Mathematical optimization

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{s.t.} & \quad f_j(x) \leq 0, \quad j = 1, \ldots, m, \\
& \quad x \in S
\end{align*}
\]  

(1)

\(x\). \((x_1, \ldots, x_n)\). optimization variable.

\(f_0\). \(\mathbb{R}^n \rightarrow \mathbb{R}\). objective function.

\(f_j\). \(\mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \ldots, m\). constraint functions.\(^1\)

\(x^*\). optimal or solution of the problem \(^1\).

\(^1\). The constraint can be various and is not limited to \(\leq 0\).
$S$. basic feasible set.

$Q$. $\{x \in S | f_j(x) \leq 0, j = 1, ..., m \}$. feasible set.

In general optimization problems are unsolvable.
Classification of optimization problems

2. Unconstrained problem.
5. Linearly constrained problem.
   a. Linear optimization problem.
   b. Quadratic optimization problem.
7. Feasible. Strictly feasible......

They are just constrains on the type of $f_0$, $f_j$, $S$ and $x^*$!
Classification of solutions

1. Global solution.
2. Local solution.
Analyze a method $\mathcal{M}$

$\mathcal{M}$. numerical method

$\mathcal{P}$. a class of problems

$\Sigma$. Model. A known “part” of problem $\mathcal{P}$.

$\mathcal{O}$. Oracle. A unit answers the successive question of the method.

To solve the problem means to find an approximate solution to $\mathcal{M}$ with an accuracy $\varepsilon > 0$. Thus we need

$\mathcal{T}_\varepsilon$. a stopping criterion

Thus our problem class is

$$\mathcal{F} \equiv (\Sigma, \mathcal{O}, \mathcal{T}_\varepsilon).$$
Performance: The total amount of computational efforts required by $M$ to solve $P$.

1. *analytical complexity*

2. *arithmetical complexity*

---

**Optimal.** if upper complexity bounds of the method are proportional to the lower complexity bound of the problem class
Example: Complexity bounds for global optimization

Consider a constrained minimization problem without functional constraints,

\[
\min_{x \in B_n} f(x).
\]  
(2)

The basic feasible set of this problem is \(B_n\), which is an \(n\)-dimensional box in \(\mathbb{R}^n\):

\[
B_n = \{x \in \mathbb{R}^n | 0 \leq x(i) \leq 1, i = 1, \ldots, n\}.
\]  
(3)

The distance is measured using \(l_\infty\)-norm.
We further assume that the function is $L$-Lipschitzian.

We can construct an optimal method for this problem and show that it is not solvable by computers.
Main fields

- Goals of the methods
- Classes of functional components
  - General global optimization: Let’s wait for quantum computers!
  - Nonlinear optimization: We can find a local minimum under restrictions[1].
  - Convex optimization: We can find the global minimum under restrictions.
  - Interior-point polynomial-time methods: We can find the global minimum under restrictions for convex sets and functions with explicit structure.
• Description of the oracle

Convex optimization also benefits nonconvex optimization.
Convex sets

**Line.** Points of the form \( y = \theta x_1 + (1 - \theta) x_2 \), where \( \theta \in \mathbb{R} \), form the *line* passing through \( x_1 \) and \( x_2 \).

**Line segment.** \( y = x_2 + \theta (x_1 - x_2) \) where \( 0 \leq \theta \leq 1 \).

**Affine set.** A set \( C \subseteq \mathbb{R}^n \) is *affine* if the line through any two distinct points in \( C \) lies in \( C \).

**Affine combination.** \( \theta_1 x_1 + \ldots + \theta_k x_k \) where \( \theta_1 + \ldots + \theta_k = 1 \) is an affine combination of the points \( x_1, \ldots, x_k \).

**Theorem 1.** *If* \( C \) *is an affine set, \( x_1, \ldots, x_k \in C \), then the affine combination of \( x_1, \ldots, x_k \) also belongs to \( C \).*
Theorem 2. If $C$ is an affine set and $x_0 \in C$, then the set

$$V = C - x_0 = \{x - x_0 | x \in C\}$$

is a subspace.

**Dimension of an affine set.** Defined to be the dimension of the subspace $V = C - x_0$, where $x_0$ is any element of $C$.

**Affine dimension.** The affine dimension of a set $C$ is the dimension of its affine hull\(^2\).

**Affine hull.** The set of all affine combinations for points of some set $C$. Denoted by $\text{aff } C$.

---

\(^2\) Unit circle in 2-d space has affine dimension 2 but by most definitions of dimension it has dimension 1.
Relative interior. The *relative interior* of the set $C$ is the interior relative to $\text{aff } C$:

$$\text{relint } C = \{ x \in C | B(x, r) \cap \text{aff } C \subseteq C \text{ for some } r > 0 \}.$$ 

Relative boundary. The *relative boundary* of a set $C$ is $\text{cl } C \setminus \text{relint } C$, where $\text{cl } C$ is the clojure of $C$.

The affine hull of $C$ is the smallest affine set that contains $C$.

Convex set. A set $C$ is *convex* if the line segment between any two points in $C$ lies in $C$.

Convex combination. A *convex combination* of the points $x_1, \ldots, x_k$ is $\theta_1 x_1 + \cdots + \theta_k x_k$, where $\theta_1 + \cdots + \theta_k = 1$ and $\theta_i \geq 0$, $i = 1, \ldots, k$. 

14
Theorem 3. A set is convex iff it contains every convex combination of its points.

Convex hull. A convex hull of a set $C$, denoted $\text{conv } C$, is the set of all convex combinations of points in $C$. It is the smallest convex set that contains $C$.

Theorem 4. Suppose $C \subseteq \mathbb{R}^n$ is convex and $x$ is a random vector with $x \in C$ with probability one. Then $\mathbb{E}x \in C$.

Cone. A set $C$ is called a cone, or nonnegative homogeneous, if for every $x \in C$ and $\theta \geq 0$ we have $\theta x \in C$. 
**Convex cone.** A set $C$ is a *convex cone* if it is convex and a cone, which means for any $x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$, we have

$$\theta_1 x_1 + \theta_2 x_2 \in C.$$

**Conic combination.** or *nonnegative linear combination*. A point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$ with $\theta_1, \ldots, \theta_k \geq 0$.

**Theorem 5.** A set $C$ is a convex cone iff it contains all conic combinations of its elements.

**Conic hull.** A set of all conic combinations of points in $C$. This is the smallest convex cone that contains $C$. 

16
Examples

Hyperplane. A set of the form \( \{x | a^T x = b \} \), where \( a \in \mathbb{R}^n \), \( a \neq 0 \) and \( b \in \mathbb{R} \). (Normal vector is \( a \).)

Halfspaces. A hyperplane divides \( \mathbb{R}^n \) into two halfspaces. A (closed) halfspace is a set of the form

\[
\{ x | a^T x \leq b \},
\]

where \( a \neq 0 \). The open halfspace uses a strict inequality.

Euclidean ball. \( B(x_c, r) = \{ x | \| x - x_c \|_2 \leq r \} = \{ x_c + r u | \| u \|_2 \leq 1 \} \).
Ellipsoid. $\mathcal{E} = \{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\} = \{x_c + A u \mid \|u\|_2 \leq 1\}$, where $P = P^T \succ 0$ and $A$ is square and nonsingular.

Norm ball. $\{x \mid \|x - x_c\| \leq r\}$.

**Theorem 6.** *Norm balls are convex.*

Norm cone. $\{(x, t) \mid \|x\| \leq t\} \subseteq \mathbb{R}^{n+1}$.

3. For Euclidean norm:

$$\{(x, t) \in \mathbb{R}^{n+1} \mid \|x\|_2 \leq t\} = \left\{ \begin{pmatrix} x \\ t \end{pmatrix} \mid \begin{pmatrix} x \\ t \end{pmatrix}^T \begin{pmatrix} I & -I \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} \leq 0, t \geq 0 \right\}.$$
Figure 1. An example of norm cone.

This is also called Lorentz cone.
**Polyhedron.** The solution set of a finite number of linear equalities and inequalities and thus the intersection of a finite number of half spaces and hyperplanes.

\[
P = \{ x | a_j^T x \leq b_j, \ j = 1, \ldots, m, \ c_j^T x = d_j, \ j = 1, \ldots, p \} = \{ x | A x \preceq b, \ C x = d \}.
\]

**Theorem 7.** *The intersection of convex sets is a convex set.*

**Theorem 8.** *Polyhedra are convex sets.*

**Nonnegative orthant.** The set of points with nonnegative components, i.e.,

\[
\mathbb{R}_+^n = \{ x \in \mathbb{R}^n | x_i \geq 0, \ i = 1, \ldots, n \}.
\]
**Affinely independent.** $k + 1$ points $v_0, \ldots, v_k$ are affinely independent if $v_1 - v_0, \ldots, v_k - v_0$ are linearly independent.

**Simplexes.** The simplex determined by $k + 1$ affinely independent points $v_0, \ldots, v_k \in \mathbb{R}^n$ is

$$C = \text{conv}\{v_0, \ldots, v_k\} = \{\theta_0 v_0 + \cdots + \theta_k v_k | \theta \succeq 0, 1^T \theta = 1\}.$$

The affine dimension of this simplex is $k$.

**Probability simplex.** $x \succeq 0, 1^T x = 1$.

**Unit simplex.** $x \succeq 0, 1^T x \leq 1$.

The above simplex can be described using polyhedron.
**Theorem 9.** A generalization of convex hull is

\[ \{ \theta_1 v_1 + \cdots + \theta_k v_k | \theta_1 + \cdots + \theta_m = 1, \theta_i \geq 0, i = 1, \ldots, k \} , \]

where \( m \leq k^4 \).

This defines a polyhedron, and conversely, every polyhedron can be represented in this form.

**Positive semidefinite cone.** \( S^+_n = \{ X \in S^n | X \succeq 0 \} \), where \( S^n \) denotes the set of symmetric \( n \times n \) matrices.

---

4. We can interpret it as the convex hull of the points \( v_1, \ldots, v_m \), plus the conic hull of the points \( v_{m+1}, \ldots, v_k \).
Operations that preserve convexity

Theorem 10. Convexity is preserved under intersection.

Theorem 11. Every closed convex set $S$ is a (usually infinite) intersection of halfspaces. In fact, a closed convex set $S$ is the intersection of all halfspaces that contain it:

$$S = \bigcap \{\mathcal{H} | \mathcal{H} \text{ halfspace, } S \subseteq \mathcal{H} \}.$$

Affine function. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is affine if it is a sum of a linear function and a constant, i.e., $f(x) = A x + b$.

Theorem 12. Convexity is preserved under affine function, i.e., if $S$ is convex and $f$ is an affine function, then $f(S)$ is convex.
Note that the inverse function of an affine function is an affine function. Examples of affine function include *translation* and *projection*.

**Corollary 13.** The partial sum of two convex sets \( S_1, S_2 \in \mathbb{R}^n \times \mathbb{R}^m \) defined as

\[
S = \{(x, y_1 + y_2) | (x, y_1) \in S_1, (x, y_2) \in S_2\},
\]

where \( x \in \mathbb{R}^n \) and \( y_i \in \mathbb{R}^m \) is convex.

**Perspective function.** \( P: \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n \), with domain \( \text{dom } P = \mathbb{R}^n \times \mathbb{R}_{++} \), as \( P(z, t) = z/t \). The *perspective function* scales or normalizes vectors so the last component is one, and then drops the last component.
Theorem 14. If $C \subseteq \text{dom } P$ is convex, then its image

$$P(C) = \{P(x) | x \in C\}$$

is convex, where $P$ is perspective function.

Proof. For any two points $(x, z_1), (y, z_2)$ in $C$ we have

$$\frac{\theta x + (1 - \theta) y}{\theta z_1 + (1 - \theta) z_2} \in P(C).$$

From this we know the perspective function maps line segments to line segments and thus it preserves the convexity of sets. □
Theorem 15. The inverse image of a convex set under the perspective function is also convex. If $C \subseteq \mathbb{R}^n$ is convex, then

$$P^{-1}(C) = \{(x, t) \in \mathbb{R}^{n+1} | x/t \in C, t > 0\}$$

is convex.

Linear-fractional function. Compose the perspective function with an affine function. Suppose $g: \mathbb{R}^n \to \mathbb{R}^{m+1}$ is affine:

$$g(x) = \left( \begin{array}{c} A \\ c^T \end{array} \right) x + \left( \begin{array}{c} b \\ d \end{array} \right),$$

then

$$f = P \circ g = (A x + b) / (c^T x + d), \quad \text{dom } f = \{x | c^T x + d > 0\}$$

is a linear-fractional function. Affine function is a special case of linear-fractional function.

Example 17. (Conditional probabilities) Let $p_{i,j}$ denote $\text{prob}(u = 1, v = j)$. Then the conditional probability $f_{*,j} = \text{prob}(u = \ast | v = j) = \frac{p_{*,j}}{\sum_{k=1}^{n} p_{k,j}}$ is obtained by a linear-fractional mapping from $p_{*,j}$. It follows that if $C$ is a convex set of joint probabilities for $(u, v)$, then the associated set of conditional probabilities of $u$ given $v$ is also convex.
Generalized inequalities

Proper cone. A cone \( \subseteq \mathbb{R}^n \) that

1. convex
2. closed
3. solid, which means it has nonempty interior
4. pointed, which means that it is oriented, \( x \in K, -x \in K \implies x = 0 \).

Proper cone can be used to define a partial ordering on \( \mathbb{R}^n \).

Generalized inequality. \( x \preceq_K y \iff y - x \in K \), \( x \prec_K y \iff y - x \in \text{int } K \).
**Linear ordering.** Any two points are *comparable*. $\leq$ on $\mathbb{R}$ is *linear ordering* but generalized inequality generally does not have this property\(^5\).

**Example 18.** When $K = \mathbb{R}_+$, $\preceq_K$ is the usual ordering $\leq$ on $\mathbb{R}$.

**Properties of generalized inequalities**

1. Preserved under addition.
2. Transitive.
3. Preserved under nonnegative scaling.
4. Reflexive.
5. Antisymmetric.

---

\(^5\) This makes concepts like minimum and maximum more complicated.
6. Preserved under limits.

**Minimum and minimal elements**

**Minimum.** \(x \in S\) is the *minimum* element of \(S\) with respect to the generalized inequality \(\preceq_K\) if for every \(y \in S\) we have \(x \preceq_K y\).

**Minimal.** \(x \in S\) is a *minimal* element of \(S\) with respect to the generalized inequality \(\preceq_K\) if \(y \in S, y \preceq_K x\) only if \(y = x\).

**Maximum/Maximal.** Defined in a similar way.

**Theorem 19.** If a set has a minimum element, then it is unique. A set can have many different minimal elements.

**Theorem 20.** A point \(x \in S\) is the minimum element of \(S\) iff \(S \subseteq x + K\).
Theorem 21. A point \( x \in S \) is a minimal element iff \( (x - K) \cap S = \{x\} \).

Figure 2. Left: minimum. Right: minimal.
Example 22. (Minimum and minimal elements of a set of symmetric matrices) We associate with each \( A \in S_{++}^n \) an ellipsoid centered at the origin, given by \( \mathcal{E}_A = \{ x | x^T A^{-1} x \leq 1 \} \). We have \( A \preceq B \) iff \( \mathcal{E}_A \subseteq \mathcal{E}_B \).

Let \( v_1, \ldots, v_k \in \mathbb{R}^n \) be given and define

\[
S = \{ P \in S_{++}^n | v_i^T P^{-1} v_i \leq 1, \ i = 1, \ldots, k \},
\]

which corresponds to the set of ellipsoids that contain the points \( v_1, \ldots, v_k \). The set does not have a minimum element but have minimal elements.
Figure 3. $\mathcal{E}_2$ is a minimal element.
Separating and supporting hyperplanes

Theorem 23. (Separating hyperplane theorem) Suppose $C$ and $D$ are two convex sets that do not intersect, i.e., $C \cap D = \emptyset$. Then there exists $a \neq 0$ and $b$ such that $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$. The hyperplane $\{x | a^T x = b\}$ is called a separating hyperplane for the sets $C$ and $D$.

Proof. Construct a plane orthogonal to the line formed by two points achieving the $\text{dist}(C, D)$.

Strict separation. If the inequalities in separation become strict, it is called a strict separation. Strict separation is not always
possible even for closed convex sets.

**Theorem 24.** Any two convex sets $C$ and $D$, at least one of which open, are disjoint iff there exists a separating hyperplane.

**Supporting hyperplane.** Suppose $C \subseteq \mathbb{R}^n$, and $x_0$ is a point in its boundary $\text{bd } C = \text{cl } C \setminus \text{int } C$. If $a \neq 0$ satisfies $a^T x \leq a^T x_0$ for all $x \in C$, then the hyperplane $\{x | a^T x = a^T x_0\}$ is called a supporting hyperplane to $C$ at the point $x_0$.

---

6. This is equivalent to saying that the point $x_0$ and the set $C$ are separated by the hyperplane.
Theorem 25. (Supporting hyperplane theorem) For any non-empty convex set $C$, and any $x_0 \in \text{bd } C$, there exists a supporting hyperplane to $C$ at $x_0$. 

Figure 4. The supporting hyperplane.
Theorem 26. If a set is closed, has nonempty interior, and has a supporting hyperplane at every point in its boundary, then it is convex.
Dual cones and generalized inequalities

**Dual cone.** Let $K$ be a cone. The set $K^* = \{ y | x^T y \geq 0 \text{ for all } x \in K \}$ is called the *dual cone* of $K$.

*Figure 5.* $y$ is in $K^*$ and $z$ is not. Geometrically, $y \in K^*$ iff $y$ is the normal of a hyperplane that supports $K$ at the origin.
Theorem 27. $K^*$ is a cone, and is always convex, even when the original cone is not.

Example 28. (Positive semidefinite cone) The positive semidefinite cone $S^+_n$ is self-dual under the standard inner product
\[ \text{tr}(XY) = \sum_{i,j=1}^{n} X_{ij} Y_{ji} = \sum_{i,j=1}^{n} X_{ij} Y_{ij}. \]

Dual norm. $\|u\|_* = \sup \{ u^T x \|x\| \leq 1 \}$.

Example 29. (Dual of a norm cone) The dual of the norm cone $K = \{(x, t) \in \mathbb{R}^{n+1} ||x|| \leq t \}$ is the cone defined by the dual norm,
i.e.,

\[ K^* = \{(u, v) \in \mathbb{R}^{n+1} \mid \|u\|_* \leq v\} \].

The properties of dual cone

1. \( K^* \) is closed and convex.
2. \( K_1 \subseteq K_2 \implies K_2^* \subseteq K_1^* \).
3. If \( K \) has nonempty interior, then \( K^* \) is pointed.
4. If the closure of \( K \) is pointed then \( K^* \) has nonempty interior.
5. \( K^{**} \) is the closure of the convex hull of \( K \). (Hence if \( K \) is convex and closed, \( K^{**} = K \).)
Theorem 30. If $K$ is a proper cone, then so is its dual $K^*$. 

Dual generalized inequalities

**Dual generalized inequality.** $\preceq_{K^*}$. Note that if the convex cone $K$ is proper, then $K^*$ is proper. Since for a proper cone $K = K^{**}$, the dual generalized inequality associated with $\preceq_{K^*}$ is $\preceq_K$. 

Theorem 31. 

1. $x \preceq_K y$ iff $\lambda^T x \leq \lambda^T y$ for all $\lambda \succeq_{K^*} 0$. 

2. $x \prec_K y$ iff $\lambda^T x < \lambda^T y$ for all $\lambda \succeq_{K^*} 0, \lambda \neq 0$. 

The geometric interpretation is easy to get.
Example 32. (Theorem of alternatives for linear strict generalized inequalities) Suppose \( K \subseteq \mathbb{R}^m \) is a proper cone. Consider the strict generalized inequality

\[
A x \prec_K b,
\]

where \( x \in \mathbb{R}^n \).

An alternative is there exists a \( \lambda \) such that

\[
\lambda^T A = 0, \quad \lambda^T b \leq 0, \quad \lambda \succeq_K 0, \quad \lambda \neq 0.
\]

Theorem 33. \( x \) is the minimum element of \( S \), with respect to the generalized inequality \( \preceq_K \), iff for all \( \lambda \succeq_K 0 \), \( x \) is the unique minimizer of \( \lambda^T z \) over \( z \in S \).
Figure 6. Dual characterization of minimum element. The point \( x \) is the minimum element of the set \( S \) with respect to \( \mathbb{R}^2_+ \). This is equivalent to: for every \( \lambda \succ 0 \), the hyperplane \( \{ z \mid \lambda^T (z - x) = 0 \} \) strictly supports \( S \) at \( x \), i.e., contains \( S \) on one side, and touches it only at \( x \).
Theorem 34. If $\lambda \succ K^* 0$ and $x$ minimizes $\lambda^T z$ over $z \in S$, then $x$ is minimal. Note that the converse is generally false.

Theorem 35. Provided that the set $S$ is convex, we can say that for any minimal element $x$ there exists a nonzero $\lambda \succ K^* 0$ such that $x$ minimizes $\lambda^T z$ over $z \in S$. 
Figure 8. Why the converse is not true and why $\succeq$ instead of $\succ$. $K = \mathbb{R}_+^2$ for example. Left. The point $x_1 \in S_1$ is minimal, but is not a minimizer of $\lambda^T z$ over $S_1$ for any $\lambda \succ 0$. Right. The point $x_2 \in S_2$ is not minimal, but it does minimize $\lambda^T z$ over $z \in S_2$ for $\lambda = (0, 1) \succeq 0$. 

Bibliography