1 Convex Optimization

Some Basics and Terminology

- A standard form for general optimization problems:
  \[
  \begin{align*}
  \min & \quad f_0(x) \\
  \text{s.t.} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
  & \quad h_i(x) = 0, \quad i = 1, \ldots, p
  \end{align*}
  \]
  where \(f_0\) is the objective function, \(f_i, i = 1, \ldots, m,\) are the inequality constraint functions, and \(h_i(x) i = 1, \ldots, p,\) are the equality constraint functions.

- The set
  \[
  D = \bigcap_{i=0}^{m} \text{dom} f_i \cap \bigcap_{i=1}^{p} \text{dom} h_i
  \]
  is called the problem domain.

- The set
  \[
  C = \{ x \mid f_i(x) \leq 0, \quad i = 1, \ldots, m, \quad h_i(x) = 0, \quad i = 1, \ldots, p \}
  \]
  is called the feasible set, or the constraint set. A point \(x\) is feasible if \(x \in C\), and infeasible otherwise.

- The inequality constraint \(f_i(x)\) is active at \(x \in C\) if \(f_i(x) = 0\).

- A point \(x\) is strictly feasible if
  \[
  f_i(x) < 0, \quad i = 1, \ldots, m, \quad h_i(x) = 0, \quad i = 1, \ldots, p
  \]
  i.e., all inequality constraints are inactive.

- A problem is feasible if there exists \(x \in C\); infeasible if \(C = \emptyset\); and strictly feasible if there is a strictly feasible point.

- A problem is unconstrained if \(C = \mathbb{R}^n\).

- The quantity
  \[
  p^* = \inf \{ f_0(x) \mid x \in C \}
  \]
  is called the optimal value (of the associated problem). We use \(p^* = +\infty\) to represent the case where the problem is infeasible. If \(p^* = -\infty\) (which does not have a solution), the problem is said to be unbounded below.

- A point \(x^*\) is globally optimal or simply optimal if \(x^* \in C \& f_0(x^*) = p^*\). A point \(x\) is locally optimal if there is an \(R > 0\) such that
  \[
  f_0(x) = \inf \{ f_0(\tilde{x}) \mid \tilde{x} \in C, \quad \|\tilde{x} - x\|_2 \leq R \}
  \]

- A maximization problem can be reformulated as a minimization problem since
  \[
  \max_{x \in C} f_0(x) = - \min_{x \in C} -f_0(x)
  \]
• A feasibility problem

\[
\begin{align*}
\text{find } & \quad x \\
\text{s.t. } & \quad x \in C
\end{align*}
\]

is an optimization problem where \( f_0(x) = 0 \). In this case \( p^* = 0 \) if \( C \neq \emptyset \), and \( p^* = +\infty \) if \( C = \emptyset \).

**Convex Optimization Problems**

• A standard problem

\[
\begin{align*}
\min & \quad f_0(x) \\
\text{s.t. } & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

is convex if \( f_0, \ldots, f_m \) are convex and \( h_1, \ldots, h_p \) are affine. Thus, a convex optimization problem may be expressed in a more convenient form:

\[
\begin{align*}
\min & \quad f_0(x) \\
\text{s.t. } & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

where \( A \in \mathbb{R}^{p \times n}, \ b \in \mathbb{R}^p \).

• For convex problems, any locally optimal point is globally optimal.

• An optimality criterion for convex problems: Suppose that \( f_0 \) is differentiable. A point \( x \in C \) is optimal if and only if

\[
\nabla f_0(x)^T(y - x) \geq 0, \quad \forall y \in C
\]

• Convex problems with generalized inequalities:

\[
\begin{align*}
\min & \quad f_0(x) \\
\text{s.t. } & \quad f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, L \\
& \quad Ax = b
\end{align*}
\]

where \( K_i \subset \mathbb{R}^{m_i} \) are proper cones, \( \preceq_{K_i} \) are generalized inequalities with respect to \( K_i \), and \( f_i : \mathbb{R}^n \to \mathbb{R}^{m_i} \) are \( K_i \)-convex.

• Quasi-convex optimization problems:

\[
\begin{align*}
\min & \quad f_0(x) \\
\text{s.t. } & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

where \( f_0(x) \) is quasiconvex, and \( f_i(x) \) convex for \( i \geq 1 \). A quasi-convex problem may be solved by a bisection method, in which the solution is obtained by solving a sequence of problems that are convex.
Important Classes of Convex Optimization Problems

- Linear program (LP): The general form of LP is
  \[
  \begin{align*}
  \min c^T x \\
  \text{s.t. } Gx &\preceq h \\
  Ax &\preceq b
  \end{align*}
  \]

  Other popular forms include the standard form
  \[
  \begin{align*}
  \min c^T x \\
  \text{s.t. } x &\succeq 0, \ Ax = b
  \end{align*}
  \]

  and the inequality form
  \[
  \begin{align*}
  \min c^T x \\
  \text{s.t. } Gx - h &\preceq 0
  \end{align*}
  \]

- Quadratic program (QP):
  \[
  \begin{align*}
  &\min \frac{1}{2} x^T P x + q^T x + r \\
  &\text{s.t. } Ax = b, \ Gx \preceq h
  \end{align*}
  \]
  where \( P \succeq 0 \).

- Quadratically constrained QP (QCQP):
  \[
  \begin{align*}
  &\min \frac{1}{2} x^T P_0 x + q_0^T x + r_0 \\
  &\text{s.t. } \frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0, \ i = 1, \ldots, m \\
  &Ax = b
  \end{align*}
  \]
  where \( P_i \succeq 0 \) for \( i = 0, 1, \ldots, m \).

- Geometric program (GP):
  \[
  \begin{align*}
  &\min \sum_{k=1}^{K_0} c_{0,k} x_1^{a_{0,1,k}} x_2^{a_{0,2,k}} \cdots x_n^{a_{0,n,k}} \\
  &\text{s.t. } \sum_{k=1}^{K_i} c_{i,k} x_1^{a_{i,1,k}} x_2^{a_{i,2,k}} \cdots x_n^{a_{i,n,k}} \leq 1, \ i = 1, \ldots, m \\
  &d_1 x_1^{g_{1,1}} \cdots x_n^{g_{1,n}} = 1, \ i = 1, \ldots, p
  \end{align*}
  \]
  where \( D = \mathbb{R}^n_{++} \). GP is nonconvex, but can be reformulated as a convex problem in the form of
  \[
  \begin{align*}
  &\min \log \sum_{k=1}^{K_0} e^{c_{0,k} y + b_{0,k}} \\
  &\text{s.t. } \log \sum_{k=1}^{K_i} e^{c_{i,k} y + b_{i,k}} \leq 0, \ i = 1, \ldots, m \\
  &g_i^T y + h_i = 0, \ i = 1, \ldots, p
  \end{align*}
  \]
  where \( y_i = \log x_i, \ b_{i,k} = \log c_{i,k} \), and \( h_i = \log d_i \).

- Second order cone program (SOCP):
  \[
  \begin{align*}
  &\min c^T x \\
  &\text{s.t. } \| A_i x + b_i \|_2 \leq f_i^T x + d_i, \ i = 1, \ldots, L \\
  &Gx = h
  \end{align*}
  \]
• Semidefinite program (SDP): Standard form:

\[
\begin{align*}
\min & \; \text{tr}(CX) \\
\text{s.t.} & \; X \succeq 0 \\
& \; \text{tr}(A_i X) = b_i, \quad i = 1, \ldots, m
\end{align*}
\]

where \( A_i \in \mathbb{S}^{n \times n} \), and \( C \in \mathbb{S}^{n \times n} \). Inequality form:

\[
\begin{align*}
\min & \; c^T x \\
\text{s.t.} & \; F(x) \preceq 0
\end{align*}
\]

where \( F(x) = F_0 + x_1 F_1 + \ldots + x_m F_m \), \( F_i \in \mathbb{S}^{p \times p} \).

Some Reformulation Tricks

• Epigraph reformulation: The standard form problem can be reformulated as

\[
\begin{align*}
\min & \; t \\
\text{s.t.} & \; f_0(x) \leq t \\
& \; f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \; h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

Moreover, if the objective function is a sum of some functions; say \( f_0(x) = \sum_{i=1}^k g_i(x) \), the problem can be reformulated as

\[
\begin{align*}
\min & \; \sum_{i=1}^k t_i \\
\text{s.t.} & \; g_i(x) \leq t_i, \quad i = 1, \ldots, k \\
& \; f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \; h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

Some examples:

− Piecewise linear function minimization

\[
\begin{align*}
\min_{x \in \mathcal{C}} \max_{i=1,\ldots,m} \{a_i^T x + b_i\} \iff \min_{x \in \mathcal{C}, t \in \mathbb{R}} \quad & \quad t \\
\text{s.t.} & \; \max_{i=1,\ldots,m} \{a_i^T x + b_i\} \leq t,
\end{align*}
\]

\[
\begin{align*}
\iff \min_{x \in \mathcal{C}, t \in \mathbb{R}} \quad & \quad t \\
\text{s.t.} & \; a_i^T x + b_i \leq t, \; i = 1, \ldots, m
\end{align*}
\]

− \( \ell_\infty \)-norm (Chebychev) approximation

\[
\begin{align*}
\min \; \|Ax - b\|_\infty & \iff \min_{x \in \mathbb{R}^n, r \in \mathbb{R}^m} \quad & \quad t \\
\text{s.t.} & \; \max_{i=1,\ldots,m} |r_i| \leq t \\
& \; r = Ax - b \\
\iff \min_{x \in \mathbb{R}^n, r \in \mathbb{R}^m} \quad & \quad t \\
\text{s.t.} & \; -t1 \preceq r \preceq t1 \\
& \; r = Ax - b
\end{align*}
\]
\(-\ell_1\)-norm approximation

\[
\min \|Ax - b\|_1 \iff \min_{x \in \mathbb{R}^n, r \in \mathbb{R}^m} \sum_{i=1}^{m} |r_i|
\]

s.t. \(r = Ax - b\)

\[
\iff \min_{x \in \mathbb{R}^n, r \in \mathbb{R}^m} \sum_{i=1}^{m} t_i
\]

s.t. \(-t_i \leq r_i \leq t_i, \quad i = 1, \ldots, m\)

\(r = Ax - b\)

- Schur complements: Let \(X \in \mathbb{S}^n\) and partition

\[
X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}.
\]

The matrix \(S = C - B^T A^{-1} B\) is called the Schur complement of \(A\) in \(X\) (provided \(A \succ 0\)).

Important facts:

- \(X \succeq 0\) if and only if \(A \succ 0\) and \(S \succeq 0\).
- Suppose \(A \succ 0\). Then, \(X \succeq 0\) if and only if \(S \succeq 0\).
- \(X \succeq 0\) if and only if \(A \succeq 0\), \((I - AA^T)B = 0\), and \(S \succeq 0\).
- Example: The convex quadratic inequality

\[
(Ax + b)^T(Ax + b) - c^T x - d \leq 0
\]

is equivalent to

\[
\begin{bmatrix} I & Ax + b \\ (Ax + b)^T & c^T x + d \end{bmatrix} \succeq 0.
\]

2 Duality

Lagrange Dual Function and Problem

- The Lagrange associated with a standard (not necessarily convex) problem

\[
\min f_0(x)
\]

s.t. \(f_i(x) \leq 0, \quad i = 1, \ldots, m\)

\(h_i(x) = 0, \quad i = 1, \ldots, p\)

is

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

The variables \(\lambda\) and \(\nu\) are called the dual variables or Lagrange multiplier.

- Lagrange dual function:

\[
g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu)
\]

The dual function \(g\) is concave in \((\lambda, \nu)\), even when the original or primal problem is nonconvex.

- A pair \((\lambda, \nu)\) is said to be dual feasible if \(\lambda \succeq 0\) and \((\lambda, \nu) \in \text{dom } g\) [i.e., note that \((\lambda, \nu) \in \text{dom } g\)

implies \(g(\lambda, \nu) > -\infty\)].
• Lower bound on the optimal value: For any $\lambda \succeq 0$ and $\nu$,

$$g(\lambda, \nu) \leq p^\star.$$ 

• Dual problem:

$$\begin{align*}
\max & \quad g(\lambda, \nu) \\
\text{s.t.} & \quad \lambda \succeq 0
\end{align*}$$

The dual problem is convex, even when the primal is not. The dual optimal value

$$d^\star = \sup\{g(\lambda, \nu) \mid \lambda \succeq 0, \nu \in \mathbb{R}^p\}$$

serves as the best lower bound for the primal optimal value $p^\star$, in a Lagrangian sense.

• The result

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

is called weak duality. If the problem is such that

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

then we say that strong duality holds. The quantity

$$p^\star - d^\star$$

is called the duality gap.

**Strong Duality and the Implications**

• Convex problems: Strong duality usually (but not always) holds for convex problems. A sufficient condition that covers many instances of strong duality is the Slater’s condition (or Slater’s constraint qualification): If there exist $\tilde{x} \in \text{relint} \ D$ with $f_i(\tilde{x}) < 0$ for $i = 1, \ldots, m$ and $h_i(\tilde{x}) = 0$ for $i = 1, \ldots, p$, then strong duality holds.

• Strong duality generally does not hold for nonconvex problems, except for some special cases such as the following:

  – The smallest eigenvalue problem:

    $$\begin{align*}
    \min & \quad x^T C x \\
    \text{s.t.} & \quad x^T x = 1
    \end{align*}$$

    for any $C \in \mathbb{S}^n$. Its dual problem is

    $$\begin{align*}
    \max & \quad -\nu \\
    \text{s.t.} & \quad C + \nu I \succeq 0
    \end{align*}$$

    Strong duality holds because the two problems achieve an optimal value that is $\lambda_{\min}(C)$.

  – Indefinite QP with one indefinite quadratic constraints:

    $$\begin{align*}
    \min & \quad x^T A_0 x + 2b_0^T x + c_0 \\
    \text{s.t.} & \quad x^T A_1 x + 2b_1^T x + c_1 \leq 0
    \end{align*}$$
where $A_0, A_1 \in \mathbf{S}^n$ are not necessarily PSD. Its dual problem is

$$\max \gamma$$

$$\text{s.t. } \begin{bmatrix} A_0 & b_0 \\ b_0^T & c_0 - \gamma \end{bmatrix} + \lambda \begin{bmatrix} A_1 & b_1 \end{bmatrix} \geq 0$$

$$\lambda \geq 0$$

In this problem, strong duality holds when there exist an $\hat{x}$ such that $\hat{x}^T A_1 \hat{x} + 2b_1^T \hat{x} + c_1 < 0$. (the proof is based on the $\mathbf{S}$-procedure)

- Max-min characterization: Suppose that there is no equality constraint. We have that

$$\sup_{\lambda \geq 0} \inf_x L(x, \lambda) \leq \inf_x \sup_{\lambda \geq 0} L(x, \lambda).$$

When strong duality holds,

$$\sup_{\lambda \geq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geq 0} L(x, \lambda).$$

- Complementary slackness: Let $x^*$ and $(\lambda^*, \nu^*)$ be the primal and dual optimal points, respectively. When strong duality holds, the following condition must be satisfied:

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m$$

This condition implies that

$$\lambda_i^* > 0 \implies f_i(x^*) = 0 \quad f_i(x^*) < 0 \implies \lambda_i^* = 0$$

- Solving the primal problem using the dual: Suppose that strong duality holds, and that an optimal $(\lambda^*, \nu^*)$ is known. Also, suppose that the solution of

$$\min_x L(x, \lambda^*, \nu^*)$$

is unique. If the solution of the above problem is primal feasible, then it is primal optimal; otherwise no primal optimal solution exists.

- Karush-Kuhn-Tucker (KKT) optimality conditions: Suppose that $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ are differentiable. The following conditions are known as the KKT conditions:

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

$$f_i(x^*) \leq 0, \quad i = 1, \ldots, m$$

$$h_i(x^*) = 0, \quad i = 1, \ldots, p$$

$$\lambda_i^* \geq 0, \quad i = 1, \ldots, m$$

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m$$

- For problems with strong duality, the KKT conditions are necessary optimality conditions; i.e., if $x^*$ and $(\lambda^*, \nu^*)$ are primal and dual optimal, then the KKT conditions hold.

- For problems without strong duality, the KKT conditions are necessary for local optimality under a mild assumption: If $x^*$ is locally optimal and $x^*$ is regular (see nonlinear programming textbooks for the definition of regularity), then there exists $(\lambda^*, \nu^*)$ such that the KKT conditions hold.

- For convex problems with strong duality (e.g., when the Slater’s condition is satisfied), the KKT conditions are the sufficient and necessary optimality conditions; i.e., $x^*$ and $(\lambda^*, \nu^*)$ are primal and dual optimal if and only if the KKT conditions hold.
Case of Generalized Inequality

- Consider a primal problem based on generalized inequalities

\[
\begin{align*}
\min & \quad f_0(x) \\
\text{s.t.} & \quad f_i(x) \preceq K_i, \quad 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

where \( K_i \) are the proper cones. The duality results described above essentially apply in this case.

- The Lagrangian:

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

where \( \lambda = (\lambda_1, \ldots, \lambda_m) \).

- Dual function:

\[
g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)
\]

The dual function is concave in \((\lambda, \nu)\).

- Dual problem:

\[
\begin{align*}
\max & \quad g(\lambda, \nu) \\
\text{s.t.} & \quad \lambda_i \succeq K_i^*, \quad 0, \quad i = 1, \ldots, m
\end{align*}
\]

- Slater’s condition. Suppose that the (primal) problem is convex. Then, strong duality holds if there exist \( x \in \text{relint} \ D \) with \( f_i(x) \prec K_i^* 0 \) for \( i = 1, \ldots, m \) and \( h_i(x) = 0 \) for \( i = 1, \ldots, p \).

- Complementary slackness: Suppose strong duality holds. Then,

\[
(\lambda_i^*)^T f_i(x^*) = 0, \quad i = 1, \ldots, m
\]

which we can conclude that

\[
\begin{align*}
\lambda_i^* & \succeq K_i^* 0 \implies f_i(x^*) = 0 \\
f_i(x^*) & \prec K_i 0 \implies \lambda_i^* = 0
\end{align*}
\]

- KKT optimality conditions:

\[
\nabla f_0(x^*) + \sum_{i=1}^{m} Df_i(x^*)\lambda_i^* + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0
\]

\[
\begin{align*}
& f_i(x^*) \prec K_i 0, \quad i = 1, \ldots, m \\
& h_i(x^*) = 0, \quad i = 1, \ldots, p \\
& \lambda_i^* \succeq K_i^* 0, \quad i = 1, \ldots, m
\end{align*}
\]

\[
(\lambda_i^*)^T f_i(x^*) = 0, \quad i = 1, \ldots, m
\]

Here, \( Df_i(x) \in \mathbb{R}^{m_i \times n} \) is the derivative of \( f_i \) at \( x \).
Examples

- Standard form LP:

\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad x \succeq 0, \ Ax = b
\end{align*}
\]

Its dual problem is

\[
\begin{align*}
\max & \quad -b^T \nu \\
\text{s.t.} & \quad A^T \nu + c \succeq 0
\end{align*}
\]

Note that the dual problem is an inequality form LP.

- QCQP:

\[
\begin{align*}
\min & \quad \frac{1}{2} x^T P_0 x + q_0^T x + r_0 \\
\text{s.t.} & \quad \frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

Its dual problem is an SDP

\[
\begin{align*}
\max & \quad t \\
\text{s.t.} & \quad \begin{bmatrix} 2P(\lambda) & q(\lambda) \\ q(\lambda)^T & r(\lambda) - t \end{bmatrix} \succeq 0 \\
\lambda & \succeq 0
\end{align*}
\]

where \( P(\lambda) = P_0 + \sum_{i=1}^{m} \lambda_i P_i \), \( q(\lambda) = q_0 + \sum_{i=1}^{m} \lambda_i q_i \), and \( r(\lambda) = r_0 + \sum_{i=1}^{m} \lambda_i r_i \).

- Entropy maximization:

\[
\begin{align*}
\min & \quad \sum_{i=1}^{n} x_i \log x_i \\
\text{s.t.} & \quad 1^T x = 1
\end{align*}
\]

where the problem domain is \( \mathcal{D} = \mathbb{R}^n_{++} \). The dual is

\[
\max \quad -\nu - ne^{-\nu - 1}
\]

- Standard form SDP:

\[
\begin{align*}
\min & \quad \text{tr}(CX) \\
\text{s.t.} & \quad \text{tr}(A_i X) = b_i, \quad i = 1, \ldots, p \\
& \quad X \succeq 0
\end{align*}
\]

The dual problem is

\[
\begin{align*}
\max & \quad -b^T \nu \\
\text{s.t.} & \quad C + \sum_{i=1}^{p} \nu_i A_i \succeq 0
\end{align*}
\]

The dual of the standard form SDP is reminiscent of the inequality form SDP.