Learning objectives.

1. Define a convex set using containment of line segments.
2. Prove convexity directly and recognize convexity from sublevel sets, intersections, and affine transformations.
3. Disprove convexity with two feasible points whose line segment leaves the set.
4. Compute elementary Euclidean projections from a formula sheet.
5. Perform one projected-gradient step when the projection is available in closed form.

1 Recap: Convex optimization

Last time, we defined convex optimization as

$$\min f_0(x) \text{ s.t. } f_1(x) \le 0, \dots, f_m(x) \le 0.$$

If $f_0, f_1, \dots, f_m$ are *convex*, then **every local minimum is global**. In many cases, the global minima can be efficiently found.

Question 1. What is the definition of a convex function?

Solution:

$$\begin{aligned} & f(x) \text{ is convex} \\ \iff & f(tx + (1-t)y) \le tf(x) + (1-t)f(y) \\ & \quad \text{for all } 0 \le t \le 1, \text{ all } x, y \\ \iff & \nabla^2 f(x) \succeq 0 \quad \text{for all } x \\ \iff & \lambda_i(\nabla^2 f(x)) \ge 0 \quad \text{for all } i, \text{ all } x. \end{aligned}$$

Recall the *positive semidefinite* notation

$$A \succeq 0 \iff z^T A z \ge 0 \text{ for all } z.$$

2 Convexity of Feature Regression

The convexity of polynomial regression is a special case of the following more general result with polynomial features $\phi_j(x) = x^j$.

Theorem 1. Feature regression. Given data (x_i, y_i) for $i = 1, \ldots, m$, define the regression function

$$p_\theta(x) = c_0\phi_0(x) + c_1\phi_1(x) + \cdots + c_d\phi_d(x),$$

where $\theta = (c_0, c_1, \ldots, c_d)$ are the fit coefficients, and define the least-squares fit

$$f(\theta) = \frac{1}{2}\sum_{i=1}^{m}(p_\theta(x_i) - y_i)^2.$$

Then, f is a convex function of θ.

Proof: We observe that p_θ is *linear* with respect to θ, so each

$$p_\theta(x_i) = \begin{bmatrix}\phi_0(x_i) & \phi_1(x_i) & \cdots & \phi_d(x_i)\end{bmatrix}\begin{bmatrix}c_0 \\ c_1 \\ \vdots \\ c_d\end{bmatrix} = a_i^T\theta.$$

Substituting and taking derivatives

$$f(\theta) = \frac{1}{2}\sum_{i=1}^{m}(a_i^T\theta - y_i)^2,$$

$$\nabla f(\theta) = \sum_{i=1}^{m} a_i(a_i^T\theta - y_i), \qquad \nabla^2 f(\theta) = \sum_{i=1}^{m} a_i a_i^T.$$

We conclude that f is convex via the second-order test

$$z^T\nabla^2 f(\theta) z = \sum_{i=1}^{m} z^T a_i a_i^T z = \sum_{i=1}^{m}(a_i^T z)^2 \geq 0.$$

$\square$

Example 2. The is the following problem convex?

$$\min \quad x_1 + x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 \le 1$$

Solution: Yes. We have $\min\{f_0(x) : f_1(x) \le 0\}$ where

$$f_0(x_1, x_2) = x_1 + x_2$$
$$f_1(x_1, x_2) = x_1^2 + x_2^2 - 1$$

and by the second-order test, we have

$$\nabla^2 f_0(x_1, x_2) = 0, \qquad \nabla^2 f_1(x_1, x_2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so $\lambda_i(\nabla^2 f_0) = 0$ and $\lambda_i(\nabla^2 f_1) = 1 \ge 0$.

3 Solving constrained as unconstrained

We have so far focused on unconstrained. But convex optimization is interesting and powerful primarily because of constraints. Such problems can be converted back into unconstrained. Consider

$$\min_{x_1,x_2} f(x_1,x_2) \text{ s.t. } x_1^2+x_2^2 \le 1.$$

Option 1. Quadratic penalty method

$$F_\mu(x) \coloneqq f(x) + \frac{1}{2\mu}[(x_1^2+x_2^2-1)_+]^2$$

where $(t)_+ = \max\{t,0\}$.

Option 2. Logarithmic barrier method

$$F_\mu(x) \coloneqq f(x) - \mu \log(1-x_1^2-x_2^2).$$

Question 2. What is the difference?

Solution:

- Penalty is more permissive / optimistic; allows constraint violation.
- Barrier is stricter / pessimistic; requires strict feasibility and prevents crossing the constraint boundary.

Question 3. What does the parameter μ do?

Solution:

- Larger μ gives a stronger barrier; smaller μ gives a weaker barrier.
- As $\mu \to 0$, the barrier solution can approach the constraint boundary.

Question 4. Why doesn't this "solve" constrained optimization?

Solution:

- For every $\mu > 0$, the barrier solution remains strictly feasible and generally does not equal a boundary optimum. To approach the constrained optimum, decrease $\mu \to 0$.
- How do we find a strictly feasible starting point?

4 Today's motivation: Projected GD

More abstractly, we consider

$$\min_x f(x) \text{ subject to } x \in C.$$

Our previous definition used

$$C = \{x \in \mathbb{R}^n : f_1(x) \le 0, \ldots, f_m(x) \le 0\}.$$

Sometimes, we can efficiently project onto C. Then, can use projected GD

$$x_+ = \text{proj}_C(x - \alpha \nabla f(x))$$

to find a provable local min.

Question 5. Why could ProjGD be better than penalty/barrier?

Solution: ProjGD always satisfies constraints exactly. Penalty and barrier methods never perfectly satisfy the constraints. Penalty always slightly violate them, while barrier are always more conservative.

5 Convex sets

The constrained problem is *convex* if f is convex and the set C is also *convex*. In that case, the provable local min is also global min.

A set $C \subseteq \mathbb{R}^n$ is convex if

$$x, y \in C, \quad 0 \le t \le 1 \qquad \Longrightarrow \qquad tx + (1-t)y \in C.$$

In words, the line segment between any two points in the set stays inside the set.

Example 3. Prove directly that

$$C = \{x \in \mathbb{R}^2 : x_1 + x_2 \le 1\} \text{ is convex.}$$

Solution: Take arbitrary $u, v \in C$ and $0 \le t \le 1$. Feasibility gives

$$u_1 + u_2 \le 1, \qquad v_1 + v_2 \le 1.$$

For the convex combination,

$$\begin{aligned}[tu + (1-t)v]_1 + [tu + (1-t)v]_2 &= t(u_1 + u_2) + (1-t)(v_1 + v_2) \\ &\le t + (1-t) = 1.\end{aligned}$$

Therefore $tu + (1-t)v \in C$, so C is convex.

Theorem 4. If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then every sublevel set

$$C_\alpha = \{x \in \mathbb{R}^n : f(x) \le \alpha\} \text{ is convex.}$$

Converse is false; functions with convex sublevel set are called *quasi-convex*, and are not necessarily convex.

Proof: Indeed, if $f(x), f(y) \le \alpha$, then

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y) \le \alpha.$$

□

Example 5. The same set as a sublevel set. Prove again that

$$C = \{x \in \mathbb{R}^2 : x_1 + x_2 \le 1\} \text{ is convex,}$$

now using the sublevel-set rule.

Solution: Let $f(x) = x_1 + x_2$. This function is affine, $\nabla^2 f(x) = 0 \succeq 0$, so it is convex. Since

$$C = \{x : f(x) \le 1\},$$

the set C is a sublevel set of a convex function and is therefore convex.

When is the structural proof preferable to the direct proof? Use the structural proof when the set is visibly assembled from known convex functions or known convex sets. Use the direct proof when no recognition rule applies cleanly.

6 Projection onto a convex set

For a nonempty closed convex set C, the Euclidean projection of y is the unique closest point

$$\mathrm{proj}_C(y) = \arg\min_{x \in C} \frac{1}{2} \|x - y\|_2^2 .$$

For this lecture, projection is a formula to execute, not an algorithm to analyze.

Box $[\ell, u]$. Projection clamps each coordinate:

$$[\mathrm{proj}_{[\ell,u]}(y)]_i = \min\{u_i, \max\{\ell_i, y_i\}\}.$$

Nonnegative orthant.

$$[\mathrm{proj}_{\mathbb{R}^n_+}(y)]_i = \max\{y_i, 0\}.$$

Euclidean ball $\|x\|_2 \le r$.

$$\mathrm{proj}_C(y) = \begin{cases} y, & \|y\|_2 \le r, \\ r\, y / \|y\|_2 , & \|y\|_2 > r. \end{cases}$$

Question 6. What does projection do when the tentative point is already feasible?

Solution: Nothing: if $y \in C$, then $\mathrm{proj}_C(y) = y$.

7 Projected gradient descent

For $\min_{x \in C} f(x)$, one projected-gradient step is

$$\underbrace{y = x - \alpha \nabla f(x),}_{\text{ordinary gradient step}} \qquad \underbrace{x^+ = \operatorname{proj}_C(y)}_{\text{project back}}.$$

The procedure is always: gradient, tentative step, projection.

Example 6. One projected-gradient step. Let

$$f(x) = \frac{1}{2} \|x - c\|_2^2, \qquad c = (2, -1), \qquad C = [0, 1]^2.$$

Starting from $x_0 = (1/2, 1/2)$, take one projected-gradient step with $\alpha = 1$.

Solution: The gradient is

$$\nabla f(x) = x - c.$$

At x_0,

$$\nabla f(x_0) = (1/2, 1/2) - (2, -1) = (-3/2, 3/2).$$

The tentative step is

$$y = x_0 - \nabla f(x_0) = (2, -1).$$

Clamping both coordinates to $[0, 1]$ gives

$$x_1 = \operatorname{proj}_{[0,1]^2}(2, -1) = (1, 0).$$

Example 7. Nonnegative matrix factorization. Suppose that we have a symmetric $n \times n$ matrix $M = M^T$ that we believe factors as xx^T with $x \geq 0$ elementwise. Derive a projected gradient descent method for computing these nonnegative factors.

Solution: The objective is

$$f(x) = \frac{1}{2}\|xx^T - M\|_F^2.$$

We compute the new candidate step

$$x - \eta \nabla f(x) = x - 2\eta(xx^T - M)x,$$

and project it onto the nonnegative orthant by clamping

$$x_+ = \max\{x - 2\eta(xx^T - M)x, 0\}.$$

Question 7. Why don't we use projected GD everywhere?

Solution: Most sets don't have tractable projections. The projection might be as difficult as the original problem.

8 Basic rules for convex sets

The following constructions preserve convexity:

1. **Intersections.** If every C_i is convex, then $\bigcap_i C_i$ is convex.
2. **Affine images.** If C is convex, then
   $AC + b = \{Ax + b : x \in C\}$ is convex.
3. **Affine preimages.** If D is convex, then
   $\{x : Ax + b \in D\}$ is convex.

Example 8. Triangular sector. Prove that

$$C = \{x \in \mathbb{R}^2 : x_1 \geq 0,\ x_2 \geq 0,\ x_1 + x_2 \leq 1\} \text{ is convex.}$$

Solution: Write

$$C = \{x : -x_1 \leq 0\} \cap \{x : -x_2 \leq 0\} \cap \{x : x_1 + x_2 \leq 1\}.$$

Each set is a halfspace, hence a sublevel set of an affine function. Each set is convex, and an intersection of convex sets is convex. Therefore C is convex.

Example 9. Polytope. For any $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$, prove that

$$C = \{x \in \mathbb{R}^n : Ax \leq b\} \text{ is convex.}$$

Solution: The set C is the intersection of halfspaces

$$C = \bigcap_{i=1}^{m} \{x : a_i^T x \leq b_i\}.$$

Each half space is convex. Alternatively, C is the affine preimage of the negative orthant $\{z \in \mathbb{R}^m : z \leq 0\}$.

Example 10. Circular sector. Prove that

$$C = \{x \in \mathbb{R}^2 : x_1 \geq 0,\ x_2 \geq 0,\ x_1^2 + x_2^2 \leq 1\} \text{ is convex.}$$

Solution: Write

$$C = \{x : -x_1 \leq 0\} \cap \{x : -x_2 \leq 0\} \cap \{x : x_1^2 + x_2^2 \leq 1\}.$$

First two sets are halfspaces. Third set is the sublevel set of $f(x) = \frac{1}{2}\|x\|^2$, whose Hessian is $\nabla^2 f(x) = I \succeq 0$. Each set is convex, and an intersection of convex sets is convex. Therefore C is convex.

Example 11. Ellipsoid. For any $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$, prove that

$$C = \{x \in \mathbb{R}^n : \|Ax - b\| \leq 1\} \text{ is convex.}$$

Solution: The set C is the affine preimage of the Euclidean ball $B = \{y \in \mathbb{R}^m : \|y\| \leq 1\}$.

Alternatively, the set C is the sublevel set of $f(x) = \frac{1}{2}\|Ax - b\|^2$, which is convex because $\nabla^2 f(x) = A^T A \succeq 0$.

Example 12. Convex hull. Given m points $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$, the convex hull

$$\text{convx}\{v_1, \ldots, v_m\} = \{\sum_{i=1}^m \lambda_i v_i : \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1\} \text{ is convex.}$$

Solution: The simplex $\{\lambda \geq 0 : \sum_i \lambda_i = 1\}$ is convex, and $\lambda \mapsto \sum_{i=1}^m \lambda_i v_i$ is linear. Therefore, the convex hull is the affine image of a convex set.

Example 13. Zonotope. For any $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$, prove that

$$C = \{Ax + b : \|x\|_\infty \leq 1\} \text{ is convex.}$$

Solution: The set C is the affine image of the ℓ_∞ ball $B = \{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}$. In turn,

$$\|x\|_\infty \leq 1 \iff -1 \leq x_i \leq 1 \text{ for all } i,$$

so $B = \{x \in \mathbb{R}^n : Gx \leq h\}$ where $G = [I; -I]$ and $h = \mathbf{1}$. The ℓ_∞ ball is just a polytope, and is therefore convex.

9 Disproving convexity

How to *disprove* convexity? Find two points $x, y \in C$ and one $0 < t < 1$ such that $tx + (1-t)y \notin C$. A midpoint, corresponding to $t = 1/2$, is often the easiest choice.

Example 14. Show that

$$C = \{x \in \mathbb{R}^2 : \|x\|_2 = 1\} \text{ is } \textbf{not} \text{ convex.}$$

Solution: The points

$$u = (1, 0), \qquad v = (-1, 0)$$

belong to C. Their midpoint is

$$\frac{u+v}{2} = (0, 0), \qquad \|(0, 0)\|_2 = 0 \neq 1.$$

Therefore the line segment leaves the set, so C is not convex.

Question 8. Why does the equality $\|x\|_2 = 1$ behave differently from $\|x\|_2 \leq 1$?

Solution: The inequality describes a **sublevel set** of the convex norm and includes the interior. The equality keeps only the **boundary**, and a chord between boundary points can pass through the missing interior.