4. Convex optimization problems

- optimization problem in standard form
- convex optimization problems
- quasiconvex optimization
- linear optimization
- quadratic optimization
- geometric programming
- generalized inequality constraints
- semidefinite programming
- vector optimization

Vector optimization

general vector optimization problem

$$\begin{array}{ll} \text{minimize (w.r.t. } K) & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

vector objective $f_0 : \mathbf{R}^n \to \mathbf{R}^q$, minimized w.r.t. proper cone $K \in \mathbf{R}^q$

convex vector optimization problem

$$\begin{array}{ll} \text{minimize (w.r.t. } K) & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

with f_0 K-convex, f_1, . . . , f_m convex

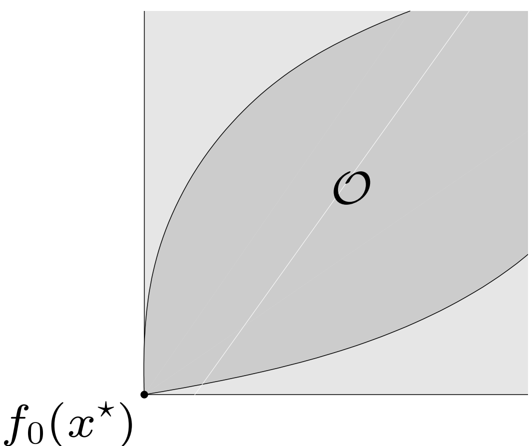
$K = \mathbb{R}^n_+$

Optimal and Pareto optimal points

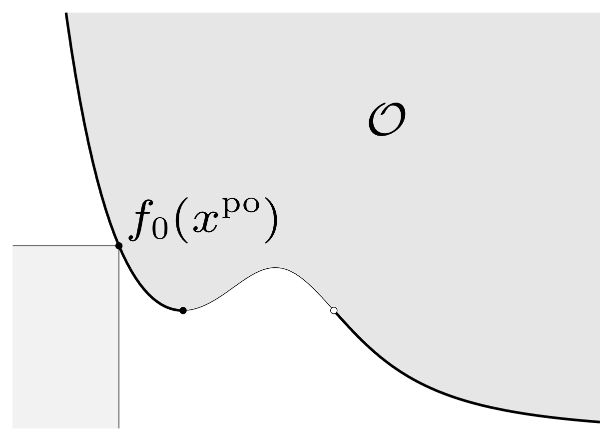
set of achievable objective values

$$\mathcal{O} = \{f_0(x) \mid x \text{ feasible}\}$$

- feasible x is **optimal** if $f_0(x)$ is the minimum value of $\mathcal{O}$
- feasible x is **Pareto optimal** if $f_0(x)$ is a minimal value of $\mathcal{O}$



$x^\star$ is optimal



$x^{\rm po}$ is Pareto optimal

Multicriterion optimization

vector optimization problem with $K = \mathbf{R}^q_+$

$$f_0(x) = (F_1(x), \ldots, F_q(x))$$

- q different objectives F_i; roughly speaking we want all F_i's to be small
- feasible $x^\star$ is optimal if

$$y \text{ feasible} \quad \Longrightarrow \quad f_0(x^\star) \preceq f_0(y)$$

  if there exists an optimal point, the objectives are noncompeting

- feasible x^{po} is Pareto optimal if

$$y \text{ feasible}, \quad f_0(y) \preceq f_0(x^{\mathrm{po}}) \quad \Longrightarrow \quad f_0(x^{\mathrm{po}}) = f_0(y)$$

  if there are multiple Pareto optimal values, there is a trade-off between the objectives

Scalarization

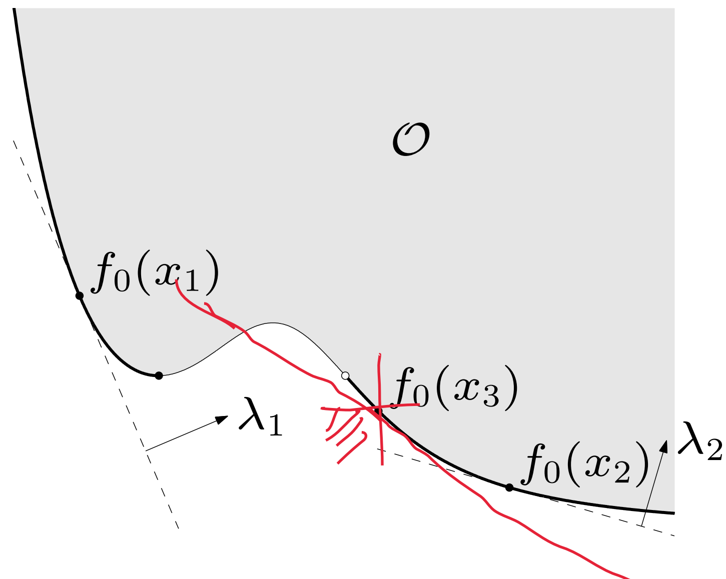
to find Pareto optimal points: choose $\lambda \succ_{K^*} 0$ and solve scalar problem

$$\begin{array}{ll} \text{minimize} & \lambda^T f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

$\Rightarrow f_0(x)$: minimal

$\lambda^T (f_0(x) - f_0(x^*)) \geq 0$

if x is optimal for scalar problem, then it is Pareto-optimal for vector optimization problem



for convex vector optimization problems, can find (almost) all Pareto optimal points by varying $\lambda \succ_{K^*} 0$

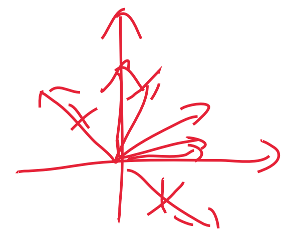
Scalarization for multicriterion problems

$(K^* = K = \mathbf{R}^n_+)$

to find Pareto optimal points, minimize positive weighted sum

$$\lambda^T f_0(x) = \lambda_1 F_1(x) + \cdots + \lambda_q F_q(x)$$

$\lambda \succ_{\mathbf{R}^n_+} 0$



examples

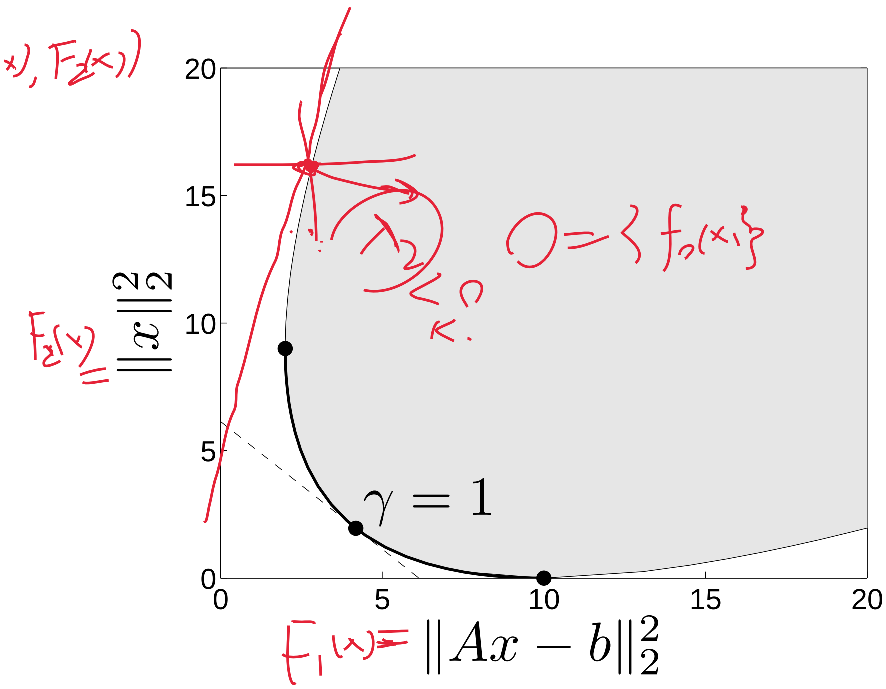
- regularized least-squares problem of page 4–43

$f_0(x) = (F_1(x), F_2(x))$

take $\lambda = (1, \gamma)$ with $\gamma > 0$

$$\text{minimize} \quad \|Ax - b\|_2^2 + \gamma \|x\|_2^2$$

for fixed γ, a LS problem



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)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & \boxed{f_i(x) \leq 0,} \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

$f_i(x_1) = -0.1.$ $f_i(x_2) = -100$

variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

Lagrangian: $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$, with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

$$L(x, \lambda, \nu) = f_0(x) + \underbrace{\sum_{i=1}^m \lambda_i f_i(x)}_{\lambda^T f(x)} + \underbrace{\sum_{i=1}^p \nu_i h_i(x)}_{\nu^T h(x)}$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

$f(x,y)$ convex . x . for $y \in A$
(concave)

$\Rightarrow g(x) = \sup_{y \in A} f(x,y)$ convex

$\inf f(x,y)$: concave

Lagrange dual function

Lagrange dual function: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\
&= \inf_{x \in \mathcal{D}} \left(f_0(x) + \underbrace{\sum_{i=1}^{m} \lambda_i f_i(x)}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \nu_i h_i(x)}_{0} \right)
\end{aligned}
$$

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^\star$ $\quad \inf_x f_0(x)$, x : feasible

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 $p^\star \geq g(\lambda, \nu)$

Least-norm solution of linear equations

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b \end{array}$$

dual function

- Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$
- to minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \Longrightarrow \quad x = -(1/2) A^T \nu$$

- 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 ν

lower bound property: $p^\star \geq -(1/4)\nu^T A A^T \nu - b^T \nu$ for all ν

$$\min_x \; a^T x + \underline{b} = \begin{cases} b, & a=0 \\ -\infty, & \text{otherwise.} \end{cases}$$

Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

dual function

- Lagrangian is

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned}$$

- 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: $p^\star \geq -b^T \nu$ if $A^T \nu + c \succeq 0$

Equality constrained norm minimization

① $u^T v \le \|u\| \cdot \|v\|_*$

② $\|v\|_{**} = \|v\|$

③ $\ell_p\text{-norm} \xrightarrow{\text{dual}} \ell_q\text{-norm}$ $\left(\frac{1}{p}+\frac{1}{q}=1,\ p,q \ge 1\right)$

(ℓ_1-norm, ℓ_∞-norm) (ℓ_2-norm, ℓ_2-norm)

$$
\begin{array}{ll}
\text{minimize} & \|x\| \\
\text{subject to} & Ax = b
\end{array}
$$

$f^*(y) = \sup_x \left(y^T x - f_0(x)\right)$

$= -\inf_x \left(f_0(x) - y^T x\right)$

dual function

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

where $\|v\|_* = \sup_{\|u\| \le 1} u^T v$ is dual norm of $\| \cdot \|$

$\|x\| \ge \|x\| \cdot \|y\|_* \ge x^T y$

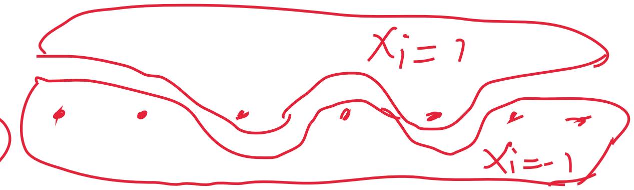
proof: follows from $\inf_x (\|x\| - y^T x) = 0$ if $\|y\|_* \le 1$, $-\infty$ otherwise

- if $\|y\|_* \le 1$, then $\|x\| - y^T x \ge 0$ for all x, with equality if $x = 0$
- if $\|y\|_* > 1$, choose $x = tu$ where $\|u\| \le 1$, $u^T y = \|y\|_* > 1$:

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

lower bound property: $p^\star \ge b^T \nu$ if $\|A^T \nu\|_* \le 1$

Two-way partitioning



$$
\begin{array}{ll}
\text{minimize} & x^T W x \quad = \sum_{ij} x_i x_j w_{ij} \\
\text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n
\end{array}
$$

$x_i = +1, -1$

- 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

dual function

$\sum_i v_i x_i^2 = x^T \mathrm{diag}(v) \cdot x$

$$
\begin{array}{rcl}
g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) & = & \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\
& = & \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}
\end{array}
$$

$$\min_x x^T P x = \begin{cases} 0, & P \succeq 0 \quad K = S^n_+ \\ -\infty, & \text{otherwise} \end{cases}$$

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

example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^\star \geq n\lambda_{\min}(W)$

Lagrange dual and conjugate function

$$\inf_x f(x) = -\sup_x (-f(x))$$

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \preceq b, \quad Cx = d \end{array}$$

dual function

$$\inf_x \left(f_0(x) + (A^T\lambda + C^T\nu)^T x\right) = -\sup_x \left((-(A^T\lambda + C^T\nu))^T x - f_0(x)\right) = -f^*(y) = -f_0^*(-(A^T\lambda + C^T\nu))$$

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbf{dom}\, f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu\right) \\ &= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu \end{aligned}$$

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

example: entropy maximization

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

Quiz: derive the dual function of entropy maximization problem.

Lagrange dual and conjugate function

Primal $\boxed{\begin{array}{ll} \min_x & \underline{\|x\|} \\ \text{s.t} & \underline{Ax=b} \end{array}} = p^*$

$\boxed{\begin{array}{l} f_0(x) = \underline{\|x\|} \\ f_0^*(y) = \begin{cases} 0, & \underline{\|y\|_* \leq 1} \\ \infty, & \text{otherwise.} \end{cases} \end{array}}$ indicator func.

$\underline{L(x,\nu)} = \|x\| + \nu^T(Ax-b)$

$= \nu^T Ax + \|x\| - \nu^T b$

$\underline{f^*(y) = \sup_x \left(y^T x - f(x) \right)}$

$\|A\|_*$

$\underline{\|aA\|_* = |a| \|A\|_*}$

Dual: $g(\nu) = \inf_x L(x,\nu) = \inf_x \left(\nu^T Ax + \|x\| \right) - \nu^T b$

$= \underline{-\sup_x \left((-\nu^T A)x - \|x\| \right)} - \nu^T b$

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

$\left(p^* \geq \underline{-\nu^T b}, \text{ if } \|\nu^T A\|_* \leq 1 \right)$

The dual problem

$p^* - d^*$: optimal (nonnegative) duality gap

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

$d^* \le p^*$

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

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

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \qquad\qquad \begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0 \end{array}$$

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

$A^T\nu + c = \lambda \ge 0$

Weak and strong duality

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

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

  for example, solving the SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

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

strong duality: $d^\star = p^\star$

- 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

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

$Cx \preceq d$

$d^\star = p^\star$

if it is strictly feasible, *i.e.*,

non-empty

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

interior

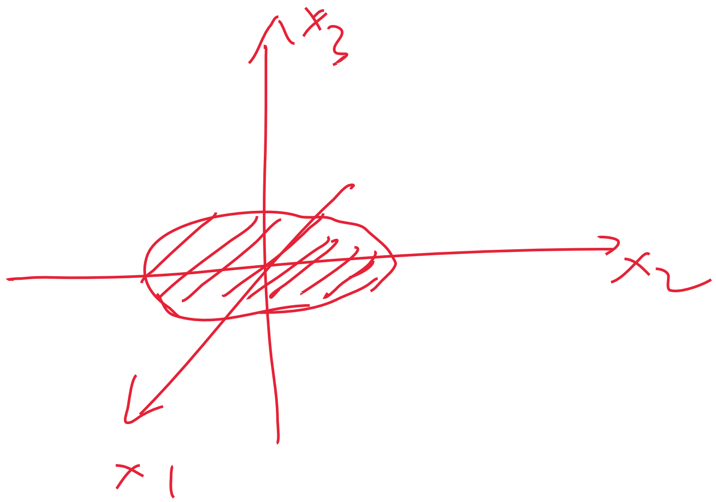
- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)
- can be sharpened: *e.g.*, 1 can replace $\mathbf{int}\,\mathcal{D}$ with $\mathbf{relint}\,\mathcal{D}$ (interior relative to affine hull); 2 linear inequalities do not need to hold with strict inequality, . . .

Slater's condition $\longrightarrow$ feasibility.

- there exist many other types of constraint qualifications

Slater's constraint qualification

$C = \{x \in R^3 \mid x_1^2 + x_2^2 \leq 1, \; \underline{x_3 = 0}\}$.



$\text{int } C = \phi$

$\underline{\text{rel int } C} = \{x \in R^3 \mid x_1^2 + x_2^2 < \underline{1}, \; x_3 = 0\}$

$\text{aff } C = C$

$\text{aff dim} = 2$.

Inequality form LP

primal problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

dual function

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

dual problem

$$\begin{array}{ll} \text{maximize} & -b^T\lambda \\ \text{subject to} & A^T\lambda + c = 0, \quad \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ except when primal and dual are infeasible

Quadratic program

primal problem (assume $P \in \mathbf{S}^n_{++}$)

$$\begin{array}{ll} \text{minimize} & x^T P x \\ \text{subject to} & Ax \preceq b \end{array}$$

dual function

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

dual problem

$$\begin{array}{ll} \text{maximize} & -(1/4)\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- from Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^\star = d^\star$ always

A nonconvex problem with strong duality

$$\begin{array}{ll} \text{minimize} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leq 1 \end{array}$$

$A \not\succeq 0$, 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: $g(\lambda) = -b^T(A + \lambda I)^\dagger b - \lambda$

dual problem and equivalent SDP:

$$\begin{array}{ll} \text{maximize} & -b^T(A + \lambda I)^\dagger b - \lambda \\ \text{subject to} & A + \lambda I \succeq 0 \\ & b \in \mathcal{R}(A + \lambda I) \end{array}$$

$$\begin{array}{ll} \text{maximize} & -t - \lambda \\ \text{subject to} & \left[\begin{array}{cc} A + \lambda I & b \\ b^T & t \end{array} \right] \succeq 0 \end{array}$$

strong duality although primal problem is not convex (not easy to show)

Geometric interpretation

$\min_x \; f_0(x)$
$\text{s.t. } f_1(x) \le 0$

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

interpretation of dual function:

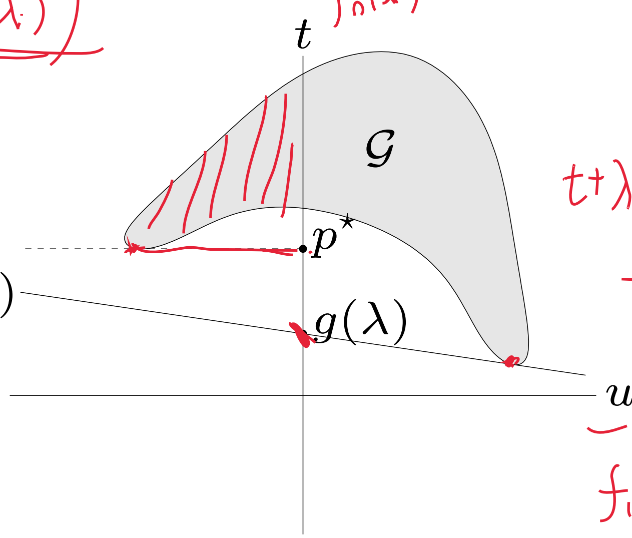
$$g(\lambda) = \inf_{(u,t)\in\mathcal{G}} (t + \lambda u), \qquad \text{where} \quad \mathcal{G} = \{(\underbrace{f_1(x)}_{u}, \underbrace{f_0(x)}_{t}) \mid x \in \mathcal{D}\}$$

$t + \lambda u \ge g(\lambda)$

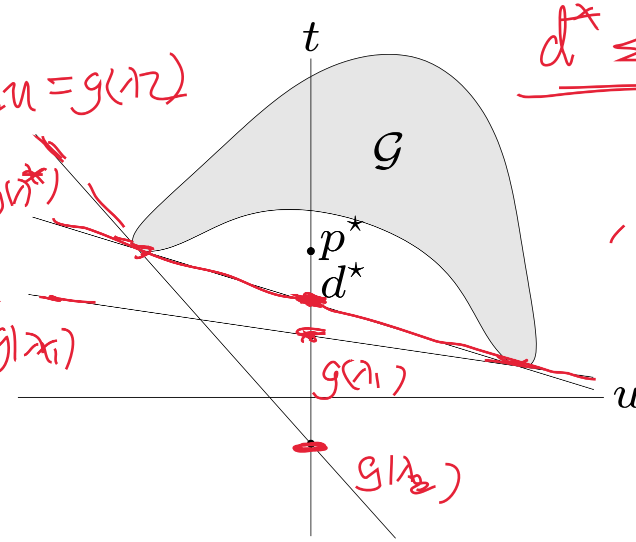
$\lambda u + t = g(\lambda)$

$\{\lambda, 1\} \cdot \begin{bmatrix} u \\ t \end{bmatrix}$

non-vertical



$d^\star \le p^\star$

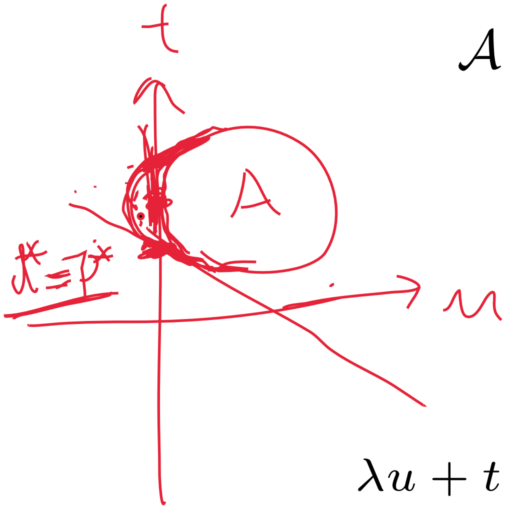


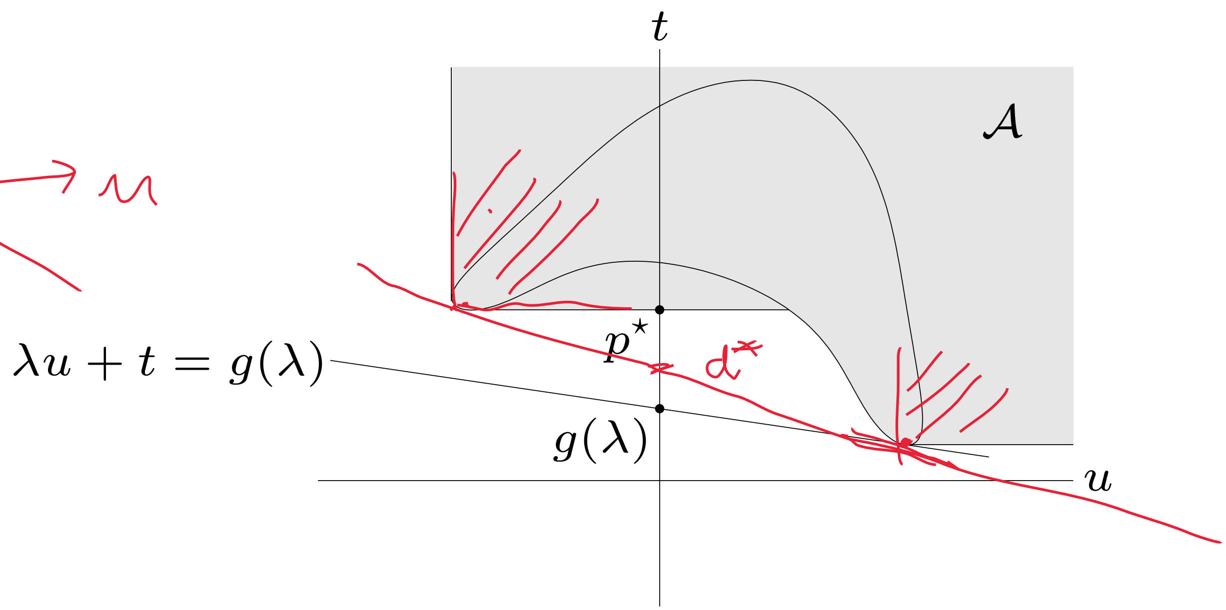
$\lambda \uparrow \quad g(\lambda) \uparrow$

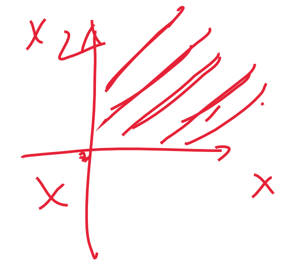
- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t-axis at $t = g(\lambda)$

epigraph variation: same interpretation if $\mathcal{G}$ is replaced with

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







$\{(u,t) \mid x_1 \ge u,\, x_2 \ge t\}$

strong duality

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

Complementary slackness

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

$p^\star = d^\star$

$$
\begin{aligned}
f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) = f_0(x^\star) \\
&\leq f_0(x^\star) + \underbrace{\sum_{i=1}^m \lambda_i^\star f_i(x^\star)}_{\leq 0} + \underbrace{\sum_{i=1}^p \nu_i^\star h_i(x^\star)}_{= 0} \\
&\leq f_0(x^\star)
\end{aligned}
$$

$f_0(x^\star) \leq f_0(x^\star) + 0 \leq f_0(x^\star)$

hence, the two inequalities hold with equality

② • $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$

① • $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0$$

Karush-Kuhn-Tucker (KKT) conditions

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

Strong duality holds + differentiable

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$
2. dual constraints: $\lambda \succeq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$\nabla_x L = 0$

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

from page 5–17: if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

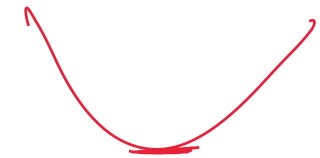
Strong duality + differentiable.

x, λ, ν are optimal $\Longrightarrow$ KKT holds

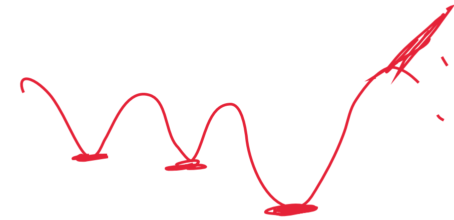
$\cancel{\Longleftarrow}$

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, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$ $\quad d^* = p^*$

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

x is optimal if and only if there exist λ, ν 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

Slater's $\Rightarrow$ (Strong duality) + differentiable

x, λ, ν optimal $\Leftrightarrow$ KKT holds

(for convex optimization)

example: water-filling (assume $\alpha_i > 0$)

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^n \log(x_i + \alpha_i) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

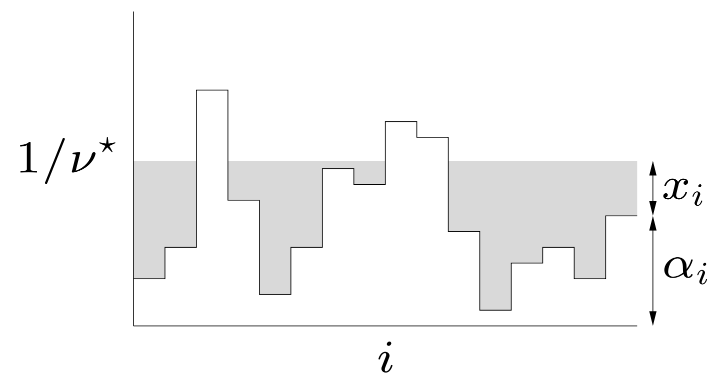
x is optimal iff $x \succeq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbf{R}^n$, $\nu \in \mathbf{R}$ such that

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

- 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 ν from $\mathbf{1}^T x = \sum_{i=1}^n \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

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



Perturbation and sensitivity analysis

(unperturbed) optimization problem and its dual

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \le 0, & i = 1, \ldots, m \\ & h_i(x) = 0, & i = 1, \ldots, p \end{array} \qquad \begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

$p^*(0,0)$

u_i >0, relax / loosen
<0, restrict / tight

perturbed problem and its dual

u: tighting force

$$\begin{array}{lll} \text{min.} & f_0(x) & \\ \text{s.t.} & f_i(x) \le u_i, & i = 1, \ldots, m \\ & h_i(x) = v_i, & i = 1, \ldots, p \end{array} \qquad \begin{array}{ll} \text{max.} & g(\lambda, \nu) - u^T\lambda - v^T\nu \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

$p^*(u,v)$

v: shifting term

- x is primal variable; u, v are parameters
- $p^\star(u, v)$ is optimal value as a function of u, v
- we are interested in information about $p^\star(u, v)$ 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 $\lambda^\star$, $\nu^\star$ are dual optimal for unperturbed problem

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

$d^* = p^*$

apply weak duality to perturbed problem:

$$
\begin{aligned}
p^\star(u, v) &\geq g(\lambda^\star, \nu^\star) - u^T\lambda^\star - v^T\nu^\star \\
&= p^\star(0, 0) - u^T\lambda^\star - v^T\nu^\star
\end{aligned}
$$

lower bound of $p^*(u,v)$

sensitivity interpretation

- if $\lambda_i^\star$ large: $p^\star$ increases greatly if we tighten constraint i ($u_i < 0$)
- if $\lambda_i^\star$ small: $p^\star$ does not decrease much if we loosen constraint i ($u_i > 0$)
- if $\nu_i^\star$ large and positive: $p^\star$ increases greatly if we take $v_i < 0$;
  if $\nu_i^\star$ large and negative: $p^\star$ increases greatly if we take $v_i > 0$
- if $\nu_i^\star$ small and positive: $p^\star$ does not decrease much if we take $v_i > 0$;
  if $\nu_i^\star$ small and negative: $p^\star$ does not decrease much if we take $v_i < 0$

local sensitivity: if (in addition) $p^\star(u,v)$ is differentiable at $(0,0)$, then

$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \qquad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}$$

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

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

$p^\star(te_i, 0) \geq p^\star(0,0) - t e_i^T \lambda_i^\star$

$t \to 0^+$ $(t>0)$

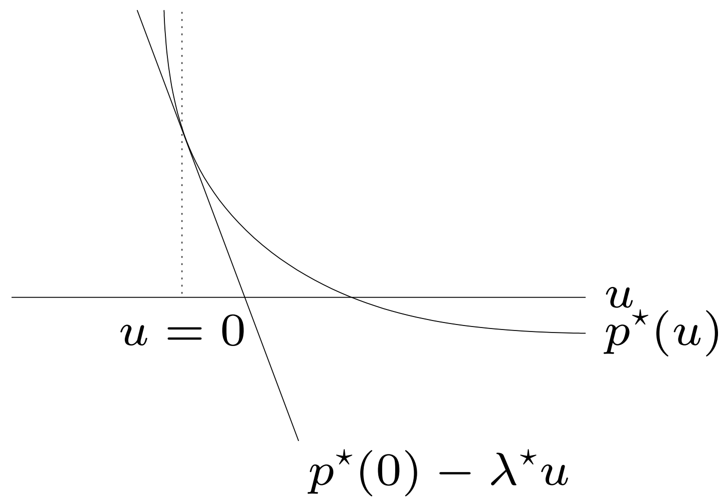
$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t \nearrow 0} \frac{p^\star(te_i,0) - p^\star(0,0)}{t} \leq -\lambda_i^\star$$

$t \to 0^-$

$(t<0)$

hence, equality

$p^\star(u)$ 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 ϕ convex, increasing

Introducing new variables and equality constraints

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

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

reformulated problem and its dual

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax+b-y=0 \end{array} \qquad \begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0 \end{array}$$

dual function follows from

$$\begin{aligned} g(\nu) &= \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T\nu) \\ &= \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

norm approximation problem: minimize $\|Ax - b\|$

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & y = Ax - b \end{array}$$

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

$$\begin{aligned} g(\nu) &= \inf_{x,y}(\|y\| + \nu^T y - \nu^T Ax + b^T\nu) \\ &= \begin{cases} b^T\nu + \inf_y(\|y\| + \nu^T y) & A^T\nu = 0 \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} b^T\nu & A^T\nu = 0, \quad \|\nu\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

(see page 5–4)

dual of norm approximation problem

$$\begin{array}{ll} \text{maximize} & b^T\nu \\ \text{subject to} & A^T\nu = 0, \quad \|\nu\|_* \leq 1 \end{array}$$

Implicit constraints

LP with box constraints: primal and dual problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T\nu - \mathbf{1}^T\lambda_1 - \mathbf{1}^T\lambda_2 \\ \text{subject to} & c + A^T\nu + \lambda_1 - \lambda_2 = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0 \end{array}$$

reformulation with box constraints made implicit

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

dual function

$$\begin{aligned} g(\nu) &= \inf_{-\mathbf{1} \preceq x \preceq \mathbf{1}} (c^T x + \nu^T(Ax - b)) \\ &= -b^T\nu - \|A^T\nu + c\|_1 \end{aligned}$$

dual problem: maximize $-b^T\nu - \|A^T\nu + c\|_1$

Problems with generalized inequalities

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

$\preceq_{K_i}$ is generalized inequality on $\mathbf{R}^{k_i}$

definitions are parallel to scalar case:

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

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

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

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

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

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

dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda_1, \ldots, \lambda_m, \nu) \\ \text{subject to} & \lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m \end{array}$$

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

Semidefinite program

primal SDP $(F_i, G \in \mathbf{S}^k)$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n \preceq G \end{array}$$

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

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

dual SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{tr}(GZ) \\ \text{subject to} & Z \succeq 0, \quad \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \end{array}$$

$p^\star = d^\star$ if primal SDP is strictly feasible ($\exists x$ with $x_1 F_1 + \cdots + x_n F_n \prec G$)