Convex Optimization

Yunlong Song (宋运龙)

Why don't scientists trust atoms anymore? Because they make up everything! — ChatGPT, tell me a joke

Convex Optimization

Convex Optimization is a subfield of optimization that deals with finding the minimum of a convex function over a convex set. We summarize the key concepts of convex optimization below by following the book by Boyd and Vandenberghe.

Convex Sets

Convex Sets are sets that contain the line segment connecting any two points in the set. A set S is convex if for any two points $x_1, x_2 \in S$ and any $\lambda \in [0, 1]$, the point $\lambda x_1 + (1 - \lambda)x_2 \in S$.

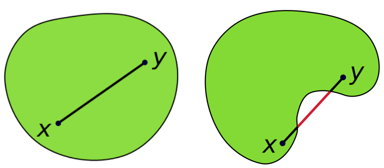

Figure 1: Convex Set vs. Non-Convex Set

Important examples of convex sets include:

- **Empty set:** The set containing no points is convex.
- **Hyperplane:** A set of the form $\{x \mid a^T x = b\}$.
- **Halfspace:** A set of the form $\{x \mid a^T x \leq b\}$.
- **Euclidean ball and ellipsoids:** A set of the form $\{x \mid \|x - x_c\|_2 \leq r\}$.
- **Norm balls and norm cones:** A set of the form $\{x \mid \|x - x_c\|_p \leq r\}$.
- **Polyhedra:** A set of the form $\{x \mid Ax \preceq b, Cx = d\}$.
- **Positive semidefinite cone:** A set of the form $\{X \in \mathbb{R}^{n \times n} \mid X \succeq 0\}$.

Operations that preserve convexity include:

- **Intersection:** The intersection of convex sets is convex.
- **Affine functions:** The image of a convex set under an affine function is convex.
- **Linear functions:** The image of a convex set under a linear function is convex.
- **Perspective functions:** The perspective of a convex function is convex.

Convex Functions

Convex Functions are functions that have a non-negative second derivative. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for any two points $x_1, x_2 \in \mathbb{R}^n$ and any $\theta \in [0, 1]$, the inequality holds:

$$f(\theta x_1 + (1 - \theta)x_2) \leq \theta f(x_1) + (1 - \theta)f(x_2)$$

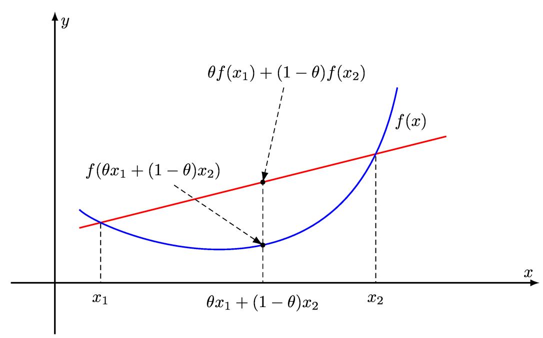

Figure 2: Convex Function

Important examples of convex functions include:

- **Affine functions:** A function of the form $f(x) = Ax + b$.
- **Quadratic functions:** A function of the form $f(x) = x^T Px + q^T x + r$.
- **Norm functions:** A function of the form $f(x) = \|x\|_p$.
- **Exponential functions:** A function of the form $f(x) = e^{a^T x + b}$.
- **Logarithmic functions:** A function of the form $f(x) = -\log(x)$.
- **Entropy functions:** A function of the form $f(x) = x \log(x)$.

Jensen's inequality: For a convex function f, the inequality holds:

$$f\left(\sum_{i=1}^{n} \theta_i x_i\right) \leq \sum_{i=1}^{n} \theta_i f(x_i)$$

where $\theta_i \geq 0$ and $\sum_{i=1}^{n} \theta_i = 1$.

First-order condition: A function f is convex if and only if for any $x, y \in \mathrm{dom}(f)$, the inequality holds:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

Second-order condition: A function f is convex if and only if its Hessian is positive semidefinite:

$$\nabla^2 f(x) \succeq 0$$

Operations that preserve convexity:

- *Nonnegative weighted sum*
- *Composition with an affine function*
- *Pointwise maximum*
- *Perspective functions*

e.g., Pointwise maximum: The pointwise maximum of m convex functions is convex:

$$f(x) = \max_{i=1}^{m} f_i(x)$$

Convex Optimization Problem

A convex optimization problem is a problem of the form:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned} \tag{1}$$

where f is a convex function, g_i are *convex functions*, and $h_i = a_i^T x - b_i$ are *affine functions*. The problem is convex if the objective function and the constraints are convex functions and the feasible set is a convex set. Convex optimization problems have the property that any local minimum is also a global minimum.

Local and global optima

- A point x is a **local minimum** if there exists a neighborhood around x such that $f(x) \leq f(y)$ for all y in the neighborhood.
- A point x is a **global minimum** if $f(x) \leq f(y)$ for all y in the feasible set.

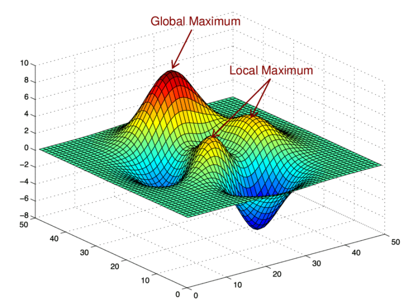

Figure 3: Local and global optima

Optimality Conditions

First-order optimality condition: If x is a local minimum of a convex function f, then the gradient of f at x is zero:

$$\nabla f(x) = 0$$

Second-order optimality condition: If x is a local minimum of a convex function f, then the Hessian of f at x is positive semidefinite:

$$\nabla^2 f(x) \succeq 0$$

Linear Programming (LP)

A linear programming problem is a convex optimization problem where the objective function and the constraints are linear functions:

$$
\begin{aligned}
\min_{x} \quad & c^T x + d \\
\text{s.t.} \quad & Gx \preceq h \\
& Ax = b
\end{aligned} \tag{2}
$$

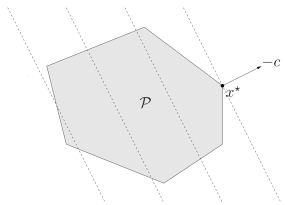

Figure 4: Geometric interpretation of a linear programming problem.

Quadratic Programming (QP)

A quadratic optimization problem is a convex optimization problem where the objective function is a quadratic function and the constraints are linear functions:

$$
\begin{aligned}
\min_{x} \quad & \frac{1}{2} x^T P x + q^T x + r \\
\text{s.t.} \quad & Gx \preceq h \\
& Ax = b
\end{aligned} \tag{3}
$$

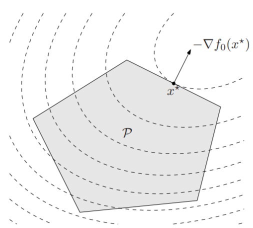

Figure 5: Geometric interpretation of a quadratic programming problem.

Geometric Programming (GP)

A geometric programming problem is a convex optimization problem where the objective function and the constraints are posynomials:

$$
\min_{x} \quad f_0(x) \tag{4}
$$

$$
\text{s.t.} \quad f_i(x) \le 1, \quad i = 1, \ldots, m \tag{5}
$$

$$
h_i(x) = 1, \quad i = 1, \ldots, p \tag{6}
$$

where $f_0, f_1, \ldots, f_m$ are *posynomials* and $h_1, \ldots, h_p$ are *monomials.*

Duality

Standard form: A optimization problem can be written in the standard form (not necessarily convex) as:

$$
\begin{aligned}
\min_{x} \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, p
\end{aligned} \tag{7}
$$

The **Lagrangian** of a convex optimization problem is defined as:

$$
L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
$$

where λ_i and ν_i are the Lagrange multipliers associated with the inequality and equality constraints, respectively.
The **dual function** is defined as:

$$
g(\lambda, \nu) = \inf_{x} L(x, \lambda, \nu)
$$

The **dual problem** is defined as:

$$
\begin{aligned}
\max_{\lambda, \nu} \quad & g(\lambda, \nu) \\
\text{s.t.} \quad & \lambda \succeq 0
\end{aligned} \tag{8}
$$

Example: Least-norm problem Consider the least-norm problem:

$$
\begin{aligned}
\min_{x} \quad & x^T x \\
\text{s.t.} \quad & Ax = b
\end{aligned} \tag{9}
$$

The Lagrangian is:

$$
L(x, \nu) = x^T x + \nu^T (Ax - b)
$$

The dual function is:

$$
g(\nu) = \inf_{x} L(x, \nu) = -\frac{1}{4} \nu^T A^T A \nu - b^T \nu
$$

The dual problem is:

$$
\begin{aligned}
\max_{\nu} \quad & -\frac{1}{4} \nu^T A^T A \nu - b^T \nu \\
\text{s.t.} \quad & \nu \succeq 0
\end{aligned} \tag{10}
$$

The **weak duality theorem** states that the optimal value of the primal problem is always greater than or equal to the optimal value of the dual problem:

$$
p^* \ge d^*
$$

The **strong duality theorem** states that if the primal problem is convex and has a feasible solution, then the optimal value of the primal problem is equal to the optimal value of the dual problem:

$$
p^* = d^*
$$

KKT Conditions

Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient conditions for a point to be a solution to a convex optimization problem. The KKT conditions for a convex optimization problem are:

$$
\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0
$$

$$
\begin{aligned}
f_i(x) \le 0, \quad & i = 1, \ldots, m \\
h_i(x) = 0, \quad & i = 1, \ldots, p \\
\lambda_i \ge 0, \quad & i = 1, \ldots, m \\
\lambda_i f_i(x) = 0, \quad & i = 1, \ldots, m
\end{aligned}
$$

where λ_i and ν_i are the Lagrange multipliers associated with the inequality and equality constraints, respectively.

Numerical Linear Algebra

Matrix Factorization/Decomposition: Matrix factorization is the process of decomposing a matrix into the product of two or more matrices. Common matrix factorizations include:

- **LU decomposition:** A matrix A is decomposed into the product of a lower triangular matrix L and an upper triangular matrix U.
- **QR decomposition:** A matrix A is decomposed into the product of an orthogonal matrix Q and an upper triangular matrix R.
- **Cholesky decomposition:** A symmetric positive definite matrix A is decomposed into the product of a lower triangular matrix L and its transpose.
- **Singular value decomposition (SVD):** A matrix A is decomposed into the product of three matrices U, Σ, and V^T.

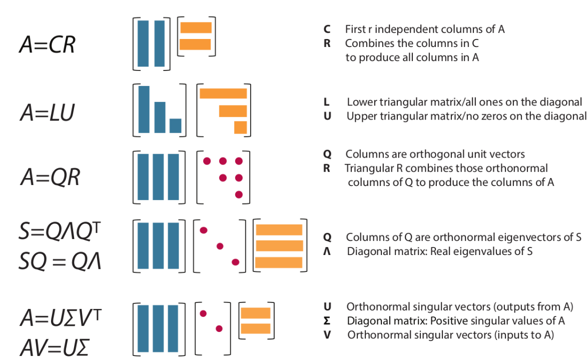

Figure 6: Matrix Factorization

Unconstrained Optimization

Unconstrained optimization problem:

$$\min_{x} \quad f(x) \tag{11}$$

we can assume that f is a convex function and twice continuously differentiable.

Gradient Descent

The gradient descent algorithm is an iterative optimization algorithm that updates the current solution x in the direction of the negative gradient of the objective function:

$$x^{(k+1)} = x^{(k)} - \alpha \nabla f(x^{(k)})$$

where α is the step size.

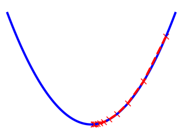

Figure 7: Gradient Descent

Newton's Method

Given a twice continuously differentiable function f, we seek to solve the optimization problem:

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{12}$$

Newton's method is an iterative optimization algorithm that updates the current solution x using the second-order Taylor expansion of the objective function:

$$f(x + \Delta x) \approx f(x) + \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(x) \Delta x$$

The update rule for Newton's method is:

$$x^{(k+1)} = x^{(k)} - \left(\nabla^2 f(x^{(k)})\right)^{-1} \nabla f(x^{(k)})$$

The geometric interpretation of Newton's method is shown in the figure below.

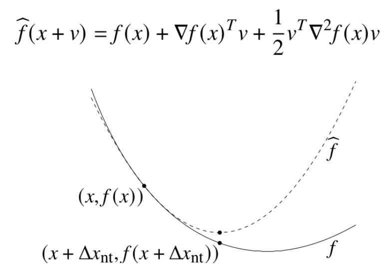

Figure 8: Newton's Method

Equality Constrained Optimization

Equality constrained optimization problem:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & Ax = b \end{aligned} \tag{13}$$

where f is a convex function and twice continuously differentiable.

Optimality conditions: The **KKT conditions** for the problem are:

$$\begin{aligned} \nabla f(x) + A^T \nu = 0 \\ Ax = b \end{aligned} \tag{14}$$

Inequality Constrained Optimization

We discuss *interior-point methods* for solving convex optimization problems with inequality constraints:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned} \tag{15}$$

where $f_0, f_1, \ldots, f_m$ are convex and twice continuously differentiable functions. We assume that the problem is feasible, i.e., there exists a point x that satisfies the constraints.

Barrier Method

The **barrier method** is an optimization algorithm that solves a sequence of barrier subproblems to approximate the solution of the original optimization problem. The barrier subproblem is defined as:

$$\begin{aligned} \min_{x} \quad & f_0(x) - \mu \sum_{i=1}^{m} \log(-f_i(x)) \\ \text{s.t.} \quad & Ax = b \end{aligned} \tag{16}$$

where $\mu > 0$ is the barrier parameter.

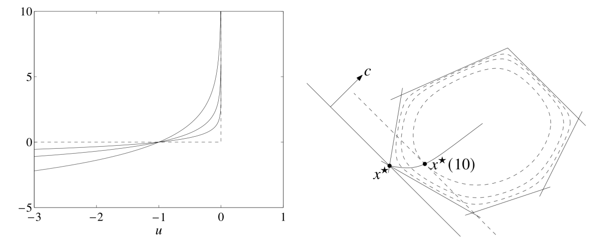

Figure 9: Barrier Method

Interior-Point Method

Interior-point methods are a class of optimization algorithms that solve convex optimization problems by iteratively moving towards the interior of the feasible set while maintaining feasibility. IPMs combine two advantages: they have a polynomial-time complexity and they can handle large-scale problems.

We denote the optimal value $f_0(x^*)$ as p^*. The **KKT conditions** for the problem are:

$$\begin{aligned} \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T \nu = 0 \\ f_i(x) \leq 0, \quad i = 1, \ldots, m \\ \lambda_i \geq 0, \quad i = 1, \ldots, m \\ \lambda_i f_i(x) = 0, \quad i = 1, \ldots, m \\ Ax = b \end{aligned} \tag{17}$$

where λ_i and ν are the Lagrange multipliers associated with the inequality and equality constraints, respectively.

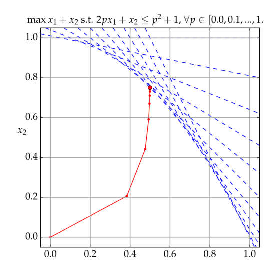

Figure 10: Interior-Point Method. Blue lines show constraints, red points show iterated solutions. (Source: Wikipedia)