CME 307: Optimization September 22, 2025

Lecture 1: Intro + Linear Algebra Review

Prof. Madeleine Udell

1 What is an Optimization Problem?

Definition (Optimization problem). An optimization problem is specified by:

- an objective function $f : \mathbb{R}^n \to \mathbb{R}$,
- a feasible set $\mathcal{X} \subseteq \mathbb{R}^n$.

The goal is to compute the *optimal value*

$$p^\star := \inf_{x \in \mathcal{X}} f(x),$$

and to find a point $x^\star \in \mathcal{X}$ attaining this value, if one exists.

Linear and Integer Optimization

We can write a linear optimization problem with equality, inequality, and bound constraints as

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & Cx \leq d \\ \text{variable} & x \in \mathbb{R}^n, \end{array}$$

with data $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m_1 \times n}$, $b \in \mathbb{R}^{m_1}$, $C \in \mathbb{R}^{m_2 \times n}$, $d \in \mathbb{R}^{m_2}$. Here,

- $c^T x$ is the linear objective to minimize,
- $Ax = b$ are linear equality constraints,
- $Cx \leq d$ are linear inequality constraints.

It is also quite common to include a *box constraint* on the optimization variable $\ell \leq x \leq u$.

If some components of x are required to be integers, we obtain a mixed-integer program (MIP):

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & Cx \leq d \\ \text{variable} & x \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2}. \end{array}$$

Example (Diet problem). We an planning a backpacking trip, and want to minimize the total weight of the food packed subject to nutritional requirements. We can write this problem as the linear program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \geq b \\ & x \geq 0 \\ \text{variable} & x \in \mathbb{R}^n, \end{array}$$

where

- $A \in \mathbb{R}^{m \times n}$ with a_{ij} = amount of nutrient i in food j,
- $b \in \mathbb{R}^m$ with b_i = required daily amount of nutrient i,
- $c \in \mathbb{R}^n$ with c_j = weight per serving of food j.

The solution $x^\star$ gives the number of servings of each food to buy.
Extensions:

- If foods are chosen in integer servings, $x \in \mathbb{Z}^n$.
- If foods come from recipes, $x = By$ where each column of B represents a recipe, with indices recording the proportion of each food in the recipe, and entries of $y \in \mathbb{R}^m$ denote the number of servings of each recipe.
- If we require diet diversity, $y \leq u$, which ensures that no recipe is used more than u times.
- If any level of a nutrient within a range $[b_{\min}, b_{\max}]$ is acceptable, we can introduce slack variables s to ensure that the nutrient levels lie in this range: $Ax + s = b$, $l \leq s \leq u$ with $b = (b_{\min} + b_{\max})/2$, $l = b_{\min} - b$, $u = b_{\max} - b$.

Nonlinear Optimization

The general nonlinear problem has the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m_1 \\ & h_j(x) = 0, \quad j = 1, \ldots, m_2 \\ \text{variable} & x \in \mathbb{R}^n \end{array}$$

where f_0, f_i, h_j may be nonlinear.

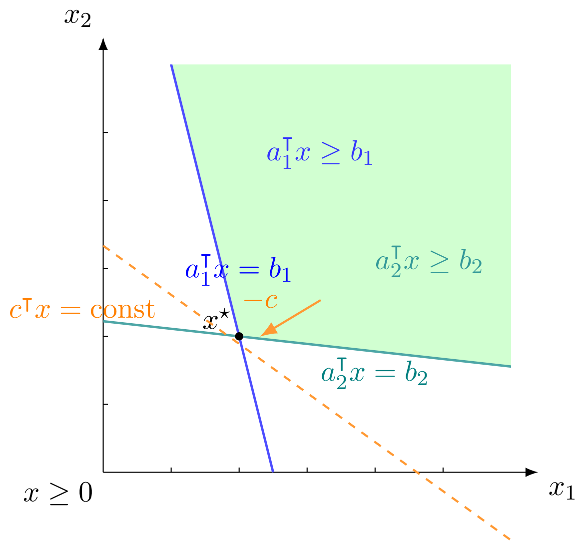

Figure 1: Feasible region for a 2D diet LP, showing halfspaces $a_i^\intercal x \geq b_i$, $x \geq 0$, and an optimal corner $x^\star$.

> **Example (Desalination plant).** Variables x control pumps, pressures, and chemical levels.
> - Objective $f_0(x)$: cost of water produced.
> - Constraints $f_i(x)$: level of impurity i in water.
> - Feasible domain: $f_i(x) \leq b_i$ for legal limits b_i.
>
> The operator asks: what setting of x minimizes cost subject to safe water quality?

Quadratic Optimization

Quadratic objectives arise frequently. A general quadratic program is

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T Hx + g^T x + c \\ \text{subject to} & Ax = b,\ Cx \leq d \\ \text{variable} & x \in \mathbb{R}^n \end{array}$$

where $H \in \mathbb{R}^{n \times n}$ is symmetric.

> **Example (Portfolio optimization).**
>
> - x_i: fraction of wealth invested in asset i,
> - Objective: minimize risk $x^T \Sigma x$ (variance),
> - Constraint: expected return $\mu^T x \geq r$,
> - Constraint: $\mathbf{1}^T x = 1$, $x \geq 0$.

Conic Optimization

A conic program generalizes LPs and QPs by allowing variables to lie in a convex cone K:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ \text{variable} & x \in K \end{array}$$

Important cones:

- Nonnegative orthant $\mathbb{R}^n_+$: recovers LP.
- Second-order cone: $\{(t, x) : \|x\|_2 \leq t\}$.
- Semidefinite cone: $\{X \succeq 0\}$.

Example (Maximum volume ellipsoid). Given points $a_1, \ldots, a_m \in \mathbb{R}^n$, find the maximum volume ellipsoid

$$E = \{x : (x - c)^T P^{-1} (x - c) \leq 1\}$$

contained in $\{x : a_i^T x \leq 1,\ i = 1, \ldots, m\}$. This can be formulated as a semidefinite program in variables (P, c).

Unconstrained Optimization

Sometimes constraints are absent:

$$\min_{x \in \mathbb{R}^n} f(x).$$

Example (Least squares). Given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, minimize

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

The solution $x^\star$ satisfies the normal equations $A^T A x^\star = A^T b$.

Finite-Sum Optimization

Finite-sum problems are ubiquitous in machine learning:

$$\min_{x \in \mathbb{R}^n} \ f(x) = \tfrac{1}{m} \sum_{i=1}^{m} f_i(x).$$

Example (Logistic regression). Given data (a_i, b_i) with $b_i \in \{-1, +1\}$, define

$$f_i(x) = \log\bigl(1 + \exp(-b_i a_i^T x)\bigr).$$

Then logistic regression solves

$$\min_{x \in \mathbb{R}^n} \ \frac{1}{m} \sum_{i=1}^{m} f_i(x) + \lambda \|x\|_2^2 .$$

Black-Box Optimization

In black-box settings, we cannot access analytic gradients or Hessians. Instead, we assume access to an oracle that returns:

- 0th-order: $f(x)$ only,
- 1st-order: $f(x)$ and $\nabla f(x)$,
- 2nd-order: $f(x)$, $\nabla f(x)$, and $\nabla^2 f(x)$.

Example (Hyperparameter tuning). We want to minimize the validation error of a neural network with respect to learning rate η and regularization α. The function is nonconvex, costly to evaluate, and no gradients are available. Solution approaches: random search, Bayesian optimization, or zeroth-order methods.

Convex vs Nonconvex Problems

Definition (Convex optimization). An optimization problem is convex if the feasible set $\mathcal{X}$ is convex and the objective f_0 is convex. Then any local optimum is also a global optimum.

Example (Support vector machine).

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^m \xi_i \\ \text{subject to} & b_i(w^T a_i + \beta) \geq 1 - \xi_i,\ \xi_i \geq 0 \end{array}$$

This is a convex quadratic program in variables (w, β, ξ).

Gotcha: Nonconvex problems (e.g. training deep neural networks) may have many local minima, saddle points, or flat regions. First-order methods can get stuck or behave unpredictably.

Modularity in Optimization

Optimization is modular:

1. Model problem mathematically.
2. Identify properties (linear? convex? integer?).
3. Use an appropriate solver or design one.
4. Iterate: approximate, reformulate, or warm-start.

Principle. The art of optimization lies as much in *modeling* and *reformulation* as in algorithm design.

Figure suggestion: A circular flow diagram labeled *Model* → *Analyze* → *Solve* → *Refine*.

2 Linear Algebra Review

Optimization problems are built on linear algebra: feasible sets are often affine subspaces, optimality conditions involve nullspaces and orthogonality, and algorithms rely on solving linear systems. This review covers the essential background.

Linear independence, span, and subspaces

Definition. Vectors $a_1, \dots, a_k \in \mathbb{R}^n$ are *linearly dependent* if there exist scalars $\lambda_1, \dots, \lambda_k$, not all zero, such that

$$\lambda_1 a_1 + \cdots + \lambda_k a_k = 0.$$

Otherwise, the vectors are *linearly independent*.
The *span* of $a_1, \dots, a_k$ is

$$\mathrm{span}\{a_1, \dots, a_k\} = \left\{ \sum_{i=1}^{k} \lambda_i a_i : \lambda_i \in \mathbb{R} \right\}.$$

A *linear subspace* $L \subseteq \mathbb{R}^n$ is a set closed under addition and scalar multiplication:

$$v, w \in L \;\Rightarrow\; v + w \in L, \;\; \lambda v \in L \;\; \forall \lambda \in \mathbb{R}.$$

An *affine subspace* is a set of the form $x_0 + L$, where $x_0 \in \mathbb{R}^n$ and L is a linear subspace.

Example. In $\mathbb{R}^2$, the vectors $(1, 0)$ and $(0, 1)$ are linearly independent; their span is $\mathbb{R}^2$. The vectors $(1, 0)$ and $(2, 0)$ are linearly dependent; their span is the x-axis, a 1-dimensional subspace of $\mathbb{R}^2$.

Span, rank, and nullspace of a matrix

Let $A \in \mathbb{R}^{m \times n}$ with columns $A_1, \dots, A_n$ and rows $a_1^\intercal, \dots, a_m^\intercal$.

Definition.

- The *span* (or column space) of A is
$$\mathrm{span}(A) = \{Ax : x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m.$$
- The *nullspace* of A is
$$\mathrm{null}(A) = \{x \in \mathbb{R}^n : Ax = 0\} \subseteq \mathbb{R}^n.$$
- The *rank* of A is
$$\mathrm{Rank}(A) = \dim(\mathrm{span}(A)).$$

Theorem (Rank–Nullity). For any $A \in \mathbb{R}^{m \times n}$,
$$\mathrm{Rank}(A) + \dim(\mathrm{null}(A)) = n.$$

Proof. Let $A = [A_1\ A_2\ \cdots\ A_n]$ with $A_j \in \mathbb{R}^m$. Choose an index set $S \subseteq \{1, \dots, n\}$ that is *minimal* such that $\{A_j : j \in S\}$ spans $\mathrm{span}(A) = \{Ax : x \in \mathbb{R}^n\}$. By minimality, $\{A_j : j \in S\}$ is linearly independent, hence $|S| = \mathrm{Rank}(A) =: r$.

Step 1 (Produce $n-r$ independent null vectors). Fix any $j \notin S$. Since $A_j \in \mathrm{span}\{A_i : i \in S\}$, there exists a vector $w^{(j)} \in \mathbb{R}^n$ supported only on S with

$$A_j = \sum_{i \in S} w_i^{(j)} A_i \quad \Longleftrightarrow \quad A\big(e_j - w^{(j)}\big) = 0.$$

Thus $z^{(j)} := e_j - w^{(j)} \in \mathrm{null}(A)$ for every $j \notin S$. These $\{z^{(j)} : j \notin S\}$ are linearly independent: if $\sum_{j \notin S} \alpha_j z^{(j)} = 0$, then looking at coordinates outside S (which only appear in the e_j parts) forces every $\alpha_j = 0$. Hence $\dim \mathrm{null}(A) \geq n - r$.

Step 2 (No room for more). Define the projection $\pi : \mathbb{R}^n \to \mathbb{R}^{n-r}$ that keeps only coordinates outside S. We claim π is *injective* on $\mathrm{null}(A)$. Indeed, if $x \in \mathrm{null}(A)$ and $\pi(x) = 0$, then x is supported on S and
$$0 = Ax = \sum_{i \in S} x_i A_i.$$

Because $\{A_i : i \in S\}$ is linearly independent, $x_i = 0$ for all $i \in S$, so $x = 0$. Therefore $\dim \mathrm{null}(A) \leq n - r$.

Combining the two steps gives $\dim \mathrm{null}(A) = n - r$, i.e., $\mathrm{Rank}(A) + \dim \mathrm{null}(A) = r + (n - r) = n$. $\square$

Example. Let

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

The columns are $A_1 = (1,0)^\intercal$, $A_2 = (0,1)^\intercal$, $A_3 = (1,1)^\intercal$. Since $A_3 = A_1 + A_2$, the rank is 2. The nullspace is the set of $x \in \mathbb{R}^3$ with $x_1 + x_3 = 0$, $x_2 + x_3 = 0$. So $\mathrm{null}(A) = \{(-t,-t,t) : t \in \mathbb{R}\}$, a 1-dimensional subspace. Rank–nullity: $2 + 1 = 3 = n$.

Solutions to linear systems

Theorem. Consider $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$.

1. The system $Ax = b$ has a solution iff $b \in \mathrm{span}(A)$.
2. If x_0 is one solution, then all solutions are

$$x_0 + v, \quad v \in \mathrm{null}(A).$$

3. The solution set is an affine subspace of dimension $n - \mathrm{Rank}(A)$.
4. The solution is unique iff $\mathrm{null}(A) = \{0\}$, i.e., iff $\mathrm{Rank}(A) = n$.

Example. Solve

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} x = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

The first equation gives $x_1 + 2x_2 = 2$. The second equation gives no condition. A particular solution is $x_0 = (2,0)^\intercal$. The nullspace is $\{(-2t,t) : t \in \mathbb{R}\}$. Hence the solution set is

$$\{(2,0)^\intercal + (-2t,t)^\intercal : t \in \mathbb{R}\},$$

an affine line in $\mathbb{R}^2$.

Invertibility

Definition. A square matrix $A \in \mathbb{R}^{n\times n}$ is *invertible* if there exists A^{-1} such that $AA^{-1} = A^{-1}A = I$.

Theorem. The following are equivalent for $A \in \mathbb{R}^{n\times n}$:

1. A is invertible.
2. $\text{Rank}(A) = n$.
3. $\text{null}(A) = \{0\}$.
4. For all $b \in \mathbb{R}^n$, the system $Ax = b$ has a unique solution.

Example. For $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, we have

$$A^{-1} = \frac{1}{-2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}.$$

Hence for any b, the system $Ax = b$ has the unique solution $x = A^{-1}b$.

Orthogonality (optional enrichment)

Definition. For a subspace $L \subseteq \mathbb{R}^n$, the *orthogonal complement* is

$$L^\perp = \{y \in \mathbb{R}^n : y^\intercal x = 0 \ \ \forall x \in L\}.$$

Fact. For any $A \in \mathbb{R}^{m\times n}$,

$$\text{null}(A) = \text{span}(A^\intercal)^\perp.$$

Example. Let $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$. The row space is spanned by $(1, 1, 0)$ and $(0, 1, 1)$. The nullspace consists of vectors $(t, -t, t)$, which are orthogonal to both rows.

Summary

Key Takeaways.

- Linear independence, span, subspaces, affine subspaces.
- Rank, nullspace, and the rank–nullity theorem.
- Solutions of $Ax = b$: existence, uniqueness, affine geometry.
- Invertibility: equivalent characterizations.
- Orthogonality: row space and nullspace are complements.