Lecture 1: Introduction and Least Squares

1. Standard Form of Optimization Problems

Every optimization problem can be written in a unified format called the **standard form**. This makes it easier to classify problems, apply solution methods, and communicate with others.

Definition 2 (Standard Form):
An optimization problem in **standard form** is:

$$\begin{aligned} \min_{\vec{x}\in\mathbb{R}^n} \quad & f_0(\vec{x}) \\ \text{subject to} \quad & f_i(\vec{x}) \le 0, \quad i = 1,\dots,m \\ & h_j(\vec{x}) = 0, \quad j = 1,\dots,p \end{aligned}$$

where $\vec{x} \in \mathbb{R}^n$ is the **decision variable**, f_0 is the **objective function**, f_i are **inequality constraint functions**, and h_j are **equality constraint functions**.

The set of all points satisfying the constraints is called the **feasible set**:

$$\Omega = \{\vec{x} \in \mathbb{R}^n : f_i(\vec{x}) \le 0 \ \forall i,\ h_j(\vec{x}) = 0 \ \forall j\}$$

A point $\vec{x}^\star \in \Omega$ is a **solution** (or **minimizer**) if $f_0(\vec{x}^\star) \le f_0(\vec{x})$ for all $\vec{x} \in \Omega$.

1.1 Components of Standard Form

Let us carefully define each piece:

Decision variable $\vec{x} \in \mathbb{R}^n$: The quantities we choose. We stack all decision variables into a single vector. For example, if we choose amounts x_1, x_2, x_3, then $\vec{x} = (x_1, x_2, x_3)^\top \in \mathbb{R}^3$.

Objective function $f_0 : \mathbb{R}^n \to \mathbb{R}$: The quantity we want to minimize. Common objectives include cost, error, distance, or negative profit.

Inequality constraints $f_i(\vec{x}) \le 0$: Restrictions that must hold with "$\le$" on the right. The standard form requires the right-hand side to be zero.

Equality constraints $h_j(\vec{x}) = 0$: Restrictions that must hold with exact equality.

Feasible set Ω: All points satisfying every constraint. If $\Omega = \emptyset$, the problem is **infeasible**.

Solution $\vec{x}^\star$: A feasible point achieving the smallest objective value.

1.2 Converting to Standard Form

Any optimization problem can be rewritten in standard form by following these steps:

Step 1: Identify decision variables.

List all quantities you can choose. Stack them into a single vector $\vec{x} = (x_1, x_2, \dots, x_n)^\top$.

Step 2: Convert maximization to minimization.

If the original problem is $\max f(\vec{x})$, rewrite as $\min(-f(\vec{x}))$. Negating the objective flips the direction.

Step 3: Rewrite inequalities as ≤ 0.

- $g(\vec{x}) \le c$ becomes $g(\vec{x}) - c \le 0$
- $g(\vec{x}) \ge c$ becomes $c - g(\vec{x}) \le 0$ (or equivalently $-g(\vec{x}) + c \le 0$)

Step 4: Rewrite equalities as $= 0$.

Move all terms to one side: $g(\vec{x}) = c$ becomes $g(\vec{x}) - c = 0$.

Step 5: Define the feasible set.

Write Ω as the intersection of all constraint sets.

Intuition: Standard form is like a common language. Once every problem is written the same way, we can develop general-purpose algorithms that work on any problem.

1.3 Solution Concepts and Notation

Optimization asks two different questions. We need notation for each.

Question 1: How good can it get?

Write $\min_{x\in S} f(x)$. This means: try all x in S, compute $f(x)$, return the smallest value. The answer is a **number**.

$$p^* = \min_{\vec{x}\in\Omega} f_0(\vec{x}) \qquad \leftarrow \text{a real number}$$

Question 2: Where does that happen?

Write $\operatorname{argmin}_{x\in S} f(x)$. This returns the **set of points** where the minimum is achieved.

$$\operatorname{argmin}_{\vec{x}\in\Omega} f_0(\vec{x}) = \{\vec{x} \in \Omega : f_0(\vec{x}) = p^*\} \qquad \leftarrow \text{a set of vectors}$$

Key distinction:
- min $\to$ **number** (the best value)
- argmin $\to$ **set** (the points achieving it)

Why is argmin a set? Because multiple points can tie:

- $f(x) = x^2$ on $\{-1, 0, 1\}$: argmin $= \{0\}$ (one point)

- $f(x) = x^2$ on $\{-1, 1\}$: argmin $= \{-1, 1\}$ (two points tie)
- $f(x) = x$ on $\mathbb{R}$: argmin $= \varnothing$ (no minimum exists)

Notation shortcut. When argmin has exactly one element, we write $\vec{x}^* = \operatorname{argmin} f_0(\vec{x})$ instead of $\vec{x}^* \in \operatorname{argmin} f_0(\vec{x})$.

1.4 (Optional) Infimum Versus Minimum

The problem. What if no minimum exists?

Example. The interval $(0, 1)$ has no minimum. Pick any $x \in (0, 1)$. Then $x/2$ is smaller and still in $(0, 1)$. So no element can be "the smallest."

But 0 feels like the answer. It is smaller than everything in $(0, 1)$. The issue: $0 \notin (0, 1)$. A minimum must be **in the set**.

The fix: infimum. The **infimum** (inf) is the greatest lower bound, whether or not it is in the set.

$$\inf(0, 1) = 0, \quad \text{but } \min(0, 1) \text{ does not exist.}$$

Why this matters. We define $p^* = \inf_{\vec{x} \in \Omega} f_0(\vec{x})$. This is always well-defined. If the inf is achieved, argmin contains those points. If not, argmin $= \varnothing$.

> **Convention:** We write min / max for simplicity, but mean inf / sup when needed.

2. Standard Form Examples

2.1 Meeting Time Problem

Example 1 (Meeting Time)**.** Alice and Bob want to schedule a meeting. Alice is free from 9am to 12pm. Bob is free from 10am to 2pm. They want to meet as early as possible.

Step 1: Identify the decision variable.

Let t = meeting start time (in hours after midnight). So $t \in \mathbb{R}^1$.

Step 2: Write the objective.

We want the earliest time, so minimize t:

$$f_0(t) = t$$

Step 3: Write the constraints.

Alice's availability: $9 \le t \le 12$. Bob's availability: $10 \le t \le 14$.

Rewriting each as ≤ 0:

$$\begin{aligned} f_1(t) &= 9 - t \le 0 \quad \text{(Alice starts at 9)} \\ f_2(t) &= t - 12 \le 0 \quad \text{(Alice ends at 12)} \\ f_3(t) &= 10 - t \le 0 \quad \text{(Bob starts at 10)} \\ f_4(t) &= t - 14 \le 0 \quad \text{(Bob ends at 14)} \end{aligned}$$

Step 4: Write the standard form.

$$\begin{aligned} \min_{t \in \mathbb{R}} \quad & t \\ \text{subject to} \quad & 9 - t \le 0 \\ & t - 12 \le 0 \\ & 10 - t \le 0 \\ & t - 14 \le 0 \end{aligned}$$

Step 5: Identify the feasible set.

The constraints $9 - t \le 0$ and $10 - t \le 0$ give $t \ge 10$.

The constraints $t - 12 \le 0$ and $t - 14 \le 0$ give $t \le 12$.

Therefore: $\Omega = [10, 12]$.

Step 6: Find the solution.

We want the smallest $t \in [10, 12]$. The answer is $t^\star = 10$.

Sanity check: At $t = 10$, both Alice (free 9–12) and Bob (free 10–14) are available. ✓

2.2 Oil & Gas Production

Example 2 (Oil & Gas)**.** A company produces oil and gas. Each barrel of oil yields \$40 profit; each unit of gas yields \$30. The refinery can process at most 100 units total. Oil requires 2 hours per barrel; gas requires 1 hour per unit. There are 120 labor hours available. How much of each should they produce to maximize profit?

Step 1: Identify decision variables.

Let x_1 = barrels of oil, x_2 = units of gas. So $\vec{x} = (x_1, x_2)^\top \in \mathbb{R}^2$.

Step 2: Write the original objective.

Maximize profit: $40x_1 + 30x_2$.

To convert to standard form, minimize the negative:

$$f_0(\vec{x}) = -40x_1 - 30x_2$$

Step 3: Write constraints in standard form.

Capacity constraint: $x_1 + x_2 \le 100$ becomes $x_1 + x_2 - 100 \le 0$.

Labor constraint: $2x_1 + x_2 \le 120$ becomes $2x_1 + x_2 - 120 \le 0$.

Nonnegativity: $x_1 \ge 0$ becomes $-x_1 \le 0$. Similarly $-x_2 \le 0$.

Step 4: Complete standard form.

$$\begin{aligned} \min_{\vec{x} \in \mathbb{R}^2} \quad & -40x_1 - 30x_2 \\ \text{subject to} \quad & x_1 + x_2 - 100 \le 0 \\ & 2x_1 + x_2 - 120 \le 0 \\ & -x_1 \le 0 \\ & -x_2 \le 0 \end{aligned}$$

Feasible set: $\Omega = \{(x_1, x_2) : x_1, x_2 \geq 0,\ x_1 + x_2 \leq 100,\ 2x_1 + x_2 \leq 120\}$.

This is a polygon in the first quadrant. The optimal solution occurs at a corner (we will prove this later for linear programs). Testing corners: $(0, 0)$, $(0, 100)$, $(60, 0)$, $(20, 80)$. The maximum profit is at $(20, 80)$ with profit $40(20) + 30(80) = 3200$.

Sanity check: At $(20, 80)$: labor used is $2(20) + 80 = 120$ hours (exactly the limit), and capacity is $20 + 80 = 100$ units (exactly the limit). Both constraints are tight at the optimum. ✓

2.3 EECS Course Sizes

Example 3 (Course Sizes LP)**.** A department offers n courses. Let x_i be the enrollment cap for course i. Each student brings revenue r_i. The total enrollment cannot exceed B (budget constraint). We want to maximize total revenue.

Step 1: Decision variable.

$\vec{x} = (x_1, x_2, \ldots, x_n)^\top \in \mathbb{R}^n$.

Step 2: Objective in vector notation.

Maximize $\sum_{i=1}^n r_i x_i = \vec{r}^\top \vec{x}$.

In standard form (minimize negative): $f_0(\vec{x}) = -\vec{r}^\top \vec{x}$.

Step 3: Constraints.

Budget: $\sum_{i=1}^n x_i \leq B$. Let $\vec{1} = (1, 1, \ldots, 1)^\top$. Then $\vec{1}^\top \vec{x} \leq B$.

Nonnegativity: $x_i \geq 0$ for all i. In vector notation: $\vec{x} \succeq \vec{0}$ (componentwise inequality).

Step 4: Standard form with vector notation.

$$\begin{aligned} \min_{\vec{x} \in \mathbb{R}^n} \quad & -\vec{r}^\top \vec{x} \\ \text{subject to} \quad & \vec{1}^\top \vec{x} - B \leq 0 \\ & -\vec{x} \preceq \vec{0} \end{aligned}$$

The notation $\vec{x} \succeq \vec{0}$ means $x_i \geq 0$ for each component i. This is called a **componentwise** or **elementwise** inequality.

Vector notation tip: Writing $-\vec{x} \preceq \vec{0}$ is equivalent to n separate constraints $-x_i \leq 0$. Compact notation makes problems with many variables easier to write and analyze.

Sanity check: With $n = 2$ courses and $\vec{r} = (5, 3)^\top$, the formulation becomes: minimize $-5x_1 - 3x_2$ subject to $x_1 + x_2 \leq B$ and $x_1, x_2 \geq 0$. The solution allocates as much as possible to the higher-revenue course. ✓

3. Least Squares

Consider the following situation: given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $\vec{y} \in \mathbb{R}^m$, we want to find $\vec{x}$ such that $A\vec{x} = \vec{y}$. But what if no exact solution exists?

This happens frequently in practice. When $m > n$ (more equations than unknowns), the system is **overdetermined** and typically has no solution. Instead of giving up, we turn this unsolvable equation into an optimization problem: find $\vec{x}$ that makes $A\vec{x}$ as close to $\vec{y}$ as possible.

3.1 Least Squares as an Optimization Problem

We measure "closeness" using the squared Euclidean distance. The **least squares problem** is:

$$\min_{\vec{x} \in \mathbb{R}^n} \|A\vec{x} - \vec{y}\|_2^2$$

Let us map this to the standard form framework:

- **Decision variable:** $\vec{x} \in \mathbb{R}^n$
- **Objective:** $f_0(\vec{x}) = \|A\vec{x} - \vec{y}\|_2^2$
- **Constraints:** none ($m = 0$ inequality, $p = 0$ equality)
- **Feasible set:** $\Omega = \mathbb{R}^n$ (all of n-dimensional space)

This is the **simplest** type of optimization problem: unconstrained minimization. Every point is feasible, so we only need to find where the objective is smallest.

3.2 Geometric Interpretation

The key insight comes from reframing the problem geometrically.

Step 1: Recognize what we are really choosing.

When we choose $\vec{x} \in \mathbb{R}^n$, we are really choosing the vector $A\vec{x} \in \mathbb{R}^m$. The set of all possible $A\vec{x}$ as $\vec{x}$ varies over $\mathbb{R}^n$ is the **column space** (or **range**) of A:

$$\mathcal{R}(A) = \{A\vec{x} : \vec{x} \in \mathbb{R}^n\}$$

Step 2: Restate the problem geometrically.

The least squares problem becomes: find the point in $\mathcal{R}(A)$ that is closest to $\vec{y}$.

Geometric view: We are projecting $\vec{y}$ onto the subspace $\mathcal{R}(A)$. The answer is the **orthogonal projection** of $\vec{y}$ onto the column space of A.

3.3 Why Projection is Optimal

Let $\vec{z} = A\vec{x}^\star$ be the orthogonal projection of $\vec{y}$ onto $\mathcal{R}(A)$. The **residual** is $\vec{e} = \vec{y} - \vec{z}$.

The defining property of orthogonal projection is:

$$\vec{e} \perp \mathcal{R}(A)$$

This means the residual is perpendicular to every vector in the column space.

Claim: The projection $\vec{z}$ minimizes $\|\vec{y} - \vec{u}\|_2$ over all $\vec{u} \in \mathcal{R}(A)$.

Proof: Let $\vec{u} \in \mathcal{R}(A)$ be any point in the column space. We want to show $\|\vec{y} - \vec{u}\|_2 \geq \|\vec{e}\|_2$.

Write the difference as:

$$\vec{y} - \vec{u} = (\vec{y} - \vec{z}) + (\vec{z} - \vec{u}) = \vec{e} + (\vec{z} - \vec{u})$$

Since $\vec{z}, \vec{u} \in \mathcal{R}(A)$, we have $\vec{z} - \vec{u} \in \mathcal{R}(A)$. But $\vec{e} \perp \mathcal{R}(A)$, so $\vec{e} \perp (\vec{z} - \vec{u})$.

By the **Pythagorean theorem** (which applies because $\vec{e}$ and $\vec{z} - \vec{u}$ are orthogonal):

$$\|\vec{y} - \vec{u}\|_2^2 = \|\vec{e}\|_2^2 + \|\vec{z} - \vec{u}\|_2^2$$

Since $\|\vec{z} - \vec{u}\|_2^2 \geq 0$, we conclude:

$$\|\vec{y} - \vec{u}\|_2^2 \geq \|\vec{e}\|_2^2$$

with equality if and only if $\vec{u} = \vec{z}$. $\square$

3.4 Normal Equations

We now translate the geometric condition $\vec{e} \perp \mathcal{R}(A)$ into algebra.

Step 1: What does orthogonality mean?

$\vec{e} \perp \mathcal{R}(A)$ means $\vec{e}$ is perpendicular to every column of A. If $\vec{a}_1, \ldots, \vec{a}_n$ are the columns of A, then:

$$\vec{a}_i^\top \vec{e} = 0 \quad \text{for } i = 1, \ldots, n$$

Step 2: Write this compactly.

Stacking these n equations into a single matrix equation:

$$A^\top \vec{e} = \vec{0}$$

Step 3: Substitute the definition of $\vec{e}$.

Since $\vec{e} = \vec{y} - A\vec{x}^\star$:

$$A^\top (\vec{y} - A\vec{x}^\star) = \vec{0}$$

Step 4: Rearrange to get the normal equations.

Normal Equations:

$$A^\top A \vec{x}^\star = A^\top \vec{y}$$

The name "normal equations" comes from the fact that the residual $\vec{e}$ is **normal** (perpendicular) to the column space.

3.5 The Least Squares Solution

Theorem 4 (Least Squares Solution):
If $A \in \mathbb{R}^{m \times n}$ has **full column rank** (i.e., $\operatorname{rank}(A) = n$), then $A^\top A$ is invertible and the unique solution to the least squares problem is:

$$\vec{x}^\star = (A^\top A)^{-1} A^\top \vec{y}$$

Why is $A^\top A$ invertible?

When A has full column rank, its columns are linearly independent. This means $A\vec{x} = \vec{0}$ implies $\vec{x} = \vec{0}$. Consider any $\vec{x}$ in the null space of $A^\top A$:

$$\begin{aligned} A^\top A \vec{x} = \vec{0} &\implies \vec{x}^\top A^\top A \vec{x} = 0 \\ &\implies \|A\vec{x}\|_2^2 = 0 \\ &\implies A\vec{x} = \vec{0} \implies \vec{x} = \vec{0} \end{aligned}$$

So $A^\top A$ has trivial null space, which means it is invertible.

Sanity check: The matrix $A^\top A$ is $n \times n$ (square), symmetric, and positive definite when A has full column rank. ✓

3.6 Linear Regression

Example 4 (Linear Regression)**.** Given data points $(t_1, y_1), (t_2, y_2), \ldots, (t_m, y_m)$, find the line $y = \alpha + \beta t$ that best fits the data.

Step 1: Identify the decision variables.

We want to choose the intercept α and slope β. Stack them: $\vec{x} = (\alpha, \beta)^\top \in \mathbb{R}^2$.

Step 2: Write the prediction for each data point.

For data point i: predicted value is $\alpha + \beta t_i$.

Step 3: Set up the matrix equation.

We want $\alpha + \beta t_i \approx y_i$ for each i. In matrix form:

$$\underbrace{\begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}_{\vec{x}} \approx \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}}_{\vec{y}}$$

Step 4: Apply the least squares formula.

The best-fit line has parameters:

$$\begin{pmatrix} \alpha^\star \\ \beta^\star \end{pmatrix} = (A^\top A)^{-1} A^\top \vec{y}$$

Why does this work?

When data is noisy, the exact system $A\vec{x} = \vec{y}$ has no solution (the points don't lie exactly on any line). Least squares finds the line that minimizes the sum of squared vertical distances from data points to the line.

Sanity check: The matrix A has dimensions $m \times 2$ (since we have m data points and 2 parameters). For A to have full column rank, we need at least 2 data points that are not at the same t-value. ✓

Key Takeaways

1. **Standard form unifies all problems:** minimize objective, all inequalities as ≤ 0, all equalities as $= 0$.
2. **Maximization becomes minimization:** negate the objective function.
3. **Feasible set Ω:** the intersection of all constraints; if empty, problem is infeasible.
4. **Vector notation:** stacking variables into $\vec{x}$ and using $\succeq$ for componentwise inequalities makes large problems compact.
5. **Least squares turns unsolvable equations into optimization:** when $A\vec{x} = \vec{y}$ has no solution, minimize $\|A\vec{x} - \vec{y}\|_2^2$.
6. **Solution is orthogonal projection:** the optimal $A\vec{x}^\star$ is the projection of $\vec{y}$ onto $\mathcal{R}(A)$.
7. **Normal equations encode orthogonality:** $A^\top A\vec{x}^\star = A^\top\vec{y}$ says the residual is perpendicular to the column space.