Convex Optimization

5. Duality

Prof. Ying Cui

Department of Electrical Engineering
Shanghai Jiao Tong University

2018
Outline

Lagrange dual function

Lagrange dual problem

Geometric interpretation

Optimality conditions

Perturbation and sensitivity analysis

Examples

Generalized inequalities
Lagrangian

standard form problem (not necessarily convex)

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

domain \( \mathcal{D} = \bigcap_{i=0}^m \text{dom} f_i \cap \bigcap_{i=1}^p \text{dom} h_i \) and optimal value \( p^* \)

▷ basic idea in Lagrangian duality: take the constraints into account by augmenting the objective function with a weighted sum of the constraint functions

Lagrangian: \( L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \), with \( \text{dom} \ L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \),

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

▷ weighted sum of objective and constraint functions

▷ \( \lambda_i \) is Lagrange multiplier associated with \( f_i(x) \leq 0 \)

▷ \( \nu_i \) is Lagrange multiplier associated with \( h_i(x) = 0 \)
Lagrange dual function (or dual function): \( g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \)

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

- \( g \) is concave even when problem is not convex, as it is pointwise infimum of a family of affine functions of \((\lambda, \nu)\)
  - pointwise minimum or infimum of concave functions is concave
- \( g \) can be \(-\infty\) when \( L \) is unbounded below in \( x \)
Lower bound property

The dual function yields lower bounds on the optimal value of the primal problem, i.e., for any \( \lambda \succeq 0 \) and any \( \nu \),

\[
g(\lambda, \nu) \leq p^*
\]

- the inequality holds but is vacuous when \( g(\lambda, \nu) = -\infty \)
- the dual function gives a nontrivial lower bound only when \( \lambda \succeq 0 \) and \( (\lambda, \nu) \in \text{dom} g \), i.e., \( g(\lambda, \nu) > -\infty \)
- refer to \( (\lambda, \nu) \) with \( \lambda \succeq 0 \), \( (\lambda, \nu) \in \text{dom} g \) as dual feasible

proof: Suppose \( \tilde{x} \) is feasible, i.e., \( f_i(\tilde{x}) \leq 0 \) and \( h_i(\tilde{x}) = 0 \), and \( \lambda \succeq 0 \). Then, we have

\[
\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq 0 \implies L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})
\]

Hence,

\[
g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})
\]

Minimizing over all feasible \( \tilde{x} \) gives \( p^* \geq g(\lambda, \nu) \).
Examples

Least-norm solution of linear equations

\[
\min_x \quad x^T x \\
\text{s.t.} \quad Ax = b
\]

dual function:

- to minimize \( L(x, \nu) = x^T x + \nu^T (Ax - b) \) over \( x \) (unconstrained convex problem), set gradient equal to zero:

\[
\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \Rightarrow \quad x = -(1/2)A^T \nu
\]

- plug in \( L(x, \nu) \) to obtain \( g \):

\[
g(\nu) = L((-1/2)A^T \nu, \nu) = (-1/4)\nu^T AA^T \nu - b^T \nu
\]

which is a concave quadratic function of \( \nu \), as \(-AA^T \preceq 0\) lower bound property:

\[
p^* \geq (-1/4)\nu^T AA^T \nu - b^T \nu, \quad \text{for all} \quad \nu
\]
Examples

Standard form LP

\[
\min_x \ c^T x \\
\text{s.t.} \ Ax = b, \ \ x \succeq 0
\]

dual function:

\[ L(x, \lambda, \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x \]
\[ = -b^T \nu + (c + A^T \nu - \lambda)^T x \]

is affine in \(x\) (bounded below only when identically zero)

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

lower bound property: nontrivial only when \(\lambda \succeq 0\) and \(A^T \nu - \lambda + c = 0\), and hence \(p^* \geq -b^T \nu\) if \(A^T \nu + c \succeq 0\)
Examples

Two-way partitioning problem \((W \in S^n)\)

\[
\begin{align*}
\min_x & \quad x^T W x \\
\text{s.t.} & \quad x_i^2 = 1, \quad i = 1, ..., n
\end{align*}
\]

- a nonconvex problem with \(2^n\) discrete feasible points
- find the two-way partition of \(\{1, ..., n\}\) with least total cost
  - \(W_{ij}\) is cost of assigning \(i, j\) to the same set
  - \(-W_{ij}\) is cost of assigning \(i, j\) to different sets

Dual function:

\[
\begin{align*}
g(\nu) &= \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) \\
&= \inf_x x^T (W + \text{diag}(\nu)) x - 1^T \nu = \begin{cases} -1^T \nu, & W + \text{diag}(\nu) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}
\end{align*}
\]

Lower bound property: \(p^* \geq -1^T \nu\) if \(W + \text{diag}(\nu) \succeq 0\)

Example: \(\nu = -\lambda_{\min}(W) 1\) gives bound \(p^* \geq n\lambda_{\min}(W)\)
Lagrange dual function and conjugate function

- **conjugate** $f^*$ of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$:
  \[ f^*(y) = \sup_{x \in \text{dom } f} (y^T x - f(x)) \]

- **dual function of**
  \[
  \begin{align*}
  &\min_x f_0(x) \\
  &\text{s.t. } x = 0
  \end{align*}
  \]
  
  \[ g(\nu) = \inf_x (f(x) + \nu^T x) = -\sup_x ((-\nu)^T x - f(x)) \]

- **relationship**:
  \[ g(\nu) = -f^*(-\nu) \]

- conjugate of any function is convex
- dual function of any problem is concave
Lagrange dual function and conjugate function

more generally (and more usefully), consider an optimization problem with linear inequality and equality constraints

\[
\min_x f_0(x)
\]

\[
s.t. \quad Ax \preceq b, \quad Cx = d
\]

dual function:

\[
g(\lambda, \nu) = \inf_{x \in \text{dom} f_0} \left( f_0(x) + \lambda^T (Ax - b) + \nu^T (Cx - d) \right)
\]

\[
= \inf_{x \in \text{dom} f_0} \left( f_0(x) + (A^T \lambda + C^T \nu)^T x \right) - b^T \lambda - d^T \nu
\]

\[
= -f_0^*(-A^T \lambda - C^T \nu) - b^T \lambda - d^T \nu
\]

domain of \( g \) follows from domain of \( f_0^* \):

\[
\text{dom} g = \{ (\lambda, \mu) | -A^T \lambda - C^T \nu \in \text{dom} f_0^* \}
\]

▷ simplify derivation of dual function if conjugate of \( f_0 \) is known
Examples

Equality constrained norm minimization

$$\min_x \|x\|$$

s.t. $$Ax = b$$

dual function:

$$g(\nu) = -b^T \nu - f_0^*(-A^T \nu) = \begin{cases} 
- b^T \nu, & \|A^T \nu\|_* \leq 1 \\
- \infty, & \text{otherwise}
\end{cases}$$

$$\triangleright$$ conjugate of $$f_0 = \| \cdot \|$$:

$$f_0^*(y) = \begin{cases} 
0, & \|y\|_* \leq 1 \\
\infty, & \text{otherwise}
\end{cases}$$

i.e., the indicator function of the dual norm unit ball, where $$\|y\|_* = \sup_{\|u\| \leq 1} u^T y$$ is dual norm of $$\| \cdot \|$$
Lagrange dual problem

$$\max_{\lambda, \nu} g(\lambda, \nu)$$

s.t. $\lambda \succeq 0$

- find best lower bound on $p^*$, obtained from Lagrange dual function
- always a convex optimization problem (maximize a concave function over a convex set), regardless of convexity of primal problem, optimal value denoted by $d^*$
- $\lambda$, $\nu$ are dual feasible if $\lambda \succeq 0$ and $g(\lambda, \nu) > -\infty$ (i.e., $(\lambda, \nu) \in \text{dom} g = \{(\lambda, \nu) | g(\lambda, \nu) > -\infty\}$)
- can often be simplified by making implicit constraint $(\lambda, \nu) \in \text{dom} g$ explicit, e.g.,
  - standard form LP and its dual

$$\min_x c^T x$$

s.t. $Ax = b$, $x \succeq 0$

$$\max_\nu - b^T \nu$$

s.t. $A^T \nu + c \succeq 0$$

SJTU Ying Cui 12 / 46
Weak duality and strong duality

**weak duality:** $d^* \leq p^*$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems, e.g.,
  - solving the SDP

\[
\max_{\nu} - \mathbf{1}^T \nu \\
\text{s.t. } W + \text{diag}(\nu) \succeq 0
\]

gives a lower bound for the two-way partitioning problem

**strong duality:** $d^* = p^*$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**
  - there exist many types of constraint qualifications
Slater’s constraint qualification

One simple constraint qualification is Slater’s condition (Slater’s constraint qualification): convex problem is strictly feasible, i.e., there exists an $x \in \text{int} D$ such that

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

- can be refined, e.g.,
  - can replace $\text{int} D$ with $\text{relint} D$ (interior relative to affine hull)
  - affine inequalities do not need to hold with strict inequality
  - reduce to feasibility when the constraints are all affine equalities and inequalities
- implies strong duality for convex problems
- implies that the dual value is attained when $d^* > -\infty$, i.e., there exists a dual feasible $(\lambda^*, \nu^*)$ with $g(\lambda^*, \nu^*) = d^* = p^*$
Examples

Inequality form LP

primal problem:

$$\min_x \ c^T x$$

$$s.t. \ Ax \leq b$$

dual function:

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

dual problem:

$$\max_{\lambda} \ -b^T \lambda$$

$$s.t. \ A^T \lambda + c = 0, \ \lambda \succeq 0$$

▶ from weaker form of Slater’s condition: strong duality holds for any LP provided the primal problem is feasible, implying strong duality holds for LPs if the dual is feasible

▶ in fact, \( p^* = d^* \) except when primal and dual are infeasible
Examples

**Quadratic program:** $P \in S_{++}^{n}$

$$
\begin{align*}
\min_{x} & \quad x^T Px \\
\text{subject to} & \quad Ax \preceq b
\end{align*}
$$

dual function:

$$
g(\lambda) = \inf_{x} (x^T Px + \lambda^T (Ax - b)) = -\frac{1}{4} \lambda^T AP^{-1} A^T \lambda - b^T \lambda
$$

dual problem:

$$
\begin{align*}
\max_{\lambda} & \quad -(1/4) \lambda^T AP^{-1} A^T \lambda - b^T \lambda \\
\text{subject to} & \quad \lambda \succeq 0
\end{align*}
$$

▶ from weaker form of Slater’s condition: strong duality holds provided the primal problem is feasible

▶ in fact, $p^* = d^*$ always holds
Examples

A nonconvex problem with strong duality: $A \not\succeq 0$

$$\min\limits_x x^T Ax + 2b^T x$$

$s.t.$ $x^T x \leq 1$

dual function:

$$g(\lambda) = \inf_x (x^T (A + \lambda I)x + 2b^T x - \lambda)$$

$$= \begin{cases} -b^T (A + \lambda I)^\dagger b - \lambda, & A + \lambda I \succeq 0, \ b \in \mathcal{R}(A + \lambda I) \\ -\infty, & \text{otherwise} \end{cases}$$

dual problem and equivalent SDP:

$$\max\limits_\lambda - b^T (A + \lambda I)^\dagger b - \lambda$$

$$\max\limits_{\lambda, t} - t - \lambda$$

$s.t.$ $A + \lambda I \succeq 0, \ b \in \mathcal{R}(A + \lambda I)$

$s.t.$ $\begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0$

$\blacktriangleright$ strong duality holds although primal problem is nonconvex (difficult to show)
Geometric interpretation

geometric interpretation via set of values

- set of values taken on by the constraint and objective functions: \( G = \{(f_1(x), \cdots, f_m(x), h_1(x), \cdots, h_p(x), f_0(x)) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} | x \in D \} \)

- optimal value: \( p^* = \inf \{t | (u, v, t) \in G, u \leq 0, v = 0 \} \)

- dual function: \( g(\lambda, \nu) = \inf \{(\lambda, \nu, 1)^T (u, v, t) | (u, v, t) \in G \} \)
  
  - if the infimum is finite, then \( (\lambda, \nu, 1)^T (u, v, t) \geq g(\lambda, \nu) \)
  
  defines a nonvertical supporting hyperplane to \( G \)

- weak duality: for all \( \lambda \geq 0 \),

\[
p^* = \inf \{t | (u, v, t) \in G, u \leq 0, v = 0 \} \\
\geq \inf \{(\lambda, \nu, 1)^T (u, v, t) | (u, v, t) \in G, u \leq 0, v = 0 \} \\
\geq \inf \{(\lambda, \nu, 1)^T (u, v, t) | (u, v, t) \in G \} \\
= g(\lambda, \nu)
\]
Geometric interpretation

Example

consider a simple problem with one constraint

$$\min_x f_0(x) \quad \quad p^* = \inf \{ t | (u, t) \in \mathcal{G}, u \leq 0 \}$$

$$\text{s.t.} \quad f_1(x) \leq 0 \quad \quad g(\lambda) = \inf_{(u, t) \in \mathcal{G}} (t + \lambda u)$$

where $\mathcal{G} = \{(f_1(x), f_0(x)) | x \in \mathcal{D} \}$

- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- hyperplane intersects $t$-axis at $t = g(\lambda)$

Figure 5.3 Geometric interpretation of dual function and lower bound $g(\lambda) \leq p^*$, for a problem with one (inequality) constraint. Given $\lambda$, we minimize $(\lambda, 1)^T (u, t)$ over $\mathcal{G} = \{(f_1(x), f_0(x)) | x \in \mathcal{D} \}$. This yields a supporting hyperplane with slope $-\lambda$. The intersection of this hyperplane with the $u = 0$ axis gives $g(\lambda)$.

Figure 5.4 Supporting hyperplanes corresponding to three dual feasible values of $\lambda$, including the optimum $\lambda^*$. Strong duality does not hold; the optimal duality gap $p^* - d^*$ is positive.
Geometric interpretation

geometric interpretation via epigraph

- epigraph form of \( G \): \( \mathcal{A} = G + (\mathbb{R}_+^m \times \{0\} \times \mathbb{R}_+) \)
  \[ = \{(u, v, t)|\exists x \in \mathcal{D}, f_i(x) \leq u_i, i = 1, \cdots, m, h_i(x) = v_i, i = 1, \cdots, p, f_0(x) \leq t\} \]
  includes all points with larger objective or inequality constraint function values

- optimal value: \( p^* = \inf\{t|(0, 0, t) \in \mathcal{A}\} \)

- dual function: if \( \lambda \succeq 0 \), then
  \[ g(\lambda, \nu) = \inf\{(\lambda, \nu, 1)^T(u, v, t)|(u, v, t) \in \mathcal{A}\} \]
  if the infimum is finite, then \( (\lambda, \nu, 1)^T(u, v, t) \geq g(\lambda, \nu) \)
  defines a nonvertical supporting hyperplane to \( \mathcal{A} \)

- weak duality: \( p^* = (\lambda, \nu, 1)^T(0, 0, p^*) \geq g(\lambda, \nu) \)

- strong duality: holds iff there exists a nonvertical supporting hyperplane to \( \mathcal{A} \) at its boundary point \((0, 0, p^*)\)
  - for convex problem, \( \mathcal{A} \) is convex, hence has a supporting hyperplane at \((0, 0, p^*)\)
  - Slater’s condition guarantees the supporting hyperplane to be nonvertical
Geometric interpretation

Example

Consider a simple problem with one constraint

\[
\begin{align*}
\min_x & \quad f_0(x) \\
\text{s.t.} & \quad f_1(x) \leq 0
\end{align*}
\]

\[
p^* = \inf \{ t | (0, t) \in \mathcal{A} \}
\]

\[
g(\lambda) = \inf \{ (\lambda, 1)^T (u, t) | (u, t) \in \mathcal{A} \}
\]

Where \( \mathcal{A} = \{ (u, t) | \exists x \in \mathcal{D}, f_1(x) \leq u, f_0(x) \leq t \} \)

\[
\lambda u + t = g(\lambda)
\]

\[
(0, p^*)
\]

\[
(0, g(\lambda))
\]

\[
(0, g(\lambda))
\]

Figure 5.5 Geometric interpretation of dual function and lower bound \( g(\lambda) \leq p^* \), for a problem with one (inequality) constraint. Given \( \lambda \), we minimize \( (\lambda, 1)^T (u, t) \) over \( \mathcal{A} = \{ (u, t) \mid \exists x \in \mathcal{D}, f_0(x) \leq t, f_1(x) \leq u \} \). This yields a supporting hyperplane with slope \( -\lambda \). The intersection of this hyperplane with the \( u = 0 \) axis gives \( g(\lambda) \).
Certificate of suboptimality and stopping criteria

do not assume the primal problem is convex, and let \( x \) and \((\lambda, \nu)\) be a primal feasible point and a dual feasible point, respectively

- \((\lambda, \nu)\) provides a proof or certificate that \( p^* \geq g(\lambda, \nu) \)
- \((\lambda, \nu)\) bounds how suboptimal \( x \) is without knowing \( p^* \):

\[
 f_0(x) - p^* \leq f_0(x) - g(\lambda, \nu)
\]

- provide nonheuristic stopping criteria in optimization alg
- \( x, (\lambda, \nu) \) localizes \( p^*, d^* \) to an interval:

\[
 p^*, d^* \in [g(\lambda, \nu), f_0(x)]
\]

with the width being the duality gap \( f_0(x) - g(\lambda, \nu) \) associated with \( x \) and \((\lambda, \nu)\)

- if \( f_0(x) = g(\lambda, \nu) \), then \( x \) is primal optimal and \((\lambda, \nu)\) is dual optimal
  - \((\lambda, \nu)\) is a certificate that proves \( x \) is optimal
  - \( x \) is a certificate that proves \((\lambda, \nu)\) is dual optimal
Complementary slackness

Let $x^*$ and $(\lambda^*, \nu^*)$ be any primal optimal and dual optimal points. Assume strong duality holds. Then,

$$f_0(x^*) = g(\lambda^*, \nu^*) = \inf_x \left( f_0(x) + \sum_{i=1}^m \lambda^*_i f_i(x) + \sum_{i=1}^p \nu^*_i h_i(x) \right)$$

$$\leq f_0(x^*) + \sum_{i=1}^m \lambda^*_i f_i(x^*) + \sum_{i=1}^p \nu^*_i h_i(x^*)$$

$$\leq f_0(x^*)$$

Hence, the two inequalities hold with equality implying:

- $x^*$ minimizes $L(x, \lambda^*, \nu^*)$ over $x$ ($L(x, \lambda^*, \nu^*)$ can have other minimizers)
- complementary slackness: $\lambda^*_i f_i(x^*) = 0$, $i = 1, \ldots, m$, i.e.,
  $$\lambda^*_i > 0 \Rightarrow f_i(x^*) = 0, \quad f_i(x^*) < 0 \Rightarrow \lambda^*_i = 0$$

$\lambda^*_i = 0$ unless the $i$th constraint is active at the optimum
Karush-Kuhn-Tucker (KKT) conditions

Consider any optimization problem with differentiable objective and constraint functions.

The following four conditions are called KKT conditions:

◮ primal constraints:

\[ f_i(x) \leq 0, \quad i = 1, \ldots, m, \quad h_i(x) = 0, \quad i = 1, \ldots, p \]

◮ dual constraints:

\[ \lambda \succeq 0 \]

◮ complementary slackness:

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

◮ gradient of \( L(x, \lambda, \nu) \) with respect to \( x \) vanishes:

\[ \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0 \]
Karush-Kuhn-Tucker (KKT) conditions

consider any optimization problem with differentiable objective and constraint functions

**KKT conditions for nonconvex/convex problems**

- for any optimization problem, if strong duality holds, any pair of primal and dual optimal points \( x^* , (\lambda^* , \nu^* ) \) must satisfy the KKT conditions

- proof: The first and second conditions hold obviously. The third condition is shown on page 23. The fourth condition follows from the fact that \( x^* = \arg \min_x L(x, \lambda^* , \nu^* ) \) (shown on page 23) and \( L(x, \lambda^* , \nu^* ) \) is differentiable.
Karush-Kuhn-Tucker (KKT) conditions

KKT conditions for convex problems

- for any convex optimization problem, any points $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap

  - proof: The first and second conditions state that $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primary and dual feasible, respectively. By noting that $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in $x$ (as $\tilde{\lambda} \succeq 0$), the fourth condition implies that $\tilde{x} = \arg\min_x L(x, \tilde{\lambda}, \tilde{\nu})$, i.e.,

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

    where the last equality is due to the first and third conditions.

    $g(\tilde{\lambda}, \tilde{\nu}) = f_0(\tilde{x})$ means zero duality gap, implying that $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal and dual optimal.

  - if a convex optimization problem satisfies Slater’s condition, then the KKT conditions provide necessary and sufficient conditions for optimality, i.e.,

    - $x$ is optimal iff there are $(\lambda, \nu)$ that, together with $x$, satisfy the KKT conditions
Karush-Kuhn-Tucker (KKT) conditions

KKT conditions play an important role in optimization

- In a few special cases, it is possible to solve the KKT conditions (and therefore, the optimization problem) analytically.

- More generally, many algorithms for convex optimization are conceived as, or can be interpreted as, methods for solving the KKT conditions.
Example

**water-filling** (assume $\alpha_i > 0$)

$$\min_x - \sum_{i=1}^{n} \log(x_i + \alpha_i)$$

s.t. $x \succeq 0, \quad 1^T x = 1$

$x$ is optimal iff $x \succeq 0, \quad 1^T x = 1$, and there exist $\lambda \in \mathbb{R}^n, \nu \in \mathbb{R}$ such that

$$\lambda \succeq 0, \quad \lambda_i x_i = 0, \quad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine $\nu$ from $1^T x = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$

thus, the optimal point is given by

$$x_i^* = \max\{0, 1/\nu^* - \alpha_i\}$$

where $\nu^*$ satisfies $\sum_{i=1}^{n} \max\{0, 1/\nu^* - \alpha_i\} = 1$
Example

interpretation:

▶ $n$ patches; level of patch $i$ is at height $a_i$
▶ flood area with unit amount of water
▶ resulting level is $1/\nu^*$
▶ depth of water above patch $i$ is $x_i^*$

Figure 5.7 Illustration of water-filling algorithm. The height of each patch is given by $a_i$. The region is flooded to a level $1/\nu^*$ which uses a total quantity of water equal to one. The height of the water (shown shaded) above each patch is the optimal value of $x_i^*$. 
Perturbation and sensitivity analysis

( unperturbed) optimization problem and its dual

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

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

perturbed problem and its dual

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

\[
\begin{align*}
\max_{\lambda, \nu} & \quad g(\lambda, \nu) - u^T \lambda - v^T \nu \\
\text{s.t.} & \quad \lambda \succeq 0
\end{align*}
\]
Perturbation and sensitivity analysis

- $x$ is primal variable, and $u$, $v$ are parameters
  - tighten ($u_i < 0$) or relax ($u_i > 0$) $i$th inequality constraint by $u_i$
  - change the right-hand side of $i$th equality constraints by $v_i$
- $p^*(u, v)$ is optimal value of perturbed problem, as a function of perturbations to the right-hand sides of the constraints
  - $p^*(0, 0) = p^*$
  - when $p^*(u, v) = \infty$, perturbations of the constraints result in infeasibility
  - when unperturbed problem is convex, $p^*(u, v)$ is a convex function of $u$ and $v$
- interested in information about $p^*(u, v)$ obtained from solution of unperturbed problem and its dual
Perturbation and sensitivity analysis

**global sensitivity**
Assume that strong duality holds for unperturbed problem, and that the dual optimum is attained. Let $\lambda^*, \nu^*$ be dual optimal for unperturbed problem. Then for all $u$ and $v$,

$$p^*(u, v) \geq p^*(0, 0) - u^T \lambda^* - v^T \nu^*$$

global sensitivity interpretation:

- large $\lambda_i^*$: $p^*$ increases greatly if tightening constraint $i$ ($u_i < 0$)
- small $\lambda_i^*$: $p^*$ does not decrease much if loosening constraint $i$ ($u_i > 0$)
- large and positive $\nu_i^*$: $p^*$ increases greatly if taking $v_i < 0$
  - large and negative $\nu_i^*$: $p^*$ increases greatly if taking $v_i > 0$
- small and positive $\nu_i^*$: $p^*$ does not decrease much if taking $v_i > 0$
  - small and negative $\nu_i^*$: $p^*$ does not decrease much if taking $v_i < 0$
Perturbation and sensitivity analysis

proof: apply weak duality to perturbed problem and then strong duality to the unperturbed problem

\[ p^*(u, v) \geq g(\lambda^*, \nu^*) - u^T \lambda^* - \nu^T \nu^* = p^*(0, 0) - u^T \lambda^* - \nu^T \nu^* \]

element: \( p^*(u) \) for a problem with one inequality constraint:

![Figure 5.10 Optimal value \( p^*(u) \) of a convex problem with one constraint \( f_1(x) \leq u \), as a function of \( u \). For \( u = 0 \), we have the original unperturbed problem; for \( u < 0 \) the constraint is tightened, and for \( u > 0 \) the constraint is loosened. The affine function \( p^*(0) - \lambda^* u \) is a lower bound on \( p^* \).]
Perturbation and sensitivity analysis

**local sensitivity**
If (in addition) \( p^*(u, v) \) is differentiable at \((0, 0)\), then

\[ \lambda_i^* = -\frac{\partial p^*(0, 0)}{\partial u_i}, \quad \nu_i^* = -\frac{\partial p^*(0, 0)}{\partial v_i} \]

local sensitivity interpretation:

- optimal Lagrange multipliers are exactly the local sensitivities of the optimal value with respect to constraint perturbations
- tightening (loosening) \( i \)th inequality constraint a small amount yields an increase (a decrease) in \( p^* \) of approximately \(-\lambda_i^* u_i\) (\(\lambda_i^* u_i\))
- local sensitivity result gives us a quantitative measure of how active a constraint is at the optimum \( x^* \)
  - \( f_i(x^*) < 0 \): constraint can be tightened or loosened a small amount without affecting the optimal value, as \( \lambda_i^* = 0 \)
  - \( f_i(x^*) = 0 \): small (large) \( \lambda_i^* \) means that constraint can be loosened or tightened a bit without much (with great) effect on the optimal value
Perturbation and sensitivity analysis

proof (for $\lambda_i^*$): choosing $u = te_i$ and $v = 0$, from global sensitivity result,

\[
\begin{align*}
\frac{p^*(te_i,0) - p^*(0,0)}{t} &\geq -\lambda_i^*, \quad t > 0 \\
\frac{p^*(te_i,0) - p^*(0,0)}{t} &\leq -\lambda_i^*, \quad t < 0 
\end{align*}
\]

\[
\lim_{t \to 0^+} \frac{p^*(te_i,0) - p^*(0,0)}{t} \geq -\lambda_i^* \\
\lim_{t \to 0^-} \frac{p^*(te_i,0) - p^*(0,0)}{t} \leq -\lambda_i^*
\]

Thus, $\frac{\partial p^*(0,0)}{\partial u_i} = -\lambda_i^*$. 

Duality and problem reformulations

- equivalent formulations of a problem can lead to very different dual problems
- reformulating the primal problem can be useful when the dual problem is difficult to derive, or uninteresting

**common reformulations**

- introduce new variables and associated equality constraints
- replacing the objective with an increasing function of the original objective
- make explicit constraints implicit (i.e., incorporating them into the domain of objective) or vice-versa
Introducing new variables and equality constraints

**unconstrained problem:** \( \min_x f_0(Ax + b) \)

- dual function is constant: \( g = \inf_x f_0(Ax + b) = p^* \)
- strong duality holds, i.e., \( p^* = d^* \), but dual is not useful

reformulated problem and its dual:

\[
\begin{align*}
\min_{x,y} & \quad f_0(y) \\
\max & \quad b^T \nu - f_0^*(\nu) \\
\text{s.t.} & \quad Ax + b - y = 0 \\
& \quad A^T \nu = 0
\end{align*}
\]

dual function follows from

\[
g(\nu) = \inf_{x,y} (f_0(y) + \nu^T (Ax + b - y)) = \inf_{x,y} (f_0(y) - \nu^T y + \nu^T Ax) + b^T \nu
\]

\[
= \begin{cases} 
\inf_y (f_0(y) - \nu^T y) + b^T \nu, & A^T \nu = 0 \\
-\infty, & \text{otherwise}
\end{cases}
\]

\[
= \begin{cases} 
-f_0^*(\nu) + b^T \nu, & A^T \nu = 0 \\
-\infty, & \text{otherwise}
\end{cases}
\]
Introducing new variables and equality constraints

**minimum norm problem:** \( \min_x ||Ax - b|| \)

- dual function is constant: \( g = \inf_x ||Ax - b|| = p^* \)
- strong duality holds, i.e., \( p^* = d^* \), but dual is not useful

reformulated problem and its dual:

\[
\begin{align*}
\min_{x, y} & \quad ||y|| \\
\text{s.t.} & \quad y = Ax - b
\end{align*}
\]

\[
\begin{align*}
\max_{\nu} & \quad b^T \nu \\
\text{s.t.} & \quad A^T \nu = 0, \quad ||\nu||^* \leq 1
\end{align*}
\]

Dual function follows from

\[
g(\nu) = \inf_{x, y} (||y|| + \nu^T (y - Ax + b))
\]

\[
= \begin{cases} 
\inf_y (||y|| + \nu^T y) + b^T \nu, & A^T \nu = 0 \\
-\infty, & \text{otherwise}
\end{cases}
\]

\[
= \begin{cases} 
b^T \nu, & A^T \nu = 0, \quad ||\nu||^* \leq 1 \\
-\infty, & \text{otherwise}
\end{cases}
\]
Transforming the objective

replacing the objective with an increasing function of the original objective

**minimum norm problem:** $\min_x \|Ax - b\|$

reformulated problem and its dual:

$$\begin{align*}
\min_{x, y} & \quad \frac{1}{2}\|y\|^2 \\
\max_{\nu} & \quad -\frac{1}{2}\|\nu\|_*^2 + b^T \nu \\
\text{s.t.} & \quad y = Ax - b \\
\text{s.t.} & \quad A^T \nu = 0
\end{align*}$$

dual function follows from

$$g(\nu) = \inf_{x, y} \left(\frac{1}{2}\|y\|^2 + \nu^T (y - Ax + b)\right)$$

$$= \begin{cases} 
\inf_y (\frac{1}{2}\|y\|^2 + \nu^T y) + b^T \nu, & A^T \nu = 0 \\
-\infty, & \text{otherwise}
\end{cases}$$

$$= \begin{cases} 
-\frac{1}{2}\|\nu\|_*^2 + b^T \nu, & A^T \nu = 0 \\
-\infty, & \text{otherwise}
\end{cases}$$

last inequality: conjugate of $\frac{1}{2}\|\cdot\|^2$ is $\frac{1}{2}\|\cdot\|_*^2$ (Ex.3.27, pp.93)
Implicit constraints

make explicit constraints implicit (i.e., incorporating them into the domain of objective) or vice-versa

**LP with box constraints:** primal and dual problem

\[
\begin{align*}
\min_x & \quad c^T x \\
\text{s.t.} & \quad Ax = b \\
& \quad -1 \leq x \leq 1
\end{align*}
\]

\[
\begin{align*}
\max_{\lambda, \nu} & \quad -b^T \nu - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2 \\
\text{s.t.} & \quad c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\
& \quad \lambda_1 \geq 0, \quad \lambda_2 \geq 0
\end{align*}
\]

reformulated problem with box constraints made implicit and its dual:

\[
\begin{align*}
\max_x & \quad f_0(x) = \begin{cases} 
  c^T x, & -1 \leq x \leq 1 \\
  \infty, & \text{otherwise}
\end{cases} \\
\text{s.t.} & \quad Ax = b
\end{align*}
\]

dual function follows from:

\[
\begin{align*}
g(\nu) = \inf_{-1 \leq x \leq 1} (c^T x + \nu^T (A x - b)) &= -b^T \nu - \|A^T \nu + c\|_1
\end{align*}
\]
Problems with generalized inequality constraints

do not assume convexity of problem

\[ p^* \triangleq \min_x f_0(x) \]

\[ \text{s.t. } f_i(x) \preceq_{K_i} 0, \ i = 1, \ldots, m \]

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

\( K_i \subseteq R^{k_i} \) is a proper cone; \( \preceq_{K_i} \) is a generalized inequality on \( R^{k_i} \)

- Lagrange multiplier vector associated with \( f_i(x) \preceq_{K_i} 0 \):
  \( \lambda_i \in R^{k_i} \), Lagrange multiplier associated with \( h_i(x) = 0 \):
  \( \nu_i \in R \)

- Lagrangian \( L : R^n \times R^{k_1} \times \ldots \times R^{k_m} \times R^p \rightarrow R \):

  \[
  L(x, \lambda_1, \ldots, \lambda_m, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
  \]

- (concave) dual function \( g : R^{k_1} \times \ldots \times R^{k_m} \times R^p \rightarrow R \):

  \[
  g(\lambda_1, \ldots, \lambda_m, \nu) = \inf_{x \in D} L(x, \lambda_1, \ldots, \lambda_m, \nu)
  \]
Lower bound property

The dual function yields lower bounds on the optimal value of the primal problem, i.e., for any $\lambda_i \succeq_{K_i} 0$ and any $\nu$,

$$p^* \geq g(\lambda_1, \ldots, \lambda_m, \nu)$$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq_{K_i} 0$, then

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

$$\geq \inf_{x \in D} L(x, \lambda_1, \ldots, \lambda_m, \nu) = g(\lambda_1, \ldots, \lambda_m, \nu)$$

where the first inequality follows from the definition of the dual cone. Minimizing over all feasible $\tilde{x}$ gives $p^* \geq g(\lambda_1, \ldots, \lambda_m, \nu)$.

dual problem:

$$d^* \triangleq \max_{\lambda_1, \ldots, \lambda_m, \nu} g(\lambda_1, \ldots, \lambda_m, \nu)$$

s.t. $\lambda_i \succeq_{K_i} 0, \quad i = 1, \ldots, m$
Weak duality and strong duality

**weak duality:** $d^* \leq p^*$
- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

**strong duality:** $d^* = p^*$
- does not hold in general
- holds for convex problem with constraint qualification, e.g.,
  - Slater’s condition: primal problem is strictly feasible
Examples

Semidefinite program
primal SDP \((F_i, G \in S^k, \text{positive semidefinite cone } K_1 = S^k_+)\):
\[
\begin{align*}
\min_x & \quad c^T x \\
\text{s.t.} & \quad x_1 F_1 + \ldots + x_n F_n \preceq G
\end{align*}
\]

Lagrange multiplier is matrix \(Z \in S^k\) and Lagrangian is
\[
L(x, Z) = c^T x + \text{tr}(Z(x_1 F_1 + \ldots + x_n F_n - G))
\]
\[
= x_1(c_1 + \text{tr}(F_1 Z)) + \cdots + x_n(c_n + \text{tr}(F_n Z)) - \text{tr}(GZ)
\]

dual function
\[
g(Z) = \inf_x L(x, Z) = \begin{cases}
-\text{tr}(GZ) & \text{tr}(F_i Z) + c_i = 0, \ i = 1, \ldots, n \\
-\infty & \text{otherwise}
\end{cases}
\]
dual SDP:
\[
\begin{align*}
\max_Z & \quad -\text{tr}(GZ) \\
\text{s.t.} & \quad Z \succeq 0, \ \text{tr}(F_i Z) + c_i = 0, \ i = 1, \ldots, n
\end{align*}
\]
Examples

Cone program in standard form
primal CP: (proper cone $K \subseteq \mathbb{R}^n$):

$$\min_x c^T x$$

s.t. $x \succeq_K 0$, $Ax = b$

- Lagrange multipliers $\lambda \in \mathbb{R}^n$, $\nu \in \mathbb{R}^m$ and Lagrangian is

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

- dual function

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

dual SDP:

$$\max_{\nu} -b^T \nu$$

s.t. $A^T \nu + c \succeq_{K^*} 0$
KKT conditions

differentiable \( f_i, h_i \)

- primal constraints:
  \( f_i(x) \leq K_i \ 0, \ i = 1, \ldots, m, \ h_i(x) = 0, \ i = 1, \ldots, p \)

- dual constraints: \( \lambda \succeq K_i^* \ 0 \)

- complementary slackness: \( \lambda_i^T f_i(x) = 0, \ i = 1, \ldots, m, \) implying
  \( \lambda_i \succ K_i^* \ 0 \Rightarrow f_i(x) = 0, \ f_i(x) \prec K_i \ 0 \Rightarrow \lambda_i = 0 \)

- gradient of Lagrangian with respect to \( x \) vanishes:
  \[
  \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i^T \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0
  \]

**KKT conditions for nonconvex/convex problems**

if strong duality holds, any primal optimal and any dual optimal must satisfy the KKT conditions

**KKT conditions for convex problems**

if strong duality holds, the KKT conditions provide necessary and sufficient conditions for optimality