

# Optimization for the information and data sciences

Mahdi Soltanolkotabi

Duality

Ming Hsieh Department of Electrical Engineering



**USC** University of  
Southern California

# 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)

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

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

**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:**  $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ ,

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

$g$  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

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

## dual function

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

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

- plug in in  $L$  to obtain  $g$ :

$$g(\nu) = L((-1/2)A^T \nu, \nu) = -\frac{1}{4}\nu^T AA^T \nu - b^T \nu$$

a concave function of  $\nu$

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

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

# Equality constrained norm minimization

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

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

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  $x = 0$
- if  $\|y\|_* > 1$ , choose  $x = tu$  where  $\|u\| \leq 1$ ,  $u^T y = \|y\|_* > 1$ :

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

**lower bound property:**  $p^* \geq b^T \nu$  if  $\|A^T \nu\|_* \leq 1$

# Two-way partitioning

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

- a nonconvex problem; feasible set contains  $2^n$  discrete points
- interpretation: partition  $\{1, \dots, n\}$  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

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

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

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



# Lagrange dual and conjugate function

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

## dual function

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

- 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

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

# The dual problem

## Lagrange dual problem

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

- 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  $\lambda \succeq 0, (\lambda, \nu) \in \mathbf{dom}g$
- often simplified by making implicit constraint  $(\lambda, \nu) \in \mathbf{dom}g$  explicit

**example:** standard form LP and its dual (page 6)

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

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

# 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

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

gives a lower bound for the two-way partitioning problem on page 8

**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

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

if it is strictly feasible, i.e.,

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

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

# Inequality form LP

## primal problem

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

## dual function

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

## dual problem

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

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

# Quadratic program

**primal problem** (assume  $P \in \mathcal{S}_{++}^n$ )

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

**dual function**

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

**dual problem**

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

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

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

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

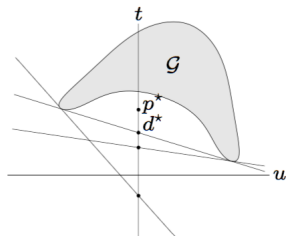
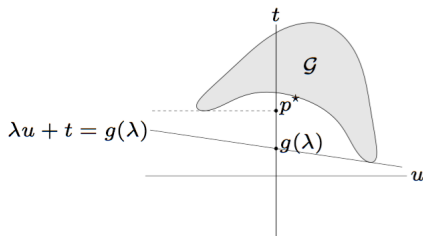
strong duality although primal problem is not convex (not easy to show)

# Geometric interpretation

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

**interpretation of dual function:**

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

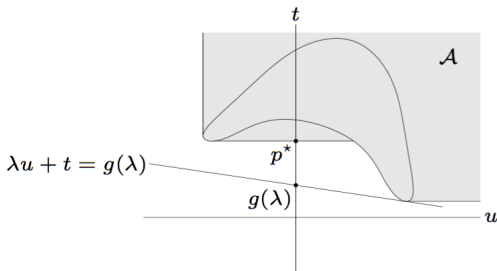


- $\lambda u + t = g(\lambda)$  is (non-vertical) supporting hyperplane to  $\mathcal{G}$
- hyperplane intersects  $t$ -axis at  $t = g(\lambda)$



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

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



### 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

# Complementary slackness

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

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

hence, the two inequalities hold with equality

- $x^*$  minimizes  $L(x, \lambda^*, \nu^*)$
- $\lambda_i^* f_i(x^*) = 0$  for  $i = 1, \dots, m$  (known as complementary slackness):

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

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

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

from page 18: if strong duality holds and  $x, \lambda, \nu$  are optimal, then they must satisfy the KKT conditions

# 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  $\alpha_i > 0$ )

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

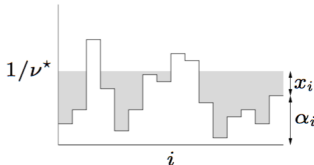
$x$  is optimal iff  $x \succeq 0$ ,  $\mathbf{1}^T x = 1$ , and there exist  $\lambda \in \mathbb{R}^n$ ,  $\nu \in \mathbb{R}$  such that

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

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

## interpretation

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



# Perturbation and sensitivity analysis

## (unperturbed) optimization problem and its dual

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

## perturbed problem and its dual

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

- $x$  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

## 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:

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

## 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

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

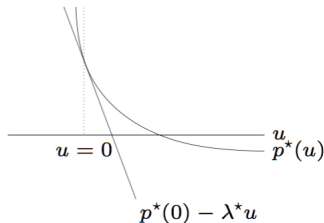
proof (for  $\lambda_i^*$ ): from global sensitivity result,

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

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

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., 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

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax + b - y = 0 \end{array}$$

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

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)

**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  $-b^T \nu - \|A^T \nu + c\|_1$

# Feasibility problems

**feasibility problem A** (variables  $x \in \mathbb{R}^n$ )

$$f_i(x) < 0, \quad i = 1, \dots, m, \quad h_i(x) = 0, \quad i = 1, \dots, p$$

**feasibility problem B** (variables  $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$ )

$$\lambda \succeq 0, \quad \lambda \neq 0, \quad g(\lambda, \nu) \geq 0$$

where  $g(\lambda, \nu) = \inf_x (\sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x))$

- feasibility problem B is convex ( $g$  is concave), even if problem A is not
- A and B are always **weak alternatives**: at most one is feasible  
proof: assume  $\tilde{x}$  satisfies A,  $\lambda, \nu$  satisfy B

$$0 \leq g(\lambda, \nu) \leq \sum_{i=1}^m \lambda_i f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) < 0$$

# Theorem of alternatives

A and B are **strong alternatives** if exactly one of the two is feasible

- can prove infeasibility of A by producing solution of B and vice-versa

# Problems with generalized inequalities

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

$\preceq_{K_i}$  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  $L : \mathbb{R}^n \times \mathbb{R}^{k_1} \times \dots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \rightarrow \mathbb{R}$ , is defined as

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

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

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

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

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

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

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

**dual problem**

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

- 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** ( $F_i, G \in \mathcal{S}^k$ )

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

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

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

**dual SDP**

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

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