x_{out} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\bar{b}}$$

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

OK, \bar{A} is full rank and wide.

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

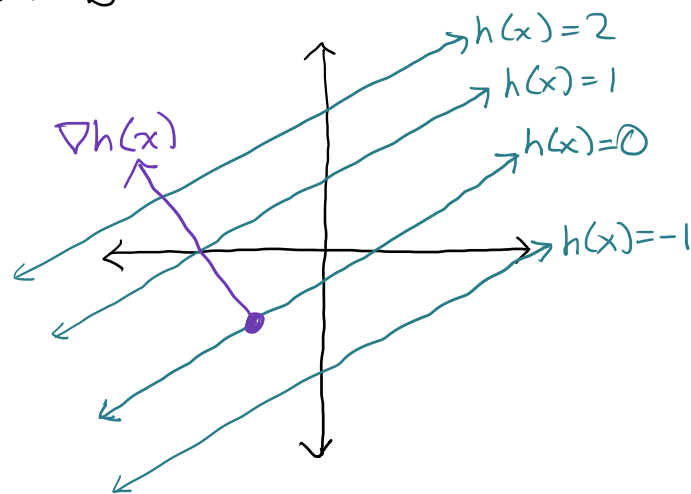
Last time: for unconstrained optimization with $f \in \mathcal{C}^2$,

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

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

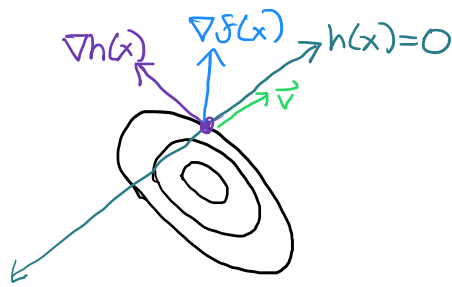
consider level surfaces of $h(x)$:

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



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^* ?

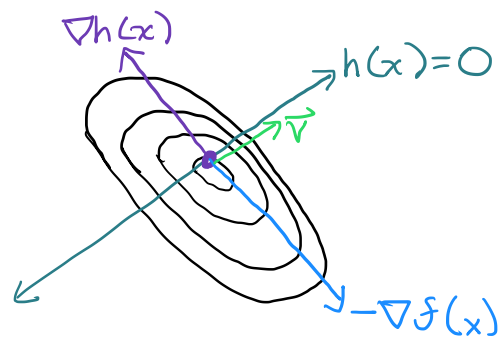


To remain along the constraint set $h(x)=0$, we cannot move along $\nabla h(x)$, which necessarily moves along direction of greatest increase, and can only move along the tangent direction \vec{v} to $h(x)$.

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

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

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



Thus, if x^* is a local optimizer,

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

$$\rightarrow \nabla f(x^*) + \lambda \nabla h(x^*) = 0$$

where the sign on λ has been swapped for convenience

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

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

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 succinctly rewrite the N.C.O.:

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