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

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

If $f_0, f_1, \ldots, f_m$ 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?

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.

Example 2. The is the following problem convex?

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

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

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

Question 2. What is the difference?

Question 3. What does the parameter μ do?

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

4 Today's motivation: Projected GD

More abstractly, we consider

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

Our previous definition used

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

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

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

to find a provable local min.

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

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

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

Example 3. Prove directly that

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

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.

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.

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?

7 Projected gradient descent

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

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

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 $\alpha = 1$.

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.

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

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

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

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

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

Example 10. Circular sector. Prove that

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

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

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

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

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

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

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

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

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

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