Topic 4: Duality

This topic introduces duality theory as a tool to analyze an optimization problem. We first discuss this theory for a general problem (without assuming convexity), then apply it to a convex optimization problem.

1. The Lagrange dual function:

Consider a general (not necessarily convex) problem,

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

called the "primal" problem.

where the domain \( D \) is nonempty

\[
D = \bigcap_{i=0}^{m} \text{dom } f_i \cap \bigcap_{i=1}^{p} \text{dom } g_i
\]

Denote the optimal value as \( p^* \).

2) Lagrange duality takes into account the constraints by augmenting the objective function with a weighted sum of the constraint functions: Define the Lagrangian as:

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

\[
\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \quad \text{dom} \mathcal{L} = D \times \mathbb{R}^m \times \mathbb{R}^p.
\]

\( \lambda_i \): Lagrange multiplier associated with \( i \)th inequality constraint

\( \nu_i \): " \( i \)th equality constraint

Vectors \( \lambda \) and \( \nu \) (\( \lambda \in \mathbb{R}^m \), \( \nu \in \mathbb{R}^p \)) are called dual variables (or Lagrange multiplier vectors).

3) Lagrange dual function:

This is a function of only the dual variables \((\lambda, \nu)\) obtained by finding the minimum of the Lagrangian over \( x \).
Lagrangian $L(x, \lambda, \nu)$: function of all primal & dual variables

$x$: primal variables

$\lambda, \nu$: dual variables

Dual function:

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

$$= \inf_{x \in D} \left( f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

Since $g(\lambda, \nu)$ is the infimum of a family of affine functions of $(\lambda, \nu)$, it is always concave even if the original problem is non-convex.

1) Lower bound on the optimal value:

$$\forall \lambda \geq 0, \forall \nu = g(\lambda, \nu) \leq p^*$$

This property can be easily verified:

Suppose $x$ is feasible $\rightarrow f_i(x) \leq 0, i = 1, \ldots, m$

$\sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \leq 0 \forall \lambda_i \geq 0.$

Then $L(x, \lambda, \nu) \leq f_0(x) \forall x$ feasible.

Thus $g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu) \leq f_0(x) \forall x$ feasible.

$$\rightarrow g(\lambda, \nu) \leq p^*.$$
Example: (i) Inequality form LP.

\[ \begin{align*}
\min \quad & c^T x \\
\text{st.} \quad & a_i^T x - b_i \leq 0, \quad i = 1, \ldots, m
\end{align*} \]

\[ L(x, \lambda) = c^T x + \sum_{i=1}^{m} \lambda_i (a_i^T x - b_i) \]

= \text{-}b^T \lambda + (A^T \lambda + c)^T x

Then

\[ g(\lambda) = \inf_{x} L(x, \lambda) = \begin{cases} -b^T \lambda & \text{if } A^T \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases} \]

We obtain a non-trivial lower bound only when \( \lambda \) satisfies \( A^T \lambda + c = 0 \); then \( -b^T \lambda \) is a lower bound on the optimal value of the LP.

(ii) Two-way partitioning problem: (nonconvex)

\[ \begin{align*}
\min \quad & x^T W x \\
\text{st.} \quad & x_i^2 = 1, \quad i = 1, \ldots, n, \quad W \in S^n
\end{align*} \]

This problem is nonconvex, combinatoric in \( x \). The set of feasible points is finite (contains \( 2^n \) points).

We can find the optimal point by exhaustive search if \( n \leq 50 \), but larger \( n \) (\( n > 50 \)) the search is prohibitive (\( 2^n \) operations).

Two-way partitioning interpretation: \( x_i = 1 \) or \( x_i = -1 \).

\( W_{ij} \): cost of having elements \( i \) and \( j \) in the same partition

\( -W_{ij} \): cost of having elements \( i \) and \( j \) in different partitions

We want to minimize total cost.

The Lagrangian of this problem is

\[ L(x, \nu) = x^T W x + \sum_{i=2}^{n} \nu_i (x_i^2 - 1) \]

= \( x^T (W + \text{diag}(\nu)) x - 1^T \nu \)
The dual function is
\[ g(\nu) = \begin{cases} -\nu^T \nu & \text{if } \ W + \text{diag}(\nu) \geq 0 \\ -\infty & \text{otherwise} \end{cases} \]

From the dual function we can obtain a lower bound by taking a specific value of the dual variable \( \nu \), such as
\[ \nu = -\lambda_{\min}(W), 1, \]
then
\[ W + \text{diag}(\nu) = W - \lambda_{\min}(W), I \geq 0, \]
which yields the lower bound
\[ p^* \geq -\nu^T \nu = n \lambda_{\min}(W). \]

1) Lagrange dual and the conjugate function:
The Lagrange dual is closely related to the conjugate function
\[ f^*(y) = \sup_{x \in \mathbb{R}^n} (y^T x - f(x)) \]
Consider
\[ \min_{x \in \mathbb{R}^n} f(x) \]
s.t.
\[ Ax \leq b \]
\[ Cx = d \]
then
\[ g(\lambda, \nu) = \frac{d}{x} (f(x) + \lambda^T (Ax - b) + \nu^T (Cx - d)) \]
\[ = -\lambda^T b - \nu^T d + \frac{d}{x} f(x) + (\lambda^T A + \nu^T C)x \]
\[ = -\lambda^T b - \nu^T d - f^* (-\lambda^T A - \nu^T C) \]

This relation is useful if we readily know the conjugate function of \( f(x) \).
The dual function is
\[ g(\nu) = \begin{cases} -1^T\nu & \text{if } W + \text{diag}(\nu) \preceq 0 \\ -\infty & \text{otherwise} \end{cases} \]

From the dual function we can obtain a lower bound by taking a specific value of the dual variable \( \nu \), such as
\[ \nu = -\lambda_{\min}(W), \]
then
\[ W + \text{diag}(\nu) = W - \lambda_{\min}(W).I \succeq 0, \]
which yields the lower bound
\[ p^* \geq -1^T \nu = n\lambda_{\min}(W). \]

+) Lagrange dual and the conjugate function:

The Lagrange dual is closely related to the conjugate function
\[ f^*(y) = \sup_{x \in \text{dom} f} (y^T x - f(x)) \]

Consider
\[ \min f_0(x) \]
subject to
\[ Ax \leq b \]
\[ Cx = d \]
then
\[ g(\lambda, \nu) = \frac{1}{\nu} \left( f_0(x) + \lambda^T (Ax - b) + \nu^T (Cx - d) \right) \]
\[ = -\lambda^T b - \nu^T d + \frac{1}{\nu} \left( f_0(x) + (\lambda^T A + \nu^T C)x \right) \]
\[ = -\lambda^T b - \nu^T d + f^* (\lambda^T A + \nu^T C) \]

This relation is useful if we readily know the conjugate function of \( f_0(x) \).
Example: Minimum volume convex ellipsoid
\[ \min \quad f_0(x) = \log \det(x^{-1}) \]
\[ \text{s.t.} \quad a_i^T x a_i \leq 1, \quad i = 1, \ldots, m \]
\[ \text{dom} f_0 = S^{n+}. \]
The conjugate function of \( f_0(x) \) is:
\[ f_0^*(y) = \log \det(y^{-1}) - n. \]
The original problem has inequality constraints that are linear in \( x \), which can be expressed as
\[ \text{tr}(a_i a_i^T x) \leq 1 \]
Then the dual function is:
\[ g(\lambda) = \inf \left\{ \log \det(x^{-1}) + \sum_{i=1}^{m} \lambda_i [\text{tr}(a_i a_i^T x) - 1] \right\} \]
\[ = -f_0^* \left( -\sum_{i=1}^{m} \lambda_i (a_i a_i^T) \right) - 1^T \lambda \]
\[ = \begin{cases} 
\log \det (+ \sum_{i=1}^{m} \lambda_i (a_i a_i^T)) + n - 1^T \lambda, & \text{if } \sum_{i=1}^{m} \lambda_i (a_i a_i^T) > 0 \\
-\infty, & \text{otherwise}
\end{cases} \]
Thus for any \( \lambda \geq 0 \) such that \( \sum_{i=1}^{m} (a_i a_i^T) \lambda_i > 0 \), we have a lower bound for \( p^* \) as
\[ \log \det \left( \sum_{i=1}^{m} \lambda_i (a_i a_i^T) \right) + n - 1^T \lambda. \]

2. The Lagrange dual problem:
We want to know what is the best lower bound that can be obtained from the Lagrange dual function?
Example: Minimum volume convex ellipsoid
\[ \min f_0(X) = \log \det X^{-1} \]
\[ \text{s.t. } a_i^T X a_i \leq 1, \ i = 1 \ldots n \]
\[ \text{dom } f_0 = S^n_{++} \]
The conjugate function of \( f_0(X) \) is:
\[ f^*_0(Y) = \log \det (Y)^{-1} - n \]

The original problem has inequality constraints that are linear in \( X \), which can be expressed as
\[ \text{tr} (a_i a_i^T X) \leq 1 \]
Then the dual function is:
\[ g(\lambda) = \inf \left( \log \det (X^{-1}) + \sum_{i=1}^n \lambda_i \left[ \text{tr} (a_i a_i^T X) - 1 \right] \right) \]
\[ = -f^*_0 \left( -\sum_{i=1}^n \lambda_i (a_i a_i^T) \right) - 1^T \lambda \]
\[ = \begin{cases} 
\log \det \left( + \sum_{i=1}^n \lambda_i (a_i a_i^T) \right) + n - 1^T \lambda & \text{if } \sum_{i=1}^n \lambda_i (a_i a_i^T) > 0 \\
-\infty & \text{otherwise}
\end{cases} \]

Thus for any \( \lambda \geq 0 \) such that \( \sum_{i=1}^n (a_i a_i^T)^T \lambda_i > 0 \), we have a lower bound for \( p^* \) as
\[ \log \det \left( \sum_{i=1}^n \lambda_i (a_i a_i^T) \right) + n - 1^T \lambda \]

Exercise 13:

2. The Lagrange dual problem:

We want to know what is the best lower bound that can be obtained from the Lagrange dual function?
This best bond lead to another opt problem:
\[ \max \ g(A, \nu) \]
\[ \text{s.t.} \quad \lambda \geq 0 \]

This problem is called the Lagrange dual problem. The original problem is then called the primal problem.

- Dual feasible: a pair \((\lambda, \nu)\) such that \(\lambda \geq 0\) and \(g(A, \nu) \geq -\infty\) is called dual feasible.

- \((\lambda^*, \nu^*)\): dual optimal (or optimal Lagrange multipliers) if they are optimal for the dual problem.

- The dual problem is always convex regardless of the primal problem being caused or not.

- We often simplify the dual problem by making any implicit constraints that \((\lambda, \nu)\) belong to be explicit.

Example: \[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad a_i^T x \leq b \\
\end{align*}
\] \[\iff\]
\[
\begin{align*}
\max & \quad -b^T \lambda \\
\text{s.t.} & \quad A^T \lambda + c = 0, \\
\lambda & \geq 0 \\
\end{align*}
\]

1) Optimal dual value: \(d^*\)
   - We always have \(d^* \leq p^*\) → weak duality.
   - The difference \(p^* - d^*\) is called the duality gap.

Example: two-way partitioning
\[
\begin{align*}
\min & \quad x^T W x \\
\text{s.t.} & \quad x_i^2 = 1, \quad i = 1, \ldots, n \\
\text{difficult to solve} & \quad (n < 30) \\
\end{align*}
\] \[\iff\]
\[
\begin{align*}
\max & \quad -v^T u \\
\text{s.t.} & \quad W + \text{diag}(v) \succeq 0, \\
\text{easy to solve, even for} & \quad \text{large } n \quad (n \approx 1000) \\
\end{align*}
\]
+ Strong duality:
  
  • If \( d^* = p^* \) then we have strong duality. With strong duality, then \((\lambda^*, \nu^*)\) serves as certificate of optimality for optimal \( x^* \).

  • For convex problems, we usually (but not always) have strong duality.

  There are many conditions that guarantee strong duality for convex problems. These conditions are called "constraints qualifications." One of them is Slater's condition.

+ Slater’s condition: \( \exists x \) "strictly feasible" such that

\[
 f_i(x) < 0, \quad i = 1, \ldots, m, \quad Ax = b.
\]

For affine inequalities, Slater’s condition can be refined (relaxed) so that the affine inequalities do not have to hold with strict inequalities. Specifically, suppose \( f_1, \ldots, f_k \) are affine, then strong duality holds if

\[
 \begin{align*}
 f_1(x) &\leq 0, \quad i = 1, \ldots, k, \\
 f_i(x) &< 0, \quad i = k+1, \ldots, m \\
 Ax & = b
\end{align*}
\]

Slater's theorem: Strong duality holds if the problem is convex and Slater's condition holds (i.e. there exists a strictly feasible point).

Slater's condition also implies that if \( d^* < \infty \) then it is attained (achievable). That is, there exists a dual feasible point \((\lambda^*, \nu^*)\) such that \( g(\lambda^*, \nu^*) = d^* = p^* \).
Example: (i) LP: Strong duality always holds as long as
the problem is feasible ($p^* < \infty$).
\[
\min c^T x \\
\text{s.t. } Ax \leq b \iff \max -b^T \lambda \\
\text{s.t. } A^T \lambda + c = 0 \\
\lambda \geq 0
\]
$p^* = d^*$ except when $p^* = \infty$, $d^* = -\infty$.
(both primal and dual are infeasible).

(ii) Minimum covering ellipsoid: Strong duality always holds.
(as the inequality constraint is linear in $x$: \exists x \text{ s.t. } a_i^T x a_i \leq 1).

(iii) Non-convex problem with strong duality.
\[
\min x^T A x + 2b^T x \\
\text{s.t. } x^T x \leq 1
\]
$A \in S^n$ but $A \not\succeq 0$ so the primal problem is non-convex.
\[
\begin{align*}
L(x, \lambda) &= x^T A x + 2b^T x + \lambda(x^T x - 1) \\
&= x^T (A + \lambda I)x + 2b^T x - \lambda
\end{align*}
\]
\[
g(A) = \begin{cases} -b^T (A + \lambda I)^\dagger b - \lambda & \text{if } A + \lambda I \succeq 0 \\
-\infty & \text{otherwise}
\end{cases}
\]
where $(A + \lambda I)^\dagger$ is the pseudoinverse of $A + \lambda I$.

Dual problem:
\[
\max -b^T (A + \lambda I)^\dagger b - \lambda \\
\text{s.t. } A + \lambda I \succeq 0 \\
b \in \mathbb{R}(A + \lambda I)
\]
\[
\text{SDP reformulate s.t. } \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0
\]
Strong duality holds (not easy to show).
Geometric interpretation and sketch of a proof for strong duality.

Define a set of function values:

\[ A = \{ (u, v, t) \mid \exists x \in D: u_i \geq f_i(x), \forall i = 1, \ldots, m \}
\]

\[ v_i \geq h_i(x), \forall i = 1, \ldots, p \]

\[ t \geq g_0(x) \}

where \[ u \in \mathbb{R}^m_+ \]

\[ v \in \mathbb{R}^p \]

\[ t \in \mathbb{R}_+ \]

The set \( A \) is convex for convex problems.

Then the set of feasible values in \( A \) is \( \{ (0, 0, t) \} \), and the optimal value can be expressed as

\[ p^* = \min \{ t \mid (0, 0, t) \in A \} \]

Also, the dual function at \( (\lambda, \nu) \) for \( \lambda \geq 0 \) can be obtained as

\[ g(\lambda, \nu) = \min \{ (\lambda, \nu, 1)^T (u, v, t) \mid (u, v, t) \in A \} \]

since

\[ g(\lambda, \nu) = \min \left( \sum_i \lambda_i f_i(x) + \sum_i \nu_i h_i(x) + f_0(x) \right) \]

\[ = \min \left( \sum_i \lambda_i u_i + \sum_i \nu_i v_i + t \right) \]

\[ = \min \left( \sum_i (\lambda_i, \nu_i, 1) \right) \]

\[ = \min \left( \sum_i \right) \]

Thus, \( g(\lambda, \nu) \) is a hyperplane supporting set \( A \), which is on the form

\[ a^T w \geq b \quad \forall w \in A \]

where \( \{ a = (\lambda, \nu, 1) \} \)

\[ b = g(\lambda, \nu) \].
Plot the set $A$ with the last element ($t$) as the "vertical axis. It is easier to visualize in a 2D plane, so let's consider a problem with only one inequality constraint, so that $u \in \mathbb{R}$, $u \neq 0$.

This plot is for a generic (non-convex) problem, so set $A$ is not convex.

Hyperplane $a^T w = b$ or $A u + t = g(A)$ for a given $A$.

Weak duality: We already have for each $(A, u)$ then $g(A, u) \leq (A, u, 1) \mathbf{T} (u, v, s, t)$ for $(u, v, s, t) \in A$ but $(0, 0, p^*) \in A$ thus $g(A, u) \leq (A, u, 1) \mathbf{T} (0, 0, p^*) = p^* + A, u$

Strong duality: For a convex problem, set $A$ is convex.

This plot is for a convex set $A$ resulting from a convex problem.
The idea is then to say that since \( A \) is convex, then the point \((0,0,p^*)\) on the boundary of \( A \) must have a supporting hyperplane at it (we saw that this was not necessarily true for a non-convex set \( A \)).

Then Slater's condition is used to ensure that this supporting hyperplane at \((0,0,p^*)\) is not vertical (the same as the \(t\) axis), which ensures that there's an intersection between this supporting hyperplane and the vertical axis \( t \), so that \( p^* > -\infty \).

Once \( p^* > -\infty \) then the intersection is both \( p^* \) and \( d^* \) therefore we get \( p^* = d^* \).

+ ) Max-min and Saddle-point Interpretation:

Consider a problem with only inequality for simplicity (but analysis can be extended to problems with equality constraints).

The Lagrangian is \( L(x, \lambda) = \sum_{i=1}^{m} \lambda_i f_i(x) + f_0(x) \).

Now note that since \( \lambda \geq 0 \), we have
\[
\sup_{\lambda \geq 0} L(x, \lambda) = \sup_{\lambda \geq 0} \left( f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right)
\]
\[
= \begin{cases} 
  f_0(x) & \text{if } f_i(x) \leq 0, \ i = 1, \ldots, m \\
  -\infty & \text{otherwise}
\end{cases}
\]

Thus for all feasible \( x \), then
\[
f_0(x) = \sup_{\lambda \geq 0} L(x, \lambda)
\]

Hence
\[
p^* = \min_{x \in F} f_0(x) = \min_{x} \sup_{\lambda \geq 0} L(x, \lambda)
\]
The idea is then to say that since $A$ is convex, then the point $(0, 0, p^*)$ on the boundary of $A$ must have a supporting hyperplane at it (we saw that this was not necessarily true for a non-convex set $A$).

Then Slater's condition is used to ensure that this supporting hyperplane at $(0, 0, p^*)$ is not vertical (the same as the $t$-axis), which ensures that there is an intersection between this supporting hyperplane and the vertical axis $t \geq 0$ so that $p^* > -\infty$.

Once $p^* > -\infty$ then the intersection is both $p^*$ and $d^*$, therefore we get: $p^* = d^*$.

1) Max-min and saddle point interpretation:

Consider a problem with only inequality constraints for simplicity (but analysis can be extended to problems with equality constraints).

The Lagrangian is $L(\lambda, x) = \sum_{i=1}^{m} \lambda_i f_i(x) + f_0(x)$.

Now note that since $\lambda \geq 0$, we have

$$\sup_{\lambda \geq 0} L(x, \lambda) = \sup_{\lambda \geq 0} (f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x))$$

Thus for all feasible $x$, then

$$f_0(x) = \sup_{\lambda \geq 0} L(x, \lambda) \quad \forall x \in \mathcal{F} \text{ (feasible set)}$$

Hence

$$p^* = \inf_{x \in \mathcal{F}} f_0(x) = \inf_{x \in \mathcal{F}} \sup_{\lambda \geq 0} L(x, \lambda)$$
By the definition of the dual function, we also have
\[ d^* = \sup_{\lambda \geq 0} g(\lambda) = \sup_{\lambda \geq 0} \inf_x L(x, \lambda). \]

Thus weak duality implies
\[ \sup_{\lambda \geq 0} \inf_x L(x, \lambda) \leq \inf_x \sup_{\lambda \geq 0} L(x, \lambda). \]

In fact this inequality holds in general for every function \( L(x, \lambda) \). (max-min inequality)

Strong duality holds when we have equality:
\[ \sup_{\lambda \geq 0} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda \geq 0} L(x, \lambda). \]

**Saddle point:** For a given function \( f(w, z) \), the pair \( (\bar{w}, \bar{z}) \) is called the saddle point for \( f \) if
\[ f(\bar{w}, \bar{z}) \leq f(\tilde{w}, \bar{z}) \leq f(w, \bar{z}) \quad \forall w, z. \]

\( \bar{w} \) minimizes \( f \) over \( w \):
\[ f(\bar{w}, \bar{z}) = \inf_w f(w, \bar{z}) \]
\( \bar{z} \) maximizes \( f \) over \( z \):
\[ f(\bar{w}, \bar{z}) = \sup_z f(\bar{w}, z) \]

\[ \implies \inf_w \sup_z f(w, z) = \sup_z \inf_w f(w, z) = f(\bar{w}, \bar{z}) \]
Thus strong duality holds and the saddle point is the common value.

Strong duality can also be interpreted as a zero-sum game between two players, one wants to minimize the payoff to the other, and the other wants to maximize this payoff.
3. Optimality conditions:

Again, for this part we do not assume the original (primal) problem to be convex. So the theory applies to all optimization problems.

4) Certificate of (sub) optimality and stopping criterion:

- If we can always find a dual feasible point \((\lambda, \nu)\), then we can establish a lower bound on \(p^*\):
  \[
p^* \geq g(\lambda, \nu).
  \]
  \((\lambda, \nu)\) provides a proof or certificate that \(p^* \geq g(\lambda, \nu)\).

- If strong duality holds \((\lambda^*, \nu^*)\) provides a certificate of optimality:
  \[
p^* = g(\lambda^*, \nu^*)
  \]

- Duality gap: define \(\Delta = f_0(\mathbf{x}) - g(\lambda, \nu)\), \(\mathbf{x}, \lambda, \nu\) feasible
  then
  \[
f_0(\mathbf{x}) - p^* \leq \Delta
  \]

  This establishes that point \(\mathbf{x}\) is \(\Delta\)-suboptimal.

- We can use this property to establish a non-heuristic stopping criterion for algorithms, so that we stop when
  \[
f(\mathbf{x}(k)) - g(\lambda(k), \nu(k)) \leq \varepsilon \text{ tolerance}
  \]
  where \(k\) indicates the iteration number (much more on this later).

4) Complementary slackness:

Suppose that strong duality holds. Let \(x^*\) and \((\lambda^*, \nu^*)\) be the primal & dual optimal points.
We have:
\[
\begin{align*}
 f_0(x^*) &= g(A^*, \nu^*) \quad \text{(strong duality)} \\
 &= \inf_{x} \left( f_0(x) + \sum_{i=1}^{m} A_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \\
 &\leq f_0(x^*) + \sum_{i=1}^{m} A_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \quad \geq 0 \leq 0 \\
 &\leq f_0(x^*)
\end{align*}
\]
which implies we must have equality through. Thus
\[
\sum_{i=1}^{m} A_i^* f_i(x^*) = 0
\]
Since each term is non-positive (\( \leq 0 \)), this implies
\[
A_i^* f_i(x_i^*) = 0 \quad \forall i = 1, \ldots, m.
\]
This condition is called complementary slackness.

Complementary slackness can also be expressed as:
\[
A_i^* > 0 \rightarrow f_i^*(x) = 0 \quad \text{(tight constraint)}
\]
\[
or \quad f_i^*(x) < 0 \rightarrow A_i^* = 0 \quad \text{(loose constraint)}.
\]

\[\text{KKT Conditions:}\] Assume that \( f_0, f_1, \ldots, f_m, h_1, \ldots, h_p \) are all differentiable, but we do not assume they are convex.

Let \( x^* \) and \((A^*, \nu^*)\) be any primal and dual optimal pair with zero duality gap. Since \( x^* \) minimizes \( \mathcal{L}(x, A^*, \nu^*) \) over \( x \), it follows that
\[
\nabla_x \mathcal{L}(x^*, A^*, \nu^*) \mid_{x=x^*} = 0
\]
\[
\Leftrightarrow \quad \nabla f_0(x^*) + \sum_{i=1}^{m} A_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) = 0.
\]
Coupled with the fact that \( x^* \) is feasible, we then have the following set of conditions:

\[
\begin{align*}
    f_i(x^*) & \leq 0 & i = 1 \ldots m \\
    h_i(x^*) & = 0 & i = 1 \ldots p \\
    \lambda_i & > 0 & i = 1 \ldots m \\
    \lambda_i \cdot f_i(x^*) & = 0 & i = 1 \ldots m \\
    \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i \nabla f_i(x^*) + \sum_{i=1}^p \nu_i h_i(x^*) & = 0
\end{align*}
\]

This set of conditions is called the KKT conditions.

The KKT condition holds for any optimization problem that obtains strong duality and has differentiable objectives and constraint functions.

**KKT for convex problems**

For a primal convex problem, the KKT condition is also sufficient for \( x^* \), \( (\lambda^*, \nu^*) \) to be primal and dual optimal.

So for convex problems, KKT is both necessary and sufficient.

That is, if the point \( \tilde{x} \), \( (\tilde{\lambda}, \tilde{\nu}) \) that satisfy the KKT condition for a convex problem (\( f_i(x) \) convex, \( h_i(x) \) affine):

\[
\begin{align*}
    f_i(x) & \leq 0 & i = 1 \ldots m \\
    h_i(x) & = 0 & i = 1 \ldots p \\
    \lambda_i & > 0 & i = 1 \ldots m \\
    \lambda_i \cdot f_i(x) & = 0 & i = 1 \ldots m \\
    \nabla f_0(x) + \sum \tilde{\lambda}_i \nabla f_i(x) + \sum \tilde{\nu}_i h_i(x) & = 0
\end{align*}
\]

then \( \tilde{x} \), \( (\tilde{\lambda}, \tilde{\nu}) \) are primal and dual optimal with zero duality gap.
This sufficiency condition is easy to show:

- The first 2 conditions imply \( \tilde{x} \) feasible.
- The last condition implies \( \tilde{x} \) minimizes \( L(x, \lambda, \nu) \).

\[
g(\tilde{x}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) = f(\tilde{x}) + \sum \lambda_i f_i(\tilde{x}) + \sum \nu_i h_i(\tilde{x})
\]

\[
= f(\tilde{x}) + \sum \lambda_i f_i(\tilde{x}) + \sum \nu_i h_i(\tilde{x})
\]

\[
\rightarrow \text{strong duality holds for } \tilde{x}, (\tilde{\lambda}, \tilde{\nu}) \text{ (with weak duality as)}
\]

Convexity is used in the second argument: \( \tilde{x} \) minimizes \( L(x, \lambda, \nu) \).

This is based on the fact that

\( \lambda_i \geq 0 \rightarrow L(x, \lambda, \nu) \) is convex in \( x \)

(Non-negative weighted sum, and \( h_i(x) \))

are affine.

Then the last KKT condition which states the gradient of \( L(x, \lambda, \nu) \) vanishes at \( \tilde{x} = \tilde{x} \) implies \( \tilde{x} \) minimizes \( L(x, \lambda, \nu) \).

This conclusion may not hold if the functions \( f_i \) are non-convex or \( h_i \) are not affine.

**Example:** Water filling in achieving the communication capacity

\[
\min \quad \frac{1}{n} \sum_{i=1}^{n} \log(x_i + x_i^*)
\]

S.T. \( x_i \geq 0 \), \( 1^Tx = 1 \),

where \( x_i > 0 \).

\( x_i \): transmit power on channel \( i \)

\( x_i \): noise power on channel \( i \).

\[
L(x, \lambda, \nu) = \frac{1}{n} \sum_{i=1}^{n} \log(x_i + x_i^*) - \lambda^T x + \nu (1^T x - 1)
\]

The KKT conditions reduce to

\[
x^* \geq 0 \rightarrow 1^T x^* = 1 \Rightarrow \lambda^* = 0, \lambda^* = 0
\]

\[
\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* = 0, \quad i = 1 \ldots n
\]
This sufficiency condition is easy to show:

- The first 2 conditions imply \( \bar{x} \) feasible.
- The last condition implies \( \bar{x} \) minimizes \( L(x, \lambda, \nu) \).

\[
g(\bar{x}, \lambda, \nu) = L(x, \lambda, \nu) = f_0(\bar{x}) + \sum \lambda_i f_i(\bar{x}) + \sum \nu_i h_i(\bar{x})
\]

\[
\Rightarrow \text{Strong duality holds for } \bar{x}, (\lambda, \nu) \text{ (with weak duality \( \geq \))}
\]

Convexity is used in the second argument: \( \bar{x} \) minimizes \( L(x, \lambda, \nu) \).

This is based on the fact that \( \lambda_i \geq 0 \Rightarrow L(x, \lambda, \nu) \) is convex in \( \bar{x} \)

(non-negative weighted sum, and \( h_i(\bar{x}) \) are affine)

Then the last KKT condition which states the gradient of \( L(x, \lambda, \nu) \) vanishes at \( \bar{x} = x \) implies \( \bar{x} \) minimizes \( L(x, \lambda, \nu) \).

This conclusion may not hold if the functions \( f_i \) are non-convex or \( h_i \) are not affine.

**Example:** Water filling in achieving the communication capacity.

\[
\begin{align*}
\min & \quad \sum_{i=1}^{n} \log(\alpha_i + x_i) \\
\text{s.t.} & \quad x_i \geq 0, \quad x^T x = 1,
\end{align*}
\]

where \( \alpha_i > 0 \).

- \( x_i \): transmit power on channel \( i \)
- \( \nu \): noise power on channel \( i \).

\[
L(x, \lambda, \nu) = -\sum_{i=1}^{n} \log(\alpha_i + x_i) - \lambda^T x + \nu(1^T x - 1)
\]

The KKT conditions reduce to

\[
\begin{align*}
x^* \geq 0, \quad 1^T x^* = 1, \quad \lambda_i^* x_i^* = 0, \quad \lambda^* \geq 0 \\
\frac{1}{x_i^* + \alpha_i} - \lambda_i^* + \nu^* = 0, \quad i = 1 \ldots n
\end{align*}
\]
Since \( x_i^* \) is a slack variable, we can eliminate it, for which we can write
\[
x_i^* \left( \frac{-1}{x_i^* + x_i^*} + \nu^* \right) = 0
\]
\[ x_i^* \geq 0, \quad \nu^* = 1 \]
\[ \nu^* \geq \nu / (x_i^* + x_i^*) \]

Solving this set of conditions leads to
\[
x_i^* = \left\{ \begin{array}{ll}
\nu^* - x_i & \nu^* < \frac{1}{x_i^*} \\
0 & \nu^* \geq \frac{1}{x_i^*}
\end{array} \right.
\]

or simply
\[
x_i^* = \left( \frac{1}{\nu^*} - x_i \right)^+ \quad \text{(notation)}
\]

To find \( \nu^* \), substitute this expression into \( \nu^* = 1 \) to get
\[
\sum_{i=1}^{n} \max \left\{ 0, \frac{1}{\nu^*} - x_i \right\} = 1
\]

---

1) Solving the primal problem via the dual:

Sometimes, the dual problem is easier to solve (for example, it has some special structures, or admits analytical solutions).

Suppose that we have strong duality and know an optimal point \((\lambda^*, \nu^*)\). Then, we can solve the following problem to find the optimal primal values. (Suppose its solution is unique)
\[
\min f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x)
\]

If the solution of the above problem is primal feasible, then it must be the primal optimal point. If that solution is not primal feasible then no primal optimal point exists, i.e., the primal optimum is not attained.

4. Perturbation & sensitivity analysis

4.1) The perturbed problem:

\[
\min f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x)
\]

\[
\text{s.t. } f_i(x) \leq u_i, \quad i = 1 \ldots m
\]

\[
h_i(x) = v_i, \quad i = 1 \ldots p.
\]

This problem is the same as the original one if \( u_i = 0, v_i = 0 \).

If \( u_i > 0 \), we relax the \( i \)th inequality constraint

\( u_i < 0 \), we tighten

Define \( p^*(\lambda, v) \) as the optimal value of the perturbed prob

\[
p^*(\lambda, v) = \min \{ f(x) \mid x \in \mathbb{E} : f_i(x) \leq u_i, \quad i = 1 \ldots m \}
\]

\[
h_i(x) = v_i, \quad i = 1 \ldots p.
\]

Note that \( p^*(0, 0) = p^* \), the optimal value of the primal prob

4.1.1) Comparison:

- **unperturbed**
  
  \[
  \begin{align*}
  \min & \quad f(x) \\
  \text{s.t.} & \quad f_i(x) \leq 0 \\
  & \quad h_i(x) = 0
  \end{align*}
  \]

- **perturbed**
  
  \[
  \begin{align*}
  \min & \quad f(x) \\
  \text{s.t.} & \quad f_i(x) \leq u_i \\
  & \quad h_i(x) = v_i
  \end{align*}
  \]

- **dual**
  
  \[
  \max g(A, \nu) \\
  \text{s.t.} \quad \lambda \geq 0
  \]

\[
\max g(A, \nu) - \lambda u - v_i
\]

\[
\text{s.t.} \quad \lambda \geq 0
\]
Assume that strong duality holds, and the dual optimal is attained.

Let \((\lambda^*, \nu^*)\) be the dual optimal point, then
\[
P^*(u, v) \geq P^*(0, 0) - \lambda^T u - \nu^T v.
\]

To see this, suppose \(x^\star\) is a feasible point of the perturbed problem. By strong duality, we have
\[
P^*(0, 0) = g(\lambda^*, \nu^*)
= \min_{x} \left( f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{n} \nu_i^* h_i(x) \right)
\leq f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^* f_i(x^\star) + \sum_{i=1}^{n} \nu_i^* h_i(x^\star)
\leq f_0(x^\star) + \lambda^T u + \nu^T v. \quad \forall x \text{ feasible}
\]

Hence for all \(x\) feasible for the perturbed problem, then
\[
f_0(x) \geq P^*(0, 0) - \lambda^T u - \nu^T v \quad \forall x \in \mathcal{F}(u, v)
\Rightarrow P^*(0, 0) - \lambda^T u - \nu^T v \leq \inf_{x \in \mathcal{F}(u, v)} f_0(x) = P^*(u, v).
\]

\(\star\) Sensitivity interpretation: When strong duality holds, we can infer various sensitivity interpretations:

- \(\lambda_i^*\) is the sensitivity factor of the \(i\)th inequality constraint.

  If \(\lambda_i^*\) is large, then tightening the \(i\)th constraint \((u_i < 0)\) will greatly increase the optimal value \(P^*(u, v)\).

  If \(\lambda_i^*\) is small and we loosen the \(i\)th constraint \((u_i > 0)\) then the optimal value will not decrease much.
If $\lambda_i^* = 0$, the problem is insensitive to the $i_{th}$ inequality constraint (see the local analysis below).

Similarly for $\nu_i^*$:

If $\nu_i^* > 0$ and is large $\Rightarrow$ $p^*$ very sensitive to $\nu_i^*$ $(\nu_i^* < 0)$

$\nu_i^* \rightarrow n$ $(\nu_i^* > 0)$.

Local sensitivity analysis: In addition, if $p^*(u,v)$ is differentiable at $(0,0)$ then:

$$\lambda_i^* = \frac{\partial p^*(u,v)}{\partial u_i} \bigg|_{u=0,v=0}$$

$$\nu_i^* = \frac{\partial p^*(u,v)}{\partial v_i} \bigg|_{u=0,v=0}$$

This is because:

$$\frac{\partial p^*(0,0)}{\partial u_i} = \lim_{t \to 0} \frac{p^*(t,e_i,0) - p^*(0,0)}{t}$$

$e_i$ is a vector with the $i_{th}$ element as 1 and the rest 0. Since we have $p^*(u,v) \geq p^*(0,0) - \lambda^* U - \nu^* V$

$\Rightarrow p^*(t,e_i,0) - p^*(0,0) \geq -\lambda_i^* t$

Thus

$$\frac{p^*(t,e_i,0) - p^*(0,0)}{t} \begin{cases} \geq -\lambda_i^* \text{ if } t > 0 \\ \leq -\lambda_i^* \text{ if } t < 0. \end{cases}$$

Taking the limit as $t \to 0$ yields

$$\frac{\partial p^*(0,0)}{\partial u_i} = -\lambda_i^*.$$
Thus $x_i^*$ gives the local sensitivity of $p^*$ to the $i$th equality constraint.

- If $x_i^* = 0$, $f_i(x^*) < 0 \rightarrow$ the constraint is inactive and can be tightened or loosened by a small amount without affecting the optimal value $p^*$.

- If $x_i^* > 0$, $f_i(x^*) = 0 \rightarrow$ the constraint is active; then $x_i^*$ tells how sensitive is the optimal value to this constraint if it is loosen or tightened a little bit. The interpretation is symmetric:

  Tightening by $u_i < 0$, $|u_i|$ small $\rightarrow$ increase in $p^*$ by $\approx x_i^*$.
  Loosening by $u_i > 0$, $|u_i|$ small $\rightarrow$ decrease $p^*$ by $\approx +x_i^*$.

1) Shadow pricing interpretation:

Let $f_i(x)$ be the cost ($-f_i(x)$ the profit), $f_i(x)$ be constraints on resources.

Then $x_i^*$ tells approximately how much more profit the firm can make for a small increase in availability of resource $i$.

Then $x_i^*$ is the equilibrium price for resource $i$. The firm can buy more resource $i$ if the market price is lower than $x_i^*$, and can sell resource $i$ if the price is higher than $x_i^*$. Based on that, the firm can make profit.

2) Theorems of alternatives:

Here we apply Lagrangean duality theory to determine the feasibility of a system of inequalities and equalities.
\[
\begin{align*}
\begin{cases}
& f_i(x) \leq 0, \quad i = 1 \ldots m \\
& h_i(x) = 0, \quad i = 1 \ldots p
\end{cases}
\end{align*}
\]

This problem can be cast as

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

Assume the domain \( D \) is non-empty.

If \( p^* = 0 \rightarrow \) set of inequalities & equalities are feasible

\( p = \infty \rightarrow \) infeasible

Dual function:

\[
g(\lambda, \nu) = \max \left( \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)
\]

and the dual problem is

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

Note that for \( \lambda > 0 \rightarrow g(\lambda \lambda, \lambda \nu) = \lambda g(\lambda, \nu) \). Thus the optimal dual value is:

\[
d^* = \begin{cases} 
\infty & \text{if } \lambda \geq 0, \quad g(\lambda, \nu) > 0 \text{ is feasible} \\
0 & \text{if } \lambda \geq 0, \quad g(\lambda, \nu) > 0 \text{ is infeasible}
\end{cases}
\]

Since \( d^* \leq p^* \), we can conclude that

if \( \lambda \geq 0, \quad g(\lambda, \nu) > 0 \) is feasible \( (d^* = \infty) \)

then \( f_i(x) \leq 0, \quad h_i(x) = 0 \) is infeasible \( (p^* = \infty) \).

Alternatively, if \( f_i(x) \leq 0, \quad h_i(x) = 0 \) is feasible \( (p^* = 0) \)

then \( \lambda \geq 0, \quad g(\lambda, \nu) = 0 \) is infeasible \( (d^* = 0) \).
Thus the two sets of inequalities and equalities:
\[ f_i(x) \leq 0, \quad h_i(x) = 0 \]
and \( \lambda \geq 0, \quad g(\lambda, y) > 0 \)
are called weak alternatives. That is, at most one of the two sets is feasible.

Weak alternatives hold for all problems (Convex not required)

Weaker alternatives with strict inequalities:
\[ f_i(x) < 0, \quad h_i(x) = 0 \]
and \( \lambda \geq 0, \quad \lambda \neq 0, \quad g(\lambda, y) > 0 \)
are weak alternatives.

This is easy to show: Suppose \( \exists x : f_i(x) < 0, h_i(x) = 0 \),
then \( \exists \lambda \geq 0, \lambda \neq 0 : \)
\[ \sum_i f_i(x) + \sum_i h_i(x) < 0 \]

Thus
\[ g(\lambda, y) = \inf \left( \sum_i f_i(x) + \sum_i h_i(x) \right) \]
\[ \leq \sum_i f_i(x) + \sum_i h_i(x) < 0 \]

Strong alternatives: Strong alternatives mean exactly one system is feasible.

Strong alternatives hold for convex feasibility problems (\( f_i \) are convex and \( h_i \) are affine), provided some qualification holds.
Two systems: \( f_i(g_i) < 0, \ i = 1 \ldots m, \ Ax = b \)
\[ \text{and} \quad x \geq 0, \ x \neq 0, \ g(x, u) \geq 0 \]
are strong alternatives if \( \exists x \in \text{the interior of } D \)
such that \( Ax = b \).
The proof is based on strong duality & Slater's condition.

Similarly, two systems
\[ f_i(g_i) \leq 0, \ i = 1 \ldots m, \ Ax = b \]
\[ \text{and} \quad x \geq 0, \ g(x, u) > 0 \]
are strong alternatives if \( \exists x \in \text{the interior of } D \)
such that \( Ax = b \) and the value \( p^* \) is attained.

**Example:** Two systems
\[ Ax \leq 0, \ c^T x \leq 0 \quad , \quad \mathbb{A} \in \mathbb{R}^{m \times n}, c \in \mathbb{R}^n \]
\[ \text{and} \quad A^Ty + c = 0, \ y \geq 0 \]
are strong alternatives. *(Fatka's lemma)*

This can be seen by analyzing the dual of the LP:
\[
\begin{align*}
\text{min } & c^T x \\
\text{s.t.} & Ax \leq 0
\end{align*}
\]
\[
\text{dual} \quad \begin{align*}
\text{max } & 0 \\
\text{s.t.} & A^Ty + c = 0 \\
& y \geq 0
\end{align*}
\]
\[ L(x, \lambda) = c^T x + \lambda^T(Ax) = (A^T \lambda + c)^T x \]
\[ g(\lambda) = \min_{\lambda} L(x, \lambda) = \begin{cases} 0 & A^T \lambda + c = 0 \\
-\infty & \text{otherwise} \end{cases} \]
Since \( x = 0 \) is feasible for the primal - strong duality holds
and \( p^* = d^* \). Thus strong alternatives hold.
+ \text{ Semidefinite program:} \\
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad \sum_{i=1}^{n} x_i F_i + G \preceq 0, \quad F_i, G \in S^k.
\end{align*}

The Lagrangian variable of this SDP is a matrix $Z \in S^k_+$ \\
$L(x, Z) = c^T x + \text{tr} \left( \left( \sum_{i=1}^{n} x_i F_i + G \right) Z \right)$ \\
$= x_1 (c_1 + \text{tr}(F_1 Z)) + x_2 (c_2 + \text{tr}(F_2 Z)) + \ldots$ \\
$+ x_n (c_n + \text{tr}(F_n Z)) + \text{tr}(G Z).$
$L(x, Z) \text{ is affine in } x$ \\
$\rightarrow g(Z) = \begin{cases} \\
\text{tr}(G Z) & \text{if } c_i + \text{tr}(F_i Z) = 0 \quad \forall i = 1, \ldots, n \\
-\infty & \text{otherwise} \end{cases}$

The dual problem can then be expressed as \\
$\begin{align*}
\min & \quad \text{tr}(G Z) \\
\text{s.t.} & \quad \text{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\
& \quad Z \succeq 0.
\end{align*}$

Strong duality holds when the primal problem is strictly feasible, 
that is, \exists \tilde{x} \text{ such that} \\
$x_1 F_1 + x_2 F_2 + \ldots + x_n F_n + G \preceq 0.$