5. Duality

- Lagrange dual problem
- weak and strong duality
- geometric interpretation
- optimality conditions
- 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: \(L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}\), 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

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

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

\( g \) 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

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

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

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

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 \( Ax = b, \quad x \succeq 0 \)

\[ \text{Ax - b = 0} \quad -x \leq 0 \]

**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

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

for any nonzero vector \( y \), we can make \( y'x \) arbitrarily small

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

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

Recall \( A'v - \lambda + c = 0 \)
Then \( A'v + c = \lambda \)
But \( \lambda \geq 0 \)
Then \( A'v + c \geq 0 \)
Equality constrained norm minimization

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

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 \(x = 0\) (Cauchy-Schwarz: \(y^\top x \leq \|y\|_* \|x\| \leq \|x\|\))
- if \(\|y\|_* > 1\), choose \(x = tu\) where \(\|u\| \leq 1\), \(u^T y = \|y\|_* > 1\): since \(\|y\|_* = \sup_{\|u\| \leq 1} u^\top y > 1\)

\[
\begin{align*}
|x| - y^\top x &= t |u| - t y^\top u = t |u| - t |y|_* \\
&= t(\|u\| - \|y\|_*) \to -\infty \quad \text{as } t \to \infty
\end{align*}
\]

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

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

- 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

Duality
The dual problem

Lagrange dual problem

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

- 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)

minimize \[ c^T x \]
subject to \[ Ax = b \]
\[ x \succeq 0 \]

maximize \[ -b^T \nu \]
subject to \[ A^T \nu + c \succeq 0 \]
A nonconvex problem with strong duality

\[
\begin{align*}
\text{minimize} & \quad x^T A x + 2 b^T x \\
\text{subject to} & \quad x^T x \leq 1 \\
& \quad x'x - 1 \leq 0
\end{align*}
\]

\(A \not\succeq 0\), 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 \(b \notin \mathcal{R}(A + \lambda I)\)
- minimized by \(x = -(A + \lambda I)^\dagger b\) otherwise: \(g(\lambda) = -b^T (A + \lambda I)^\dagger b - \lambda\)

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

\[
\begin{align*}
L(x,\lambda) &= x'Ax + 2 b'x + \lambda(x'x - 1) = x'(A + \lambda I)x + 2 b'x - \lambda \\
g(\lambda) &= \inf_x L(x,\lambda) \\
dL/dx &= 2(A + \lambda I)x + 2b = 0 \quad \Rightarrow \quad x^* = -(A + \lambda I)^{-1}b
\end{align*}
\]

Then \(g(\lambda) = L(x^*,\lambda) = -b' (A + \lambda I)^{-1} b - \lambda\)

Lagrange dual: \(\max g(\lambda) \text{ s.t. } \lambda \geq 0\)

Let \(A = U D U'\), then \(A + \lambda I = U(D + \lambda I)U' = U S(\lambda) \ U'\), where \(s_{ii}(\lambda) = d_{ii} + \lambda\)
Then \((A + \lambda I)^{-1} = U S^{-1}(\lambda) \ U'\), where \(s_{ii}^{-1}(\lambda) = 1/(d_{ii} + \lambda)\)

Let \(U = [u_1 \ldots u_n]\), where \(u_i\) are column eigenvectors
\[
\begin{align*}
g(\lambda) &= -b' \ U S^{-1}(\lambda) \ U' \ b - \lambda = -\Sigma_i b' u_i \ s_{ii}^{-1}(\lambda) \ u_i' \ b - \lambda \\
&= -\Sigma_i s_{ii}^{-1}(\lambda) \ (b' u_i)^2 - \lambda \\
dg/d\lambda &= \Sigma_i (b' u_i)^2 / (d_{ii} + \lambda)^2 - 1
\end{align*}
\]

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

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

dual function

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

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

\[
= \sup_x \{ -f_0(x) + (A^T \lambda + C^T \nu)^T x - b^T \lambda - d^T \nu \}
\]

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

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

example: entropy maximization

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

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

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, i.e.,

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

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 Refined Slater’s condition is

there is an \( x \), \( f_i(x) \leq 0 \) for \( i = 1 \ldots k \) \( f_i(x) < 0 \) for \( i = k+1 \ldots m \) \( Ax = b \)

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

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

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

minimize $x^T P x$
subject to $A x \preceq b$

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 $-(1/4) \lambda^T A P^{-1} A^T \lambda - b^T \lambda$
subject to $\lambda \succeq 0$

• 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:

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

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

Let \( t^*, u^* = \arg\inf_{(u,t) \in \mathcal{G}} (t + \lambda u) \)

Then \( t + \lambda u \geq t^* + \lambda u^* \)

It is a supporting hyperplane!

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

( you cannot go up without violating the definition of supporting hyperplane )

what if all \( u \geq 0 \)?
(i.e. \( f_1(x) \geq 0 \) for all \( x \))
Constraint is \( f_1(x) \leq 0 \)
Then solution has \( f_1(x) = 0 \)

VERTICAL supporting hyperplane
\( \lambda^{**} = \infty \)
epigraph variation: same interpretation if $\mathcal{G}$ is replaced with

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

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

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 supp. 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

(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} (Primal feasibility)

2. dual constraints: $\lambda \succeq 0$  \hspace{1cm} (Dual feasibility)

3. complementary slackness: $\lambda_i f_i(x) = 0, \ i = 1, \ldots, m$  \hspace{1cm} if $\lambda_i > 0$ then $f_i(x) = 0$

   if $f_i(x) < 0$ then $\lambda_i = 0$

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

General idea, for general possibly nonconvex primal problem: OPTIMAL $\Rightarrow$ KKT satisfied.
(subject to some technical conditions)
Complementary slackness

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

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

hence, the two inequalities hold with equality

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

since \( h_i(x^*) = 0 \) given that \( x^* \) is feasible: \( \sum_i \lambda_i^* f_i(x^*) = 0 \)
but each term in sum is nonpositive (none of the terms can be negative because there will not be a positive to make sum = 0)
KKT conditions for convex problem

General idea, for convex primal problem: KKT satisfied $\Rightarrow$ OPTIMAL and thus KKT satisfied $\iff$ OPTIMAL (subject to some technical conditions)

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

zero duality gap since $x^* = x^*, \lambda^* = \lambda^*, \nu^* = \nu^*$

Stationarity: gradient of $L(x, \lambda^*, \nu^*)$ w.r.t. $x$ vanishes,

$\Rightarrow x^*$ minimizes $L$ ... (this is why we assumed convexity otherwise stationarity does not imply that $x^*$ is the minimizer of $L$)

$\Rightarrow L(x^*, \lambda^*, \nu^*) = \inf_x L(x, \lambda^*, \nu^*)$

$= g(\lambda^*, \nu^*)$

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

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

---

**Slide 5-11:** Slater $\Rightarrow$ strong duality
**Slide 5-18:** Strong duality + OPTIMAL $\Rightarrow$ KKT satisfied
Here so far: KKT satisfied $\Rightarrow$ OPTIMAL
Therefore assume Slater: KKT satisfied $\iff$ OPTIMAL
example: water-filling (assume $\alpha_i > 0$)

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

subject to $x_i \geq 0, \quad 1^T x = 1$

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

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$ (because $\lambda_i$ cannot be negative)
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$ (because $\lambda_i x_i = 0$)
- 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 $\alpha_i$
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

\[
\minimize \quad f_0(Ax + b)
\]

- 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

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

dual function follows from

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

\[
= \inf_y \{ f_0(y) - \nu^T y \} + \inf_x \{ \nu^T Ax \} + b^T \nu
\]

\[
= - \sup_y \{ -f_0(y) + \nu^T y \} + \inf_x \{ \nu^T Ax \} + b^T \nu
\]

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

Note: if \( A^T \nu \neq 0 \), we can pick \( x \) so that \( \nu^T Ax \) is arbitrarily small
norm approximation problem: minimize $\|Ax - b\|

minimize $\|y\|
subject to $y = Ax - b$

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

maximize $b^T \nu$
subject to $A^T \nu = 0$, $\|\nu\|_* \leq 1$
Implicit constraints

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

minimize \(c^T x\) \hspace{1cm} maximize \(-b^T \nu - 1^T \lambda_1 - 1^T \lambda_2\)

subject to \(Ax = b\) \hspace{1cm} subject to \(c + A^T \nu + \lambda_1 - \lambda_2 = 0\)

\(-1 \leq x \leq 1\) \hspace{1cm} \(\lambda_1 \geq 0, \lambda_2 \geq 0\)

**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

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

\[= \inf_{\{|x| \text{ infty} \leq 1\}} \{(A'v+c)'x\} - b'v\]
\[= - \sup_{\{|x| \text{ infty} \leq 1\}} \{(-A'v-c)'x\} - b'v\]
\[= - |A'v+c|_1 - b'v \hspace{0.5cm} \ldots \text{by norm duality}\]

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

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

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

definitions 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 $\lambda_i \succeq K_i^* 0$, 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 $\lambda_i \succeq K_i^* 0, \quad i = 1, \ldots, m$

• 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

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

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad x_1 F_1 + \cdots + x_n F_n \preceq G
\end{align*}
\]

- Lagrange multiplier is matrix \(Z \in S^k\)
- Lagrangian \(L(x, Z) = c^T x + \text{tr}(Z(x_1 F_1 + \cdots + x_n F_n - G))\)
  \[
  = -\text{tr}(GZ) + \sum_{i} x_i (c_i + \text{tr}(Z F_i))
  \]
- dual function
  \[
  g(Z) = \inf_x L(x, Z) = \begin{cases} 
  -\text{tr}(GZ) & \text{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\
  -\infty & \text{otherwise}
  \end{cases}
\]

\[
\begin{align*}
\text{dual SDP} & \\
\text{maximize} & \quad -\text{tr}(GZ) \\
\text{subject to} & \quad Z \succeq 0, \quad \text{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n
\end{align*}
\]

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