CME 307: Optimization October 13, 2025

Lecture 1: Intro + Linear Algebra Review

Prof. Madeleine Udell

1 What is an Optimization Problem?

Definition 1.1. 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 1.2. 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.

Example 1.3. 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?

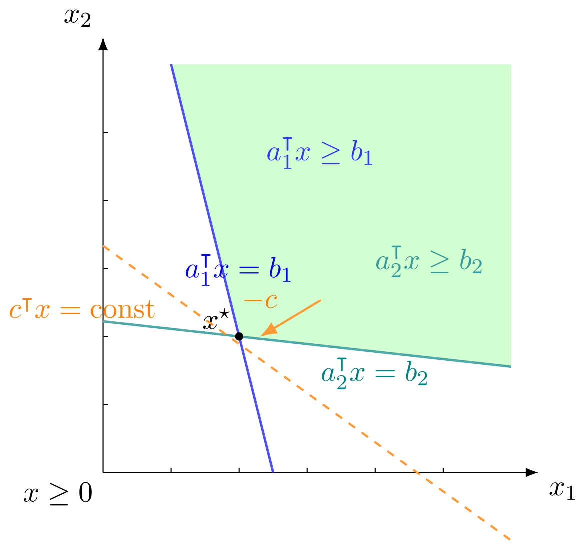

Figure 1: Feasible region for a 2D diet LP, showing halfspaces $a_i^\intercal x \geq b_i$, $x \geq 0$, and an optimal corner $x^\star$.

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.

2 Linear algebra review

2.1 Linear independence

Definition 2.1 (Span of vectors)**.** The *span* of vectors $A_1, \ldots, A_k \in \mathbb{R}^m$ is

$$\text{span}\{A_1, \ldots, A_k\} = \{\lambda_1 A_1 + \cdots + \lambda_k A_k \mid \lambda \in \mathbb{R}^k\}.$$

Vectors $A_1, \ldots, A_k$ are *linearly dependent* if there exists some nonzero $\lambda \in \mathbb{R}^k$ with $\lambda_1 A_1 + \cdots + \lambda_k A_k = 0$; otherwise, they are *linearly independent.*

If the vectors are linearly independent, none can be written as a linear combination of the others. If they are dependent, at least one can.

Example 2.2 (Quick check for dependence)**.** Let $A_1 = (1,0,1)^\top$, $A_2 = (0,1,1)^\top$, $A_3 = (1,1,2)^\top \in \mathbb{R}^3$. Then $A_3 = A_1 + A_2$, so $\{A_1, A_2, A_3\}$ is linearly dependent.

Exercise. Decide whether the set $\{(1,2,3)^\top, (2,5,8)^\top, (0,1,2)^\top\}$ is linearly independent. If not, exhibit a nontrivial linear relation.

2.2 Linear and affine subspaces

Definition 2.3 (Linear vs. affine subspace)**.** A set $L \subseteq \mathbb{R}^n$ is a *linear subspace* if it is closed under addition and scalar multiplication: $v, w \in L$ and $\lambda \in \mathbb{R}$ imply $v + w \in L$ and $\lambda v \in L$. A set $A \subseteq \mathbb{R}^n$ is *affine* if it can be written as $x_0 + L$ for some $x_0 \in \mathbb{R}^n$ and some linear subspace L.

A linear subspace always contains the origin, while an affine subspace need not.

A linear subspace contains any linear combination of points in the space. Similarly, an affine subspace contains any *affine combination* of points in the space: any combination where the coefficients sum to one.

Theorem 2.4 (Characterization of affine sets)**.** *A set $A \subseteq \mathbb{R}^n$ is affine if and only if it contains every affine combination of its points: for all $v, w \in A$ and all $\lambda \in \mathbb{R}$,*

$$\lambda v + (1 - \lambda) w \in A.$$

Proof. ($\Rightarrow$) If $A = x_0 + L$ with L a linear subspace, write $v = x_0 + \ell_v$ and $w = x_0 + \ell_w$ with $\ell_v, \ell_w \in L$. Then

$$\lambda v + (1-\lambda) w = \lambda(x_0 + \ell_v) + (1-\lambda)(x_0 + \ell_w) = x_0 + \big(\lambda \ell_v + (1-\lambda)\ell_w\big) \in x_0 + L = A,$$

since L is closed under linear combinations.

($\Leftarrow$) Fix $v \in A$ and set $L := \{w - v \mid w \in A\}$. We show L is a linear subspace. Let $u_1 = w_1 - v$ and $u_2 = w_2 - v$ with $w_1, w_2 \in A$, and $\alpha, \beta \in \mathbb{R}$. Then for any $\lambda \in \mathbb{R}$,

$$v + \lambda u_1 + (1-\lambda) u_2 = \lambda w_1 + (1-\lambda) w_2 \in A,$$

using the assumed closure under affine combinations. Taking $\lambda = \frac{\alpha}{\alpha+\beta}$ (if $\alpha + \beta \neq 0$) yields $v + \alpha u_1 + \beta u_2 \in A$, so $\alpha u_1 + \beta u_2 \in L$. If $\alpha + \beta = 0$, the same closure (e.g., with $\lambda = 1$) also implies $\alpha u_1 + \beta u_2 \in L$. Thus L is a linear subspace and $A = v + L$, i.e., A is affine. $\square$

Example 2.5. $L = \{(t, 2t) \mid t \in \mathbb{R}\}$ is a line through the origin, hence a linear subspace of $\mathbb{R}^2$. The set $A = (1,0) + L = \{(1+t, 2t) \mid t \in \mathbb{R}\}$ is a parallel line not through the origin, hence affine but not linear.

Exercise. Show that any two parallel affine subspaces in $\mathbb{R}^n$ have the same dimension. (Hint: write them as $x_0 + L$ and $y_0 + L$ for the same linear subspace L.)

2.3 Span, nullspace, and rank of a matrix

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

Definition 2.6 (Column span, nullspace, rank)**.**

$$\operatorname{span}(A) = \{Ax \mid x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m, \qquad \operatorname{null}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\} \subseteq \mathbb{R}^n,$$

$$\operatorname{Rank}(A) = \dim(\operatorname{span}(A)).$$

These objects will be the main players in describing solutions to $Ax = b$.

Theorem 2.7 (Rank-nullity)**.** *For every $A \in \mathbb{R}^{m \times n}$,*

$$\operatorname{Rank}(A) + \dim(\operatorname{null}(A)) = n.$$

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 $\operatorname{span}(A) = \{Ax : x \in \mathbb{R}^n\}$. By minimality, $\{A_j : j \in S\}$ is linearly independent, hence $|S| = \operatorname{Rank}(A) =: r$.

Step 1 (Produce $n - r$ independent null vectors). Fix any $j \notin S$. Since $A_j \in \operatorname{span}\{A_i : i \in S\}$, 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 $z^{(j)} := e_j - w^{(j)} \in \operatorname{null}(A)$ 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 $\dim \operatorname{null}(A) \geq n - r$.

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 $\operatorname{null}(A)$. Indeed, if $x \in \operatorname{null}(A)$ 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 $\dim \operatorname{null}(A) \leq n - r$.

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

Example 2.8 (Small computation)**.** For $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$, the columns span $\operatorname{span}(A) = \{(x_1 +$

$x_2,\ x_2+x_3)^\top \mid x \in \mathbb{R}^3\}$, so $\mathrm{Rank}(A) = 2$. Solving $Ax = 0$ gives $x_1 = -x_2$ and $x_3 = -x_2$, hence

$$\mathrm{null}(A) = \{(-t,\ t,\ -t)^\top \mid t \in \mathbb{R}\}, \quad \dim(\mathrm{null}(A)) = 1,$$

and rank-nullity $2 + 1 = 3 = n$ holds.

Exercise. Compute $\mathrm{Rank}(A)$ and a basis for $\mathrm{null}(A)$ for $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 1 & 1 \end{bmatrix}$. Verify rank-nullity.

2.4 Orthogonality of row space and nullspace

Definition 2.9 (Orthogonal complement)**.** For a subspace $L \subseteq \mathbb{R}^n$, the *orthogonal complement* is

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

Theorem 2.10. *For any $A \in \mathbb{R}^{m\times n}$,*

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

Proof. ($\subseteq$) If $x \in \mathrm{null}(A)$, then $Ax = 0$, so for any $y \in \mathbb{R}^m$, $(A^\intercal y)^\intercal x = y^\intercal(Ax) = 0$. Thus $x \in \mathrm{span}(A^\intercal)^\perp$.

($\supseteq$) If $x \in \mathrm{span}(A^\intercal)^\perp$, then for each row $A_i^\intercal$ of A, $(A_i^\intercal)^\intercal x = A_i x = 0$. Thus $Ax = 0$, so $x \in \mathrm{null}(A)$. □

2.5 Solution sets of linear systems

Definition 2.11 (Solution set)**.** For $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, the *solution set* of the linear system $Ax = b$ is $\{x \in \mathbb{R}^n : Ax = b\}$.

We ask: when does a solution exist, what is the dimension of the set, and when is it unique?

Proposition 2.12 (Existence, structure, and dimension)**.** *A solution to $Ax = b$ exists iff $b \in \mathrm{span}(A)$. If a solution x_0 exists, then the full solution set is the affine subspace*

$$\{x \in \mathbb{R}^n : Ax = b\} = x_0 + \mathrm{null}(A),$$

which has dimension $n - \mathrm{Rank}(A)$. In particular, the solution is unique iff $\mathrm{null}(A) = \{0\}$.

Proof. ($\Leftarrow$) If $b \in \mathrm{span}(A)$ there exists x_0 with $Ax_0 = b$, so a solution exists. ($\Rightarrow$) If $Ax = b$ has a solution x_0, then $Ax = b$ iff $A(x - x_0) = 0$, i.e., $x - x_0 \in \mathrm{null}(A)$. Thus the solution set equals $x_0 + \mathrm{null}(A)$. Its dimension is $\dim(\mathrm{null}(A)) = n - \mathrm{Rank}(A)$ by rank-nullity. Uniqueness holds iff $\mathrm{null}(A) = \{0\}$. □

Example 2.13 (Worked solution)**.** Take $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$ and $b = (1,1)^\top$. One particular solution is $x_0 = (1,0,1)^\top$ since $Ax_0 = b$. Using the nullspace from the earlier example,

$$\{x : Ax = b\} = x_0 + \mathrm{null}(A) = \{(1,0,1)^\top + t(-1,1,-1)^\top \mid t \in \mathbb{R}\},$$

an affine line of dimension $3 - \mathrm{Rank}(A) = 1$.

Exercise. For $A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 3 \end{bmatrix}$ and $b = (2,5)^\top$: (a) Decide if $b \in \mathrm{span}(A)$. (b) If solvable, find x_0 and parametrize all solutions; report the dimension. (c) State a condition on b under which $Ax = b$ would have a unique solution.

Definition 2.14. 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 2.15 (Invertibility conditions)**.** *The following are equivalent for $A \in \mathbb{R}^{n \times n}$:*

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

Proof. ($1 \Rightarrow 4$) If A is invertible, then for any $b \in \mathbb{R}^n$, $x = A^{-1}b$ is the unique solution to $Ax = b$.

($4 \Rightarrow 3$) If for all $b \in \mathbb{R}^n$, $Ax = b$ has a unique solution, then in particular $Ax = 0$ has only the trivial solution $x = 0$, so $\mathrm{null}(A) = \{0\}$.

($3 \Rightarrow 2$) If $\mathrm{null}(A) = \{0\}$, then by rank-nullity, $\mathrm{Rank}(A) + \dim(\mathrm{null}(A)) = n$ implies $\mathrm{Rank}(A) = n$.

($2 \Rightarrow 1$) If $\mathrm{Rank}(A) = n$, then the columns of A span $\mathbb{R}^n$. Thus for any $b \in \mathbb{R}^n$, there exists a solution to $Ax = b$. Since $\mathrm{Rank}(A) = n$, $\dim(\mathrm{null}(A)) = 0$, so the solution is unique. Hence (4) holds, which we already showed implies (1). □

2.6 Key concepts

- 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.