Lecture Note 1 - MATH 726

Harry Huang (using ChatGPT)

Oct.2025

1 Generic Optimization Problem

A general problem is

$$\min_{x \in X} f(x),$$

with variable $x \in \mathbb{R}^d$, feasible set $X \subseteq \mathbb{R}^d$, and objective $f : \mathbb{R}^d \to \mathbb{R}$.

2 Norms and Dual Norms

Definition 2.1 (L_p norm)**.** *For $x = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d$,*

$$\|x\|_p = \Big(\sum_{i=1}^d |x_i|^p \Big)^{1/p}.$$

Examples:

$$\|x\|_1 = \sum_{i=1}^d |x_i|, \quad \|x\|_2 = \Big(\sum_{i=1}^d x_i^2 \Big)^{1/2}, \quad \|x\|_\infty = \max_i |x_i|.$$

Definition 2.2 (Equivalent norms)**.** *Two norms $\|\cdot\|_{(1)}$, $\|\cdot\|_{(2)}$ are equivalent if*

$$\exists \lambda_1, \lambda_2 > 0 \quad s.t. \quad \lambda_2 \|x\|_{(2)} \le \|x\|_{(1)} \le \lambda_1 \|x\|_{(2)}, \ \forall x.$$

Examples:

$$\|x\|_1 \le \sqrt{d}\, \|x\|_2, \quad \|x\|_1 \ge \|x\|_2,$$
$$\|x\|_\infty \le \|x\|_2, \quad \|x\|_\infty \ge \frac{1}{\sqrt{d}}\, \|x\|_2.$$

Definition 2.3 (Dual norm)**.** *For $z \in \mathbb{R}^d$,*

$$\|z\|_* := \sup_{\|x\|=1} \langle z, x \rangle.$$

Proposition 2.1 (Hölder inequality)**.** *For all $x, z \in \mathbb{R}^d$,*

$$|\langle z, x \rangle| \le \|z\|_* \|x\|.$$

Proof. If $x = 0$ or $z = 0$, trivial. Else, let $\hat{x} = x/\|x\|$, then $\|\hat{x}\| = 1$ and

$$\|z\|_* = \sup_{\|u\|=1} \langle z, u \rangle \ge \langle z, \hat{x} \rangle = \frac{\langle z, x \rangle}{\|x\|}.$$

So $\langle z, x \rangle \le \|z\|_* \|x\|$. The same with $-\hat{x}$ gives the other direction. □

Example 2.1 (Dual pairs). • $(\|\cdot\|_2)^* = \|\cdot\|_2$,

- $(\|\cdot\|_1)^* = \|\cdot\|_\infty$,
- $(\|\cdot\|_p)^* = \|\cdot\|_q$ *if* $1/p + 1/q = 1$.

3 Feasible Sets and Constraints

- Unconstrained: $X = \mathbb{R}^d$.
- Constrained: $X = \{x : f_i(x) \le 0, h_j(x) = 0\}$.
- Example: ball $X = \{x : \|x\| \le 1\}$.
- Example: linear subspace $\{x : Ax = 0\}$.

4 Existence of Minimizers

Theorem 4.1 (Weierstrass). *If X is compact and f continuous on X, then*

$$\exists x^* \in X : f(x^*) = \min_{x \in X} f(x).$$

Proof. Let $\alpha = \inf_{x \in X} f(x)$. Choose sequence $x_k \in X$ with $f(x_k) \downarrow \alpha$. Compactness $\Rightarrow$ subsequence $x_{k_j} \to x^* \in X$. Continuity $\Rightarrow f(x^*) = \alpha$. □

Definition 4.1 (Coercivity). *f is coercive if $\|x\| \to \infty \Rightarrow f(x) \to +\infty$.*

Theorem 4.2. *If f is coercive, then $\min_{x \in \mathbb{R}^d} f(x)$ exists.*

Proof. For c large, the sublevel set $X_c = \{x : f(x) \le c\}$ is bounded and closed, hence compact. Then $\min f = \min_{x \in X_c} f(x)$, apply Weierstrass. □

Example 4.1 (Quadratic coercive). *If $f(x) = \frac{1}{2} x^\top Q x + r^\top x$ with $Q \succ 0$, then f is coercive.*

5 Lower Semicontinuity (LSC) and Indicators

Definition 5.1 (Indicator).

$$\iota_X(x) = \begin{cases} 0, & x \in X, \\ \infty, & x \notin X. \end{cases}$$

$$\min_{x \in X} f(x) = \min_{x \in \mathbb{R}^d} (f(x) + \iota_X(x)).$$

Definition 5.2 (Lower semicontinuous function). *f is LSC at $\bar{x}$ if*

$$\liminf_{x \to \bar{x}} f(x) \ge f(\bar{x}).$$

Theorem 5.1. *If f is proper, LSC, and coercive, then f attains a minimizer.*

Proof. Let $t > \min f$ and $X_t = \{x : f(x) \le t\}$.

Lemma 1. X_t is compact: nonempty (proper), bounded (coercive), closed (LSC). **Lemma 2.** $\alpha = \inf_{x \in X_t} f(x) > -\infty$ (proper).

Then pick $x_k \in X_t$ with $f(x_k) \to \alpha$. By compactness $x_k \to \bar{x} \in X_t$. LSC gives $f(\bar{x}) \le \alpha$, so $f(\bar{x}) = \alpha$. □

6 Taylor Expansions

Theorem 6.1 (First-order Taylor)**.** *If $f \in C^1$, then*

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle \, dt.$$

Also $\exists t \in [0, 1]$ such that

$$f(y) - f(x) = \langle \nabla f(x + t(y - x)), y - x \rangle.$$

Proof. Let $\phi(t) = f(x + t(y - x))$. Then $\phi'(t) = \langle \nabla f(x + t(y - x)), y - x \rangle$. So $f(y) - f(x) = \phi(1) - \phi(0) = \int_0^1 \phi'(t) \, dt$. By mean value theorem, $\phi(1) - \phi(0) = \phi'(t)$ for some t. □

Theorem 6.2 (Second-order Taylor with remainder)**.** *If $f \in C^2$, then*

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{1}{2}(y - x)^\top \nabla^2 f(x + t(y - x))(y - x),$$

for some $t \in (0, 1)$.

$$\nabla f(y) - \nabla f(x) = \int_0^1 \nabla^2 f(x + t(y - x))(y - x) \, dt.$$

7 Lipschitz Continuity and Smoothness

Definition 7.1 (Lipschitz continuity)**.** *f is M-Lipschitz if*

$$|f(y) - f(x)| \le M \|y - x\|, \quad \forall x, y.$$

Theorem 7.1. *If $f \in C^1$, then f is M-Lipschitz $\iff \|\nabla f(x)\| \le M, \ \forall x$.*

Proof. ($\Rightarrow$) Choose $u = \nabla f(x)/\|\nabla f(x)\|$. For $h > 0$,

$$f(x + hu) - f(x) = \nabla f(x + thu)^\top (hu), \quad t \in [0, 1].$$

Divide by h and let $h \to 0$. ($\Leftarrow$) Use the integral Taylor expansion. □

Definition 7.2 (L-smoothness)**.** *f is L-smooth if*

$$\|\nabla f(y) - \nabla f(x)\| \le L \|y - x\|, \ \forall x, y.$$

Lemma 7.1 (Quadratic bounds)**.** *If f is L-smooth,*

$$f(y) \le f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{L}{2} \|y - x\|^2,$$

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle - \tfrac{L}{2} \|y - x\|^2.$$

Proof. Apply the integral Taylor formula and bound the integral using Lipschitz condition on ∇f. □

Lemma 7.2 (Hessian characterization)**.** *If $f \in C^2$, then f is L-smooth (Euclidean) $\iff -LI \preceq \nabla^2 f(x) \preceq LI$ for all x.*

8 Convexity

Definition 8.1 (Convex function)**.** *f is convex if*

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y).$$

Lemma 8.1 (First-order characterization)**.** *f convex $\iff$ $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$ for all x, y.*

Theorem 8.1 (Second-order characterization)**.** *If $f \in C^2$, then*

$$f \text{ convex } \iff \nabla^2 f(x) \succeq 0.$$

Definition 8.2 (Strong convexity)**.** *f is μ-strongly convex if*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{\mu}{2}\|y - x\|^2.$$

Theorem 8.2. *f μ-strongly convex $\iff$ $f(x) - \frac{\mu}{2}\|x\|^2$ is convex.*

Definition 8.3 (Weak convexity)**.** *f is ρ-weakly convex if*

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y) + \rho\alpha(1-\alpha)\|y - x\|^2.$$

9 Optimality Conditions

Definition 9.1 (Stationary point)**.** *$x \in X$ stationary if*

$$\langle \nabla f(x), y - x \rangle \geq 0, \quad \forall y \in X.$$

Theorem 9.1 (First-order necessary condition)**.** *If $f \in C^1$, X closed convex, and x^* is a local minimizer, then*

$$\langle \nabla f(x^*), y - x^* \rangle \geq 0, \quad \forall y \in X.$$

Theorem 9.2 (Second-order conditions, unconstrained)**.** *If $f \in C^2$:*

- *If x^* local min, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succeq 0$.*
- *If $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succ 0$, then x^* is strict local min.*

Theorem 9.3 (Convex problems)**.** *If f convex, X closed convex:*

1. *Any local minimizer is global.*
2. *Solution set is convex.*
3. *If f differentiable: x^* optimal $\iff$ stationary.*

Theorem 9.4 (Strong convexity)**.** *If f is μ-strongly convex and continuous on X, then minimizer exists and is unique.*

10 Error Bounds and Sensitivity

Theorem 10.1 (Error bound under strong convexity)**.** *If f is μ-strongly convex and C^1,*

$$\|x - x^*\| \leq \frac{1}{\mu}\|\nabla f(x)\|.$$

Definition 10.1 (Polyak–Łojasiewicz condition)**.** *f satisfies PL if*

$$f(x) - f^* \leq \frac{1}{2\mu}\|\nabla f(x)\|^2.$$

Definition 10.2 (Sharpness)**.** *f has μ-sharpness if*

$$f(x) - f^* \geq \frac{\mu}{2}\|x - x^*\|^2.$$

Theorem 10.2 (Sensitivity under sharpness)**.** *If $|f(x) - \bar{f}(x)| \leq \varepsilon$ and $\bar{x}^* \in \arg\min \bar{f}$, then*

$$\|\bar{x}^* - x^*\| \leq \sqrt{\tfrac{2\varepsilon}{\mu}}.$$