Guide to Convexity in Optimization

Mikhailova Olena

20.03.2025

1 Introduction to Convexity

Convexity is one of the most important concepts in optimization because it allows us to:

- Guarantee that local minima are global minima
- Develop efficient optimization algorithms
- Analyze problems theoretically

1.1 What Does "Convex" Mean?

Imagine a simple bowl shape - this is convex. A saddle shape is non-convex. The key property is that if you draw a line between any two points on a convex object, the line stays entirely within the object.

2 Convex Sets

2.1 Basic Definition

Definition 2.1 (Convex Set). *A set $C \subseteq \mathbb{R}^n$ is **convex** if the line segment between any two points in C lies entirely in C:*

$$\forall x, y \in C, \forall \theta \in [0, 1] : \theta x + (1 - \theta)y \in C$$

2.2 Advanced Examples

- **Linear Matrix Inequality Solution Set**:

  $$\left\{ x \in \mathbb{R}^k : \sum_{i=1}^{k} x_i A_i \preceq B \right\}$$

  where $A_i, B \in \mathbb{S}^n$. Convexity follows because it's an affine pre-image of the PSD cone.

- **Conditional Probability Distributions**: The set of conditional distributions derived from convex joint distributions remains convex through linear-fractional transformations.

2.3 Key Theorems

Theorem 2.2 (Separating Hyperplane). *For disjoint convex sets C, D, there exists a, b such that:*

$$C \subseteq \{x : a^\top x \leq b\}, \quad D \subseteq \{x : a^\top x \geq b\}$$

Theorem 2.3 (Solution Set Convexity). *For convex optimization problems, the set of optimal solutions X_{opt} is convex.*

Proof. Let $x, y \in X_{\text{opt}}$. For $t \in [0, 1]$, define $z = tx + (1-t)y$. By convexity:

$$\begin{aligned} g_i(z) &\leq t g_i(x) + (1-t) g_i(y) \leq 0 \\ Az &= tAx + (1-t)Ay = b \\ f(z) &\leq t f(x) + (1-t) f(y) = f^* \end{aligned}$$

Thus $z \in X_{\text{opt}}$. □

2.4 Examples of Convex Sets

- **Norm balls**: All points within a certain distance from center

$$\{x : \|x\| \leq r\}$$

- **Hyperplanes and halfspaces**:

$$\{x : a^\top x = b\}, \quad \{x : a^\top x \leq b\}$$

- **Polyhedrons**: Solutions to systems of linear inequalities

$$\{x : Ax \leq b, Cx = d\}$$

- **Positive semidefinite cone** (for matrices):

$$\mathbb{S}^n_+ = \{X \in \mathbb{R}^{n \times n} : X = X^\top, X \succeq 0\}$$

2.5 Operations that Preserve Convexity

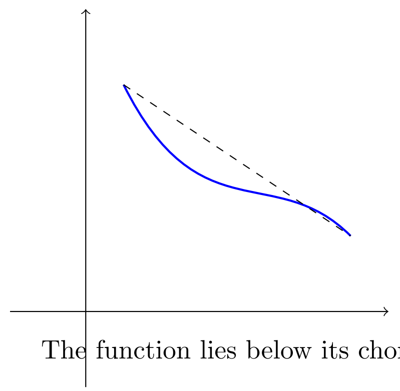
- **Intersection**: If C_1 and C_2 are convex, then $C_1 \cap C_2$ is convex
- **Affine transformations**: For convex C, the set $\{Ax + b : x \in C\}$ is convex
- **Linear-fractional transformations**: For convex C, the set $\left\{ \frac{Ax+b}{c^\top x + d} : x \in C \right\}$ is convex

3 Convex Functions

3.1 Basic Definition

Definition 3.1 (Convex Function). *A function $f : \mathbb{R}^n \to \mathbb{R}$ is* ***convex*** *if its domain is convex and:*

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \quad \forall x, y \in dom(f), \theta \in [0,1]$$



The function lies below its chords

3.2 Extended Properties

Theorem 3.2 (Partial Minimization). *If $g(x, y)$ is convex and C is convex, then $f(x) = \min_{y \in C} g(x, y)$ is convex.*

- **Composition Rules**:
  - Convex + non-decreasing $\circ$ convex = convex
  - Convex + non-increasing $\circ$ concave = convex
  - Affine composition preserves convexity: $f(Ax + b)$

3.3 Important Properties

Theorem 3.3 (Epigraph Characterization). *f is convex $\iff$ its epigraph $epi(f) = \{(x, t) : f(x) \leq t\}$ is convex.*

Theorem 3.4 (First-Order Characterization). *For differentiable f:*

$$f \textit{ convex} \iff f(y) \geq f(x) + \nabla f(x)^\top (y - x) \quad \forall x, y$$

Theorem 3.5 (Second-Order Characterization). *For twice differentiable f:*

$$f \textit{ convex} \iff \nabla^2 f(x) \succeq 0 \quad \forall x$$

3.4 Important Function Classes

- **Indicator Function**:

$$I_C(x) = \begin{cases} 0 & x \in C \\ \infty & \text{otherwise} \end{cases}$$

  Convex when C is convex

- **Support Function**:

$$I_C^*(x) = \sup_{y \in C} x^\top y$$

  Always convex (even for non-convex C)

3.5 Examples of Convex Functions

- **Affine functions**: $f(x) = a^\top x + b$
- **Quadratic functions**: $f(x) = \frac{1}{2}x^\top Qx + b^\top x + c$ (when $Q \succeq 0$)
- **Norms**: $\|x\|_p$ for $p \geq 1$
- **Exponential**: e^{ax}
- **Log-sum-exp**: $\log(\sum e^{x_i})$

3.6 Strong Convexity

Definition 3.6 (Strongly Convex Function)**.** *A function f is μ-**strongly convex** if:*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2}\|y - x\|^2$$

This means the function grows at least as fast as a quadratic function.

Theorem 3.7 (Hessian Characterization)**.** *For twice differentiable f:*

$$f \ \mu\text{-strongly convex} \iff \nabla^2 f(x) \succeq \mu I \quad \forall x$$

3.7 Algorithmic Implications

- Enables linear convergence rates
- Provides error bounds:

$$\frac{1}{2\mu}\|\nabla f(x)\|^2 \geq f(x) - f^*$$

- Guarantees unique global minimum
- Leads to faster convergence in optimization algorithms
- For twice differentiable f, equivalent to $\nabla^2 f(x) \succeq \mu I$

4 Convex Optimization Problems

4.1 Standard Form

A convex optimization problem has the form:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, ..., m \\ & a_j^\top x = b_j, \quad j = 1, ..., p \end{aligned}$$

where:

- f and g_i are convex functions
- Equality constraints are affine

4.2 Problem Transformations

- **Eliminating Equality Constraints**: Express $x = My + x_0$ where M spans nullspace of A
- **Introducing Slack Variables**: Convert $g_i(x) \leq 0$ to $g_i(x) + s_i = 0$ with $s_i \geq 0$
- **Geometric Programming**: Non-convex posynomials become convex after log-transform:

$$\min_{y} \log \sum e^{a_k^\top y + b_k} \quad \text{s.t.} \quad \log \sum e^{c_i^\top y + d_i} \leq 0$$

4.3 Key Properties

Theorem 4.1 (Global Optimality)**.** *For convex problems, any local minimum is global.*

Theorem 4.2 (Solution Uniqueness)**.** *If f is strictly convex, the solution is unique.*

Theorem 4.3 (Optimality condition)**.** *: For differentiable f, x^* is optimal iff*

$$\nabla f(x^*)^\top (y - x^*) \geq 0 \quad \forall y \textit{ feasible}$$

4.4 Solution Characteristics

- **Feasibility**: A point x is feasible if $x \in \bigcap \text{dom}(g_i)$, $g_i(x) \leq 0$, and $Ax = b$
- **Active Constraints**: Inequality g_i is active at x if $g_i(x) = 0$
- **ϵ-Suboptimality** : x is ϵ-suboptimal if $f(x) \leq f^* + \epsilon$

4.5 Important Examples

- **Linear Programs (LP)**:

$$\min_x c^\top x \quad \text{s.t.} \quad Ax \leq b$$

- **Quadratic Programs (QP)**:

$$\min_x \frac{1}{2} x^\top Q x + c^\top x \quad \text{s.t.} \quad Ax \leq b$$

- **Semidefinite Programs (SDP)**:

$$\min_X \langle C, X \rangle \quad \text{s.t.} \quad \mathcal{A}(X) = b, X \succeq 0$$

5 Applications

5.1 Lasso Regression

$$\min_\beta \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

- Convex despite non-differentiable ℓ_1-norm
- Promotes sparse solutions

5.2 Robust Lasso Variant

With Huber loss for outlier resistance:

$$\min_\beta \sum_{i=1}^n \rho(y_i - x_i^\top \beta) + \lambda \|\beta\|_1$$

where $\rho(z) = \begin{cases} \frac{1}{2}z^2 & |z| \leq \delta \\ \delta|z| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}$. Remains convex as Huber loss is convex.

5.3 Support Vector Machines

$$\min_{w,b} \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^n \max(0, 1 - y_i(w^\top x_i + b))$$

- Convex objective with piecewise linear constraints
- Can be rewritten as a quadratic program

5.4 SVM Duality

The dual SVM formulation reveals convex structure:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^\top x_j \quad \text{s.t.} \quad 0 \leq \alpha_i \leq C$$

Demonstrates convex quadratic programming structure.

6 Why Convexity Matters

- **Reliable optimization**: No need to worry about getting stuck in local minima
- **Efficient algorithms**: Specialized methods (gradient descent, interior point) work well
- **Theoretical guarantees**: Can prove convergence rates and complexity bounds
- **Recognizable structure**: Many problems can be reformulated as convex programs

7 Algorithmic Implications

7.1 Why Convexity Helps Optimization

- **No Spurious Local Minima**: Gradient descent won't get stuck in poor solutions
- **Predictable Convergence**:
  - $O(1/\epsilon)$ iterations for convex + smooth
  - Linear convergence ($O(\log 1/\epsilon)$) for strongly convex
- **Simple Stopping Criteria**: $\|\nabla f(x)\| < \epsilon$ guarantees near-optimality

7.2 Hierarchy of Convex Programs

Linear Programs (LP) → Quadratic Programs (QP) → Semidefinite Programs (SDP) → Convex Programs (CP)

Each class extends the previous one while maintaining convexity properties.

8 Recognizing Convexity

To verify if a problem is convex:

1. Check if the objective is convex (use definitions or derivative tests)
2. Verify all inequality constraints are convex
3. Ensure equality constraints are affine
4. Confirm the domain is convex

- **Significant Counterexample**: Geometric programs appear non-convex but become convex under logarithmic transformation

9 Common Mistakes

- Assuming all quadratics are convex (must check $Q \succeq 0$)
- Forgetting to verify convexity of constraints
- Treating equality constraints as inequalities
- Overlooking domain issues (e.g., log(x) requires x ¿ 0)

10 Conclusion

Convexity provides a powerful framework for optimization problems:

- Clear geometric interpretation
- Strong theoretical guarantees
- Wide range of applications
- Systematic approach to problem formulation