5. Duality

- Lagrange dual problem
- weak and strong duality
- geometric interpretation
- optimality conditions
- perturbation and sensitivity analysis
- examples
- generalized inequalities
Lagrangian

**Standard form problem** (not necessarily convex)

```
minimize \( f_0(x) \)
subject to \( f_i(x) \leq 0, \quad i = 1, \ldots, m \)
\( h_i(x) = 0, \quad i = 1, \ldots, p \)
```

variable \( x \in \mathbb{R}^n \), domain \( D \), optimal value \( p^* \)

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

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

- 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

Lagrange dual function: \( g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}, \)

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

- a concave function of \( \lambda, \nu \)
- can be \(-\infty\) for some \( \lambda, \nu \); this defines the domain of \( g \)

**Lower bound property:** if \( \lambda \geq 0 \), then \( g(\lambda, \nu) \leq p^* \)

proof: if \( x \) is feasible and \( \lambda \geq 0 \), then

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

minimizing over all feasible \( x \) gives \( p^* \geq g(\lambda, \nu) \)
Least norm solution of linear equations

\[
\begin{align*}
\text{minimize} & \quad x^T x \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

- Lagrangian is
  \[ L(x, \nu) = x^T x + \nu^T (Ax - b) \]

- to minimize \( L \) over \( x \), set gradient equal to zero:
  \[
  \nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \implies \quad x = -\frac{1}{2} A^T \nu
  \]

- plug in in \( L \) to obtain \( g \):
  \[
  g(\nu) = L(-\frac{1}{2} A^T \nu, \nu) = -\frac{1}{4} \nu^T A A^T \nu - b^T \nu
  \]
  a concave function of \( \nu \)

**Lower bound property:** \( p^* \geq -\frac{1}{4} \nu^T A A^T \nu - b^T \nu \) for all \( \nu \)
Standard form LP

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
\]

- Lagrangian is

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

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

- \( L \) is affine in \( x \), hence

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

\( g \) is linear on affine domain \( \text{dom } g = \{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\} \), hence concave

**Lower bound property:** \( p^* \geq -b^T \nu \) if \( A^T \nu + c \geq 0 \)
Equality constrained norm minimization

\[
\begin{align*}
\text{minimize} & \quad \|x\| \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

- \( \| \cdot \| \) is any norm; dual norm is defined as
  \[
  \| v \|_* = \sup_{\|u\| \leq 1} u^T v
  \]

- define Lagrangian \( L(x, \nu) = \|x\| + \nu^T (b - Ax) \)

- dual function (proof on next page):
  \[
  g(\nu) = \inf_x (\|x\| - \nu^T Ax + b^T \nu)
  = \begin{cases} 
    b^T \nu & \|A^T \nu\|_* \leq 1 \\
    -\infty & \text{otherwise}
  \end{cases}
  \]

Lower bound property: \( p^* \geq b^T \nu \) if \( \|A^T \nu\|_* \leq 1 \)
proof of expression for $g$: follows from

$$\inf_x (\|x\| - y^T x) = \begin{cases} 0 & \|y\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases} \quad (1)$$

**Case $\|y\|_* \leq 1$:**

$$\inf_x (\|x\| - y^T x) = 0$$

- $y^T x \leq \|x\| \|y\|_* \leq \|x\|$ for all $x$ (by definition of dual norm)
- $y^T x = \|x\|$ for $x = 0$

**Case $\|y\|_* > 1$:**

$$\inf_x (\|x\| - y^T x) = -\infty$$

- there exists an $\tilde{x}$ with $\|\tilde{x}\| \leq 1$ and $y^T \tilde{x} = \|y\|_* > 1$; hence $\|\tilde{x}\| - \|y\|_* < 0$
- consider $x = t\tilde{x}$ with $t > 0$:

$$\|x\| - y^T x = t (\|\tilde{x}\| - \|y\|_*) \to -\infty \text{ as } t \to \infty$$

Duality 5.7
Two-way partitioning

minimize $x^TWx$
subject to $x_i^2 = 1, \quad i = 1, \ldots, n$

- a nonconvex problem; feasible set $\{-1, 1\}^n$ contains $2^n$ discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets, $x_i \in \{-1, 1\}$ is assignment for $i$
- cost function is

$$x^TWx = \sum_{i=1}^{n} W_{ii} + 2 \sum_{i>j} W_{ij} x_i x_j$$

$$= 1^T W 1 + 2 \sum_{i>j} W_{ij} (x_i x_j - 1)$$

cost of assigning $i, j$ to different sets is $-4W_{ij}$
Lagrange dual of two-way partitioning problem

Dual function

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

\[
= \inf_x x^T (W + \text{diag}(\nu)) x - 1^T \nu
\]

\[
= \begin{cases} 
-1^T \nu & W + \text{diag}(\nu) \succeq 0 \\
-\infty & \text{otherwise}
\end{cases}
\]

Lower bound property

\[ p^* \geq -1^T \nu \quad \text{if } W + \text{diag}(\nu) \succeq 0 \]

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

minimize \( f_0(x) \)
subject to \( Ax \leq b \)
\( Cx = d \)

Dual function

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

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

• recall definition of conjugate \( f^*(y) = \sup_x (y^T x - f(x)) \)
• simplifies derivation of dual if conjugate of \( f_0 \) is known

Example: entropy maximization

\[f_0(x) = \sum_{i=1}^n x_i \log x_i, \quad f_0^*(y) = \sum_{i=1}^n e^{y_i-1}\]
The dual problem

Lagrange dual problem

maximize \( g(\lambda, \nu) \)
subject to \( \lambda \geq 0 \)

- finds best lower bound on \( p^* \), obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted by \( d^* \)
- often simplified by making implicit constraint \((\lambda, \nu) \in \text{dom } g\) explicit
- \( \lambda, \nu \) are dual feasible if \( \lambda \geq 0, (\lambda, \nu) \in \text{dom } g \)
- \( d^* = -\infty \) if problem is infeasible; \( d^* = +\infty \) if unbounded above

Example: standard form LP and its dual (page 5.5)

minimize \( c^T x \)
subject to \( Ax = b \)
\( x \geq 0 \)

maximize \( -b^T \nu \)
subject to \( A^T \nu + c \geq 0 \)
Weak 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
  for example, solving the SDP

$$\begin{align*}
\text{maximize} & \quad -1^Tv \\
\text{subject to} & \quad W + \text{diag}(v) \succeq 0
\end{align*}$$

... gives a lower bound for the two-way partitioning problem on page 5.8

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

- does not hold in general
- (usually) holds for convex problems
- sufficient conditions that guarantee strong duality in convex problems are called constraint qualifications
Slater’s constraint qualification

Convex problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

Slater’s constraint qualification: the problem is strictly feasible, \(i.e.,\)

\[
\exists x \in \text{int } \mathcal{D} : \quad f_i(x) < 0, \quad i = 1, \ldots, m, \quad Ax = b
\]

- guarantees strong duality: \(p^* = d^*\)
- also guarantees that the dual optimum is attained if \(p^* > -\infty\)
- can be sharpened: \(e.g.,\) can replace \(\text{int } \mathcal{D}\) with \(\text{relint } \mathcal{D}\) (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, \(\ldots\)
- there exist many other types of constraint qualifications
**Inequality form LP**

**Primal problem**

minimize $c^T x$

subject to $Ax \leq b$

**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**

maximize $-b^T \lambda$

subject to $A^T \lambda + c = 0$

$\lambda \geq 0$

- from Slater’s condition: $p^* = d^*$ if $A\tilde{x} < b$ for some $\tilde{x}$
- in fact, $p^* = d^*$ except when primal and dual are infeasible ($p^* = \infty$, $d^* = -\infty$)
Quadratic program

**Primal problem** (assume $P \in S_{++}^n$)

minimize $x^T P x$

subject to $A x \leq b$

**Dual function**

$$g(\lambda) = \inf_x (x^T P x + \lambda^T (A x - b)) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$$

**Dual problem**

maximize $-\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda$

subject to $\lambda \geq 0$

- from Slater's condition: $p^* = d^*$ if $A \tilde{x} < b$ for some $\tilde{x}$
- in fact, $p^* = d^*$ always
A nonconvex problem with strong duality

\[
\begin{align*}
\text{minimize} & \quad x^T A x + 2b^T x \\
\text{subject to} & \quad x^T x \leq 1
\end{align*}
\]

we allow \( A \neq 0 \), hence problem may be nonconvex

**Dual function** (derivation on next page)

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

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

**Dual problem** and equivalent SDP:

\[
\begin{align*}
\text{maximize} & \quad -b^T (A + \lambda I)^\dagger b - \lambda \\
\text{subject to} & \quad A + \lambda I \succeq 0 \\
& \quad b \in \mathcal{R}(A + \lambda I) \\
& \quad \lambda \geq 0
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad -t - \lambda \\
\text{subject to} & \quad \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0 \\
& \quad \lambda \geq 0
\end{align*}
\]

strong duality holds although primal problem is not convex (not easy to show)
proof of expression for $g$: unconstrained minimum of $f(x) = x^T P x + 2 q^T x + r$ is

$$\inf_x f(x) = \begin{cases} 
-q^T P^{-1} q + r & P > 0 \\
-q^T P^\dagger q + r & P \neq 0, P \geq 0, q \in \mathcal{R}(P) \\
-\infty & P \geq 0, q \notin \mathcal{R}(P) \\
-\infty & P \neq 0 
\end{cases}$$

- if $P \neq 0$, function $f$ is unbounded below: choose $y$ with $y^T P y < 0$ and $x = ty$

$$f(x) = t^2 (y^T P y) + 2t (q^T y) + r \to -\infty \quad \text{if } t \to \pm \infty$$

- if $P \geq 0$, decompose $q$ as $q = Pu + v$ with $u = P^\dagger q$ and $v = (I - PP^\dagger)q$

  $Pu$ is projection of $q$ on $\mathcal{R}(P)$, $v$ is projection on nullspace of $P$

- if $v \neq 0$ (i.e., $q \notin \mathcal{R}(P)$), the function $f$ is unbounded below: for $x = -tv$,

$$f(x) = t^2 (v^T P v) - 2t (q^T v) + r = -2t \|v\|^2 + r \to -\infty \quad \text{if } t \to \infty$$

- if $v = 0$, $x^* = -u$ is optimal since $f$ is convex and $\nabla f(x^*) = 2 P x^* + 2q = 0$;

$$f(x^*) = -q^T P^\dagger q + r$$
Geometric interpretation of duality

for simplicity, consider problem with one constraint $f_1(x) \leq 0$

Interpretation of dual function

$$g(\lambda) = \inf_{(u,t) \in \mathcal{G}} (t + \lambda u), \quad \text{where} \quad \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$$

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

**Epigraph variation:** same interpretation if $G$ is replaced with

$$\mathcal{A} = \{(u, t) \mid f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in \mathcal{D}\}$$

**Strong duality**

- holds if there is a non-vertical supporting hyperplane to $\mathcal{A}$ at $(0, p^*)$
- for convex problem, $\mathcal{A}$ is convex, hence has supporting hyperplane at $(0, p^*)$
- Slater's condition: if there exist $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ with $\tilde{u} < 0$, then supporting hyperplanes at $(0, p^*)$ must be non-vertical
Optimality conditions

if strong duality holds, then \( x \) is primal optimal and \((\lambda, \nu)\) is dual optimal if:

1. \( f_i(x) \leq 0 \) for \( i = 1, \ldots, m \) and \( h_i(x) = 0 \) for \( i = 1, \ldots, p \)
2. \( \lambda \geq 0 \)
3. \( f_0(x) = g(\lambda, \nu) \)

conversely, these three conditions imply optimality of \( x \), \((\lambda, \nu)\), and strong duality

next, we replace condition 3 with two equivalent conditions that are easier to use
Complementary slackness

assume $x$ satisfies the primal constraints and $\lambda \geq 0$

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

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

$$\leq f_0(x)$$

equality $f_0(x) = g(\lambda, \nu)$ holds if and only if the two inequalities hold with equality:

- first inequality: $x$ minimizes $L(\tilde{x}, \lambda, \nu)$ over $\tilde{x} \in \mathcal{D}$
- 2nd inequality: $\lambda_i f_i(x) = 0$ for $i = 1, \ldots, m$, i.e.,

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

this is known as complementary slackness
Optimality conditions

if strong duality holds, then $x$ is primal optimal and $(\lambda, \nu)$ is dual optimal if:

1. $f_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, \ldots, p$
2. $\lambda \geq 0$
3. $\lambda_i f_i(x) = 0$ for $i = 1, \ldots, m$
4. $x$ is a minimizer of $L(\cdot, \lambda, \nu)$

conversely, these four conditions imply optimality of $x$, $(\lambda, \nu)$, and strong duality
if problem is convex and the functions $f_i, h_i$ are differentiable, #4 can written as

4'. the gradient of the Lagrangian with respect to $x$ vanishes:

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

conditions 1,2,3,4' are known as Karush–Kuhn–Tucker (KKT) conditions
Convex problem with Slater constraint qualification

recall the two implications of Slater’s condition for a convex problem

- strong duality: \( p^* = d^* \)
- if optimal value is finite, dual optimum is attained: there exist dual optimal \( \lambda, \nu \)

hence, if problem is convex and Slater’s constraint qualification holds:

- \( x \) is optimal if and only if there exist \( \lambda, \nu \) such that 1–4 on p. 5.22 are satisfied
- if functions are differentiable, condition 4 can be replaced with 4’
Example: water-filling

\[
\begin{align*}
\text{minimize} & \quad - \sum_{i=1}^{n} \log(x_i + \alpha_i) \\
\text{subject to} & \quad x \geq 0 \\
& \quad 1^T x = 1
\end{align*}
\]

- we assume that \(\alpha_i > 0\)
- Lagrangian is \(L(x, \lambda, \nu) = - \sum_i \log(\tilde{x}_i + \alpha_i) - \lambda^T \tilde{x} + \nu(1^T \tilde{x} - 1)\)

**Optimality conditions:** \(x\) is optimal iff there exist \(\lambda \in \mathbb{R}^n, \nu \in \mathbb{R}\) such that

1. \(x \geq 0, 1^T x = 1\)
2. \(\lambda \geq 0\)
3. \(\lambda_i x_i = 0\) for \(i = 1, \ldots, n\)
4. \(x\) minimizes Lagrangian:

\[
\frac{1}{x_i + \alpha_i} + \lambda_i = \nu, \quad i = 1, \ldots, n
\]
Example: water-filling

Solution

- if $\nu \leq 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $x_i = 0$ and $\lambda_i = \nu - 1/\alpha_i$
- two cases may be combined as
  
  $x_i = \max\{0, \frac{1}{\nu} - \alpha_i\}, \quad \lambda_i = \max\{0, \nu - \frac{1}{\alpha_i}\}$

- determine $\nu$ from condition $\mathbf{1}^T x = 1$:
  
  $\sum_{i=1}^{n} \max\{0, \frac{1}{\nu} - \alpha_i\} = 1$

Interpretation

- $n$ patches; level of patch $i$ is at height $\alpha_i$
- flood area with unit amount of water
- resulting level is $1/\nu^*$

Duality 5.25
Example: projection on $1$-norm ball

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \| x - a \|_2^2 \\
\text{subject to} & \quad \| x \|_1 \leq 1
\end{align*}
\]

Optimality conditions

1. $\| x \|_1 \leq 1$
2. $\lambda \geq 0$
3. $\lambda (1 - \| x \|_1) = 0$
4. $x$ minimizes Lagrangian

\[
L(\tilde{x}, \lambda) = \frac{1}{2} \| \tilde{x} - a \|_2^2 + \lambda (\| \tilde{x} \|_1 - 1)
\]

\[
= \sum_{k=1}^{n} \left( \frac{1}{2} (\tilde{x}_k - a_k)^2 + \lambda |\tilde{x}_k| \right) - \lambda
\]
Example: projection on 1-norm ball

Solution

- optimization problem in condition 4 is separable; solution for \( \lambda \geq 0 \) is

\[
x_k = \begin{cases} 
  a_k - \lambda & a_k \geq \lambda \\
  0 & -\lambda \leq a_k \leq \lambda \\
  a_k + \lambda & a_k \leq -\lambda 
\end{cases}
\]

- therefore \( \|x\|_1 = \sum_k |x_k| = \sum_k \max \{0, |a_k| - \lambda\} \)

- if \( \|a\|_1 \leq 1 \), solution is \( \lambda = 0, x = a \)

- otherwise, solve piecewise-linear equation in \( \lambda \):

\[
\sum_{k=1}^{n} \max \{0, |a_k| - \lambda\} = 1
\]
Perturbation and sensitivity analysis

(Unperturbed) optimization problem and its dual

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad g(\lambda, \nu) \\
\text{subject to} & \quad \lambda \geq 0
\end{align*}
\]

Perturbed problem and its dual

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

\[
\begin{align*}
\text{maximize} & \quad g(\lambda, \nu) - u^T \lambda - v^T \nu \\
\text{subject to} & \quad \lambda \geq 0
\end{align*}
\]

- \(x\) is primal variable; \(u, \nu\) are parameters
- \(p^*(u, \nu)\) is optimal value as a function of \(u, \nu\)
- we are interested in information about \(p^*(u, \nu)\) that we can obtain from the solution of the unperturbed problem and its dual
Global sensitivity result

• assume strong duality holds for unperturbed problem, and that $\lambda^*$, $\nu^*$ are dual optimal for unperturbed problem

• apply weak duality to perturbed problem:

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

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

Sensitivity interpretation

• if $\lambda^*_i$ is large: $p^*$ increases greatly if we tighten constraint $i$ ($u_i < 0$)

• if $\lambda^*_i$ is small: $p^*$ does not decrease much if we loosen constraint $i$ ($u_i > 0$)

• if $\nu^*_i$ is large and positive: $p^*$ increases greatly if we take $v_i < 0$
  if $\nu^*_i$ is large and negative: $p^*$ increases greatly if we take $v_i > 0$

• if $\nu^*_i$ is small and positive: $p^*$ does not decrease much if we take $v_i > 0$
  if $\nu^*_i$ is small and negative: $p^*$ does not decrease much if we take $v_i < 0$
Local sensitivity result

if (in addition) $p^*(u, v)$ is differentiable at $(0, 0)$, then

$$
\lambda^*_i = -\frac{\partial p^*(0, 0)}{\partial u_i}, \quad \nu^*_i = -\frac{\partial p^*(0, 0)}{\partial v_i}
$$

proof (for $\lambda^*_i$): from global sensitivity result,

$$
\frac{\partial p^*(0, 0)}{\partial u_i} = \lim_{t \searrow 0} \frac{p^*(te_i, 0) - p^*(0, 0)}{t} \geq -\lambda^*_i
$$

$$
\frac{\partial p^*(0, 0)}{\partial u_i} = \lim_{t \nearrow 0} \frac{p^*(te_i, 0) - p^*(0, 0)}{t} \leq -\lambda^*_i
$$

hence, equality

$p^*(u)$ for a problem with one (inequality) constraint:
Duality and problem reformulations

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

Common reformulations

• introduce new variables and equality constraints
• make explicit constraints implicit or vice-versa
• transform objective or constraint functions
  e.g., replace $f_0(x)$ by $\phi(f_0(x))$ with $\phi$ convex, increasing
Introducing new variables and equality constraints

\[
\text{minimize} \quad f_0(Ax + b)
\]

- dual function is constant: \( g = \inf_x L(x) = \inf_x f_0(Ax + b) = p^* \)
- we have strong duality, but dual is quite useless

Reformulated problem and its dual

\[
\begin{align*}
\text{minimize} & \quad f_0(y) & \text{maximize} & \quad b^T \nu - f_0^*(\nu) \\
\text{subject to} & \quad Ax + b - y = 0 & \text{subject to} & \quad A^T \nu = 0
\end{align*}
\]

dual function follows from

\[
g(\nu) = \inf_{x,y} (f_0(y) - \nu^T y + \nu^T Ax + b^T \nu) = \begin{cases} 
-f_0^*(\nu) + b^T \nu & A^T \nu = 0 \\
-\infty & \text{otherwise}
\end{cases}
\]
Example: norm approximation

minimize  \( \|Ax - b\| \)  \( \longrightarrow \)  minimize  \( \|y\| \)
subject to  \( y = Ax - b \)

Dual function

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

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

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

(last step follows from (1))

Dual of norm approximation problem

maximize  \( b^T v \)
subject to  \( A^T v = 0 \)
\( \|v\|_* \leq 1 \)
Implicit constraints

LP with box constraints: primal and dual problem

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b, \quad -1 \leq x \leq 1
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad -b^T \nu - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2 \\
\text{subject to} & \quad c + A^T \nu + \lambda_1 - \lambda_2 = 0, \quad \lambda_1 \geq 0, \quad \lambda_2 \geq 0
\end{align*}
\]

Reformulation with box constraints made implicit

\[
\begin{align*}
\text{minimize} & \quad f_0(x) = \begin{cases} c^T x & -1 \leq x \leq 1 \\ \infty & \text{otherwise} \end{cases} \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

dual function

\[
g(\nu) = \inf_{-1 \leq x \leq 1} (c^T x + \nu^T (Ax - b)) = -b^T \nu - \|A^T \nu + c\|_1
\]

Dual problem: maximize \(-b^T \nu - \|A^T \nu + c\|_1\)
Problems with generalized inequalities

minimize \( f_0(x) \)
subject to \( f_i(x) \leq_{K_i} 0, \quad i = 1, \ldots, m \)
\( h_i(x) = 0, \quad i = 1, \ldots, p \)

\( \leq_{K_i} \) is generalized inequality on \( \mathbb{R}^{k_i} \)

Lagrangian and dual function: definitions are parallel to scalar case

- Lagrange multiplier for \( f_i(x) \leq_{K_i} 0 \) is vector \( \lambda_i \in \mathbb{R}^{k_i} \)
- Lagrangian \( L : \mathbb{R}^n \times \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \rightarrow \mathbb{R} \), is defined as

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

- dual function \( g : \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \rightarrow \mathbb{R} \), is defined as

\[
g(\lambda_1, \ldots, \lambda_m, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda_1, \cdots, \lambda_m, \nu)
\]
Lagrange dual of problems with generalized inequalities

**Lower bound property:** if $\lambda_i \geq_{K_i} 0$, then $g(\lambda_1, \ldots, \lambda_m, \nu) \leq p^*$

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

$$
\begin{align*}
    f_0(x) & \geq f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \\
    & \geq \inf_{\bar{x} \in D} L(\bar{x}, \lambda_1, \ldots, \lambda_m, \nu) \\
    & = g(\lambda_1, \ldots, \lambda_m, \nu)
\end{align*}
$$

minimizing over all feasible $x$ gives $p^* \geq g(\lambda_1, \ldots, \lambda_m, \nu)$

**Dual problem**

maximize \quad $g(\lambda_1, \ldots, \lambda_m, \nu)$

subject to \quad $\lambda_i \geq_{K_i} 0, \quad i = 1, \ldots, m$

- weak duality: $p^* \geq d^*$ always
- strong duality: $p^* = d^*$ for convex problem with constraint qualification (for example, Slater’s: primal problem is strictly feasible)
Semidefinite program

\[
\text{minimize } \quad c^T x \\
\text{subject to } \quad x_1 F_1 + \cdots + x_n F_n \leq G
\]

matrices $F_1, \ldots, F_n, G$ are symmetric $k \times k$

Lagrangian and dual function

- Lagrange multiplier is matrix $Z \in \mathbb{S}^k$; Lagrangian is

\[
L(x, Z) = c^T x + \text{tr} \left( Z (x_1 F_1 + \cdots + x_n F_n - G) \right) \\
= \sum_{i=1}^{n} (\text{tr}(F_i Z) + c_i) x_i - \text{tr}(GZ)
\]

- dual function

\[
g(Z) = \inf_x L(x, Z) = \begin{cases} 
-\text{tr}(GZ) & \text{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\
-\infty & \text{otherwise}
\end{cases}
\]

Duality 5.37
Dual semidefinite program

maximize \(- \text{tr}(GZ)\)
subject to \(\text{tr}(F_i Z) + c_i = 0, \ i = 1, \ldots, n\)
\(Z \succeq 0\)

Weak duality: \(p^* \geq d^*\) always

proof: for primal feasible \(x\), dual feasible \(Z\),

\[ c^T x = - \sum_{i=1}^{n} \text{tr}(F_i Z) x_i \]
\[ = - \text{tr}(GZ) + \text{tr}(Z(G - \sum_{i=1}^{n} x_i F_i)) \]
\[ \geq - \text{tr}(GZ) \]

inequality follows from \(\text{tr}(XZ) \geq 0\) for \(X \succeq 0, Z \succeq 0\)

Strong duality: \(p^* = d^*\) if primal SDP or dual SDP is strictly feasible
Complementary slackness

(P) minimize \( c^T x \) \quad \text{(D) maximize} \quad \text{− tr}(GZ)

subject to \( \sum_{i=1}^{n} x_i F_i \leq G \) \quad \text{subject to} \quad \text{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n

\[ Z \geq 0 \]

the primal and dual objective values at feasible \( x, Z \) are equal if

\[ 0 = c^T x + \text{tr}(GZ) \]

\[ = - \sum_{i=1}^{n} x_i \text{tr}(F_i Z) + \text{tr}(GZ) \]

\[ = \text{tr}(XZ) \quad \text{where} \quad X = G - x_1 F_1 - \cdots - x_n F_n \]

for \( X \geq 0, Z \geq 0 \), each of the following statements is equivalent to \( \text{tr}(XZ) = 0 \):

- \( ZX = 0 \): columns of \( X \) are in the nullspace of \( Z \)
- \( XZ = 0 \): columns of \( Z \) are in the nullspace of \( X \)

(see next page)
proof: factorize $X, Z$ as

$$X = UU^T, \quad Z = VV^T$$

- columns of $U$ span the range of $X$, columns of $V$ span the range of $Z$
- $\text{tr}(XZ)$ can be expressed as

$$\text{tr}(XZ) = \text{tr}(UU^TVV^T) = \text{tr}((U^TV)(V^TU)) = \|U^TV\|_F^2$$

- hence, $\text{tr}(XZ) = 0$ if and only if

$$U^TV = 0$$

the range of $X$ and the range of $Z$ are orthogonal subspaces
Example: two-way partitioning

recall the two-way partitioning problem and its dual (page 5.8)

(P) minimize $x^T W x$  
subject to \( x_i^2 = 1, \quad i = 1, \ldots, n \)

(D) maximize $-1^T v$  
subject to $W + \text{diag}(v) \succeq 0$

- by weak duality, $p^* \geq d^*$
- the dual problem (D) is an SDP; we derive the dual SDP and compare it with (P)
- to derive the dual of (D), we first write (D) as a minimization problem:

\[
\begin{align*}
\text{minimize} \quad & 1^T y \\
\text{subject to} \quad & W + \text{diag}(y) \succeq 0
\end{align*}
\]

(2)

the optimal value of (2) is $-d^*$
Example: two-way partitioning

Lagrangian

\[ L(y, Z) = 1^T y - \text{tr}(Z(W + \text{diag}(y))) \]
\[ = -\text{tr}(WZ) + \sum_{i=1}^{n} y_i (1 - Z_{ii}) \]

Dual function

\[ g(Z) = \inf_y L(y, Z) = \begin{cases} 
-\text{tr}(WZ) & Z_{ii} = 1, \ i = 1, \ldots, n \\
-\infty & \text{otherwise}
\end{cases} \]

Dual problem: the dual of (2) is

\[
\text{maximize} \quad -\text{tr}(WZ) \\
\text{subject to} \quad Z_{ii} = 1, \ i = 1, \ldots, n \\
\quad Z \succeq 0
\]

by strong duality with (2), optimal value is equal to \(-d^*\)
Example: two-way partitioning

replace (D) on page 5.41 by its dual

(P) minimize \( x^T W x \)  
subject to \( x_i^2 = 1, \quad i = 1, \ldots, n \)

(P') minimize \( \text{tr}(WZ) \)  
subject to \( \text{diag}(Z) = 1 \)  
\( Z \succeq 0 \)

optimal value of (P') is equal to optimal value \( d^* \) of (D)

Interpretation as relaxation

- reformulate (P) by introducing a new variable \( Z = xx^T \):

  minimize \( \text{tr}(WZ) \)  
  subject to \( \text{diag}(Z) = 1 \)  
  \( Z = xx^T \)

- replace the constraint \( Z = xx^T \) with a weaker convex constraint \( Z \succeq 0 \)