Convex Optimization Theory

Daniel P. Palomar (2025). *Portfolio Optimization: Theory and Application.* Cambridge University Press.

portfoliooptimizationbook.com

Latest update: 2025-09-02

Outline

Executive Summary

- Recent advancements in convex optimization theory and numerical methods have led to numerous applications in engineering, finance, and machine learning.
- Key references include (Bertsekas 1999; Nemirovski 2001; Bertsekas, Nedić, and Ozdaglar 2003; S. P. Boyd and Vandenberghe 2004; Nesterov 2018), with classic works by (Luenberger 1969; Rockafellar 1970) and engineering applications in (Palomar and Eldar 2009).
- Rockafellar's 1993 survey (Rockafellar 1993) emphasized that the key distinction in optimization is between **convexity** and **nonconvexity**.
- Convex problems can be solved optimally either in closed form or numerically with efficient algorithms (S. P. Boyd and Vandenberghe 2004).
- However, most practical problems are initially not convex, requiring practitioners to uncover hidden convexity.
- These slides provide a concise introduction to convex optimization theory, see also (Palomar 2025, Appendix A).
- Readers are encouraged to consult the comprehensive resource (S. P. Boyd and Vandenberghe 2004) for more details.

Outline

Optimization Problem

Mathematical Optimization Problem: Can be written in the following standard form (S. P. Boyd and Vandenberghe 2004, chap. 1):

$$\begin{array}{lll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) & \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, & i = 1, \dots, m \\ & h_i(\boldsymbol{x}) = 0, & i = 1, \dots, p \end{array}$$

Components:

- Optimization variable: $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$
- Objective function: $f_0 : \mathbb{R}^n \longrightarrow \mathbb{R}$
- Inequality constraint functions: $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}, \; i = 1, \dots, m$
- Equality constraint functions: $h_i : \mathbb{R}^n \longrightarrow \mathbb{R}, \; i = 1, \dots, p$

Unconstrained Problem: If there are no constraints, the problem is called *unconstrained*.

Goal: Find an optimal solution $\boldsymbol{x}^\star$ that minimizes f_0 while satisfying all the constraints.

Examples

Applications of Convex Optimization:

- Circuit design
- Filter design
- Communication systems (e.g., multi-antenna beamforming design, maximum likelihood detection)
- Radar systems
- Communication networks (e.g., power control in wireless networks, congestion control in the Internet)
- Financial engineering (e.g., portfolio design, index tracking)
- Model fitting (e.g., in financial data or recommender systems)
- Image processing (e.g., deblurring, compressive sensing, blind separation, inpainting)
- Robust designs under uncertainty
- Sparse regression, low-rank matrix discovery
- Machine learning, graph learning
- Biomedical applications (e.g., DNA sequencing, anti-viral vaccine design)

Examples

Basic Elements of an Optimization Problem:

- **Variables:** Differ from other fixed parameters.
- **Constraints:** Conditions that must be satisfied.
- **Objective:** The function to be minimized or maximized.

Example: Device Sizing for Electronic Circuits

- Variables: Widths and lengths of devices.
- Constraints: Manufacturing limits, timing requirements, or maximum area.
- Objective: Power consumption.

Example: Portfolio Design

- Variables: Amounts invested in different assets.
- Constraints: Budget, maximum investments per asset, or minimum return.
- Objective: Overall risk or return variance.

Definitions: Domain and Constraints

Domain of the Optimization Problem:

- Defined as the set of points for which the objective and all constraint functions are defined:

$$\mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i$$

- Interpreted as a set of implicit constraints $\boldsymbol{x} \in \mathcal{D}$.
- Contrasts with explicit constraints $f_i(\boldsymbol{x}) \leq 0$ and $h_i(\boldsymbol{x}) = 0$.

Unconstrained Problem:

- No explicit constraints.
- Example:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \log\left(a - \boldsymbol{b}^\top \boldsymbol{x}\right)$$

- Implicit constraint: $a > \boldsymbol{b}^\top \boldsymbol{x}$

Feasibility:

- A point $\boldsymbol{x}$ is **feasible** if it satisfies all the constraints, $f_i(\boldsymbol{x}) \leq 0$ and $h_i(\boldsymbol{x}) = 0$, and **infeasible** otherwise.
- A problem is **feasible** if there exists at least one feasible point and **infeasible** if no feasible points exist.

Optimal Value:

$$p^\star = \inf \{ f_0(\boldsymbol{x}) \mid f_i(\boldsymbol{x}) \leq 0,\ i = 1, \ldots, m,\ h_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, p \}$$

- If the problem is feasible, the optimal value may be achieved at an optimal solution $\boldsymbol{x}^\star$, i.e., $f_0(\boldsymbol{x}^\star) = p^\star$.
- The problem may be unbounded below, i.e., $p^\star = -\infty$.
- If the problem is infeasible, it is commonly denoted by $p^\star = +\infty$.

Optimal Points:

- A feasible point $\boldsymbol{x}$ is **optimal** if $f_0(\boldsymbol{x}) = p^\star$.
- There may be more than one optimal point; the set of optimal points is denoted by $\mathcal{X}_{\text{opt}}$.
- A feasible point $\boldsymbol{x}$ is **locally optimal** if it is optimal within a local neighborhood.

Definitions: Feasibility

Examples

- $f_0(x) = 1/x$, dom $f_0 = \mathbb{R}_{++}$: $p^\star = 0$, but no optimal point since the optimal value cannot be achieved.
- $f_0(x) = -\log x$, dom $f_0 = \mathbb{R}_{++}$: Function is unbounded below, $p^\star = -\infty$.
- $f_0(x) = x^3 - 3x$: Nonconvex function with $p^\star = -\infty$ and a local optimum at $x = 1$.

Active and Redundant Constraints

- If $\boldsymbol{x}$ is feasible and $f_i(\boldsymbol{x}) = 0$, the ith inequality constraint $f_i(\boldsymbol{x}) \leq 0$ is **active** at $\boldsymbol{x}$.
- If $f_i(\boldsymbol{x}) < 0$, the constraint $f_i(\boldsymbol{x}) \leq 0$ is **inactive**.
- Equality constraints are active at all feasible points.
- A constraint is **redundant** if deleting it does not change the feasible set.

Definitions: Feasibility Problem

Feasibility Problem:

- Goal is to find a feasible point, not necessarily to minimize or maximize any objective.
- Formulated as:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{find}} & \boldsymbol{x} \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, p \end{array}$$

Feasibility Problem as a Special Case of a General Optimization Problem:

- Can be regarded as:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & 0 \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, p \end{array}$$

- Optimal value:
  - $p^\star = 0$ if the constraints are feasible.
  - $p^\star = \infty$ if the constraints are infeasible.

Solving Optimization Problems

General Optimization Problems:

- Typically very difficult to solve.
- May require long computation times for optimal solutions or result in sub-optimal solutions in reasonable times.
- Exceptions include least squares problems, linear programming problems, and convex optimization problems.
- General nonconvex problems (nonlinear problems) are particularly challenging.

Least-squares (LS) Problems:

- Originated with Gauss in 1795.
- Formulated as:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \,.$$

- Closed-form solution: $\boldsymbol{x}^\star = \left(\boldsymbol{A}^\mathsf{T}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^\mathsf{T}\boldsymbol{b}$.
- Reliable and efficient algorithms exist.
- Considered trivial due to ease of identification and solution.

Solving Optimization Problems

Linear Programming (LP):

- Formulated as:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{a}_i^{\mathsf{T}}\boldsymbol{x} \leq b_i, \quad i = 1, \ldots, m \end{array}$$

- No closed-form solutions in general.
- Reliable and efficient algorithms and software available.
- Not as easy to recognize as LS, standard tricks can convert various problems into LPs.

Convex Optimization Problems:

- Inequality-constrained convex problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq b_i, \quad i = 1, \ldots, m \end{array}$$

- All functions are convex.
- No closed-form solutions in general.
- Reliable and efficient algorithms and software available.
- Often difficult to recognize; many tricks for transforming problems into convex form.

Solving Optimization Problems

Nonconvex Optimization

- Generally very difficult to solve.
- Requires either long computation times or compromises on optimality.
- Two main strategies:
  - **Local optimization**: Fast algorithms, but no guarantee of global optimality (depends on initial point).
  - **Global optimization**: Worst-case complexity increases exponentially with problem size, but it ensures discovery of a global solution.

Computational Complexity

- Measured in number of operations required to obtain a solution.
- Complexity is a function of the number of variables n.
- Polynomial complexity is acceptable in practice.
- Exponential complexity is not acceptable as it quickly explodes with n.

Historical Snapshot of Optimization

History of Optimization Theory (Convex Analysis):

- Extensively developed during 1900-1970.
- Computational aspects and applications came later.

Key Developments in Computational Methods:

- **Simplex Method (1947):**
  - Developed by Dantzig for linear programming (LP).
  - Very efficient in practice but has exponential worst-case complexity.
- **Ellipsoid Method (1970s):**
  - Proposed with provable polynomial worst-case complexity.
  - Can be very slow in practice.
- **Interior-Point Method (1984):**
  - Proposed by Karmarkar for LPs.
  - Polynomial-time complexity, efficient in both theory and practice.
  - Extended to quadratic programming and linear complementarity problems.

Historical Snapshot of Optimization

Interior-Point Methods For Convex Problems (1994):

- Developed by Nesterov and Nemirovskii.
- Self-concordant function theory facilitated expansion of log-barrier function-based algorithms.
- Applied to a wider array of convex problems, including semidefinite programming and second-order cone programming.

Practical Applications:

- **Linear Programming:**
  - Widely used since the 1950s for modeling real-life problems, such as allocation issues.
- **Convex Problems:**
  - Little interest in modeling real-life problems as convex problems until the mid-1990s.
  - Surge in activity related to modeling applications as convex problems following the development of interior-point methods for convex problems.

Illustrative Example: Lamp Illumination Problem

Lamp Illumination Problem:

- Goal: Achieve desired illumination I_{des} on all patches by controlling lamp powers.
- Intensity at patch k:

$$I_k = \sum_{j=1}^{m} a_{kj} p_j$$

- Coefficients: $a_{kj} = \cos\theta_{kj}/r_{kj}^2$, where θ_{kj} and r_{kj} are the angle and distance between lamp j and patch k.

Relaxed Problem: $I_k \approx I_{\text{des}}$

- Minimize the largest of the errors measured in a logarithmic scale (because the eyes perceive intensity on a log-scale):

$$\begin{array}{lll} \underset{I_1,\ldots,I_n,p_1,\ldots,p_m}{\text{minimize}} & \max_k \left|\log I_k - \log I_{\text{des}}\right| & \\ \text{subject to} & 0 \leq p_j \leq p_{\max}, & j = 1,\ldots,m \\ & I_k = \sum_{j=1}^m a_{kj} p_j, & k = 1,\ldots,n. \end{array}$$

- This problem appears complex, and we will explore various possible approaches.

Illustrative Example: Lamp Illumination Problem

Resolution Options:

1. **Heuristic guess:**
   - Use uniform power $p_j = p$ and try different values of p.
2. **Least-squares (LS) formulation:**

$$\begin{array}{lll} \underset{I_1,\ldots,I_n,p_1,\ldots,p_m}{\text{minimize}} & \sum_{k=1}^{n} \left(I_k - I_{\text{des}}\right)^2 & \\ \text{subject to} & I_k = \sum_{j=1}^{m} a_{kj} p_j, & k = 1, \ldots, n \end{array}$$

   - Clip p_j if $p_j > p_{\max}$ or $p_j < 0$ to make it feasible.
3. **Linear Programming (LP) formulation:**

$$\begin{array}{lll} \underset{I_1,\ldots,I_n,p_1,\ldots,p_m}{\text{minimize}} & \max_k \left|I_k - I_{\text{des}}\right| & \\ \text{subject to} & 0 \leq p_j \leq p_{\max}, & j = 1, \ldots, m \\ & I_k = \sum_{j=1}^{m} a_{kj} p_j, & k = 1, \ldots, n \end{array}$$

   - Can be transformed into an LP with simple manipulations.

Illustrative Example: Lamp Illumination Problem

4. **Convex Optimization Formulation:**

$$\begin{array}{lll} \underset{I_1,\ldots,I_n,p_1,\ldots,p_m}{\text{minimize}} & \max_k h\left(I_k/I_{\text{des}}\right) & \\ \text{subject to} & 0 \leq p_j \leq p_{\max}, & j = 1,\ldots,m \\ & I_k = \sum_{j=1}^{m} a_{kj} p_j, & k = 1,\ldots,n, \end{array}$$

where $h(u) = \max\{u, 1/u\}$.

Additional Constraints:

- Convex constraint example: "No more than half of total power is in any 10 lamps."
  - Can be written in convex form, keeping the problem solvable.
- Combinatorial constraint example: "No more than half of the lamps are on."
  - Appears simple but is a combinatorial constraint, making the problem exponentially complex.

Key Takeaway:

- Untrained intuition may not always work.
- Proper background and intuition are needed to discern between difficult and easy problems.

Outline

Definitions

Line Passing Through Two Points:

- Equation: $\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}$, where $\theta \in \mathbb{R}$.
- If $0 \leq \theta \leq 1$, it describes the *line segment* between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Convex Set:

- A set $\mathcal{C} \in \mathbb{R}^n$ is *convex* if the line segment between any two points in $\mathcal{C}$ lies in $\mathcal{C}$.
- Formally, for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$ and $0 \leq \theta \leq 1$:

$$\theta \boldsymbol{x} + (1-\theta)\,\boldsymbol{y} \in \mathcal{C}.$$

- For detailed information, refer to (S. P. Boyd and Vandenberghe 2004, chap. 2).

Convex Combination:

- A *convex combination* of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ is of the form $\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 + \cdots + \theta_k \boldsymbol{x}_k$, where $\theta_1 + \cdots + \theta_k = 1$ and $\theta_i \geq 0$, $i = 1, \ldots, k$.
- A set is convex if and only if it contains every convex combination of its points.

Definitions

Convex Hull:

- The *convex hull* of a set $\mathcal{C}$ is the set of all convex combinations of points in $\mathcal{C}$.
- The convex hull of $\mathcal{C}$ is always convex.
- It is the smallest convex set that contains $\mathcal{C}$.

Cone:

- A set $\mathcal{C}$ is a *cone* if for every $\boldsymbol{x} \in \mathcal{C}$ and $\theta \geq 0$, we have $\theta \boldsymbol{x} \in \mathcal{C}$.

Convex Cone:

- A set $\mathcal{C}$ is a *convex cone* if it is convex and a cone.
- Formally, for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C}$ and $\theta_1, \theta_2 \geq 0$:

$$\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 \in \mathcal{C}.$$

Hyperplanes and Halfspaces

- **Hyperplane:**
  - Set of the form:
    $$\left\{\boldsymbol{x} \mid \boldsymbol{a}^{\mathsf{T}}\boldsymbol{x} = b\right\},$$
    where $\boldsymbol{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$.
  - Geometric interpretation: Set of points orthogonal to vector $\boldsymbol{a}$ with an offset.
- **Halfspace:**
  - A hyperplane divides $\mathbb{R}^n$ into two halfspaces.
  - (Closed) halfspace:
    $$\left\{\boldsymbol{x} \mid \boldsymbol{a}^{\mathsf{T}}\boldsymbol{x} \leq b\right\}.$$

Polyhedra

- Solution set of a finite number of linear equalities and inequalities:

$$\mathcal{P} = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b},\ \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}\},$$

  where $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^m$, $d \in \mathbb{R}^p$.
- Examples:
  - Unit simplex:

$$\left\{\boldsymbol{x} \mid \boldsymbol{x} \ge \mathbf{0}, \mathbf{1}^{\mathsf{T}}\boldsymbol{x} \le 1\right\}.$$

  - Probability simplex:

$$\left\{\boldsymbol{x} \mid \boldsymbol{x} \ge \mathbf{0}, \mathbf{1}^{\mathsf{T}}\boldsymbol{x} = 1\right\}.$$

Balls and Ellipsoids

- **Euclidean ball:**
  - Center $\boldsymbol{x}_c$ and radius r:

$$\mathcal{B}(\boldsymbol{x}_c, r) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}_c\|_2 \leq r\} = \left\{\boldsymbol{x} \mid (\boldsymbol{x} - \boldsymbol{x}_c)^\mathsf{T}(\boldsymbol{x} - \boldsymbol{x}_c) \leq r^2\right\},$$

    where $\|\cdot\|_2$ is the Euclidean norm.
  - Another representation:

$$\mathcal{B}(\boldsymbol{x}_c, r) = \{\boldsymbol{x}_c + r\boldsymbol{u} \mid \|\boldsymbol{u}\|_2 \leq 1\}.$$

- **Ellipsoid:**
  - Defined as:

$$\begin{aligned}\mathcal{E}(\boldsymbol{x}_c, \boldsymbol{P}) &= \left\{\boldsymbol{x} \mid (\boldsymbol{x} - \boldsymbol{x}_c)^\mathsf{T}\boldsymbol{P}^{-1}(\boldsymbol{x} - \boldsymbol{x}_c) \leq 1\right\} \\ &= \{\boldsymbol{x}_c + \boldsymbol{A}\boldsymbol{u} \mid \|\boldsymbol{u}\|_2 \leq 1\},\end{aligned}$$

    with $\boldsymbol{P} = \boldsymbol{P}^\mathsf{T} \in \mathbb{R}^{n\times n} \succ \boldsymbol{0}$ (symmetric and positive definite), and $\boldsymbol{A}$ is the square-root matrix $\boldsymbol{P}^{1/2}$.
  - A ball is an ellipsoid with $\boldsymbol{P} = r^2 I$.

Norm Balls and Norm Cones

- **Norm ball:**
  - For any norm $\|\cdot\|$ on $\mathbb{R}^n$:
$$\mathcal{B}(\boldsymbol{x}_c, r) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}_c\| \leq r\}.$$
- **Norm cone:**
  - Convex set:
$$\mathcal{C} = \left\{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} \mid \|\boldsymbol{x}\| \leq t\right\}.$$
  - **Second-order cone (ice-cream cone):**
$$\mathcal{C} = \left\{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} \mid \|\boldsymbol{x}\|_2 \leq t\right\},$$
    where the norm is the Euclidean norm $\|\cdot\|_2$.

Positive Semidefinite Cone

- Set of symmetric positive semidefinite matrices:
$$\mathbb{S}_+^n = \left\{\boldsymbol{X} \in \mathbb{R}^{n\times n} \mid \boldsymbol{X} = \boldsymbol{X}^\mathsf{T} \succeq \boldsymbol{0}\right\}$$

Operations that Preserve Convexity

Establishing Convexity:

- Direct use of the definition of convexity can be cumbersome.
- A more practical approach is to show that the set can be obtained from simple convex sets (e.g., hyperplanes, halfspaces, balls, ellipsoids, cones) by operations that preserve convexity.

Intersection:

- Convexity is preserved under intersection.
- If $\mathcal{S}_1$ and $\mathcal{S}_2$ are convex, then $\mathcal{S}_1 \cap \mathcal{S}_2$ is convex.
- This property extends to the intersection of (possibly infinitely) multiple sets.
- Examples:
  - A **polyhedron** is the intersection of halfspaces and hyperplanes, hence convex.
  - A more sophisticated example:

$$\mathcal{S} = \left\{ \boldsymbol{x} \in \mathbb{R}^n \mid |p_{\boldsymbol{x}}(t)| \leq 1 \text{ for } |t| \leq \pi/3 \right\},$$

where $p_{\boldsymbol{x}}(t) = x_1 \cos(t) + x_2 \cos(2t) + \cdots + x_n \cos(nt)$.

Affine Composition:

- A function is *affine* if it has the form $f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$.
- If $\mathcal{S} \subseteq \mathbb{R}^n$ is convex and $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is an affine function, then the image of $\mathcal{S}$ under f,
$$f(\mathcal{S}) = \{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathcal{S}\},$$
is convex.
- Examples:
  - **Scaling** and **translation**.
  - **Projection** of a convex set onto some of its coordinates:
  $$\{\boldsymbol{x}_1 \in \mathbb{R}^m \mid (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{S} \text{ for some } \boldsymbol{x}_2 \in \mathbb{R}^n\}$$
  is convex.
  - Affine composition of the norm cone:
  $$\left\{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\| \leq \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d\right\}.$$

Perspective Function:

- The perspective function scales or normalizes vectors so the last component is one, and then drops the last component.
- Formally, the perspective function $P : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^n$, with domain dom $P = \mathbb{R}^n \times \mathbb{R}_{++}$ is defined as:

$$P(\boldsymbol{x}, t) = \boldsymbol{x}/t.$$

- Property: Images and inverse images of convex sets under perspective functions are convex.

Outline

Convex Function

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if:
  - The domain, $\operatorname{dom} f$, is a convex set
  - For all $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$ and $0 \leq \theta \leq 1$, the following holds:

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}).$$

- Geometric interpretation: the line segment (chord) between $(\boldsymbol{x}, f(\boldsymbol{x}))$ and $(\boldsymbol{y}, f(\boldsymbol{y}))$ lies above the graph of f

Strictly Convex Function

- If strict inequality holds in the convex function definition whenever $\boldsymbol{x} \neq \boldsymbol{y}$ and $0 < \theta < 1$

Concave Function

- If $-f$ is convex

For detailed information, refer to (S. P. Boyd and Vandenberghe 2004, chap. 3).

Elementary Convex and Concave Functions

Affine Function

- For an affine function, equality always holds in the convex function definition.
- All affine (and therefore also linear) functions are both convex and concave.
- Conversely, any function that is both convex and concave is affine.

Functions on $\mathbb{R}$

- **Exponential:** e^{ax} is convex on $\mathbb{R}$ for any $a \in \mathbb{R}$.
- **Powers:** x^a is convex on $\mathbb{R}_{++}$ for $a \geq 1$ or $a \leq 0$ (e.g., x^2) and concave for $0 \leq a \leq 1$.
- **Powers of absolute value:** $|x|^p$ is convex on $\mathbb{R}$ for $p \geq 1$ (e.g., $|x|$).
- **Logarithm:** $\log x$ is concave on $\mathbb{R}_{++}$.
- **Negative entropy:** $x \log x$ is convex on $\mathbb{R}_{++}$.

Elementary Convex and Concave Functions

Functions on $\mathbb{R}^n$

- **Quadratic function:** $f(\boldsymbol{x}) = \boldsymbol{x}^\mathsf{T}\boldsymbol{P}\boldsymbol{x} + 2\boldsymbol{q}^\mathsf{T}\boldsymbol{x} + r$ is convex on $\mathbb{R}^n$ if and only if $\boldsymbol{P} \succeq \boldsymbol{0}$.
- **Norms:** Every norm $\|\boldsymbol{x}\|$ is convex on $\mathbb{R}^n$ (e.g., $\|\boldsymbol{x}\|_\infty$, $\|\boldsymbol{x}\|_1$, and $\|\boldsymbol{x}\|_2$).
- **Max function:** $f(x) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbb{R}^n$.
- **Quadratic over linear function:** $f(x, y) = x^2/y$ is convex on $\mathbb{R} \times \mathbb{R}_{++}$.
- **Geometric mean:** $f(\boldsymbol{x}) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbb{R}^n_{++}$.
- **Log-sum-exp function:** $f(\boldsymbol{x}) = \log\left(e^{x_1} + \cdots + e^{x_n}\right)$ is convex on $\mathbb{R}^n$. Can be used to approximate the function $f(\boldsymbol{x}) = \max\{x_1, \ldots, x_n\}$.

Functions on $\mathbb{R}^{n\times n}$

- **Log-determinant:** The function $f(\boldsymbol{X}) = \operatorname{logdet}(\boldsymbol{X})$ is concave on $\mathbb{S}^n_{++} = \{\boldsymbol{X} \in \mathbb{R}^{n\times n} \mid \boldsymbol{X} \succ \boldsymbol{0}\}$.
- **Maximum eigenvalue:** The function

$$f(\boldsymbol{X}) = \lambda_{\max}(\boldsymbol{X}) \triangleq \sup_{\boldsymbol{y}\neq\boldsymbol{0}} \frac{\boldsymbol{y}^\mathsf{T}\boldsymbol{X}\boldsymbol{y}}{\boldsymbol{y}^\mathsf{T}\boldsymbol{y}}$$

is convex on $\mathbb{S}^n$.

Epigraph

Convexity in Sets and Functions

- The term “convex” is used to describe both sets and functions.
- These two meanings can be linked, as demonstrated below.

Graph of a Function

- Defined for a function $f : \mathbb{R}^n \to \mathbb{R}$ as the set:

$$\left\{(\boldsymbol{x}, f(\boldsymbol{x})) \in \mathbb{R}^{n+1} \mid \boldsymbol{x} \in \operatorname{dom} f\right\}$$

- This set is a subset of $\mathbb{R}^{n+1}$.

Epigraph of a Function

- Defined for a function $f : \mathbb{R}^n \to \mathbb{R}$ as the set:

$$\operatorname{epi} f = \left\{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} \mid \boldsymbol{x} \in \operatorname{dom} f,\ f(\boldsymbol{x}) \leq t\right\}$$

- Conceptualized as pouring water over the function and filling it up indefinitely.

Link Between Convex Sets and Convex Functions

A function is convex if and only if its epigraph is a convex set:

$$f \text{ is convex} \quad \Longleftrightarrow \quad \operatorname{epi} f \text{ is convex}$$

Characterization of Convex Functions: Restriction to a Line

Convex Function Property

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if it is convex when restricted to any line intersecting its domain.
- Formally, f is convex if the function $g : \mathbb{R} \to \mathbb{R}$ defined as

$$g(t) = f(\boldsymbol{x} + t\boldsymbol{v})$$

  is convex on its domain $\operatorname{dom} g = \{t \mid \boldsymbol{x} + t\boldsymbol{v} \in \operatorname{dom} f\}$, for any $\boldsymbol{x} \in \operatorname{dom} f$ and $\boldsymbol{v} \in \mathbb{R}^n$.

Utility of This Property

- Allows checking convexity by restricting the function to a line.
- Simplifies the analysis process, even for exploratory plotting.

Example

Concavity of the log-determinant function via concavity of the log function:

$$\begin{aligned} g(t) = \operatorname{logdet}(\boldsymbol{X} + t\boldsymbol{V}) &= \operatorname{logdet}(\boldsymbol{X}) + \operatorname{logdet}\left(\boldsymbol{I} + t\boldsymbol{X}^{-1/2}\boldsymbol{V}\boldsymbol{X}^{-1/2}\right) \\ &= \operatorname{logdet}(\boldsymbol{X}) + \sum_{i=1}^{n} \log(1 + t\lambda_i) \end{aligned}$$

Characterization of Convex Functions: First-order Condition

Gradient of a Differentiable Function

- For a differentiable function f, the gradient is given by:

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^n$$

- The gradient exists at each point in $\operatorname{dom} f$, which is open.

First-order Taylor Approximation

- Near $\boldsymbol{x}$, the first-order Taylor approximation of f is:

$$f(\boldsymbol{y}) \approx f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{y} - \boldsymbol{x})$$

Convexity Condition for Differentiable Functions

- Suppose f is differentiable. Then f is convex if and only if $\operatorname{dom} f$ is convex and:

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{y} - \boldsymbol{x})$$

  holds for all $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$.

Characterization of Convex Functions: First-order Condition

Geometric Interpretation

- The inequality states that for a convex function, the first-order Taylor approximation is a *global underestimator* of the function.
- Conversely, if the first-order Taylor approximation is always a global underestimator, then the function is convex.

Implications of the Inequality

- From local information about a convex function (its value and derivative at a point), we can derive global information (a global underestimator).
- This property justifies the connection between local optimality and global optimality in convex optimization problems.

Characterization of Convex Functions: Second-Order Condition

Hessian of a Twice Differentiable Function

- For a twice differentiable function f, the Hessian is given by:

$$\nabla^2 f(\boldsymbol{x}) = \left(\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} \right)_{ij} \in \mathbb{R}^{n \times n}$$

- The Hessian exists at each point in $\operatorname{dom} f$, which is open.

Second-Order Taylor Approximation

- Near $\boldsymbol{x}$, the second-order Taylor approximation of f is:

$$f(\boldsymbol{y}) \approx f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\mathsf{T}(\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^\mathsf{T} \nabla^2 f(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})$$

Convexity Condition for Twice Differentiable Functions

- Suppose f is twice differentiable. Then f is convex if and only if $\operatorname{dom} f$ is convex and its Hessian is positive semidefinite:

$$\nabla^2 f(\boldsymbol{x}) \succeq \boldsymbol{0}$$

  for all $\boldsymbol{x} \in \operatorname{dom} f$.

Characterization of Convex Functions: Second-Order Condition

Special Case for Functions on $\mathbb{R}$

- For a function on $\mathbb{R}$, this reduces to the condition $f''(x) \geq 0$ (and $\operatorname{dom} f$ convex, i.e., an interval).
- This means that the derivative is nondecreasing.

Geometric Interpretation

- The condition $\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0}$ can be interpreted as the requirement that the graph of the function have positive (upward) curvature at $\boldsymbol{x}$.

Concavity Condition

- Similarly, f is concave if and only if $\operatorname{dom} f$ is convex and:

$$\nabla^2 f(\boldsymbol{x}) \preceq \mathbf{0}$$

  for all $\boldsymbol{x} \in \operatorname{dom} f$.

Characterization of Convex Functions

Applying the Definition Directly

- Check if $f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$ and $\theta \in [0, 1]$.

Restricting the Function to a Line

- Verify if $g(t) = f(\boldsymbol{x} + t\boldsymbol{v})$ is convex for any $\boldsymbol{x} \in \operatorname{dom} f$ and $\boldsymbol{v} \in \mathbb{R}^n$.

Using the First-order Condition

- Ensure $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathsf{T}}(\boldsymbol{y} - \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$.

Employing the Second-order Condition

- Confirm $\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0}$ for all $\boldsymbol{x} \in \operatorname{dom} f$.

Operations that Preserve Convexity

- A practical approach to establish convexity is to derive the function from basic convex or concave functions through operations that preserve convexity, as explored next.

Operations that Preserve Convexity

Nonnegative Weighted Sum

- If f_1 and f_2 are both convex functions, then so is their sum $f_1 + f_2$.
- Scaling a function f with a nonnegative number $\alpha \geq 0$ preserves convexity.
- Combining nonnegative scaling and addition: a nonnegative weighted sum of convex functions (with weights $w_1, \ldots, w_m \geq 0$) $f = w_1 f_1 + \cdots + w_m f_m$ is convex.

Composition with an Affine Mapping

- Suppose $h : \mathbb{R}^m \to \mathbb{R}$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, and $\boldsymbol{b} \in \mathbb{R}^m$. Define $f : \mathbb{R}^n \to \mathbb{R}$ as the composition of h with the affine mapping $\boldsymbol{Ax} + \boldsymbol{b}$:

$$f(\boldsymbol{x}) = h(\boldsymbol{Ax} + \boldsymbol{b}),$$

  with $\operatorname{dom} f = \{\boldsymbol{x} \mid \boldsymbol{Ax} + \boldsymbol{b} \in \operatorname{dom} h\}$.
- If h is convex, so is f.
- Examples:
  - $f(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{Ax}\|$ is convex.
  - $f(\boldsymbol{X}) = \operatorname{logdet}\left(\boldsymbol{I} + \boldsymbol{HXH}^{\mathsf{T}}\right)$ is concave.

Operations that Preserve Convexity

Pointwise Maximum

- If f_1 and f_2 are convex functions, then their pointwise maximum f, defined as

$$f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\},$$

  with $\operatorname{dom} f = \operatorname{dom} f_1 \cap \operatorname{dom} f_2$, is also convex.
- This property extends to more than two functions. If $f_1, \ldots, f_m$ are convex, then their pointwise maximum

$$f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})\}$$

  is also convex.
- Example: The sum of the r largest components of $\boldsymbol{x} \in \mathbb{R}^n$, $f(\boldsymbol{x}) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$, where $x_{[i]}$ is the ith largest component of $\boldsymbol{x}$, is convex because it can be written as the pointwise maximum

$$f(\boldsymbol{x}) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}.$$

Operations that Preserve Convexity

Pointwise Supremum

- The pointwise maximum property extends to the *pointwise supremum* over an infinite set of convex functions. If for each $\boldsymbol{y} \in \mathcal{Y}$, $f(\boldsymbol{x}, \boldsymbol{y})$ is convex in $\boldsymbol{x}$, then the pointwise supremum g, defined as

$$g(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{Y}} f(\boldsymbol{x}, \boldsymbol{y}),$$

  is convex in $\boldsymbol{x}$.

- Examples:
  - Distance to farthest point in a set $\mathcal{C}$:

$$f(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{C}} \|\boldsymbol{x} - \boldsymbol{y}\|.$$

  - Maximum eigenvalue function of a symmetric matrix:

$$\lambda_{\max}(\boldsymbol{X}) = \sup_{\boldsymbol{y} \neq \boldsymbol{0}} \frac{\boldsymbol{y}^{\mathsf{T}} \boldsymbol{X} \boldsymbol{y}}{\boldsymbol{y}^{\mathsf{T}} \boldsymbol{y}}.$$

Operations that Preserve Convexity

Composition with Arbitrary Functions

- The composition of $h : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$, denoted by $f = h \circ g : \mathbb{R}^n \to \mathbb{R}$ is defined as
$$f(\boldsymbol{x}) = h(g(\boldsymbol{x})).$$
- For scalar composition $(m = 1)$:
  - f is convex if:
$$\begin{cases} h \text{ is convex nondecreasing and } g \text{ is convex} \\ h \text{ is convex nonincreasing and } g \text{ is concave} \end{cases}$$
  - f is concave if:
$$\begin{cases} h \text{ is concave nondecreasing and } g \text{ is concave} \\ h \text{ is concave nonincreasing and } g \text{ is convex} \end{cases}$$
- Examples:
  - If g is convex, then $\exp g(x)$ is convex.
  - If g is concave and positive, then $\log(g(x))$ is concave.
  - If g is concave and positive, then $1/g(x)$ is convex.
  - If g is convex and nonnegative, then $g(x)^p$ is convex for $p \geq 1$.
  - If g is convex, then $-\log(-g(x))$ is convex on $\{x \mid g(x) < 0\}$.

Operations that Preserve Convexity

Partial Minimization

- If $f(\boldsymbol{x}, \boldsymbol{y})$ is convex in $(\boldsymbol{x}, \boldsymbol{y})$ and $\mathcal{C}$ is a convex set, then the function

$$g(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in \mathcal{C}} f(\boldsymbol{x}, \boldsymbol{y})$$

  is convex in $\boldsymbol{x}$.
- Example: Distance to a set $\mathcal{C}$:

$$f(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in \mathcal{C}} \|\boldsymbol{x} - \boldsymbol{y}\| .$$

Perspective

- Suppose $f : \mathbb{R}^n \to \mathbb{R}$, then the perspective of f is the function $g : \mathbb{R}^{n+1} \to \mathbb{R}$ defined as

$$g(\boldsymbol{x}, t) = t f(\boldsymbol{x}/t),$$

  with domain $\operatorname{dom} g = \{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} \mid \boldsymbol{x}/t \in \operatorname{dom} f,\ t > 0\}$.
- The perspective operation preserves convexity.
- Example: Since $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x}$ is convex, its perspective $g(\boldsymbol{x}, t) = \boldsymbol{x}^T \boldsymbol{x}/t$ is convex for $t > 0$.

Quasiconvex Functions

α-Sublevel Set:

- Defined for a function $f : \mathbb{R}^n \to \mathbb{R}$ as:

$$\mathcal{S}_\alpha = \{\boldsymbol{x} \in \operatorname{dom} f \mid f(\boldsymbol{x}) \leq \alpha\}.$$

- Sublevel sets of a convex function are convex for any value of α.
- The converse is not true: a function can have all its sublevel sets convex but not be a convex function.
  - Example: $f(x) = -e^x$ is not convex on $\mathbb{R}$ (it is strictly concave) but all its sublevel sets are convex.

Quasiconvex and Quasiconcave Functions:

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *quasiconvex* if its domain and all its sublevel sets $\mathcal{S}_\alpha$, for all α, are convex.
- A function f is *quasiconcave* if $-f$ is quasiconvex.
- A function that is both quasiconvex and quasiconcave is called *quasilinear*.

Examples of Quasiconvex and Quasiconcave Functions:

- For a function on $\mathbb{R}$ to be quasiconvex, each sublevel set must be an interval.
- Examples:
  - $\sqrt{|x|}$ is quasiconvex on $\mathbb{R}$.
  - $\operatorname{ceil}(x) = \inf\{z \in \mathbb{Z} \mid z \geq x\}$ is quasilinear.
  - $\log x$ is quasilinear on $\mathbb{R}_{++}$.
  - $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbb{R}^2_{++}$.
  - The linear-fractional function

$$f(\boldsymbol{x}) = \frac{\boldsymbol{a}^{\mathsf{T}}\boldsymbol{x} + b}{\boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} + d}, \qquad \operatorname{dom} f = \left\{\boldsymbol{x} \mid \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} + d > 0\right\}$$

    is quasilinear.

Quasiconvex Functions

Representation of Sublevel Sets:

- Sublevel sets of a quasiconvex function f (which are convex) can be represented via inequalities of convex functions:

$$f(\boldsymbol{x}) \leq t \quad \Longleftrightarrow \quad \phi_t(\boldsymbol{x}) \leq 0,$$

where $\phi_t(\boldsymbol{x})$ is a family of convex functions in $\boldsymbol{x}$ (indexed by t).

Example of Convex Over Concave Function:

- Consider a function $f(\boldsymbol{x}) = \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}$, where $p(\boldsymbol{x}) \geq 0$ and $q(\boldsymbol{x}) > 0$. The function $f(\boldsymbol{x})$ is not convex but it is quasiconvex:

$$f(\boldsymbol{x}) \leq t \quad \Longleftrightarrow \quad p(\boldsymbol{x}) - tq(\boldsymbol{x}) \leq 0,$$

so we can take the convex function $\phi_t(\boldsymbol{x}) = p(\boldsymbol{x}) - tq(\boldsymbol{x})$ for $t \geq 0$.

Outline

Convex Optimization Problems

Definition: If the objective and inequality constraint functions of an optimization problem are convex and the equality constraint functions are linear (or affine), the problem is a *convex optimization problem* or *convex program*.

Standard Form of a Convex Optimization Problem

- A convex optimization problem can be written in standard form as:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

  where $f_0, f_1, \ldots, f_m$ are convex and the p equality constraints are affine with $\boldsymbol{A} \in \mathbb{R}^{p \times n}$ and $\boldsymbol{b} \in \mathbb{R}^p$.
- For detailed information, refer to (S. P. Boyd and Vandenberghe 2004, chap. 4).

Convex Optimization Problems

Advantages of Convex Optimization Problems

- Convex problems have a rich body of theory and algorithms with desirable convergence properties.
- Any locally optimal point in a convex optimization problem is also globally optimal.

Challenges with Nonconvex Problems

- Most problems are not convex when naturally formulated.
- Reformulating a nonconvex problem in convex form may be possible, but it is an art and there is no systematic way to do it.

Optimality Characterization

First-order Characterization of Convexity

- For a differentiable function f_0, the first-order condition for convexity is:

$$f_0(\boldsymbol{y}) \geq f_0(\boldsymbol{x}) + \nabla f_0(\boldsymbol{x})^\mathsf{T}(\boldsymbol{y} - \boldsymbol{x})$$

  for all $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}\, f_0$.

Minimum Principle

- A feasible point $\boldsymbol{x}$ is optimal if and only if:

$$\nabla f_0(\boldsymbol{x})^\mathsf{T}(\boldsymbol{y} - \boldsymbol{x}) \geq 0 \text{ for all } \boldsymbol{y} \in \mathcal{X},$$

  where $\mathcal{X}$ denotes the feasible set.
- Geometrically, this means that the gradient $\nabla f_0(\boldsymbol{x})$ defines a supporting hyperplane.

KKT Optimality Conditions

- The minimum principle may be difficult to manage in practical cases.
- A more convenient characterization of optimality, when the feasible set $\mathcal{X}$ is given in terms of constraint functions, is the *KKT optimality conditions*.

Optimality Characterization: Examples

Unconstrained Minimization Problem

- For an unconstrained problem (i.e., $m = p = 0$ with feasible set $\mathcal{X} = \mathbb{R}^n$), the optimality condition reduces to

$$\nabla f_0(\boldsymbol{x}) = 0.$$

Minimization over the Nonnegative Orthant

- Consider the problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \geq \mathbf{0}. \end{array}$$

- The optimality condition becomes:

$$\boldsymbol{x} \geq \mathbf{0}, \quad \nabla f_0(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) \geq 0 \text{ for all } \boldsymbol{y} \geq \mathbf{0}.$$

- The term $\nabla f_0(\boldsymbol{x})^T \boldsymbol{y}$ is unbounded below on $\boldsymbol{y} \geq \mathbf{0}$, unless $\nabla f_0(\boldsymbol{x}) \geq \mathbf{0}$.
- The condition reduces to $-\nabla f_0(\boldsymbol{x})^T \boldsymbol{x} \geq 0$, which further becomes $\nabla f_0(\boldsymbol{x})^T \boldsymbol{x} = 0$, due to $\boldsymbol{x} \geq \mathbf{0}$ and $\nabla f_0(\boldsymbol{x}) \geq \mathbf{0}$. More compactly:

$$\boldsymbol{x} \geq \mathbf{0}, \quad \nabla f_0(\boldsymbol{x}) \geq \mathbf{0}, \quad (\nabla f_0(\boldsymbol{x}))_i \, x_i = 0, \quad i = 1, \ldots, n.$$

Equivalent Reformulations

Equivalence of Problems:

- Two problems are considered *equivalent* if a solution to one can be easily converted into a solution for the other, and vice versa.
- A stricter form of equivalence requires a mapping between the two problems for every feasible point, not just for the optimal solutions.

Hidden Convexity:

- Most problems are not convex when naturally formulated.
- In some cases, hidden convexity can be unveiled by properly reformulating the problem.
- There is no systematic way to reformulate a problem in convex form; it is an art.

Example: Change of Variable

- Consider the problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{1+x^2} \\ \text{subject to} & x^2 \geq 1, \end{array}$$

  which is nonconvex (both the cost function and the constraint are nonconvex).
- It can be rewritten in convex form after the change of variable $y = x^2$:

$$\begin{array}{ll} \underset{y}{\text{minimize}} & \frac{1}{1+y} \\ \text{subject to} & y \geq 1, \end{array}$$

- The optimal points x can be recovered from the optimal y as $x = \pm\sqrt{y}$.

Equivalent Reformulations

Example: Transforming Functions

- Consider the problem:

$$\begin{array}{ll} \underset{x_1,x_2}{\text{minimize}} & x_1^2 + x_2^2 \\ \text{subject to} & \frac{x_1}{1+x_2^2} \leq 0 \\ & (x_1 + x_2)^2 = 0, \end{array}$$

  which is nonconvex (the inequality constraint function is nonconvex and the equality constraint function is not affine).
- It can be equivalently rewritten as the convex problem:

$$\begin{array}{ll} \underset{x_1,x_2}{\text{minimize}} & x_1^2 + x_2^2 \\ \text{subject to} & x_1 \leq 0 \\ & x_1 = -x_2. \end{array}$$

Equivalent Reformulations: Change of Variables

Reformulating a Convex Optimization Problem with a One-to-One Mapping:

- Suppose ϕ is a one-to-one mapping from $\boldsymbol{z}$ to $\boldsymbol{x}$.
- Define $\tilde{f}_i(\boldsymbol{z}) = f_i(\phi(\boldsymbol{z}))$.
- The original problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

- Can be rewritten (ignoring equality constraints) as:

$$\begin{array}{ll} \underset{\boldsymbol{z}}{\text{minimize}} & \tilde{f}_0(\boldsymbol{z}) \\ \text{subject to} & \tilde{f}_i(\boldsymbol{z}) \leq 0, \quad i = 1, \ldots, m. \end{array}$$

Preservation of Convexity:

- Convexity may or may not be preserved depending on the mapping ϕ.
- With equality constraints, the mapping ϕ has to be affine to preserve the convexity of the problem.

Equivalent Reformulations: Transformation of Functions

Reformulating a Convex Optimization Problem with Strictly Increasing Functions:

- Suppose ψ_i are strictly increasing functions satisfying $\psi_i(0) = 0$.
- Define $\tilde{f}_i(\boldsymbol{x}) = \psi_i\left(f_i(\boldsymbol{x})\right)$.
- The original problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

- Can be rewritten (ignoring equality constraints) as:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \tilde{f}_0(\boldsymbol{x}) \\ \text{subject to} & \tilde{f}_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m. \end{array}$$

Preservation of Convexity:

- Convexity may or may not be preserved depending on the mappings ψ_i.

Equivalent Reformulations: Slack Variables

Transformation Using Slack Variables:

- Observation: $f_i(\boldsymbol{x}) \leq 0$ if and only if there is an $s_i \geq 0$ that satisfies $f_i(\boldsymbol{x}) + s_i = 0$.
- By introducing nonnegative *slack variables* $s_i \geq 0$, we can transform linear (or affine) inequalities $\boldsymbol{a}_i^\top \boldsymbol{x} \leq b_i$ into linear equalities $\boldsymbol{a}_i^\top \boldsymbol{x} + s_i = b_i$.

Benefits of Using Slack Variables:

- Converts inequality constraints into equality constraints, which can simplify the problem formulation.
- Helps in revealing hidden convexity or making the problem more tractable.

Equivalent Reformulations: Eliminating Equality Constraints

Affine Equality Constraints in Convex Problems

- Equality constraints in convex problems must be affine, i.e., of the form $\boldsymbol{Ax} = \boldsymbol{b}$.
- From linear algebra, the subspace of points satisfying such affine constraints can be written as:

$$\boldsymbol{x} = \boldsymbol{Fz} + \boldsymbol{x}_0,$$

  where:
  - $\boldsymbol{x}_0$ is any solution to $\boldsymbol{Ax} = \boldsymbol{b}$.
  - $\boldsymbol{F}$ is a matrix whose range is the nullspace of $\boldsymbol{A}$, i.e., $\boldsymbol{AF} = \mathbf{0}$.
  - $\boldsymbol{z}$ is any vector of appropriate dimensions.
- Affine equality constraints can be eliminated by expressing the solution space in terms of a new variable $\boldsymbol{z}$.

Equivalent Reformulations: Eliminating Equality Constraints

Reformulating the Problem

- Original problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{array}$$

- Equivalent problem after eliminating equality constraints:

$$\begin{array}{ll} \underset{\boldsymbol{z}}{\text{minimize}} & f_0(\boldsymbol{F}\boldsymbol{z} + \boldsymbol{x}_0) \\ \text{subject to} & f_i(\boldsymbol{F}\boldsymbol{z} + \boldsymbol{x}_0) \leq 0, \quad i = 1, \ldots, m, \end{array}$$

  with variable $\boldsymbol{z}$.

Preservation of Convexity

Since the composition of a convex function with an affine function is convex, eliminating equality constraints preserves the convexity of the problem.

Epigraph Form of a Convex Problem

The epigraph form of the convex problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

Can be written as:

$$\begin{array}{ll} \underset{t, \boldsymbol{x}}{\text{minimize}} & t \\ \text{subject to} & f_0(\boldsymbol{x}) - t \leq 0 \\ & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

with variables $\boldsymbol{x} \in \mathbb{R}^n$ and $t \in \mathbb{R}$.

Equivalence to the Original Problem

- The epigraph form is equivalent to the original problem
- $(\boldsymbol{x}, t)$ is optimal for the epigraph form if and only if $\boldsymbol{x}$ is optimal for the original problem and $t = f_0(\boldsymbol{x})$

Equivalent Reformulations: Minimizing over Some Variables

Nested Minimization

- The following holds:

$$\inf_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x},\boldsymbol{y}) = \inf_{\boldsymbol{x}} \tilde{f}(\boldsymbol{x}),$$

  where $\tilde{f}(\boldsymbol{x}) = \inf_{\boldsymbol{y}} f(\boldsymbol{x},\boldsymbol{y})$.
- If $f(\boldsymbol{x},\boldsymbol{y})$ is jointly convex in $\boldsymbol{x}$ and $\boldsymbol{y}$, then $\tilde{f}(\boldsymbol{x})$ is convex.

Explanation

- We can always minimize a function by first minimizing over some set of variables, and then minimizing over the remaining ones
- This is a *nested minimization*, meaning that as $\boldsymbol{x}$ changes, the $\boldsymbol{y}$ that minimizes $f(\boldsymbol{x},\boldsymbol{y})$ to obtain $\tilde{f}(\boldsymbol{x})$ changes as well
- Do not confuse this with an alternate minimization method, where one optimizes alternately over $\boldsymbol{x}$ and $\boldsymbol{y}$ until convergence is achieved

Equivalent Reformulations: Minimizing over Some Variables

Partitioning Variables: Partition the variable $\boldsymbol{x} \in \mathbb{R}^n$ into two blocks as $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$.
The convex problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ \text{subject to} & f_i(\boldsymbol{x}_1) \leq 0, \quad i = 1, \ldots, m_1 \\ & \tilde{f}_i(\boldsymbol{x}_2) \leq 0, \quad i = 1, \ldots, m_2, \end{array}$$

Is equivalent to the convex problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}_1}{\text{minimize}} & \tilde{f}_0(\boldsymbol{x}_1) \\ \text{subject to} & f_i(\boldsymbol{x}_1) \leq 0, \quad i = 1, \ldots, m_1, \end{array}$$

where

$$\tilde{f}_0(\boldsymbol{x}_1) = \inf_{\boldsymbol{z}} \{ f_0(\boldsymbol{x}_1, \boldsymbol{z}) \mid \tilde{f}_i(\boldsymbol{z}) \leq 0, \quad i = 1, \ldots, m_2 \}.$$

Summary

- Nested minimization simplifies a convex problem by first minimizing over some variables.
- This approach preserves convexity and can make the problem more tractable.
- Partitioning variables and applying nested minimization can simplify a complex problem.

Approximate Reformulations

Approximating Nonconvex Problems: When a formulated optimization problem remains nonconvex despite attempts to unveil hidden convexity, one can resort to approximations to form an *approximated problem*, possibly convex, that is easier to solve:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \tilde{f}_0(\boldsymbol{x}) \\ \text{subject to} & \tilde{f}_i(\boldsymbol{x}) \leq 0, \quad i = 1, \dots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

where $\tilde{f}_i(\boldsymbol{x}) \approx f_i(\boldsymbol{x})$.

Types of Approximations:

1. **Conservative approximation (tightened approximation):** $\tilde{f}_i(\boldsymbol{x}) \geq f_i(\boldsymbol{x})$
    - Guarantees the feasibility of the approximated solution.
2. **Relaxed approximation (relaxation):** $\tilde{f}_i(\boldsymbol{x}) \leq f_i(\boldsymbol{x})$
    - Does not guarantee the feasibility of the approximated solution and may require an additional step to enforce feasibility.
3. **Approximation without guarantees:** $\tilde{f}_i(\boldsymbol{x}) \approx f_i(\boldsymbol{x})$

Approximate Reformulations

Examples:

- Relaxing a problem by removing some constraints (typically the more difficult ones).
- Nonconvex constraint $x^2 = 1$ or, equivalently, $x \in \{\pm 1\}$, which is a nonconvex discrete set.
  - **Relaxation:** enlarge the feasible set by using the interval $-1 \leq x \leq 1$. This relaxation defines a feasible set that is a superset of the original feasible set.
  - **Tightening:** reduce the feasible set by using $x = 1$. This tightening defines a feasible set that is a subset of the original feasible set.

Summary:

- Approximations can transform a nonconvex problem into a convex one, making it easier to solve.
- Conservative approximations guarantee feasibility, while relaxed approximations may require additional steps to enforce feasibility.
- Approximations are a practical approach to handle nonconvex problems when exact solutions are difficult to obtain.

Quasi-Convex Optimization

Quasiconvex Optimization Problem: Standard form:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

where the inequality constraint functions $f_1, \ldots, f_m$ are convex, and the objective f_0 is quasiconvex.

Difference Between Convex and Quasiconvex Optimization

- A quasiconvex optimization problem can have locally optimal solutions that are not globally optimal.
- This can occur when the function becomes flat before reaching the optimal value.

Sublevel Sets Representation

The sublevel sets of a quasiconvex function can be represented via a family of convex inequalities:

$$f(\boldsymbol{x}) \leq t \quad \Longleftrightarrow \quad \phi_t(\boldsymbol{x}) \leq 0,$$

where $\phi_t(\boldsymbol{x})$ is a family of convex functions in $\boldsymbol{x}$ (indexed by t).

Quasi-Convex Optimization via Feasibility Problems

- Let $p^\star$ denote the optimal value of the original quasiconvex optimization problem.
- Observation: If the convex feasibility problem

$$
\begin{array}{ll}
\underset{\boldsymbol{x}}{\text{find}} & \boldsymbol{x} \\
\text{subject to} & \phi_t(\boldsymbol{x}) \leq 0 \\
& f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\
& \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},
\end{array}
$$

is feasible, then $p^\star \leq t$. Conversely, if it is infeasible, then $p^\star > t$.

Quasi-Convex optimization

- Starts with an interval $[l, u]$ known to contain the optimal value $p^\star$ and sequentially halves the interval.
- The length of the interval after k iterations is $2^{-k}(u - l)$.
- Number of iterations required to achieve a tolerance of ϵ is $\lceil \log_2((u - l)/\epsilon) \rceil$.

Bisection method (aka "sandwich technique")

Initialization:

- Initialize l and u such that $p^\star \in [l, u]$.

Repeat while $(u - l) > \epsilon$**:**

- Compute midpoint of interval: $t = (l + u)/2$.
- Solve the convex feasibility problem for t.
- If feasible, set $u = t$; otherwise set $l = t$.

Outline

Taxonomy of Convex Problems

Problems can be classified into:

- LP
- QP
- QCQP
- SOCP
- SDP
- CP
- FP
- LFP
- GP

This classification is beneficial for both theoretical and algorithmic purposes.

Solvers are designed to handle specific types of problems.

Linear Programming (LP)

Linear Program (LP): A problem is called a *linear program* or *linear problem* (LP) when the objective and constraint functions are all affine:

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d \\ \text{subject to} \quad & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned}$$

where the parameters $\boldsymbol{A}$, $\boldsymbol{b}$, $\boldsymbol{c}$, d, $\boldsymbol{G}$, and $\boldsymbol{h}$ are of appropriate size.
Linear programs are convex optimization problems.

Geometric Interpretation of an LP

- Visualized as a polyhedron on an inclined flat surface.
- An optimal solution is always located at a corner of the polyhedron.
- This observation forms the basis of the popular simplex method, developed by Dantzig in 1947, for solving LPs.

Linear Programming (LP): Norm Miminization

ℓ_∞-norm minimization as an LP:

- Original convex problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \|\boldsymbol{x}\|_\infty \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

- Equivalent LP:

$$\begin{array}{ll} \underset{t,\boldsymbol{x}}{\text{minimize}} & t \\ \text{subject to} & -t\mathbf{1} \leq \boldsymbol{x} \leq t\mathbf{1} \\ & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

Linear Programming (LP): Norm Miminization

ℓ_1-norm minimization as an LP:

- Original convex problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \|\boldsymbol{x}\|_1 \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

- Equivalent LP:

$$\begin{array}{ll} \underset{\boldsymbol{t},\boldsymbol{x}}{\text{minimize}} & \sum_i t_i \\ \text{subject to} & -\boldsymbol{t} \leq \boldsymbol{x} \leq \boldsymbol{t} \\ & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

Summary

- Linear programs are a special class of convex problems where all functions are affine.
- The geometric interpretation of LPs helps in understanding the simplex method.
- Some norm minimization problems can be rewritten as LPs, making them easier to solve using linear programming techniques.

Linear-Fractional Program (LFP): The problem of minimizing a ratio of affine functions over a polyhedron is called a *linear-fractional program* (LFP):

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \dfrac{\boldsymbol{c}^\mathsf{T}\boldsymbol{x}+d}{\boldsymbol{e}^\mathsf{T}\boldsymbol{x}+f} \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

with $\operatorname{dom} f_0 = \left\{\boldsymbol{x} \mid \boldsymbol{e}^\mathsf{T}\boldsymbol{x}+f > 0\right\}$.

Properties of LFP

- An LFP is not a convex problem, but it is quasiconvex.
- Therefore, it can be solved via bisection by sequentially solving a series of feasibility LPs:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{find}} & \boldsymbol{x} \\ \text{subject to} & t(\boldsymbol{e}^\mathsf{T}\boldsymbol{x}+f) \geq \boldsymbol{c}^\mathsf{T}\boldsymbol{x}+d \\ & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{array}$$

Charnes-Cooper Transform

Alternatively, the LFP can be transformed into an LP via the Charnes-Cooper transform:

$$\begin{array}{ll} \underset{\boldsymbol{y},t}{\text{minimize}} & \boldsymbol{c}^\mathsf{T}\boldsymbol{y} + dt \\ \text{subject to} & \boldsymbol{G}\boldsymbol{y} \leq \boldsymbol{h}t \\ & \boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}t \\ & \boldsymbol{e}^\mathsf{T}\boldsymbol{y} + ft = 1 \\ & t > 0, \end{array}$$

with variables $\boldsymbol{y}, t$, related to the original variable $\boldsymbol{x}$ as:

$$\boldsymbol{y} = \frac{\boldsymbol{x}}{\boldsymbol{e}^\mathsf{T}\boldsymbol{x} + f} \quad \text{and} \quad t = \frac{1}{\boldsymbol{e}^\mathsf{T}\boldsymbol{x} + f}.$$

The original variable can be recovered from $\boldsymbol{y}, t$ as $\boldsymbol{x} = \boldsymbol{y}/t$.

Summary

- LFPs are quasiconvex problems that can be solved using the bisection method by solving a series of feasibility LPs.
- Alternatively, the Charnes-Cooper transform can convert an LFP into an LP.

Quadratic Programming

Quadratic Program (QP): A convex optimization problem is called a *quadratic program* (QP) if the objective function is (convex) quadratic, and the constraint functions are affine:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^\mathsf{T}\boldsymbol{x} + r \\ \text{subject to} & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

where $\boldsymbol{P} \succeq \mathbf{0}$. QPs include LPs as a special case when $\boldsymbol{P} = \mathbf{0}$.

Geometric Interpretation of a QP:

- Visualized as an elliptical surface intersecting a polyhedron.
- The optimal solution does not necessarily coincide with a vertex of the polyhedron.

Quadratic Programming

Examples:

- **Least squares (LS):** The LS problem:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$$

  is an unconstrained QP.

- **Box-constrained LS:** The following regression problem with upper and lower bounds on the variables:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \\ \text{subject to} & l_i \leq x_i \leq u_i, \quad i = 1, \ldots, n \end{array}$$

  is a QP.

Quadratic Programming

Quadratically Constrained Quadratic Program (QCQP): If the objective function as well as the inequality constraints are (convex) quadratic, then the problem is called a *quadratically constrained quadratic program* (QCQP):

$$
\begin{array}{ll}
\underset{\boldsymbol{x}}{\text{minimize}} & \frac{1}{2}\boldsymbol{x}^{\mathsf{T}}\boldsymbol{P}_0\boldsymbol{x} + \boldsymbol{q}_0^{\mathsf{T}}\boldsymbol{x} + r_0 \\
\text{subject to} & \frac{1}{2}\boldsymbol{x}^{\mathsf{T}}\boldsymbol{P}_i\boldsymbol{x} + \boldsymbol{q}_i^{\mathsf{T}}\boldsymbol{x} + r_i \leq 0, \quad i = 1, \dots, m \\
& \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},
\end{array}
$$

where $\boldsymbol{P}_i \succeq \boldsymbol{0}$. In this case, the feasible region is the intersection of ellipsoids. QCQPs include QPs as a special case when $\boldsymbol{P}_i = \boldsymbol{0}$ for $i = 1, \dots, m$.

Summary:

- QPs are convex optimization problems with a quadratic objective function and affine constraints.
- QCQPs extend QPs by allowing quadratic inequality constraints, resulting in feasible regions that are intersections of ellipsoids.
- Both QPs and QCQPs are important classes of convex optimization problems with numerous applications.

Second-order Cone Program (SOCP): A convex optimization problem is called a *second-order cone program* (SOCP) if it has the form:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{f}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i\| \leq \boldsymbol{c}_i^\mathsf{T}\boldsymbol{x} + d_i, \quad i = 1, \ldots, m \\ & \boldsymbol{F}\boldsymbol{x} = \boldsymbol{g}, \end{array}$$

where the constraints $\|\boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i\| \leq \boldsymbol{c}_i^\mathsf{T}\boldsymbol{x} + d_i$ are called *second-order cone* (SOC) constraints since they are affine compositions of the (convex) SOC:

$$\left\{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} \mid \|\boldsymbol{x}\| \leq t\right\}.$$

Relationship to Other Problems

- An SOCP reduces to a QCQP when $\boldsymbol{c}_i = 0$ for $i = 1, \ldots, m$ (by squaring both sides of the inequalities).
- If each $\boldsymbol{A}_i$ is a row-vector (or $\boldsymbol{A}_i = 0$), then an SOCP reduces to an LP.

Semidefinite Program (SDP): A more general convex problem than an SOCP is the *semidefinite program* (SDP), formulated as:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} \\ \text{subject to} & x_1\boldsymbol{F}_1 + x_2\boldsymbol{F}_2 + \cdots + x_n\boldsymbol{F}_n + \boldsymbol{G} \preceq \boldsymbol{0} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

which has *linear matrix inequality* (LMI) constraints of the form:

$$x_1\boldsymbol{F}_1 + \ldots + x_n\boldsymbol{F}_n + \boldsymbol{G} \preceq \boldsymbol{0},$$

where $\boldsymbol{F}_1, \cdots, \boldsymbol{F}_n, \boldsymbol{G} \in \mathbb{S}^k$ ($\mathbb{S}^k$ is the set of symmetric $k \times k$ matrices) and $\boldsymbol{A} \succeq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite.

Reduction to LP

When the matrix in the LMI inequality is diagonal:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \operatorname{Diag}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}) \preceq \boldsymbol{0}, \end{array}$$

the SDP is equivalent to the LP:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}. \end{array}$$

Reduction to SOCP

When the matrix in the LMI has a specific 2×2 block structure:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{f}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \begin{bmatrix} (\boldsymbol{c}_i^\mathsf{T}\boldsymbol{x} + d_i)\boldsymbol{I} & \boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i \\ (\boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i)^\mathsf{T} & \boldsymbol{c}_i^\mathsf{T}\boldsymbol{x} + d_i \end{bmatrix} \succeq \boldsymbol{0}, \quad i = 1, \ldots, m, \end{array}$$

the SDP is equivalent to the SOCP:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{f}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i\| \leq \boldsymbol{c}_i^\mathsf{T}\boldsymbol{x} + d_i, \quad i = 1, \ldots, m. \end{array}$$

Schur Complement

- The equivalence can be shown via the *Schur complement*:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^\mathsf{T} & \boldsymbol{C} \end{bmatrix} \succeq \boldsymbol{0} \quad \Longleftrightarrow \quad \boldsymbol{S} = \boldsymbol{C} - \boldsymbol{B}^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{B} \succeq \boldsymbol{0},$$

where we have tacitly assumed $\boldsymbol{A} \succ \boldsymbol{0}$.

Example: Eigenvalue Minimization

The maximum eigenvalue minimization problem:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \lambda_{\max}(\boldsymbol{A}(\boldsymbol{x})),$$

where $\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{A}_0 + x_1\boldsymbol{A}_1 + \cdots + x_n\boldsymbol{A}_n$, is equivalent to the SDP:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & t \\ \text{subject to} & \boldsymbol{A}(\boldsymbol{x}) \preceq t\boldsymbol{I}. \end{array}$$

This follows from:

$$\lambda_{\max}(\boldsymbol{A}(\boldsymbol{x})) \leq t \quad \Longleftrightarrow \quad \boldsymbol{A}(\boldsymbol{x}) \preceq t\boldsymbol{I}.$$

Summary

- SDPs are a general class of convex optimization problems with LMI constraints.
- They generalize LPs and SOCPs, providing a powerful modeling framework.
- The Schur complement is a useful tool for SDPs.
- For more details on SDPs, refer to (Vandenberghe and Boyd 1996).

Generalized Convex Optimization Problem:
A useful generalization of the standard convex optimization problem can be achieved by allowing the inequality constraints to be vector-valued and incorporating generalized inequalities into the constraints:

$$
\begin{array}{ll}
\underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\
\text{subject to} & \boldsymbol{f}_i(\boldsymbol{x}) \preceq_{\mathcal{K}_i} \mathbf{0}, \\
& \boldsymbol{h}_i(\boldsymbol{x}) = \mathbf{0}, \\
& 1 \leq i \leq m, \\
& 1 \leq i \leq p,
\end{array}
$$

where the generalized inequalities $\preceq_{\mathcal{K}_i}$ are defined by the proper cones $\mathcal{K}_i$ (note that $\boldsymbol{a} \preceq_{\mathcal{K}} \boldsymbol{b} \Leftrightarrow \boldsymbol{b} - \boldsymbol{a} \in \mathcal{K}$) and f_i are $\mathcal{K}_i$-convex.

Generalized Inequalities:

- A generalized inequality is a partial ordering on $\mathbb{R}^n$ that has many of the properties of the standard ordering on $\mathbb{R}$.
- Example: The matrix inequality defined by the cone of positive semidefinite $n \times n$ matrices $\mathbb{S}_+^n$.

$\mathcal{K}$-convex Functions:

- A function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^{k_i}$ is $\mathcal{K}_i$-convex if the domain is a convex set and, for all $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$ and $\theta \in [0, 1]$,

$$\boldsymbol{f}(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \preceq_{\mathcal{K}_i} \theta \boldsymbol{f}(\boldsymbol{x}) + (1-\theta)\boldsymbol{f}(\boldsymbol{y}).$$

Cone Programs (CP): Among the simplest convex optimization problems with generalized inequalities are *cone programs* (CP) (or *conic-form problems*), which have a linear objective and one inequality constraint function:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^T \boldsymbol{x} \\ \text{subject to} & \boldsymbol{F}\boldsymbol{x} + \boldsymbol{g} \preceq_{\mathcal{K}} \boldsymbol{0} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{array}$$

Special Cases of CPs:

- **Linear Programs (LPs):** If $\mathcal{K} = \mathbb{R}^n_+$ (nonnegative orthant), the partial ordering $\preceq_{\mathcal{K}}$ is the usual componentwise inequality between vectors and the CP reduces to an LP.
- **Second-order Cone Programs (SOCPs):** If $\mathcal{K} = \mathcal{C}^n$ (second-order cone), $\preceq_{\mathcal{K}}$ corresponds to a constraint of the form:

$$\left\{(\boldsymbol{x}, t) \in \mathbb{R}^{n+1} \mid \|\boldsymbol{x}\| \leq t\right\},$$

  and the CP becomes an SOCP.
- **Semidefinite Programs (SDPs):** If $\mathcal{K} = \mathbb{S}^n_+$ (positive semidefinite cone), the generalized inequality $\preceq_{\mathcal{K}}$ reduces to the usual matrix inequality:

$$x_1 \boldsymbol{F}_1 + \ldots + x_n \boldsymbol{F}_n + \boldsymbol{G} \preceq \mathbf{0},$$

  and the CP simplifies to an SDP.

Summary:

- Generalized convex optimization problems extend the standard form by incorporating vector-valued inequality constraints and generalized inequalities.
- Cone programs (CPs) are a special class of these problems

Fractional Programming (FP) involves optimization problems that include ratios.

- The simplest form of a *fractional program* (FP) is:

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{maximize}} \quad & \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}, \end{aligned}$$

  where $f(\boldsymbol{x}) \geq 0$, $g(\boldsymbol{x}) > 0$, and $\mathcal{X}$ denotes the feasible set.
- One particular case is the LFP, where both f and g are linear functions.

Applications and Extensions:

- FPs have been widely studied and extended to deal with multiple ratios, such as:

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{maximize}} \quad & \min_{i} \frac{f_i(\boldsymbol{x})}{g_i(\boldsymbol{x})} \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}. \end{aligned}$$

- FPs are nonconvex problems, making them challenging to solve.

Concave-Convex FP:

In the case known as *concave-convex FP*, where f is a concave function and g is a convex function, they can be solved relatively easily using different methods.

Methods for Solving Concave-Convex FP:

- **Bisection method:** Similar to the linear case of LFP, the bisection method involves solving a sequence of convex feasibility problems of the form:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{find}} & \boldsymbol{x} \\ \text{subject to} & t g(\boldsymbol{x}) \leq f(\boldsymbol{x}) \\ & \boldsymbol{x} \in \mathcal{X}. \end{array}$$

- **Dinkelbach's transform:** This approach eliminates the fractional objective by solving a sequence of simpler convex problems of the form:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{maximize}} & f(\boldsymbol{x}) - y^k g(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \end{array}$$

  where the weight y^k is updated as $y^k = f(\boldsymbol{x}^k)/g(\boldsymbol{x}^k)$.

Methods for Solving Concave-Convex FP: (cont'd)

- **Schaible transform:** This is a more general case of the Charnes-Cooper transform (used for LFPs). The original concave-convex FP is transformed into an equivalent convex problem:

$$\begin{array}{ll} \underset{\boldsymbol{y},t}{\text{maximize}} & tf\left(\frac{\boldsymbol{y}}{t}\right) \\ \text{subject to} & tg\left(\frac{\boldsymbol{y}}{t}\right) \leq 1 \\ & t > 0 \\ & \boldsymbol{y}/t \in \mathcal{X}, \end{array}$$

  with variables $\boldsymbol{y}, t$, related to the original variable $\boldsymbol{x}$ by:

$$\boldsymbol{y} = \frac{\boldsymbol{x}}{g(\boldsymbol{x})} \quad \text{and} \quad t = \frac{1}{g(\boldsymbol{x})}.$$

  The original variable can be easily recovered from $\boldsymbol{y}, t$ by $\boldsymbol{x} = \boldsymbol{y}/t$.

Monomial and Posynomial Functions:

- A *monomial function*, or simply a *monomial*, is a function $f : \mathbb{R}^n \to \mathbb{R}$ with $\operatorname{dom} f = \mathbb{R}^n_{++}$ defined as:

$$f(\boldsymbol{x}) = c x_1^{a_1} x_2^{a_2} \dots x_n^{a_n},$$

  where $c > 0$ and $a_i \in \mathbb{R}$.

- A *posynomial function*, or simply a *posynomial*, is a sum of monomials:

$$f(\boldsymbol{x}) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \dots x_n^{a_{nk}},$$

  where $c_k > 0$.

Geometric Programming

Geometric Program (GP) is a (nonconvex) problem of the form:

$$
\begin{array}{ll}
\underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\
\text{subject to} & f_i(\boldsymbol{x}) \leq 1, \\
 & h_i(\boldsymbol{x}) = 1, \\
 & i = 1, \ldots, m, \\
 & i = 1, \ldots, p,
\end{array}
$$

where $f_0, \ldots, f_m$ are posynomials and $h_1, \ldots, h_p$ are monomials. The domain of this problem is $\mathcal{D} = \mathbb{R}^n_{++}$, i.e., the constraint $\boldsymbol{x} > \mathbf{0}$ is implicit.

Transformation to Convex Form:

- Apply the change of variables $y_i = \log x_i$ and $b = \log c$ on the monomial $f(\boldsymbol{x}) = c x_1^{a_1} x_2^{a_2} \dots x_n^{a_n}$.
- The log of the monomial is:

$$\tilde{f}(\boldsymbol{y}) = \log f(e^{\boldsymbol{x}}) = b + a_1 y_1 + a_2 y_2 + \cdots + a_n y_n = b + \boldsymbol{a}^\mathsf{T} \boldsymbol{y},$$

  which is an affine function.
- For a posynomial, we obtain:

$$\tilde{f}(\boldsymbol{y}) = \log \sum_{k=1}^{K} e^{b_k + \boldsymbol{a}_k^\mathsf{T} \boldsymbol{y}},$$

  which is the so-called log-sum-exp function, a convex function.

Transformed GP in Convex Form:

- The resulting transformed GP in convex form is:

$$
\begin{array}{ll}
\underset{\mathbf{y}}{\text{minimize}} & \log \sum_{k=1}^{K_0} e^{b_{0k}+\mathbf{a}_{0k}^T \mathbf{y}} \\
\text{subject to} & \log \sum_{k=1}^{K_i} e^{b_{ik}+\mathbf{a}_{ik}^T \mathbf{y}} \leq 0, \\
& h_i + \mathbf{g}_i^T \mathbf{y} = 0, \\
& i = 1, \ldots, m, \\
& i = 1, \ldots, p.
\end{array}
$$

- Comprehensive monographs on GP include (S. Boyd et al. 2007) and (Chiang 2005).

Summary:

- Geometric programming involves optimization problems with monomial and posynomial functions.
- These problems are not naturally convex but can be converted into convex optimization problems through a logarithmic change of variables and transformation of the objective and constraint functions.

Outline

Lagrange Duality

- Links the original minimization problem (primal problem) with a maximization problem (dual problem).
- Sometimes the dual problem is simpler to solve than the primal one.
- Fundamental results: duality gap and Karush-Kuhn-Tucker (KKT) optimality conditions.
- KKT optimality conditions:
  - may help obtain a closed-form solution to the original problem
  - allow to characterize properties of the solutions
  - are key in developing primal-dual interior-point methods.

Optimization Problem in Standard Form

$$
\begin{array}{lll}
\underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) & \\
\text{subject to} & f_i(\boldsymbol{x}) \leq 0, & i = 1, \dots, m \\
& h_i(\boldsymbol{x}) = 0, & i = 1, \dots, p,
\end{array}
$$

where variable $\boldsymbol{x} \in \mathbb{R}^n$, domain $\mathcal{D}$, and optimal value $p^\star$

Lagrangian Definition: $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$:

$$
L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x})
$$

Lagrange Multipliers

- λ_i and ν_i are the Lagrange multipliers for inequality and equality constraints, respectively
- vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are called dual variables or Lagrange multiplier vectors

Lagrange Duality

Lagrange Dual Function:

- Defined as the minimum value of the Lagrangian over $\boldsymbol{x}$ for a given $(\boldsymbol{\lambda}, \boldsymbol{\nu})$:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in \mathcal{D}} \left\{ f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x}) \right\}.$$

- Infimum is with respect to all $\boldsymbol{x} \in \mathcal{D}$ (not necessarily feasible points).
- Dual function is concave, even if the original problem is not convex.

Primal and Dual Variables/Functions:

- Original optimization variable $\boldsymbol{x}$ is the primal variable.
- Lagrange multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are the dual variables
- Original objective function $f_0(\boldsymbol{x})$ is the primal function.
- The infimum of the Lagrangian is the dual function.

Lower Bounds on the Optimal Value:

- Dual function $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ provides lower bounds on the optimal value $p^\star$:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$$

- Holds even if the original problem is not convex.
- Verified through inequalities for any feasible $\boldsymbol{x}$:

$$\begin{aligned} f_0(\boldsymbol{x}) &\geq f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x}) \\ &\geq \inf_{\boldsymbol{z} \in \mathcal{D}} \left\{ f_0(\boldsymbol{z}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{z}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{z}) \right\} \\ &= g(\boldsymbol{\lambda}, \boldsymbol{\nu}), \end{aligned}$$

- Primal-dual feasible pair $(\boldsymbol{x}, (\boldsymbol{\lambda}, \boldsymbol{\nu}))$ localizes the optimal value within an interval:

$$p^\star \in \left[g\left(\boldsymbol{\lambda}, \boldsymbol{\nu}\right), f_0(\boldsymbol{x})\right].$$

- Utilized in optimization algorithms for non-heuristic stopping criteria.

Lagrange Dual Problem

Tightest Lower Bound

- Lagrange dual function gives a lower bound on the optimal value $p^\star$: $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$.
- Tightest lower bound leads to the Lagrange dual problem:

$$
\begin{array}{ll}
\underset{\boldsymbol{\lambda}, \boldsymbol{\nu}}{\text{maximize}} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\
\text{subject to} & \boldsymbol{\lambda} \geq \mathbf{0}.
\end{array}
$$

Dual Problem and Primal Problem

- Original problem is called the primal problem.
- Variables $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ are dual feasible if $\boldsymbol{\lambda} \geq \mathbf{0}$ and $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$.
- $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ are dual optimal if they are optimal for the dual problem.
- Optimal value of the dual problem is denoted by $d^\star$.

Example: Least-norm Solution of Linear Equations

- Problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{x}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{array}$$

- Lagrangian:

$$L(\boldsymbol{x}, \boldsymbol{\nu}) = \boldsymbol{x}^\mathsf{T}\boldsymbol{x} + \boldsymbol{\nu}^\mathsf{T}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}).$$

- Dual function:

$$g(\boldsymbol{\nu}) = L\left(-\frac{1}{2}\boldsymbol{A}^\mathsf{T}\boldsymbol{\nu}, \boldsymbol{\nu}\right) = -\frac{1}{4}\boldsymbol{\nu}^\mathsf{T}\boldsymbol{A}\boldsymbol{A}^\mathsf{T}\boldsymbol{\nu} - \boldsymbol{b}^\mathsf{T}\boldsymbol{\nu}.$$

- Lower bound property:

$$p^\star \geq -\frac{1}{4}\boldsymbol{\nu}^\mathsf{T}\boldsymbol{A}\boldsymbol{A}^\mathsf{T}\boldsymbol{\nu} - \boldsymbol{b}^\mathsf{T}\boldsymbol{\nu} \quad \text{for all } \boldsymbol{\nu}.$$

- Dual problem (QP):

$$\underset{\boldsymbol{\nu}}{\text{maximize}} \quad -\tfrac{1}{4}\boldsymbol{\nu}^\mathsf{T}\boldsymbol{A}\boldsymbol{A}^\mathsf{T}\boldsymbol{\nu} - \boldsymbol{b}^\mathsf{T}\boldsymbol{\nu}.$$

Lagrange Dual Problem

Example: Standard Form LP

- Problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \geq \boldsymbol{0}. \end{array}$$

- Lagrangian:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + \boldsymbol{\nu}^\mathsf{T}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{\lambda}^\mathsf{T}\boldsymbol{x} = \left(\boldsymbol{c} + \boldsymbol{A}^\mathsf{T}\boldsymbol{\nu} - \boldsymbol{\lambda}\right)^\mathsf{T}\boldsymbol{x} - \boldsymbol{b}^\mathsf{T}\boldsymbol{\nu}.$$

- Dual function:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\boldsymbol{b}^\mathsf{T}\boldsymbol{\nu} & \boldsymbol{c} + \boldsymbol{A}^\mathsf{T}\boldsymbol{\nu} - \boldsymbol{\lambda} = \boldsymbol{0} \\ -\infty & \text{otherwise,} \end{cases}$$

- Lower bound property:

$$p^\star \geq -\boldsymbol{b}^\mathsf{T}\boldsymbol{\nu} \quad \text{if } \boldsymbol{c} + \boldsymbol{A}^\mathsf{T}\boldsymbol{\nu} \geq \boldsymbol{0}.$$

- Dual problem (LP):

$$\begin{array}{ll} \underset{\boldsymbol{\nu}}{\text{maximize}} & -\boldsymbol{b}^\mathsf{T}\boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{c} + \boldsymbol{A}^\mathsf{T}\boldsymbol{\nu} \geq \boldsymbol{0}. \end{array}$$

Lagrange Dual Problem

Example: Two-way partitioning

- Nonconvex problem (due to matrix $\boldsymbol{W} \not\succeq \mathbf{0}$ and quadratic equality constraints):

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{x}^T \boldsymbol{W} \boldsymbol{x} \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n. \end{array}$$

- Lagrangian:

$$L(\boldsymbol{x}, \boldsymbol{\nu}) = \boldsymbol{x}^T \boldsymbol{W} \boldsymbol{x} + \sum_{i=1}^{n} \nu_i \left(x_i^2 - 1 \right) = \boldsymbol{x}^T \left(\boldsymbol{W} + \operatorname{Diag}(\boldsymbol{\nu}) \right) \boldsymbol{x} - \mathbf{1}^T \boldsymbol{\nu}.$$

- Dual function:

$$g(\boldsymbol{\nu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{1}^T \boldsymbol{\nu} & \boldsymbol{W} + \operatorname{Diag}(\boldsymbol{\nu}) \succeq \mathbf{0} \\ -\infty & \text{otherwise.} \end{cases}$$

- Dual problem (SDP):

$$\begin{array}{ll} \underset{\boldsymbol{\nu}}{\text{maximize}} & -\mathbf{1}^T \boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{W} + \operatorname{Diag}(\boldsymbol{\nu}) \succeq \mathbf{0}. \end{array}$$

Lagrange Dual Problem

Example: Introducing new variables

- Original problem:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2 \,.$$

- Introduce dummy variables:

$$\begin{array}{ll} \underset{\boldsymbol{x},\boldsymbol{y}}{\text{minimize}} & \|\boldsymbol{y}\|_2 \\ \text{subject to} & \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}. \end{array}$$

- Dual problem:

$$\begin{array}{ll} \underset{\boldsymbol{\nu}}{\text{maximize}} & \boldsymbol{b}^{\mathsf{T}}\boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{A}^{\mathsf{T}}\boldsymbol{\nu} = \boldsymbol{0}, \quad \|\boldsymbol{\nu}\|_2 \leq 1. \end{array}$$

Example: Implicit constraints

- Original problem (LP with box constraints):

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & -\mathbf{1} \leq \boldsymbol{x} \leq \mathbf{1}. \end{array}$$

- Dual problem:

$$\begin{array}{ll} \underset{\boldsymbol{\nu},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}{\text{maximize}} & -\boldsymbol{b}^\mathsf{T}\boldsymbol{\nu} - \mathbf{1}^\mathsf{T}\boldsymbol{\lambda}_1 - \mathbf{1}^\mathsf{T}\boldsymbol{\lambda}_2 \\ \text{subject to} & \boldsymbol{c} + \boldsymbol{A}^\mathsf{T}\boldsymbol{\nu} + \boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2 = 0 \\ & \boldsymbol{\lambda}_1 \geq \mathbf{0}, \quad \boldsymbol{\lambda}_2 \geq \mathbf{0}. \end{array}$$

- Rewriting the primal problem with implicit constraints:

$$f_0(\boldsymbol{x}) = \begin{cases} \boldsymbol{c}^\mathsf{T}\boldsymbol{x} & -\mathbf{1} \leq \boldsymbol{x} \leq \mathbf{1} \\ \infty & \text{otherwise} \end{cases}$$

- More insightful dual problem:

$$\underset{\boldsymbol{\nu}}{\text{maximize}} \quad -\boldsymbol{b}^\mathsf{T}\boldsymbol{\nu} - \left\|\boldsymbol{A}^\mathsf{T}\boldsymbol{\nu} + \boldsymbol{c}\right\|_1.$$

Weak and Strong Duality

Weak Duality:

- Optimal value $d^\star$ of the Lagrange dual problem is the tightest lower bound on $p^\star$:

$$d^\star \leq p^\star$$

- Difference $\Gamma = p^\star - d^\star$ is called the optimal duality gap (always nonnegative).
- Useful for establishing a lower limit on the optimal value of a challenging problem.

Strong Duality:

- Equality in weak duality:

$$d^\star = p^\star$$

- Implies the duality gap is zero.
- Strong duality is desirable and facilitates solving difficult problems via the dual.
- Does not hold for general optimization problems but often holds for convex problems under certain conditions (constraint qualifications).

Weak and Strong Duality

Constraint Qualifications: (there are many types)

Slater's Condition:

- Requires the existence of a strictly feasible point:

$$\boldsymbol{x} \in \operatorname{relint} \mathcal{D} \text{ such that } f_i(\boldsymbol{x}) < 0,\ i = 1, \ldots, m, \text{ and } h_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, p.$$

- Strong duality holds if Slater's condition is met and the problem is convex.

Example: Inequality Form LP

- Problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^{\mathsf{T}} \boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}. \end{array}$$

- Dual problem:

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}}{\text{maximize}} & -\boldsymbol{b}^{\mathsf{T}} \boldsymbol{\lambda} \\ \text{subject to} & \boldsymbol{A}^{\mathsf{T}} \boldsymbol{\lambda} + \boldsymbol{c} = \boldsymbol{0}, \quad \boldsymbol{\lambda} \geq 0. \end{array}$$

- Strong duality holds if $\boldsymbol{A}\tilde{\boldsymbol{x}} < \boldsymbol{b}$ for some $\tilde{\boldsymbol{x}}$.
- In this case, $p^\star = d^\star$ always holds (except when both primal and dual problems are infeasible).

Weak and Strong Duality

Example: Convex QP

- Problem (with $\boldsymbol{P} \succeq \mathbf{0}$):

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{x}^{\mathsf{T}}\boldsymbol{P}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}. \end{array}$$

- Dual problem:

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}}{\text{maximize}} & -\frac{1}{4}\boldsymbol{\lambda}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{P}^{-1}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{\lambda} - \boldsymbol{b}^{\mathsf{T}}\boldsymbol{\lambda} \\ \text{subject to} & \boldsymbol{\lambda} \geq \mathbf{0}. \end{array}$$

- Strong duality holds if $\boldsymbol{A}\tilde{\boldsymbol{x}} < \boldsymbol{b}$ for some $\tilde{\boldsymbol{x}}$.
- In this case, $p^\star = d^\star$ always holds.

Example: Nonconvex QP

- Problem (with $\boldsymbol{P} \not\succeq \mathbf{0}$):

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{x}^\mathsf{T}\boldsymbol{A}\boldsymbol{x} + 2\boldsymbol{b}^\mathsf{T}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{x}^\mathsf{T}\boldsymbol{x} \leq 1. \end{array}$$

- Dual problem (SDP):

$$\begin{array}{ll} \underset{t,\lambda}{\text{maximize}} & -t - \lambda \\ \text{subject to} & \begin{bmatrix} \boldsymbol{A} + \lambda\boldsymbol{I} & \boldsymbol{b} \\ \boldsymbol{b}^\mathsf{T} & t \end{bmatrix} \succeq 0. \end{array}$$

- Strong duality holds even though the original problem is nonconvex (not trivial to show).

Complementary Slackness

Strong Duality and Optimality

If strong duality holds, primal and dual optimal values are equal: $d^\star = p^\star$.

Let $\boldsymbol{x}^\star$ be a primal optimal point and $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ be a dual optimal point:

$$\begin{aligned}
f_0(\boldsymbol{x}^\star) &= g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) \\
&= \inf_{\boldsymbol{x}\in\mathcal{D}} \left\{ f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^\star f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^\star h_i(\boldsymbol{x}) \right\} \\
&\le f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{p} \nu_i^\star h_i(\boldsymbol{x}^\star) \\
&\le f_0(\boldsymbol{x}^\star),
\end{aligned}$$

where

- the first line comes from the zero duality gap,
- the second line is the definition of the dual function,
- the third line follows from the definition of infimum,
- the fourth line results from feasibility ($\lambda_i^\star \ge 0$, $f_i(\boldsymbol{x}^\star) \le 0$, and $h_i(\boldsymbol{x}^\star) = 0$).

Complementary Slackness

Complementary Slackness Conditions

The two inequalities in the chain must hold with equality.

- Equality in the first inequality means $\boldsymbol{x}^\star$ minimizes $L(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ over $\boldsymbol{x}$.
- Equality in the second inequality implies:

$$\sum_{i=1}^{m} \lambda_i^\star f_i(\boldsymbol{x}^\star) = 0.$$

- This leads (because each term is nonpositive) to the complementary slackness conditions:

$$\lambda_i^\star f_i(\boldsymbol{x}^\star) = 0, \quad i = 1, \ldots, m.$$

- They hold for any primal optimal $\boldsymbol{x}^\star$ and dual optimal $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ (under strong duality).

Interpretation of Complementary Slackness

- If a Lagrange multiplier is active ($\lambda_i^\star > 0$), then the constraint is at the boundary of the feasible set $f_i\left(x^\star\right) = 0$.
- If a constraint is strictly feasible ($f_i(x^\star) < 0$), then $\lambda_i^\star = 0$ (the Lagrange multiplier is not necessary).

Karush-Kuhn-Tucker (KKT) Optimality Conditions

Preliminaries

- assume functions $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ are differentiable
- let $\boldsymbol{x}^\star$ and $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ be any primal and dual optimal points with zero duality gap
- since $\boldsymbol{x}^\star$ minimizes $L(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ over $\boldsymbol{x}$, its gradient must vanish:

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

KKT Conditions: For any optimization problem (not necessarily convex), any pair of optimal and dual points, $\boldsymbol{x}^\star$ and $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$, must satisfy:

$$\begin{aligned}
f_i(\boldsymbol{x}^\star) &\leq 0, \quad i = 1, \ldots, m && \text{(primal feasibility)} \\
h_i(\boldsymbol{x}^\star) &= 0, \quad i = 1, \ldots, p && \\
\lambda_i^\star &\geq 0, \quad i = 1, \ldots, m && \text{(dual feasibility)} \\
\lambda_i^\star f_i(\boldsymbol{x}^\star) &= 0, \quad i = 1, \ldots, m && \text{(complementary slackness)} \\
\nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(\boldsymbol{x}^\star) &= \mathbf{0} && \text{(zero Lagrangian gradient)}
\end{aligned}$$

Karush-Kuhn-Tucker (KKT) Optimality Conditions

Sufficiency of KKT Conditions for Convex Problems

- for convex optimization problems, the KKT conditions are also sufficient for optimality
- if $\tilde{\boldsymbol{x}}$ achieves a zero gradient in the Lagrangian, it minimizes the Lagrangian:

$$\begin{aligned} g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) &= L(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) \\ &= f_0(\tilde{\boldsymbol{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\boldsymbol{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{\boldsymbol{x}}) \\ &= f_0(\tilde{\boldsymbol{x}}), \end{aligned}$$

- this shows zero duality gap and, therefore, primal and dual optimality

Importance of KKT Conditions

- Play a key role in optimization.
- Can sometimes characterize the solution analytically.
- Many algorithms are conceived or can be interpreted as methods for solving the KKT conditions.

Karush-Kuhn-Tucker (KKT) Optimality Conditions

KKT conditions:

$$
\begin{aligned}
f_i(\boldsymbol{x}^\star) &\le 0, \quad i = 1, \dots, m & &\text{(primal feasibility)} \\
h_i(\boldsymbol{x}^\star) &= 0, \quad i = 1, \dots, p & & \\
\lambda_i^\star &\ge 0, \quad i = 1, \dots, m & &\text{(dual feasibility)} \\
\lambda_i^\star f_i(\boldsymbol{x}^\star) &= 0, \quad i = 1, \dots, m & &\text{(complementary slackness)} \\
\nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(\boldsymbol{x}^\star) &= \mathbf{0} & &\text{(zero Lagrangian gradient)}
\end{aligned}
$$

Theorem: KKT optimality conditions

For any optimization problem with differentiable functions:

- For any optimization problem (not necessarily convex) with strong duality, the KKT conditions are necessary for optimality.
- For a convex optimization problem satisfying Slater's condition, strong duality follows, and the KKT conditions are necessary and sufficient for optimality.

- Consider the original optimization problem:

$$\begin{array}{lll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) & \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, & i = 1, \ldots, m \\ & h_i(\boldsymbol{x}) = 0 & i = 1, \ldots, p, \end{array}$$

  and its dual problem:

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}, \boldsymbol{\nu}}{\text{maximize}} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{\lambda} \geq \mathbf{0}. \end{array}$$

- We now add a **perturbation** on the constraints:

$$f_i(\boldsymbol{x}) \leq u_i$$

  and

$$h_i(\boldsymbol{x}) = v_i.$$

Perturbation and Sensitivity Analysis

- The perturbed optimization problem is defined as

$$\begin{array}{lll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) & \\ \text{subject to} & f_i(\boldsymbol{x}) \leq u_i, & i = 1, \ldots, m \\ & h_i(\boldsymbol{x}) = v_i, & i = 1, \ldots, p \end{array}$$

and its corresponding perturbed dual problem becomes

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}, \boldsymbol{\nu}}{\text{maximize}} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) - \boldsymbol{u}^{\mathsf{T}} \boldsymbol{\lambda} - \boldsymbol{v}^{\mathsf{T}} \boldsymbol{\nu} \\ \text{subject to} & \boldsymbol{\lambda} \geq \mathbf{0}. \end{array}$$

- We define $p^{\star}(\boldsymbol{u}, \boldsymbol{v})$ as the optimal value of the perturbed problem as a function of $\boldsymbol{u}$ and $\boldsymbol{v}$.

Perturbation and Sensitivity Analysis

- **Local sensitivity:** If strong duality holds for the original problem and $p^\star(\boldsymbol{u}, \boldsymbol{v})$ is differentiable at $(\mathbf{0}, \mathbf{0})$, then

$$\frac{\partial p^\star(\mathbf{0}, \mathbf{0})}{\partial u_i} = -\lambda_i^\star, \qquad \frac{\partial p^\star(\mathbf{0}, \mathbf{0})}{\partial v_i} = -\nu_i^\star.$$

- **Global sensitivity:** If strong duality holds for the original problem, then (from weak duality)

$$\begin{aligned} p^\star(\boldsymbol{u}, \boldsymbol{v}) &\geq g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) - \boldsymbol{u}^\mathsf{T}\boldsymbol{\lambda}^\star - \boldsymbol{v}^\mathsf{T}\boldsymbol{\nu}^\star \\ &= p^\star(\mathbf{0}, \mathbf{0}) - \boldsymbol{u}^\mathsf{T}\boldsymbol{\lambda}^\star - \boldsymbol{v}^\mathsf{T}\boldsymbol{\nu}^\star. \end{aligned}$$

- **Interpretation:**
  - if $\lambda_i^\star$ large: $p^\star$ increases a lot if we tighten constraint i ($u_i < 0$)
  - if $\lambda_i^\star$ small: $p^\star$ does not decrease much if we loosen constraint i ($u_i > 0$)
  - if $\nu_i^\star$ large and positive: $p^\star$ increases a lot if we take $v_i < 0$
  - if $\nu_i^\star$ large and negative: $p^\star$ increases a lot if we take $v_i > 0$.

Outline

Generalized Inequalities

Definition of a Cone: A set $\mathcal{K} \subset \mathbb{R}^n$ such that for every $\boldsymbol{x} \in \mathcal{K}$ and $\theta \geq 0$, we have $\theta \boldsymbol{x} \in \mathcal{K}$.

Proper Cone: A cone $\mathcal{K}$ is called a proper cone if it is:

- convex
- closed
- solid (i.e., has a nonempty interior)
- pointed (i.e., contains no line)

Generalized Inequality

A partial ordering on $\mathbb{R}^n$ with properties similar to the standard ordering on $\mathbb{R}$:

- $\boldsymbol{x} \preceq_{\mathcal{K}} \boldsymbol{y} \quad \Longleftrightarrow \quad \boldsymbol{y} - \boldsymbol{x} \in \mathcal{K}$
- also written as $\boldsymbol{y} \succeq_{\mathcal{K}} \boldsymbol{x}$

Example: Nonnegative Orthant and Componentwise Inequality

- The nonnegative orthant $\mathcal{K} = \mathbb{R}^n_+$ is a proper cone.
- Associated generalized inequality $\preceq_{\mathcal{K}}$ corresponds to componentwise inequality: $\boldsymbol{x} \preceq_{\mathcal{K}} \boldsymbol{y}$ means $x_i \leq y_i$ for $i = 1, \ldots, m$.

Generalized Inequalities

Example: Positive Semidefinite Cone and Matrix Inequality

- The positive semidefinite cone $\mathbb{S}_+^n$ is a proper cone in the set of $n \times n$ symmetric matrices $\mathbb{S}^n$.
- Associated generalized inequality $\preceq_\mathcal{K}$ is the usual matrix inequality: $\boldsymbol{X} \preceq_\mathcal{K} \boldsymbol{Y}$ means $\boldsymbol{Y} - \boldsymbol{X}$ is positive semidefinite (typically written as $\boldsymbol{Y} \succeq \boldsymbol{X}$).

Properties of Generalized Inequality

- Meant to suggest an analogy to ordinary inequality on $\mathbb{R}$ (i.e., $\leq$).
- Many properties of ordinary inequality hold for generalized inequalities.
- Important differences:
  - $\leq$ on $\mathbb{R}$ is a total ordering: any two points are comparable ($x \leq y$ or $y \leq x$).
  - Generalized inequalities define a partial ordering: not all points are comparable.
  - Concepts like minimum and maximum are more complicated.

Generalized Inequalities and Convex Functions

- Extend the definition of a convex function to the vector case.
- A function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^q$ is $\mathcal{K}$-convex if: For all $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} \boldsymbol{f}$ and $\theta \in [0, 1]$: $\boldsymbol{f}(\theta \boldsymbol{x} + (1 - \theta)\boldsymbol{y}) \preceq_\mathcal{K} \theta \boldsymbol{f}(\boldsymbol{x}) + (1 - \theta)\boldsymbol{f}(\boldsymbol{y})$.

Vector Optimization Problem:

$$
\begin{array}{lll}
\underset{\boldsymbol{x}}{\text{minimize}} \text{ (with respect to } \mathcal{K}) & \boldsymbol{f}_0(\boldsymbol{x}) & \\
\text{subject to} & f_i(\boldsymbol{x}) \leq 0, & 1 \leq i \leq m \\
 & h_i(\boldsymbol{x}) = 0, & 1 \leq i \leq p,
\end{array}
$$

where:

- $\mathcal{K} \subset \mathbb{R}^q$ is a proper cone
- $\boldsymbol{f}_0 : \mathbb{R}^n \to \mathbb{R}^q$ is the vector-valued objective function
- $f_i : \mathbb{R}^n \to \mathbb{R}$ are the inequality constraint functions
- $h_i : \mathbb{R}^n \to \mathbb{R}$ are the equality constraint functions.

Comparison with Standard Optimization Problem:

- The objective function takes values in $\mathbb{R}^q$.
- Includes a proper cone $\mathcal{K}$ to compare objective values.

Convex Vector Optimization Problem:

- The problem is convex if:
  - The objective function $\boldsymbol{f}_0$ is $\mathcal{K}$-convex.
  - The inequality constraint functions $f_1, \ldots, f_m$ are convex.
  - The equality constraint functions $h_1, \ldots, h_p$ are affine (can be written as $\boldsymbol{Ax} = \boldsymbol{b}$).

Interpretation of the Problem:

- The general inequality $\preceq_{\mathcal{K}}$ is a partial ordering, not a total ordering.
- This means:
  - We may encounter two points, $\boldsymbol{x}$ and $\boldsymbol{y}$, that are not comparable.
  - Neither $\boldsymbol{f}_0(\boldsymbol{x}) \preceq_{\mathcal{K}} \boldsymbol{f}_0(\boldsymbol{y})$ nor $\boldsymbol{f}_0(\boldsymbol{y}) \preceq_{\mathcal{K}} \boldsymbol{f}_0(\boldsymbol{x})$ holds.
  - This situation cannot occur in the standard optimization problem.

Pareto Optimality

Achievable Objective Values: Set of objective values of feasible points:

$$\mathcal{O} = \{\boldsymbol{f}_0(\boldsymbol{x}) \mid \boldsymbol{x} \text{ is feasible}\}$$

Minimum Element:

- The set $\mathcal{O}$ has a minimum element if there is a feasible $\boldsymbol{x}^\star$ such that:

$$\boldsymbol{f}_0(\boldsymbol{x}^\star) \preceq_{\mathcal{K}} \boldsymbol{f}_0(\boldsymbol{x}) \text{ for all feasible } \boldsymbol{x}$$

- Then $\boldsymbol{x}^\star$ is optimal for the problem and $\boldsymbol{f}_0(\boldsymbol{x}^\star)$ is the optimal value.
- Using set notation, $\boldsymbol{x}^\star$ is optimal if and only if it is feasible and

$$\mathcal{O} \subseteq \boldsymbol{f}_0(\boldsymbol{x}^\star) + \mathcal{K}$$

General Case:

- Most vector optimization problems do not have an optimal point because of incomparable points via the cone $\mathcal{K}$.
- When $\mathcal{O}$ lacks a minimum element, we discuss *minimal elements*.

Pareto Optimality

Minimal Elements and Pareto Optimality:

- Minimal elements are the best among the objective values that can be compared.
- Points $\boldsymbol{x}$ achieving minimal elements in $\mathcal{O}$ are *Pareto optimal* points.
- $\boldsymbol{f}_0(\boldsymbol{x})$ is a *Pareto optimal value* for the vector optimization problem.

Pareto Optimal Point:

- A point $\boldsymbol{x}^{\text{po}}$ is Pareto optimal if it is feasible and, for any other feasible $\boldsymbol{x}$,

$$\boldsymbol{f}_0(\boldsymbol{x}) \preceq_{\mathcal{K}} \boldsymbol{f}_0(\boldsymbol{x}^{\text{po}}) \text{ implies } \boldsymbol{f}_0(\boldsymbol{x}) = \boldsymbol{f}_0(\boldsymbol{x}^{\text{po}}).$$

- Using set notation, $\boldsymbol{x}^{\text{po}}$ is Pareto optimal if and only if it is feasible and

$$(\boldsymbol{f}_0(\boldsymbol{x}^{\text{po}}) - \mathcal{K}) \cap \mathcal{O} = \boldsymbol{f}_0(\boldsymbol{x}^{\text{po}}).$$

- In words, $\boldsymbol{x}^{\text{po}}$ cannot be in the cone of points worse than any other point.

Efficient Frontier:

- A vector optimization problem usually has many Pareto optimal values (and points).
- These values lie on the boundary of the set of achievable objective values, termed the *efficient frontier*.

Multi-Objective Optimization Problem:

- Based on the cone $\mathcal{K} = \mathbb{R}^q_+$ (nonnegative orthant).
- Also called *multicriterion* optimization problem.
- Components of the vector objective function $\boldsymbol{f}_0$, denoted by $F_1, \ldots, F_q$, represent q different scalar objectives to be minimized.
- Simplest case: *bi-objective* or *bi-criterion* optimization problems with two objectives $F_1(\boldsymbol{x})$ and $F_2(\boldsymbol{x})$.

Convex Multi-Objective Optimization Problem:

The problem is convex if:

- Inequality constraint functions $f_1, \ldots, f_m$ are convex.
- Equality constraint functions $h_1, \ldots, h_p$ are affine (denoted as $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$).
- Objectives $F_1, \ldots, F_q$ are convex.

Comparison of Feasible Points:

- For two feasible points $\boldsymbol{x}$ and $\boldsymbol{y}$:
  - $F_i(\mathbf{x}) \leq F_i(\mathbf{y})$ means $\boldsymbol{x}$ is at least as good as $\boldsymbol{y}$ according to the ith objective.
  - $F_i(\mathbf{x}) < F_i(\mathbf{y})$ means $\boldsymbol{x}$ is better than $\boldsymbol{y}$ according to the ith objective.
- $\boldsymbol{x}$ dominates $\boldsymbol{y}$ if:
  - $F_i(\mathbf{x}) \leq F_i(\mathbf{y})$ for $i = 1, \ldots, q$.
  - For at least one j, $F_j(\mathbf{x}) < F_j(\mathbf{y})$.
- In words, $\boldsymbol{x}$ is better than $\boldsymbol{y}$ if $\boldsymbol{x}$ meets or beats $\boldsymbol{y}$ on all objectives and beats it in at least one objective.

Multi-Objective Optimization

Optimality in Multi-Objective Optimization

- For a point $\boldsymbol{x}^\star$ to be considered optimal, it must be simultaneously optimal for each scalar problem:
$$F_i(\boldsymbol{x}^\star) \leq F_i(\boldsymbol{y}), \quad i = 1, \ldots, q.$$
- Generally, this cannot happen unless the objectives are *noncompeting* (no trade-offs among objectives).
- In most practical problems, there is a trade-off among objectives, leading to no single optimal solution.

Pareto Optimality

- A point $\boldsymbol{x}^{\mathrm{po}}$ is Pareto optimal if no objective can be improved without degrading at least one other objective.
- The set of Pareto optimal values is called the *optimal trade-off surface* or, when $q = 2$, the *optimal trade-off curve*.
- Also referred to as the *efficient frontier* in other contexts.

Scalarization

Scalarization Technique

- Standard method for finding Pareto optimal points in a vector optimization problem.
- For a multi-objective optimization problem, the scalarized problem is:

$$\begin{array}{lll} \underset{\boldsymbol{x}}{\text{minimize}} & \sum_{i=1}^{q} \lambda_i F_i(\boldsymbol{x}) & \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, & 1 \leq i \leq m \\ & h_i(\boldsymbol{x}) = 0, & 1 \leq i \leq p \end{array}$$

- Weights λ_i are associated with the objectives and must satisfy $\lambda_i \geq 0$.

Properties of Scalarization

- An optimal point of the scalarized problem is Pareto optimal for the multi-objective optimization problem.
- Different weights yield different Pareto optimal solutions.
- Some Pareto optimal points may not be obtainable via scalarization.

Scalarization

Convex Multi-Objective Optimization:

- For convex problems, the scalarized problem is also convex.
- Yields all Pareto optimal points for different weights.
- For every Pareto optimal point, there are weights such that it is optimal in the scalarized problem.

Example: Regularized Least Squares (LS)

- Modified least squares problem with two objectives:
  - $F_1(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$: measure of regression error.
  - $F_2(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$: measure of the energy of $\boldsymbol{x}$.
- Multi-objective optimization problem formulation:

$$\underset{\boldsymbol{x}}{\text{minimize}} \text{ (with respect to } \mathbb{R}_+^2) \quad f_0(\boldsymbol{x}) = (F_1(\boldsymbol{x}), F_2(\boldsymbol{x}))$$

- Corresponding scalarization:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \gamma \|\boldsymbol{x}\|_2^2 \ ,$$

  where $\gamma \geq 0$ is the weight indicating the preference in the trade-off.

Outline

Summary

- **Optimization** has a long history, with theory developed over the past century and algorithms evolving from the 1947 simplex method to mid-1990s interior-point methods.
- **Optimization problems** are generally hard to solve with exponential time complexity, but convex problems have manageable polynomial time complexity, making convex optimization appealing.
- **Convex problems** consist of convex functions and sets, supported by rich theory and efficient algorithms, with numerous solvers available in most programming languages.
- **Lagrange duality** offers powerful theoretical results, including the **KKT optimality conditions** for characterizing optimal solutions.
- The standard problem formulation can be extended with **multi-objective** and **robust formulations**.

References I

Bertsekas, D. P. 1999. *Nonlinear Programming*. Athena Scientific.

Bertsekas, D. P., A. Nedić, and A. E. Ozdaglar. 2003. *Convex Analysis and Optimization*. Belmont, MA, USA: Athena Scientific.

Boyd, S. P., and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Boyd, S., S. J. Kim, L. Vandenberghe, and A. Hassibi. 2007. *A Tutorial on Geometric Programming*. Optimization and Engineering, Springer.

Chiang, M. 2005. "Geometric Programming for Communication Systems." *Foundations and Trends in Communications and Information Theory*.

Luenberger, D. G. 1969. *Optimization by Vector Space Methods*. New York: Wiley.

Nemirovski, A. 2001. "Lectures on Modern Convex Optimization." In *Society for Industrial and Applied Mathematics (SIAM)*.

References II

Nesterov, Y. 2018. *Lectures on Convex Optimization*. 2nd ed. Springer.

Palomar, D. P. 2025. *Portfolio Optimization: Theory and Application*. Cambridge University Press.

Palomar, D. P., and Y. C. Eldar. 2009. *Convex Optimization in Signal Processing and Communications*. Cambridge University Press.

Rockafellar, R. T. 1970. *Convex Analysis*. 2nd ed. Princeton, NJ: Princeton Univ. Press.

———. 1993. "Lagrange Multipliers and Optimality." *SIAM Review* 35 (2): 183–238.

Vandenberghe, L., and S. Boyd. 1996. "Semidefinite Programming." *SIAM Review* 38 (1): 49–95.