MAT 3007 Optimization: Tutorial 10

Guxin DU

July 8, 2025

Recap: Optimality Conditions for Unconstrained Problems

Theorem 1 (First-Order Necessary Condition).

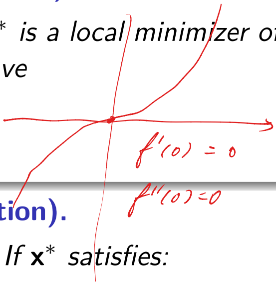
If f is continuously differentiable and $\mathbf{x}^*$ is a local minimizer of $f(\cdot)$ for an unconstrained problem, then we must have $\nabla f(\mathbf{x}^*) = 0$.

Theorem 2 (Second-Order Necessary Condition).

If f is second-order continuously differentiable $\mathbf{x}^*$ is a local minimizer of $f(\cdot)$ for an unconstrained problem, then we must have

1. $\nabla f(\mathbf{x}^*) = 0$;
2. $\nabla^2 f(\mathbf{x}^*)$ *is positive semi-definite.*

$f(x) = x^3$



Theorem 3 (Second-Order Sufficient Condition).

If f is second-order continuously differentiable. If $\mathbf{x}^*$ satisfies:

1. $\nabla f(\mathbf{x}^*) = 0$;
2. $\nabla^2 f(\mathbf{x}^*)$ *is positive definite.*

$\lambda_{min}(\nabla^2 f(x^*)) > 0$

Then $\mathbf{x}^*$ is a local minimizer of f.

Recap: Optimality Conditions for Unconstrained Problems

$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is 2×2 Matrix

$\det A = \lambda_1 \lambda_2 \geq 0$

① $\lambda_1, \lambda_2 \geq 0$

② $\lambda_1, \lambda_2 \leq 0$

$trA = \lambda_1 + \lambda_2$

Tips:

1. Check if a matrix is positive definite/semi-definite; $det(A) = \prod_i \lambda_i; tr(A) = \sum_i \lambda_i$
2. When to check sufficiency/necessity
   - Find candidates for optimal solutions or prove a point x is not a local optimum
   - Prove a point x is a local optimum
3. Understand the insights behind those conditions.

Exercise 1

$$\nabla f = \begin{pmatrix} 4x_1^3 + 6x_1^2 - 4x_1x_2 \\ -2x_1^2 + 8x_2 \end{pmatrix} = 0 \quad \begin{cases} ① & x_1 = x_2 = 0 \\ ② & x_1 = -2, \ x_2 = 1 \end{cases}$$

$$\nabla^2 f = \begin{pmatrix} 12x_1^2 + 12x_1 - 4x_2 & -4x_1 \\ -4x_1 & 8 \end{pmatrix}$$

Find all local minimizer, local maximizer and saddle points of f.

$$f(x) = x_1^4 + 2(x_1 - x_2)x_1^2 + 4x_2^2$$

$$\nabla^2 f(0,0) = \begin{pmatrix} 0, & 0 \\ 0, & 8 \end{pmatrix} \qquad \nabla^2 f(-2,1) = \begin{pmatrix} 20 & 8 \\ 8 & 8 \end{pmatrix} \quad PD$$

$\det = 20 \times 8 - 8 \times 8 = 160 - 64 > 0$

$tr = 28 > 0$

$(-2,1)$ is minimizer

Solution for exercise 1

$$\nabla f(x) = \begin{pmatrix} 4x_1^3 + 6x_1^2 - 4x_1x_2 \\ -2x_1^2 + 8x_2 \end{pmatrix}, \nabla^2 f(x) = \begin{pmatrix} 12x_1^2 + 12x_1 - 4x_2 & -4x_1 \\ -4x_1 & 8 \end{pmatrix}$$

- **Step 1** Calculate all stationary points of f by solving $\nabla f(x) = 0$:
  $\nabla f(x) = 0 \Rightarrow x_1^* = (0,0), x_2^* = (-2,1)$.
- **Step 2** Determine the definiteness of the Hessian $\nabla^2 f(x^*)$ to decide whether the stationary points x^* are local minima, maxima or saddle points.

$$\nabla^2 f(x_1^*) = \begin{pmatrix} 0 & 0 \\ 0 & 8 \end{pmatrix} \nabla^2 f(x_2^*) = \begin{pmatrix} 20 & 8 \\ 8 & 8 \end{pmatrix}$$

$tr(\nabla^2 f(x_2^*)) = 28 > 0$, $det(\nabla^2 f(x_2^*)) = 160 - 64 > 0$, $\nabla^2 f(x_2^*)$ is PD and x_2^* is a local minimum.

$\nabla f(x_1^*)$ has eigen value $0, 8$, is a PSD matrix. So we can not use the second-order optimality conditions to decide whether it is a saddle point or local minimum.
We consider the function f directly around x^*:

$$f(|t|, 0) = t^4 + 2(|t|)t^2 = |t|^3(|t| + 2) > f(0, 0) = 0$$

$$f(-|t|, 0) = t^4 + 2(-|t|)t^2 = |t|^3(|t| - 2) < f(0, 0) = 0$$

So x_1^* is a saddle point.

Recap: Optimality Conditions for Linear Constrained Problems

The linear constrained optimization problem is

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & A\mathbf{x} \geq (=)\ b. \end{aligned}$$

Theorem 4 (Equality Constraint).

If $\mathbf{x}^*$ is the minimizer, then there must exist $\mathbf{y}$ such that $A^\top \mathbf{y} = \nabla f(\mathbf{x}^*)$.

Theorem 5 (Inequality Constraint).

If $\mathbf{x}^*$ is the minimizer, then there must exist $\mathbf{y} \geq 0$ such that

$$\begin{aligned} & A^\top \mathbf{y} = \nabla f(\mathbf{x}^*), \\ & y_i(a_i^\top \mathbf{x}^* - b_i) = 0, \forall i. \end{aligned}$$

where $a_i^\top$ is the ith row of A.

$$\min f(x) \quad \text{st } Ax = b \qquad A = \begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{pmatrix}$$

Orthogonally Complement.

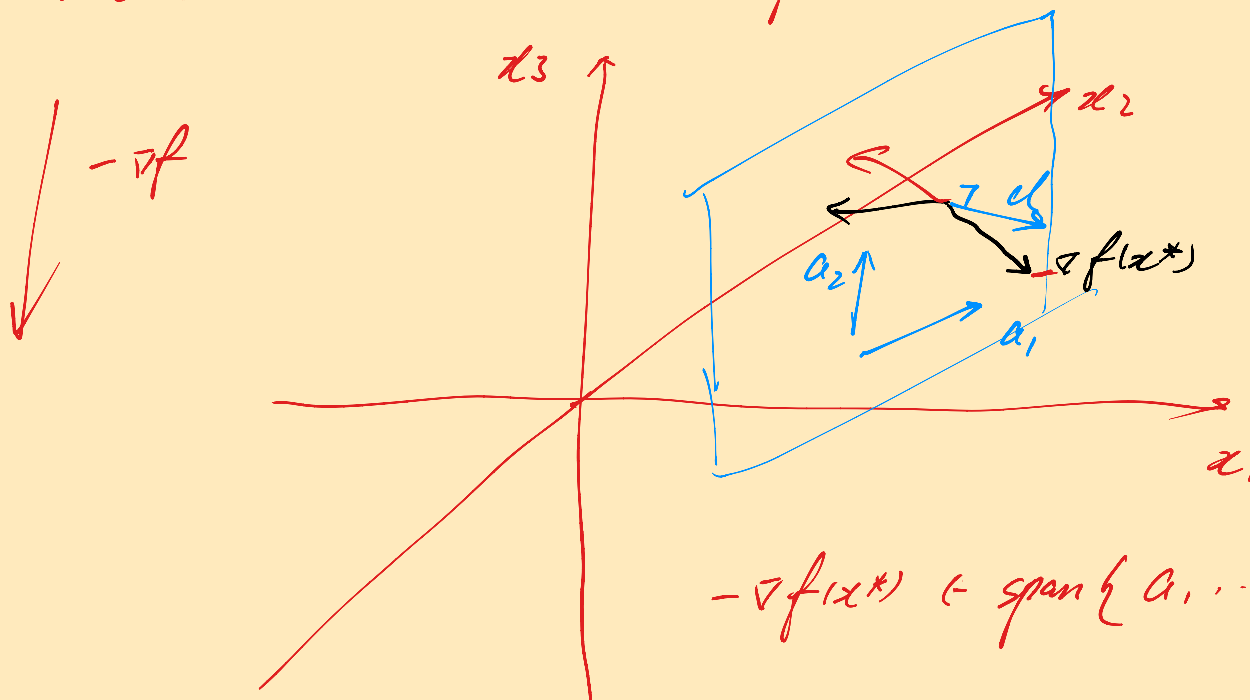
$$S_F := \{ d \in \mathbb{R}^n : Ad = 0 \} \perp \operatorname{span}\{a_1, \dots, a_m\}$$

$\operatorname{span}\{(1,0)\}$ $\quad$ $\operatorname{span}\{(0,1)\}$

$$S_D := \{ d \in \mathbb{R}^n : \nabla f(x^*)^T d < 0 \}$$

If x^* opt $\quad S_F \cap S_D = \phi$ $\quad$ Given x^* optimal.

We want to show that if $d \in S_D \Rightarrow d \perp S_F \overset{\text{Trivial}}{\Longrightarrow} d \in \operatorname{span}\{a_1 \dots a_m\}$



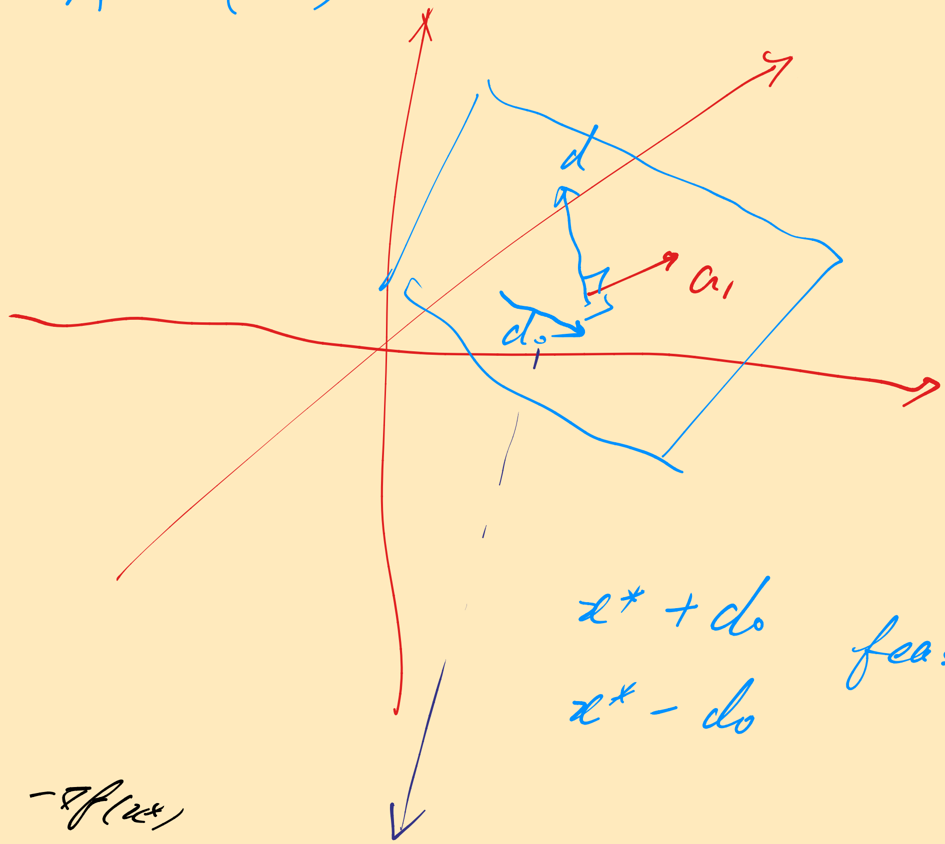
$\forall d \in S_F$

$$-\nabla f(x^*) \in \operatorname{span}\{a_1, \dots, a_m\}$$

$$y_1 a_1 + \dots + y_m a_m = -\nabla f(x^*)$$

$A = (a_1^T)$

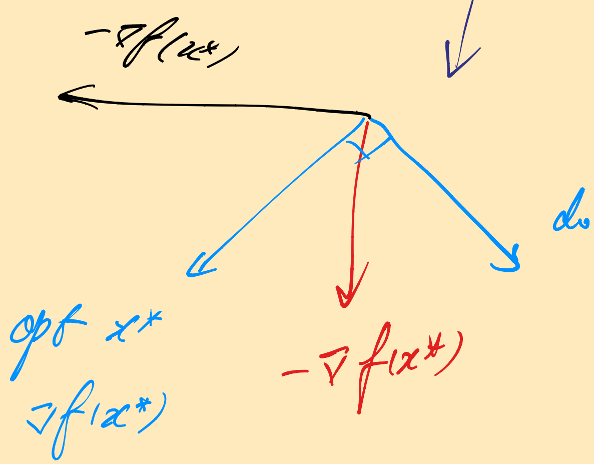
3d



$x^* + d_0$

$x^* - d_0$

feasible



$\nabla f(x^*) = \lambda a_1$

$a \in \mathbb{R}^n$

Tips:

1. Check the sign in the constraint ($=$ or $\geq$);
2. Carefully check the matrix transpose;
3. Understand the derivation of the two theorems:
   - Find the feasible direction set S_F and descent direction set S_D;
   - Write the condition $S_F \bigcap S_D = \emptyset$;
   - Use dual feasibility to construct the alternative systems;

Exercise 2

Find the distance from the origin $(0,0)^\top$ to the polyhedron $S = \{(x_1, x_2)^\top | x_1 + x_2 \geq 4, 2x_1 + x_2 \geq 5\}$.
Note that the problem is equivalent to solve the following problem:

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & x_1 + x_2 \geq 4 \\ & 2x_1 + x_2 \geq 5 \end{aligned}$$

Solution to exercise 2

$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$, $b = (4,5)^\top$, $\nabla f(x) = (2x_1, 2x_2)^\top$. So the solution x^* satisfies there exist $y \geq 0$ such that

$$A^\top y = \nabla f(x^*)$$
$$y_i(a_i^\top x^* - b_i) = 0, i = 1, 2,$$

which are

Step 1. Solve, and obtain some solutions x^1, x^2, ...

$$y_1 + 2y_2 = 2x_1, y_1 + y_2 = 2x_2$$
$$y_1(x_1 + x_2 - 4) = 0, y_2(2x_1 + x_2 - 5) = 0$$
$$y_1, y_2 \geq 0, x_1 + x_2 \geq 4, 2x_1 + x_2 \geq 5$$

Step 2: Check if x^1, x^2 ... feasible or not

Solution to exercise 2

$$
\begin{aligned}
y_1 + 2y_2 = 2x_1, y_1 + y_2 = 2x_2 \\
y_1(x_1 + x_2 - 4) = 0, y_2(2x_1 + x_2 - 5) = 0 \\
y_1, y_2 \geq 0, x_1 + x_2 \geq 4, 2x_1 + x_2 \geq 5
\end{aligned}
$$

From the first two equialities, we have $y_1 = -2x_1 + 4x_2, y_2 = 2x_1 - 2x_2$.

1. If $y_1 = y_2 = 0$, we can conclude that $x_1 = x_2 = 0$, which is infeasible.
2. If $y_1 > 0, y_2 = 0$, we have $x_1 + x_2 = 4$, $2x_1 - 2x_2 = y_2 = 0$. The solution is $x_1 = x_2 = 2$, which is feasible.
3. If $y_1 = 0, y_2 > 0$, we have $-2x_1 + 4x_2 = y_1 = 0$, $2x_1 + x_2 = 5$, whose solution is $x_1 = 2, x_2 = 1$, also infeasible.
4. If $y_1, y_2 > 0$, we have $x_1 + x_2 = 4, 2x_1 + x_2 = 5$ whose solution is $x_1 = 1, x_2 = 3$, while $y_2 = 2x_1 - 2x_2 = -4 < 0$.

So the optimal solution is $x^* = (2, 2)^\top$, and the distance is $2\sqrt{2}$.

Exercise 3

Consider the following problem:

$$\begin{aligned} \min_{x_1, x_2} \quad & (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{s.t.} \quad & 2x_1 + x_2 - 6 \le 0 \\ & x_1 + 2x_2 - 6 \le 0. \end{aligned} \tag{1}$$

- Prove that $x_1 = 11/5$ and $x_2 = 8/5$ is a local minimizer.

Solution for Exercise 3

From the necessary condition, if x_1^*, x_2^* is a local minimum of the problem, then there exists some $\mathbf{y} \geq 0 = [y_1;\ y_2]$ satisfying

$A^T y = \nabla f(x^*)$
$$[2x_1 - 6;\ 2x_2 - 4] = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix} [y_1;\ y_2],$$

$y_i(a_i^T x - b_i) = 0$
$$y_1(2x_1 + x_2 - 6) = 0,$$
$$y_2(x_1 + 2x_2 - 6) = 0.$$

Given that $x_1 = 11/5$ and $x_2 = 8/5$, we have the following linear system on $\mathbf{y}$:

$$[-8/5;\ -4/5] = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix} [y_1;\ y_2],$$
$$-3y_2/5 = 0.$$

Thus $y_1 = 4/5$ and $y_2 = 0$. We obtain the desired result.

Triangle Inequality for p Norm

p norm: For $p \geq 1$, the p-norm of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$ is defined as

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

Particularly, for $p = \infty$, the infinity norm is defined as

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Triangle Inequality: For any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and any $p \geq 1$ (∞ included), the triangle inequality states that

$$\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p.$$

The inequality above is also known as Minkowski's inequality.

Proof of Triangle Inequality for p Norm

For $p = 1$, the triangle inequality follows directly from $|x_i + y_i| \leq |x_i| + |y_i|$ for each $1 \leq i \leq n$.
For $p = \infty$, we have

$$\max_{1 \leq i \leq n} |x_i + y_i| \leq \max_{1 \leq i \leq n} (|x_i| + |y_i|) \leq \max_{1 \leq i \leq n} |x_i| + \max_{1 \leq i \leq n} |y_i|,$$

For $1 < p < \infty$, we first assume that $x \neq 0$ and $y \neq 0$. Otherwise the inequality would be trivial. Then we consider the function $\phi(t) = t^p$ defined on $t \in (0, \infty)$. $\phi(t)$ is convex as $\phi''(t) = p(p-1)t^{p-2} > 0$ for $t > 0$. By the convexity of ϕ, we have for any $s, t > 0$ and $\lambda \in [0, 1]$,

$$(\lambda s + (1-\lambda)t)^p \leq \lambda s^p + (1-\lambda)t^p.$$

Proof of Triangle Inequality for p Norm (continued)

For any $1 \leq i \leq n$, we let $s = \frac{|x_i|}{\|\mathbf{x}\|_p}$, $t = \frac{|y_i|}{\|\mathbf{y}\|_p}$, $\lambda = \frac{\|\mathbf{x}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p}$ and $1 - \lambda = \frac{\|\mathbf{y}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p}$, by the above inequality, we have

$$\left(\frac{|x_i| + |y_i|}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p}\right)^p \leq \frac{\|\mathbf{x}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p}\left(\frac{|x_i|}{\|\mathbf{x}\|_p}\right)^p + \frac{\|\mathbf{y}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p}\left(\frac{|y_i|}{\|\mathbf{y}\|_p}\right)^p.$$

Sum the above inequality over all i from 1 to n, we obtain

$$\frac{\sum_{i=1}^n (|x_i| + |y_i|)^p}{(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)^p} \leq \frac{\|\mathbf{x}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p}\frac{\|\mathbf{x}\|_p^p}{\|\mathbf{x}\|_p^p} + \frac{\|\mathbf{y}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p}\frac{\|\mathbf{y}\|_p^p}{\|\mathbf{y}\|_p^p} = 1.$$

Therefore, we have

$$\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p.$$

Coercive

Definition 6.

A continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be **coercive** if

$$\lim_{\|x\| \to \infty} f(x) = +\infty.$$

i.e. , $\forall B > 0$, $\exists r > 0$ such that if $\|x\| \geq r$ then $f(x) > B$.

Theorem 7.

Let f be a continuous and coercive function. Then for all $\alpha > 0$, the level set

$$L_{\leq \alpha} := \{x : f(x) \leq \alpha\}$$

is compact and f has at least one global minimizer.