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)

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

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

**Lagrangian:** \( L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \), 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

Lagrange 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, can be \(-\infty\) for some \( \lambda, \nu \)

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

**proof:** if \( \tilde{x} \) is feasible and \( \lambda \succeq 0 \), then

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

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

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

**dual function**

- Lagrangian is \( L(x, \nu) = x^T x + \nu^T (Ax - b) \)
- to minimize \( L \) over \( x \), set gradient equal to zero:
  \[
  \nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \implies \quad x = -(1/2)A^T \nu
  \]
- plug in in \( L \) to obtain \( g \):
  \[
  g(\nu) = L((-1/2)A^T \nu, \nu) = -\frac{1}{4} \nu^T A A^T \nu - b^T \nu
  \]
  a concave function of \( \nu \)

**lower bound property:** \( p^* \geq -(1/4)\nu^T A A^T \nu - b^T \nu \) for all \( \nu \)
**Standard form LP**

minimize \( c^T x \)

subject to \( Ax = b, \quad x \geq 0 \)

\( Ax - b = 0 \quad -x \leq 0 \)

**dual function**

- 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 \( \{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\} \), hence concave

**lower bound property:** \( p^* \geq -b^T \nu \) if \( A^T \nu + c \geq 0 \)

---

Recall \( A^T \nu - \lambda + c = 0 \)
Then \( A^T \nu + c = \lambda \)
But \( \lambda \geq 0 \)
Then \( A^T \nu + c \geq 0 \)
Equality constrained norm minimization

\[
\begin{align*}
\text{minimize} & \quad \|x\| \\
\text{subject to} & \quad Ax = b \\
& \quad -Ax + b = 0
\end{align*}
\]

dual function

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

where \(\|\nu\|_* = \sup_{\|u\| \leq 1} u^T \nu\) is dual norm of \(\| \cdot \|\)

Let \(y = A^T \nu\), proof: follows from \(\inf_x (\|x\| - y^T x) = 0\) if \(\|y\|_* \leq 1\), \(-\infty\) otherwise

- if \(\|y\|_* \leq 1\), then \(\|x\| - y^T x \geq 0\) for all \(x\), with equality if \(x = 0\)

  \(\text{Cauchy-Schwarz: } y^T x \leq \|y\|_* \|x\| \leq \|x\|\)

- if \(\|y\|_* > 1\), choose \(x = tu\) where \(\|u\| \leq 1\), \(u^T y = \|y\|_* > 1\): since \(\|y\|_* = \sup_{\|u\| \leq 1} u^T y > 1\)

  \[
  |x| - y^T x = t |u| - y^T u = t |u| - t |y|_* \\
  = t (\|u\| - \|y\|_*) \to -\infty \quad \text{as } t \to \infty
  \]

lower bound property: \(p^* \geq b^T \nu\) if \(\|A^T \nu\|_* \leq 1\)
Two-way partitioning

minimize \( x^T W x \)
subject to \( x_i^2 = 1, \quad i = 1, \ldots, n \)
\( x_i \) is -1 or +1

- a nonconvex problem; feasible set contains \( 2^n \) discrete points

- interpretation: partition \( \{1, \ldots, n\} \) in two sets; \( W_{ij} \) is cost of assigning \( i, j \) to the same set; \(-W_{ij}\) is cost of assigning to different sets
  (one set is all i's where \( x_i = -1 \), the second set is all i's where \( x_i = +1 \))

dual function

\[
g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) = \inf_x x^T (W + \text{diag}(\nu)) x - 1^T \nu
\]

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

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

if \( W + \text{diag}(\nu) \) has at least one negative eigenvalue we can make \( x'(W + \text{diag}(\nu))x \) arbitrarily small
The dual problem

Lagrange dual problem

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

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

**example:** standard form LP and its dual (page 5–5)

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

\[
\begin{align*}
\text{maximize} & \quad -b^T \nu \\
\text{subject to} & \quad A^T \nu + c \succeq 0
\end{align*}
\]
A nonconvex problem with strong duality

minimize $x^T Ax + 2b^T x$
subject to $x^T x \leq 1$

$A \not\succeq 0$, hence nonconvex

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

- unbounded below if $A + \lambda I \not\succeq 0$ or if $A + \lambda I \succeq 0$ and $b \not\in \mathcal{R}(A + \lambda I)$
- minimized by $x = -(A + \lambda I)^\dagger b$ otherwise: $g(\lambda) = -b^T (A + \lambda I)^\dagger b - \lambda$

For simplicity assume $(A+\lambda I) > 0$

**dual problem**

maximize $-b^T (A + \lambda I)^\dagger b - \lambda$
subject to $A + \lambda I \succeq 0$

$L(x,\lambda) = x^T Ax + 2b^T x + \lambda(x^T x - 1) = x^T (A + \lambda I)x + 2b^T x - \lambda$

$g(\lambda) = \inf_x L(x,\lambda)$

d$L/dx = 2(A+\lambda I)x + 2b = 0$ $=>$ $x^\star = -(A+\lambda I)^{-1} b$

Then $g(\lambda) = L(x^\star,\lambda) = -b^T (A+\lambda I)^{-1} b - \lambda$

Lagrange dual: max $g(\lambda)$ s.t. $\lambda \geq 0$

Let $A = U\Sigma U'$, then $A + \lambda I = U(D + \lambda I)U' = U\Sigma(\lambda)U'$, where $s_{ii}(\lambda) = d_{ii} + \lambda$

Then $(A+\lambda I)^{-1} = U\Sigma^{-1}(\lambda)U'$, where $s_{ii}^{-1}(\lambda) = 1/(d_{ii} + \lambda)$

Let $U = [u_{i1} \ldots u_{in}]$, where $u_{i}$ are column eigenvectors

$g(\lambda) = -b^T U\Sigma^{-1}(\lambda)U' b - \lambda = -\Sigma_i b^T u_i s_{ii}^{-1}(\lambda)u_i^T b - \lambda$

$= -\Sigma_i s_{ii}^{-1}(\lambda)(b^T u_i)^2 - \lambda$

d$g/d\lambda = \Sigma_i (b^T u_i)^2 / (d_{ii} + \lambda)^2 - 1$

It is easy to use a ONE-DIMENSIONAL gradient ascent or Newton method!
Lagrange dual and conjugate function

\[\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad Ax \preceq b, \quad Cx = d
\end{align*}\]

dual function

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

\[\begin{align*}
&= \inf_x \left\{ f_0(x) + (A^T \lambda + C^T \nu)^T x \right\} - b^T \lambda - d^T \nu \\
&= \sup_x \left\{ -(A^T \lambda + C^T \nu)^T x \right\} - b^T \lambda - d^T \nu \\
&= -f_0^*(-A^T \lambda - C^T \nu) - b^T \lambda - d^T \nu
\end{align*}\]

- recall definition of conjugate \(f^*(y) = \sup_{x \in \text{dom} f} f(y^T x - f(x))\)
- simplifies derivation of dual if conjugate of \(f_0\) is known

example: entropy maximization

\[\begin{align*}
f_0(x) &= \sum_{i=1}^{n} x_i \log x_i, \\
f_0^*(y) &= \sum_{i=1}^{n} e^{y_i - 1}
\end{align*}\]
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

\[
\text{maximize } -1^T \nu \\
\text{subject to } W + \text{diag}(\nu) \succeq 0
\]

gives a lower bound for the two-way partitioning problem on page 5–7

**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**
Slater’s constraint qualification

strong duality holds for a convex problem

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

if it is strictly feasible, \textit{i.e.},

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

\textbf{strict inequality}

- also guarantees that the dual optimum is attained (if \( p^* > -\infty \))
- can be sharpened:

Assume \( f_1(x) \ldots f_k(x) \) are affine and \( \text{dom}(f_0) \) open, then the \textit{REFINED} Slater’s condition is

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

Thus, if all inequalities are affine (\( k=m \)) then strict inequality is not necessary!

- there exist many other types of constraint qualifications
Inequality form LP

primal problem

\[ \text{minimize } \quad c^T x \]
\[ \text{subject to } \quad Ax \leq b \]

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

\[ \text{maximize } \quad -b^T \lambda \]
\[ \text{subject to } \quad A^T \lambda + c = 0, \quad \lambda \succeq 0 \]

• from Slater’s condition: \( p^* = d^* \) if \( A\tilde{x} \prec b \) for some \( \tilde{x} \)

• in fact, \( p^* = d^* \) except when primal and dual are infeasible (refined Slater’s)
Quadratic program

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

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

**dual function**

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

**dual problem**

\[
\begin{align*}
\text{maximize} & \quad -(1/4) \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\
\text{subject to} & \quad \lambda \succeq 0
\end{align*}
\]

- from Slater’s condition: $p^* = d^*$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^* = d^*$ always (refined Slater’s)
**Geometric interpretation**

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

**interpretation of dual function:**

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

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

Dual: $\lambda^{**} = \arg\max_\{\lambda \geq 0\} g(\lambda)$

$d^{**} = g(\lambda^{**})$ is the “tightest” supporting hyperplane

(If you cannot go up without violating the definition of supporting hyperplane)
epigraph variation: same interpretation if $G$ is replaced with

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

strong duality

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

(explained in previous slide)
Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable $f_i, h_i$):

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$  \hspace{1cm} \text{(Primal feasibility)}
2. dual constraints: $\lambda \succeq 0$  \hspace{1cm} \text{(Dual feasibility)}
3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$  \hspace{1cm} \text{if } \lambda_i > 0 \text{ then } f_i(x) = 0$
   \hspace{1cm} \text{if } f_i(x) < 0 \text{ then } \lambda_i = 0$
4. gradient of Lagrangian with respect to $x$ vanishes: \hspace{1cm} \text{(Stationarity)}

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

from page 5–17: if strong duality holds and $x$, $\lambda$, $\nu$ are optimal, then they must satisfy the KKT conditions
Complementary slackness

assume strong duality holds, \( x^\star \) is primal optimal, \((\lambda^\star, \nu^\star)\) is dual optimal

\[
\begin{align*}
f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left( f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \\
&\leq f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^* f_i(x^\star) + \sum_{i=1}^{p} \nu_i^* h_i(x^\star) \\
&\leq f_0(x^\star)
\end{align*}
\]

hence, the two inequalities hold with equality

- \( x^\star \) minimizes \( L(x, \lambda^\star, \nu^\star) \)
- \( \lambda_i^* f_i(x^\star) = 0 \) for \( i = 1, \ldots, m \) (known as complementary slackness):
  \[
  \lambda_i^* > 0 \implies f_i(x^\star) = 0, \quad f_i(x^\star) < 0 \implies \lambda_i^* = 0
  \]
KKT conditions for convex problem

if \( \tilde{x}, \tilde{\lambda}, \tilde{\nu} \) satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: \( f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \)
- from 4th condition (and convexity): \( g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \)

hence, \( f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu}) \)

if **Slater’s condition** is satisfied:

\( x \) is optimal if and only if there exist \( \lambda, \nu \) that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition \( \nabla f_0(x) = 0 \) for unconstrained problem
example: water-filling (assume $\alpha_i > 0$)

minimize $-\sum_{i=1}^{n} \log(x_i + \alpha_i)$
subject to $x \succeq 0$, $1^T x = 1$

$x$ is optimal iff $x \succeq 0$, $1^T x = 1$, 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 $1^T x = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

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

(unperturbed) optimization problem and its dual

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

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

perturbed problem and its dual

min. \( f_0(x) \)
s.t. \( f_i(x) \leq u_i, \quad i = 1, \ldots, m \)
\( h_i(x) = v_i, \quad i = 1, \ldots, p \)

max. \( g(\lambda, \nu) - u^T \lambda - v^T \nu \)
s.t. \( \lambda \succeq 0 \)

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

Duality
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 $\lambda_i^*$ large: $p^*$ increases greatly if we tighten constraint $i$ ($u_i < 0$)
• if $\lambda_i^*$ small: $p^*$ does not decrease much if we loosen constraint $i$ ($u_i > 0$)
• if $\nu_i^*$ large and positive: $p^*$ increases greatly if we take $v_i < 0$;
  if $\nu_i^*$ large and negative: $p^*$ increases greatly if we take $v_i > 0$
• if $\nu_i^*$ small and positive: $p^*$ does not decrease much if we take $v_i > 0$;
  if $\nu_i^*$ small and negative: $p^*$ does not decrease much if we take $v_i < 0$
**local sensitivity:** if (in addition) \( p^*(u,v) \) is differentiable at \((0,0)\), then

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

**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
  
  \[ e.g., \ \text{replace } f_0(x) \text{ by } \phi(f_0(x)) \]  
  with \( \phi \) convex, increasing
Introducing new variables and equality constraints

\[
\begin{align*}
\text{minimize} & \quad f_0(Ax + b) \\
\text{dual function is constant:} & \quad g = \inf_x L(x) = \inf_x f_0(Ax + b) = p^* \\
\text{we have strong duality, but dual is quite useless}
\end{align*}
\]

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

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

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

can look up conjugate of \( \| \cdot \| \), or derive dual directly

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

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

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

(see page 5–4)

dual of norm approximation problem

\[
\begin{align*}
\text{maximize} & \quad b^T \nu \\
\text{subject to} & \quad A^T \nu = 0, \quad \| \nu \|_* \leq 1
\end{align*}
\]
Implicit constraints

LP with box constraints: primal and dual problem

minimize $c^T x$ 
subject to $Ax = b$ 
$-1 \leq x \leq 1$

maximize $-b^T \nu - 1^T \lambda_1 - 1^T \lambda_2$
subject to $c + A^T \nu + \lambda_1 - \lambda_2 = 0$ 
$\lambda_1 \geq 0, \quad \lambda_2 \geq 0$

reformulation with box constraints made implicit

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

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

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

\preceq_{K_i} \text{ is generalized inequality on } \mathbb{R}^{k_i}

definitions \text{ are parallel to scalar case:}

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

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

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

\[
g(\lambda_1, \ldots, \lambda_m, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu)
\]
lower bound property: if $\lambda_i \succeq_{K_i^*} 0$, then $g(\lambda_1, \ldots, \lambda_m, \nu) \leq p^*$

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

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

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

$$= g(\lambda_1, \ldots, \lambda_m, \nu)$$

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

dual problem

maximize $g(\lambda_1, \ldots, \lambda_m, \nu)$
subject to $\lambda_i \succeq_{K_i^*} 0$, $i = 1, \ldots, m$

- 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

primal SDP \((F_i, G \in S^k)\)

minimize \(c^T x\)
subject to \(x_1 F_1 + \cdots + x_n F_n \preceq G\)

- Lagrange multiplier is matrix \(Z \in S^k\)
- Lagrangian \(L(x, Z) = c^T x + \text{tr} (Z(x_1 F_1 + \cdots + x_n F_n - G))\)
- dual function

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

dual SDP

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

\(\rho^* = d^*\) if primal SDP is strictly feasible (\(\exists x\) with \(x_1 F_1 + \cdots + x_n F_n < G\))