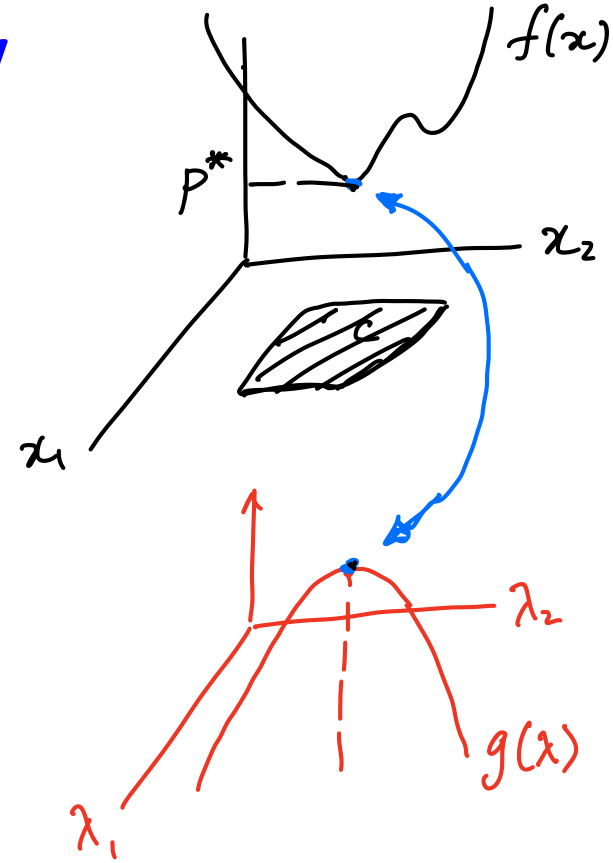
7. Duality



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

Lagrangian

standard form problem (not necessarily convex)

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

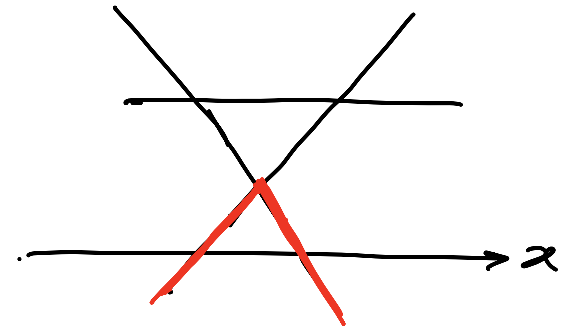
variable $x \in \mathbf{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

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

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

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

Lagrange dual function



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

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

pointwise min of family of convex (affine) functions

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

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

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

$f_i(\tilde{x}) \leq 0, \quad h_i(\tilde{x}) = 0$

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

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

Least-norm solution of linear equations

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^T x = \|x\|_2^2 \\ \text{subject to} & Ax = b \end{array}$$

$L(x, \nu) = x^T x + \sum \nu_i (a_i^T x - b_i)$

$\nu^T(Ax - b)$

dual function

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

$g(\nu) = \inf_{x \in \mathbb{R}^n} L(x, \nu)$

- 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:

$$g(\nu) = L(\underbrace{(-1/2)A^T \nu}, \nu) = -\frac{1}{4}\nu^T \overbrace{A A^T}^{\succeq 0} \nu - b^T \nu$$

concave quadratic fct of ν, dom $g = \mathbb{R}^n$

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

Standard form LP

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

$-x \preceq 0$

dual function

- Lagrangian is

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + \underbrace{(c + A^T \nu - \lambda)}_{=0}{}^T x \end{aligned}$$

- L is linear in x, hence

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

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

$, \lambda \succeq 0$ $\Big\}$ $A^T\nu + c \succeq 0$

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 $A^T \nu + c \succeq 0$

$p^\star \geq g(\lambda, \nu)$ for any $\lambda, \nu \in \mathrm{dom}\, g$

Equality constrained norm minimization

general norm

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

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

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 $\|\cdot\|$

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 $\|u\| \le 1$, $u^T y = \|y\|_* > 1$:

$|t|\,\|u\| - t u^T y = t\|u\| - t\|y\|_*$

$$\|x\| - y^T x = t(\underbrace{\|u\| - \|y\|_*}_{\le 0}) \to -\infty \quad \text{as } t \to \infty$$

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

$$L(x,\nu) = \|x\| - (A^T\nu)^T x + b^T\nu$$

$$g(\nu) = \inf_x L(x,\nu) = b^T\nu + \inf_x \left(\widetilde{\|x\|} - \overbrace{(\underbrace{A^T\nu}_{y})^T x} \right) = b^T\nu + \begin{cases} 0, & \|A^T\nu\|_* \le 1 \\ -\infty, & \text{other} \end{cases}$$

<u>dual norm of $\|x\|$:</u> $\|y\|_* = \sup_{\|x\| \le 1} y^T x$

$|x^T y| \le \|x\| \underbrace{\|y\|_*}_{\le 1}$ (generalization of Cauchy-Schwarz & Holder ineq)

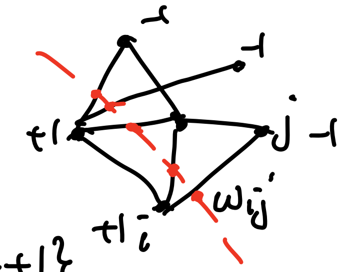
$x^T y \le \|x\|$ if $\|y\|_* \le 1$

$\|x\| - x^T y \ge 0$

Goemans & Williamson on max-cut (late 90's)

- max-cut problem
- Ising model "

Example: two-way partitioning



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

$x_i \in \{-1, +1\}$

- 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

$$= -\mathbf{1}^T \nu + x^T \begin{bmatrix} \nu_1 & & 0 \\ & \ddots & \\ 0 & & \nu_n \end{bmatrix} x$$

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

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

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

$$W - \begin{bmatrix} \lambda_{\min}(W) & & \\ & \ddots & \\ & & \lambda_{\min}(W) \end{bmatrix} = W - \lambda_{\min}(W) I$$

$$\lambda_i(W - \lambda_{\min}(W) I) = \lambda_i(W) - \lambda_{\min}(W) \geq 0$$

bound: $-1^T \nu = n\,\lambda_{min}(W)$

$$\begin{cases} \min\limits_{x} & x^T W x \\ \text{s.t.} & x_i^2 = 1 \end{cases}$$

relax

$$\begin{cases} \min\limits_{x} & x^T W x \\ \text{s.t.} & \sum x_i^2 = n \Leftrightarrow \|x\|_2 = \sqrt{n} \end{cases}$$

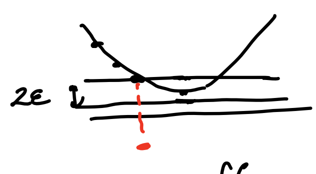
$$\begin{cases} \min\limits_{x} & x^T W x = n\,\lambda_{min}(W) \\ \text{s.t.} & \|x\|_2 = \sqrt{n} \end{cases}$$

Announcements

- For midterm, bring a 'blue book' or some sheets of paper (that you clip or staple). Submit your cheat sheets.
- Main results from lecture can be quoted & used without re-proving (e.g. properties that preserve convexity, convexity of basic sets / fcts shown in lecture, key theorems, etc.)

importance & use of duality: - physical interpretation

- lower bd on hard non-convex problems
- primal/dual algorithm for convex
  solve primal & dual prob's together; good stopping criterion



$x_{t+1} = f(x_t, \cdots)$

The dual problem

primal problem

Lagrange dual problem

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

- 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 7–5)

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

dual LP:

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

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

  $$g(\nu) = \begin{cases} -1^T\nu, & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty, & \text{other} \end{cases}$$

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

  (Goemans-Williamson)

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

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

duality gap $= p^* - d^*$

- 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

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

if it is strictly feasible, *i.e.*, $\mathcal{D}$ = intersection of domains of $f_0, f_1, \ldots, f_m$

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

- also guarantees that the dual optimum is attained (if $p^\star > -\infty$)

FYI

- can be sharpened: (*e.g.*, can replace $\mathbf{int}\,\mathcal{D}$ with relative interior; linear inequalities do not need to hold with strict inequality, etc. . .)
- there exist many other types of constraint qualifications

Inequality form LP

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

primal problem

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

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

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

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

- 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 $P \in \mathbf{S}^n_{++}$)

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

dual function

$$g(\lambda) = \inf_x \left(x^T P x + \underbrace{\lambda^T (Ax - b)}_{(A^T\lambda)^T x - b^T\lambda} \right) = -\frac{1}{4} \lambda^T \underbrace{A \overbrace{P^{-1}}^{\succ 0} A^T}_{\succeq 0} \lambda - b^T \lambda$$

dual problem $\quad \nabla_x L(x,\lambda) = 0 \qquad 2Px + A^T\lambda = 0 \Rightarrow x = -\frac{1}{2} P^{-1} A^T \lambda$

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

- 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

Economic (price) interpretation

e.g., $a_1x_1 + x_2 \le 10$ ← resource limit

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

$f_0(x)$ is cost of operating firm at operating condition x; constraints give resource limits.

suppose:

- can violate $f_i(x) \le 0$ by paying additional cost of λ_i (in dollars per unit violation), *i.e.*, incur cost $\lambda_i f_i(x)$ if $f_i(x) > 0$
- can sell unused portion of ith constraint at same price, *i.e.*, gain profit $\lambda_i f_i(x)$ if $f_i(x) < 0$

total cost to firm to operate at x at constraint prices λ_i: $\lambda_i \geq 0$

$$L(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$$

interpretations:

dual function: $g(\lambda) = \inf_x L(x, \lambda)$ is optimal cost to firm at constraint prices λ_i

weak duality: cost can be lowered if firm allowed to pay for violated constraints (and get paid for non-tight ones)

duality gap: advantage to firm under this scenario

strong duality: $\lambda^\star$ give prices for which firm has no advantage in being allowed to violate constraints . . .

market-clearing prices

$$\left[\begin{array}{ll} \max & g(\lambda) \\ \text{s.t.} & \lambda \succeq 0 \end{array}\right. \Rightarrow \lambda^*$$

Min-max & saddle-point interpretation

$$\begin{bmatrix} \min_{x} & f_0(x) \\ \text{s.t.} & f_i(x) \le 0 \end{bmatrix}$$

$i=1,\dots,m$

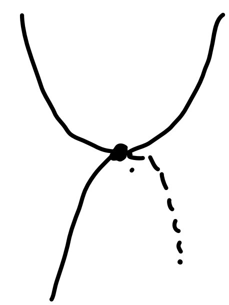
can express primal and dual problems in a more symmetric form:

$$\sup_{\lambda \succeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) = \begin{cases} \overbrace{f_0(x)}^{\text{obj}} & \overbrace{f_i(x) \le 0}^{\text{constraints}}, \\ +\infty & \text{otherwise,} \end{cases}$$

so $p^\star = \inf_x \sup_{\lambda \succeq 0} L(x, \lambda)$.

also by definition, $d^\star = \sup_{\lambda \succeq 0} \underbrace{\inf_x L(x, \lambda)}_{g(\lambda)}$.

$L(x,\lambda)$



weak duality can be expressed as

$$\sup_{\lambda \succeq 0} \inf_x L(x, \lambda) \le \inf_x \sup_{\lambda \succeq 0} L(x, \lambda)$$

strong duality when equality holds. means can switch the order of minimization over x and maximization over $\lambda \succeq 0$.

if $x^\star$ and $\lambda^\star$ are primal and dual optimal and strong duality holds, they form a saddle-point for the Lagrangian (converse also true).

game theory

Example problems for review:

1. Is this set convex:

$$S = \{ (a,b,c) \in \mathbb{R}^3 \mid ax^p + bx + c = 0 \text{ has no real solutions}, c > 0 \}$$

$p \geq 2$ (integer)

$ax^p + bx + \underset{>0}{c} = 0$ has no real solutions $\Leftrightarrow$ $ax^p + bx + c > 0 \quad \forall x \in \mathbb{R}$

for every x, a half-space in a, b, c

$\Rightarrow$ S = intersection of half-spaces $\Rightarrow$ convex.

2. Is $f(X) = \det(X)$ $\quad \mathrm{dom} f = S^n_{++}$ convex, concave, neither?

$\log\det X$ concave $\rightarrow$ $f(X) = \exp(\underbrace{\log\det X}_{\text{concave}})$ $\Rightarrow$ composition rules don't work here!

try to find a counter-example:

$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$ even simpler: $X = \begin{bmatrix} x & 0 \\ 0 & z \end{bmatrix}$ $\quad \det X = f(x,z) = xz$

$x > 0$
$z > 0$

$f(x,z) = xz$, $x > 0$, $z > 0$ not convex nor concave: Hessian $\nabla^2 f(x)$ is indefinite

$\begin{bmatrix} & \\ & \end{bmatrix} \overset{?}{\gtrless} 0$

determinant
Trace
$\Rightarrow$ eig's are both positive & negative

($f(X) = (\det(X))^{1/n}$ isn't helpful either)

Complementary slackness

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

$$\begin{aligned} f_0(x^\star) = g(\lambda^\star,\nu^\star) &= \inf_x \left(f_0(x)+\sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x)\right) \rightarrow L(x,\lambda,\nu)\Big|_{\lambda^*,\nu^*} \\ &\overset{=}{\le} f_0(x^\star) + \underbrace{\sum_{i=1}^m \underbrace{\lambda_i^\star}_{\ge 0} \underbrace{f_i(x^\star)}_{\le 0}}_{\le 0} + \sum_{i=1}^p \nu_i^\star \overset{0}{\cancel{h_i(x^\star)}} \rightarrow L(x^*,\lambda^*,\nu^*) \\ &\overset{=}{\le} f_0(x^\star) \end{aligned}$$

hence, the two inequalities hold with equality

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

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

$$\left[\begin{array}{l} \text{min. } f_0(x) \\ \text{s.t. } f_1(x) \le 0 \rightarrow \lambda_1 > 0 \Rightarrow f_1(x) = 0 \text{, tight} \\ \quad\;\; f_2(x) \le 0 \end{array}\right.$$

Announcements

- HW6 due tonight, HW7 will be posted tonight
- Midterms being graded (will post sol's & grades by end of week)
- CVX / CVXPY (examples from the book on CVX website)
- Final exam: ~10-hr take home,
  can be taken on Sat March 9 or Sun March 10.

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: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$
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:

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

from page 7–16: if strong duality holds and x, λ, ν 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): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

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

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

$\nabla_x L(x, \tilde{\lambda}, \tilde{\nu})\Big|_{x=\tilde{x}} = 0$ & L is convex in x $\Rightarrow$ $\tilde{x}$ minimize $L(x, \tilde{\lambda}, \tilde{\nu})$

$L(x, \tilde{\lambda}, \tilde{\nu}) = f_0(x) + \sum \tilde{\lambda}_i f_i(x) + \sum \tilde{\nu}_i h_i(x)$

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

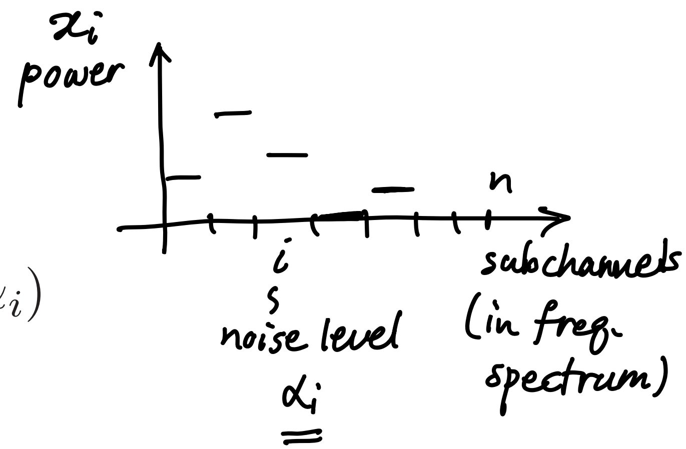
- if **Slater's condition** is satisfied:

  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

KKT conditions

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



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

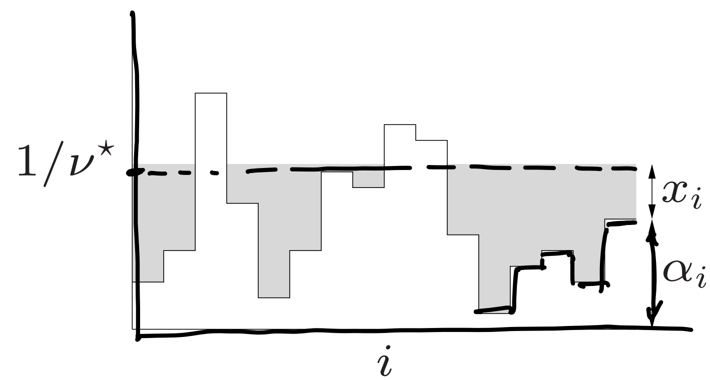
x is optimal iff $x \succeq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbf{R}^n$, $\nu \in \mathbf{R}$ such that

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

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

interpretation

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



KKT conditions:

1) x^* primal feasible : $x^* \geq 0 \qquad 1^T x^* = 1$

2) λ^*, ν^* dual " : $\lambda^* \geq 0$

$\rightarrow$ 3) $\lambda_i^* x_i^* = 0$

4) $\nabla_x L(x, \lambda^*, \nu^*) = 0$

$$L(x, \lambda^*, \nu^*) = -\sum_{i=1}^{n} \log(x_i + \alpha_i) - \lambda^T x + \nu^T (1^T x - 1)$$

$$\nabla_{x_i} L(x, \lambda^*, \nu^*) = -\frac{1}{x_i + \alpha_i} - \lambda_i^* + \nu^*$$

$$\nabla_{x_i} L \Big|_{x_i = x_i^*} = 0 \implies \nu^* = \lambda_i^* + \frac{1}{x_i^* + \alpha_i}$$

from (3) : $\lambda_i^* x_i^* = 0$

- $x_i^* = 0 \rightarrow \nu^* = \lambda_i^* + \frac{1}{\alpha_i} \implies \lambda_i^* = \nu^* - 1/\alpha_i$

  from $\lambda_i^* \geq 0$: $\nu^* \geq 1/\alpha_i$

- $\lambda_i^* = 0 \Rightarrow \nu^* = \frac{1}{x_i^* + \alpha_i} \implies x_i^* = \frac{1}{\nu^*} - \alpha_i$

  from $x_i^* \geq 0$: $\nu^* \leq 1/\alpha_i$

So x_i^* is either zero or equal to $1/\nu^* - \alpha_i$: $x_i^* = \max\{0, 1/\nu^* - \alpha_i\}$

$1^T x^* = 1$: $\displaystyle\sum_{i=1}^{n} \min\{0, 1/\nu^* - \alpha_i\} = 1 \rightarrow$ scalar valued equation in ν^*

Perturbation and sensitivity analysis

(unperturbed) optimization problem and its dual

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

perturbed problem and its dual

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

- 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

apply weak duality to perturbed problem:

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

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

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

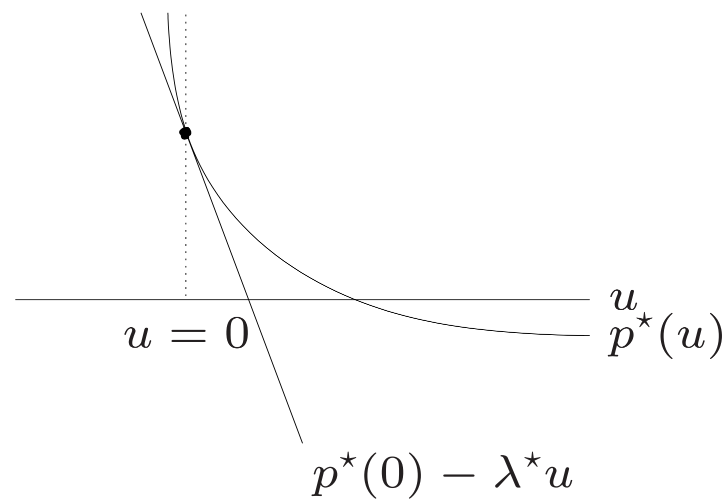
proof (for $\lambda_i^\star$): from global sensitivity result,

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

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

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

$$\left[\begin{array}{ll}\underset{x}{\text{minimize}} & f_0(\underbrace{Ax+b}_{=y})\end{array}\right.$$

- 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

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

dual function follows from

$$L(x,y,\nu) = \nu^T A x + (f_0(y) - \nu^T y) + b^T\nu$$

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

norm approximation problem: $\text{minimize}_{x}\ \|Ax - b\|$

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

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

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

(see page 7–4)

dual of norm approximation problem

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

Implicit constraints

LP with box constraints: primal and dual problem

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

reformulation with box constraints made implicit (box constraints are interpreted as being part of domain of obj)

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

dual function

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

$(c + A^T\nu)_i > 0 \to x_i = -1$

$(c + A^T\nu)_i < 0 \to x_i = 1$

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

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

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

- 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

$$K^* = \{ y \mid \underbrace{\langle y, x \rangle \geq 0}, \forall\, x \in K \}$$

$$\begin{aligned} f_0(\tilde{x}) &\geq f_0(\tilde{x}) + \sum_{i=1}^{m} \underbrace{\overset{K^*}{\lambda_i^T} f_i(\tilde{x})}_{\leq 0} + \sum_{i=1}^{p} \nu_i \cancelto{0}{h_i(\tilde{x})} \\ &\geq \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu) \\ &= g(\lambda_1, \ldots, \lambda_m, \nu) \end{aligned}$$

$\lambda_i \in K_i^*$

$\lambda_i \succeq_{K_i^*} 0$

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

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

$$-G + \sum x_i F_i \preceq_{S^n_+} 0$$

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

$$= \sum c_i x_i + \sum x_i \,\mathrm{tr}(Z F_i) - \mathrm{tr}\, Z G$$

$$= \sum (c_i + \mathrm{tr}(Z F_i)) x_i - \mathrm{tr}(Z G)$$

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

dual SDP

$$Z \in K^* \Rightarrow Z \succeq 0$$

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

Slater cond.:

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

A nonconvex problem with strong duality

skip

appendix B of textbook

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^T A x + 2b^T x \\ \text{subject to} & \underline{x^T x \le 1} \quad \|x\| \le 1 \end{array}$$

aka:
- lossless S-procedure (control) (lossless convexification)
- trust-region subprob. (optimization)

$\underline{A \not\succeq 0}$, 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:

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

strong duality *although primal problem is not convex* (not easy to show; if interested, see appendix B in book)