Duality

A supplementary note to Chapter 5 of *Convex Optimization* by S. Boyd and L. Vandenberghe

Optimization Lab.

IE department
Seoul National University

8th October 2009

Recall our optimization, $\min\{f_0(x)|\ f_i(x) \leq 0, i = 1, \cdots, m,\ h_j(x) = 0, j = 1, \ldots, p\}$ and $\mathcal{D} = \bigcap_{i=0}^{n} \mathbf{dom} f_i \cap \bigcap_{i=1}^{p} \mathbf{dom} h_i$.

Definition

Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is

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

where λ and ν called *dual variables* or *Lagrange multipliers*.

Definition

Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x) \right).$$

Thus g is pointwise infimum of affine functions of (λ, ν).

Lemma

For any $\lambda \geq 0$ and ν, $g(\lambda, \nu) \leq p^*$.

Proof For any feasible x, $\lambda \geq 0$, and ν, we have $\sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p \nu_j h_j(x) \leq 0$, and hence $L(x, \lambda, \nu) \leq f_0(x)$. Therefore, we have

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(x, \lambda, \nu) \leq f_0(x). \square$$

Dual variables (λ, ν) are called *dual feasible* when $\lambda \geq 0$, and $g(\lambda, \nu) > -\infty$.

Linear approximation interpretation

Notice our optimization is equivalent to

$$\min f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{j=1}^{p} I_0(h_j(x)),$$

if I_- and I_0 satisfy $I_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases}$, and $I_0(u) = \begin{cases} 0 & u = 0 \\ \infty & u \neq 0 \end{cases}$.
If we replace $I_-(u)$ and $I_0(u)$ with $\lambda_i u$ and $\mu_i u$ respectively, then we get Lagrange dual function

$$\min L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x).$$

Since $\lambda_i u \leq I_-(u)$ and $\nu_i u \leq I_0(u)$ for all u, dual function yields a lower bound on optimal value.

Least-squares solution of linear system

$$\begin{array}{ll} \min & x^T x \\ \text{s.t.} & Ax = b \end{array} \rightarrow L(x,\nu) = x^T x + \nu^T (Ax - b).$$

$L(x,\nu)$ is convex quadratic in x, and infimum attains when $\nabla_x L(x,\nu) = 2x + A^T \nu = 0$ or $x = -(1/2)A^T \nu$. Thus Lagrange dual is

$$g(\nu) = \inf_x L(x,\nu) = L(-(1/2)A^T\nu, \nu) = -(1/4)\nu^T A A^T \nu - b^T \nu.$$

Thus $-(1/4)\nu^T A A^T \nu - b^T \nu \leq \inf\{x^T x | Ax = b\}$ for all feasible pair (x,ν).

Standard form LP

$$\begin{array}{ll} \min & c^Tx \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array} \rightarrow \begin{array}{rcl} L(x,\lambda,\nu) & = & c^Tx - \sum_{i=1}^{n} \lambda_i x_i + \nu^T(Ax - b) \\ & = & -b^T\nu + (c + A^T\nu - \lambda)x. \end{array}$$

Since $g(\lambda,\nu)$ is pointwise infimum of x, we have the following:

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

Two-way partitioning

$$\begin{array}{ll} \min & x^T W x \\ \text{s.t.} & x_j^2 = 1,\ j = 1, \cdots, n \end{array}$$

W_{ij} and $-W_{ij}$, resp. are costs of having i and j in same set and different sets in partition. An NP-hard problem.
Can obtain lower bounds on optimal value from Lagrange dual:

$$\begin{array}{rcl} L(x,\nu) & = & x^T W x + \sum_{j=1}^{n} \nu_j (x_j^2 - 1) = x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu \\ \to \quad g(\nu) & = & \inf_x x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu \\ & = & \begin{cases} -\mathbf{1}^T \nu & W + \text{diag}(\nu) \geq 0 \\ -\infty & \textit{otherwise} \end{cases} \end{array}$$

For example, $\nu = -\lambda_{min}(W)\mathbf{1}$ is dual feasible yields a bound,

$$p^* \geq -\mathbf{1}^T \nu = n\lambda_{min}(W).$$

- Recall conjugate function $f^*(y) = \sup_{x \in \mathrm{dom} f} \{y^T x - f(x)\}$.
- Dual and conjugate: a simple case

$$\begin{array}{lll} \min & f(x) & \Rightarrow \\ \text{s.t.} & x = 0 & \end{array} \quad \begin{array}{rcl} L(x,\nu) & = & f(x) + \nu^T x \\ g(\nu) & = & \inf_x\{f(x) + \nu^T x\} = -\sup_x\{(-\nu)^T x - f(x)\} \\ & = & -f^*(-\nu). \end{array}$$

- Dual and conjugate: for linear constraints

$$\begin{array}{lll} \min & f_0(x) & \Rightarrow \\ \text{s.t.} & Ax \le b & \\ & Cx = d & \end{array} \quad \begin{array}{rcl} g(\lambda,\nu) & = & \inf_x\{f_0(x) + \lambda^T(Ax - b) + \nu^T(Cx - d)\} \\ & = & -b^T\lambda - d^T\nu + \inf_x\{f_0(x)(A^T\lambda + C^T\nu)^T x\} \\ & = & -b^T\lambda - d^T\nu + f_0^*(-A^T\lambda - C^T\nu), \end{array}$$

where $\mathrm{dom} g = \{(\lambda,\nu) | -A^T\lambda - C^T\nu \in \mathrm{dom} f_0^*\}$.

- Equality constrained norm minimization

$$\begin{array}{ll} \min & \|x\| \\ \text{s.t.} & Ax = b, \end{array} \quad \text{Then } f_0^*(y) = \sup_x \{y^T x - f(x)\} = \begin{cases} 0 & \text{if } \|y\|_* \leq 1, \\ \infty & \text{otherwise.} \end{cases}$$

Therefore, we have the following dual function:

$$g(\nu) = -b^T\nu - f_0^*(-A^T\nu) = \begin{cases} -b^T\nu & \|A^T\nu\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

- Entropy maximization

$$\begin{array}{ll} \min & f_0(x) = \sum_i x_i \log x_i \\ \text{s.t.} & Ax \leq b, \mathbf{1}^T x = 1, \end{array} \quad \text{where } \mathbf{dom} f_0 = \mathbb{R}^n_{++}$$

Since $f_0^*(y) = \sum_i e^{y_i - 1}$, we have dual function

$$g(\lambda, \nu) = -b^T\lambda - \nu - \sum_{i=1}^n e^{-a_i^T\lambda - \nu - 1} = -b^T\lambda - \nu - e^{-\nu-1}\sum_{i=1}^n e^{-a_i^T\lambda}.$$

- Minimum volume covering ellipsoid

$$\begin{array}{ll} \min & f_0(X) = \log\det X^{-1} \\ \text{s.t.} & a_i^T X a_i \leq 1, i = 1, \cdots, m, \end{array} \quad \text{where } \operatorname{dom} f_0 = \mathbb{S}_{++}^n$$

When a solid S is linearly transformed into AS,

$$\operatorname{vol}(AS) = \det(A^T A)^{-1/2} \operatorname{vol}(S).$$

Consider ellipsoid $E_X = \{z | z^T X z \leq 1\}$, image of linear transform X of unit circle. Volume of E_X is proportional to $(\det X^{-1})^{1/2}$. Therefore, via this optimization, can obtain a min vol ellipsoid including a_1, ..., a_n.

- What is the *best* lower bound from Lagrange dual function?

$$\begin{array}{ll} \text{max} & g(\lambda,\nu) \\ \text{s.t.} & \lambda \geq 0. \end{array}$$

  Similarly, we can define the dual feasibility and the dual optimality, and we have

$$\text{dom}g = \{(\lambda,\nu)|g(\lambda,\nu) > -\infty\}.$$

- Lagrange dual problem is a convex optimization problem regardless of convexity of original problem (or primal problem).

Lagrange dual of LP

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, x \geq 0 \end{array} \rightarrow g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Unless $A^T \nu - \lambda + c = 0$, g is infeasible, so Lagrange dual problem can be represented as

$$\begin{array}{ll} \max & -b^T \nu \\ \text{s.t.} & A^T \nu - \lambda + c = 0 \\ & \lambda \geq 0, \end{array} \quad \text{or equvalently,} \quad \begin{array}{ll} \max & -b^T \nu \\ \text{s.t.} & A^T \nu + c \geq 0. \end{array}$$

Similarly, if LP has inequality constraints, then Lagrange dual problem is given as

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax \leq b \end{array} \quad \Rightarrow \quad \begin{array}{ll} \max & -b^T \lambda \\ \text{s.t.} & A^T \lambda + c = 0 \\ & \lambda \geq 0. \end{array}$$

Since Lagrange dual provides lower bound for primal,

$$d^* \leq p^*,$$

where p^* and d^*, resp. are optimal values of primal and dual problems.

From this weak duality, we have

$$\begin{array}{lcl} \text{Primal unbounded below} & \Rightarrow & \text{dual infeasible,} \\ \text{Dual unbounded above} & \Rightarrow & \text{primal infeasible.} \end{array}$$

Dual problem, due to convexity, is solvable efficiently in many cases. For example, lower bound on two-way partitioning problem can be computed via following SDP:

$$\begin{array}{ll} \max & -\mathbf{1}^T \nu \\ \text{s.t.} & W + \text{diag}(\nu) \geq 0. \end{array}$$

- We say *strong duality* holds if $d^* = p^*$.
- Strong duality does not hold in general. To guarantee it we need some constraint qualification such as a Slater-type condition.
- Slater's condition: relint$\mathcal{D} \neq \emptyset$. Namely, $\exists\ x \in \mathcal{D}$ meeting every inequality constraint *strictly*.

Theorem

Suppose primal is convex and satisfies Slater's condition. Then strong duality holds and dual optimum is attained.

Refined Slater's condition

- If the first k constraint functions are affine, then the following weaker condition holds: There exists an $x \in \mathcal{D}$ with

$$f_i(x) \leq 0, i = 1, \cdots, k, \quad f_i(x) < 0, i = k+1, \cdots, m, \quad Ax = b.$$

  In other words, x need not have to affine inequalities strictly.

- Slater's condition implies that the dual optimal value is attained when $d^* > -\infty$, that is, there exists a dual feasible (λ^*, ν^*) with $g(\lambda^*, \nu^*) = d^* = p^*$.

- Least-squares solution of linear equations

$$\begin{array}{ll} \min & x^T x \\ \text{s.t.} & Ax = b \end{array} \quad \Rightarrow \quad \begin{array}{ll} \max & -(1/4)\nu^T AA^T \nu - b^T \nu \\ \text{s.t.} & \end{array}$$

Since primal is convex and meets Slater's condition, $p^* = d^*$ if primal is feasible. (Actually, feasibility assumption is not necessary.)

- Lagrange dual of LP
Since every constraint in LP is affine, strong duality holds if primal is feasible. Similarly, if dual is feasible, then strong duality holds.
Strong duality of LP may fail when both primal and dual are infeasible.

Lagrange dual of QCQP

Recall QCQP,

$$\begin{array}{ll} \min & (1/2)x^TP_0x + q_0^Tx + r_0 \\ \text{s.t.} & (1/2)x^TP_ix + q_i^Tx + r_i \leq 0, i = 1, \cdots, m, \end{array}$$

where $P_0 \in \mathbb{S}^n_{++}, P_i \in \mathbb{S}^n_+, \forall i$.

Lagrangian is $L(x, \lambda) = (1/2)x^TP(\lambda)x + q(\lambda)^Tx + r(\lambda)$, where

$$P(\lambda) = P_0 + \textstyle\sum_i \lambda_i P_i, \quad q(\lambda) = q_0 + \sum_i \lambda_i q_i, \quad r(\lambda) = r_0 + \sum_i \lambda_i r_i.$$

Hence dual is

$$\begin{array}{ll} \max & -(1/2)q(\lambda)^TP(\lambda)^{-1}q(\lambda) + r(\lambda) \\ \text{s.t.} & \lambda \geq 0. \end{array}$$

Inequality constraint functions of QCQP are not affine, so strong duality holds when $(1/2)x^TP_ix + q_i^Tx + r_i < 0$ for all i.

- Entropy maximization

$$\begin{array}{ll} \min & \sum_i x_i \log x_i \\ \text{s.t.} & Ax \le b \\ & \mathbf{1}^T x = 1 \end{array} \quad \xrightarrow{\text{dual}} \quad \begin{array}{ll} \max & -b^T\lambda - \nu - e^{-\nu-1}\sum_i e^{-a_i^T\lambda} \\ \text{s.t.} & \lambda \ge 0 \end{array}$$

$$\xrightarrow{\text{max'zed over } \nu} \quad \begin{array}{ll} \max & -b^T\lambda - \log\left(\sum_i e^{-a_i^T\lambda}\right) \\ \text{s.t.} & \lambda \ge 0 \end{array}$$

- Minimum volume covering ellipsoid

$$\begin{array}{ll} \min & \log\det X^{-1} \\ \text{s.t.} & a_i^T X a_i \le 1, \forall i \end{array} \quad \Rightarrow \quad \begin{array}{ll} \max & \log\det(\sum_i \lambda_i a_i a_i^T) - \mathbf{1}^T\lambda + n \\ \text{s.t.} & \lambda \ge 0 \end{array}$$

Inequality constraint functions in primal are affine for X, so strong duality holds when $\exists\ X \in \mathbb{S}^n_{++}$ such that $a_i^T X a_i \le 1\ \forall i$, which is always true. Thus, entropy maximization always has strong duality.

A geometric interpretation of dual in terms of the set

$$\mathcal{G} = \{(f_1(x), \cdots, f_m(x), h_1(x), \cdots, h_p(x), f_0(x)) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} | x \in \mathcal{D}\}.$$

For optimization

$$\begin{array}{lll} \min & f_0(x), & \\ \text{s.t.} & f_i(x) \leq 0, & i = 1, \cdots, m \\ & h_j(x) = 0, & j = 1, \cdots, p, \end{array}$$

its optimal value p^* can be represented as

$$p^* = \inf\{t | (u, v, t) \in \mathcal{G}, u \leq 0, v = 0\}.$$

Dual function at (λ, ν) is

$$g(\lambda,\nu) = \inf\left\{ \sum_{i=1}^m \lambda_i u_i + \sum_{j=1}^p \nu_i v_i + t | (u,v,t) \in \mathcal{G} \right\}$$
$$= \inf\{(\lambda,\nu,1)^T(u,v,t) | (u,v,t) \in \mathcal{G}\}.$$

Thus for any $(u, v, t) \in \mathcal{G}$ we have

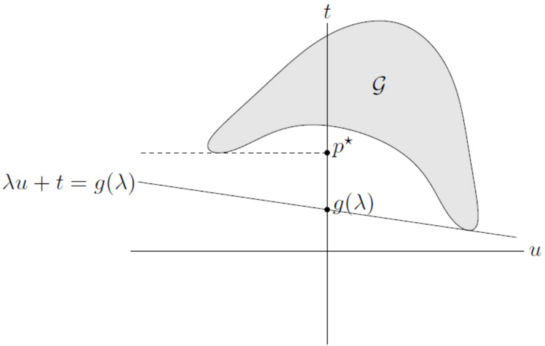
$$(\lambda,\nu,1)^T(u,v,t) \geq g(\lambda,\nu),$$

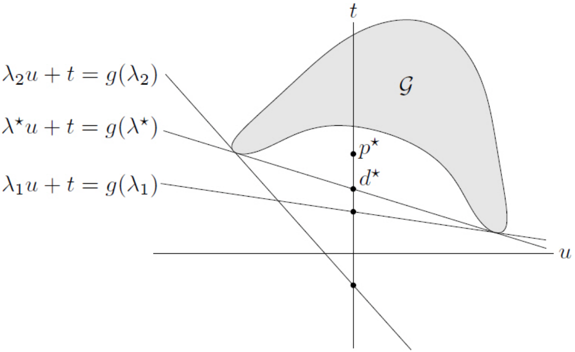
nonvertical supporting hyperplane in sense of last nonzero coordinate 1. Suppose $\lambda \geq 0$. Then, $t \geq (\lambda,\nu,1)^T(u,v,t)$ if $u \leq 0$ and $v = 0$. Thus,

$$\begin{aligned} p^* &= \inf\{t | (u,v,t) \in \mathcal{G}, u \leq 0, v = 0\} \\ &\geq \inf\{(\lambda,\nu,1)^T(u,v,t) | (u,v,t) \in \mathcal{G}, u \leq 0, v = 0\} \\ &\geq \inf\{(\lambda,\nu,1)^T(u,v,t) | (u,v,t) \in \mathcal{G}\} \\ &= g(\lambda,\nu), \end{aligned}$$

weak duality!

Consider case $m = 1$ so that $\mathcal{G}$ can be illustrated in $\mathbb{R}^2$. Given λ, minimizing $(\lambda, 1)^T(u, t)$ over $\mathcal{G}$ yields a supporting hyperplane with slope $-\lambda$:





For three dual feasible values of λ, including optimum λ^*, strong duality does not hold; duality gap $p^* - d^*$ is positive.

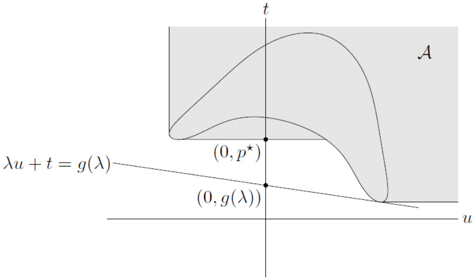
Consider following epigraph variation of $\mathcal{G}$,

$$\begin{aligned}\mathcal{A} := \{(u, v, t) | f_i(x) \le u_i,\ \forall i,\ h_i(x) = v_i,\ \ \forall i,\ f_0(x) \le t,\ \text{for some } x \in \mathcal{D}\} \\ = \mathcal{G} + (\mathbb{R}_+^m \times \{0\} \times \mathbb{R}_+).\end{aligned}$$

Then, easy to see

$$p^* = \inf\{t | (0, 0, t) \in \mathcal{A}\},$$
$$\text{For any } \lambda \ge 0,\ g(\lambda, \nu) = \inf\{(\lambda, \nu, 1)^T (u, v, t) | (u, v, t) \in \mathcal{A}\},\ \text{and}$$
$$\text{Since } (0, 0, p^*) \in \text{bd}\mathcal{A},\ p^* = (\lambda, \nu, 1)^T (0, 0, p^*) \ge g(\lambda, \nu).$$

Thus strong duality holds iff for some (λ, ν), $p^* = (\lambda, \nu, 1)^T(0, 0, p^*) = g(\lambda, \nu)$, i.e. $\exists$ non-vertical supporting hyperplane to $\mathcal{A}$ at $(0, p^*)$.



Consider primal with Slater's condition: $\exists\ \tilde{x} \in \text{relint}\mathcal{D}$ with $f_i(\tilde{x}) < 0$, and $A\tilde{x} = b$.

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, \quad i = 1, \cdots, m \quad f_0, \cdots, f_m \text{ convex} \\ & Ax = b. \end{array}$$

Theorem

Suppose primal is convex and satisfies Slater's condition. Then strong duality holds and dual optimum is attained.

Proof For a simpler proof, introduce little stronger assumptions:

1. Domain $\mathcal{D}$ has nonempty interior, i.e. $\text{relint}\mathcal{D} = \text{int}\mathcal{D}$, and
2. $\text{rank}\ A = p$.

Slater's cond implies feasibility. Hence case $p^* = +\infty$ is excluded. If $p^* = -\infty$, then weak duality implies $d^* = -\infty$, and theorem holds vacuously. Hence we assume throughout $p^* > -\infty$.

Primal convexity implies $\mathcal{A} = \mathcal{G} + (\mathbb{R}_+^m \times \{0\} \times \mathbb{R}_+)$ is convex. We define second convex set

$$\mathcal{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} | s < p^*\}.$$

Then $\mathcal{A}$ and $\mathcal{B}$ are disjoint. By separating hyperplane thm, $\exists\ (\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ and α s.t.

$$\begin{aligned} (u, v, t) \in \mathcal{A} \quad &\Rightarrow \quad \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \geq \alpha, \quad (1) \\ (u, v, t) \in \mathcal{B} \quad &\Rightarrow \quad \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \leq \alpha. \quad (2) \end{aligned}$$

From (1), we conclude that $\tilde{\lambda} \geq 0$ and $\mu \geq 0$, and (2) implies that $\mu t \leq \alpha$ for all $t < p^*$ or that $\mu p^* \leq \alpha$. Therefore, we have the following:

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) + \mu f_0(x) \geq \alpha \geq \mu p^*. \quad (3)$$

Assume $\mu > 0$. Then, from (3),

$$L(x, \tilde{\lambda}/\mu, \tilde{\nu}/\mu) \geq p^*, \ \forall x \in \mathcal{D}.$$

Hence, by minimizing over x, it follows that $g(\lambda, \nu) \geq p^*$ for $\lambda = \tilde{\lambda}/\mu, \nu = \tilde{\nu}/\mu$. By weak duality, $g(\lambda, \nu) \leq p^*$, so $g(\lambda, \nu) = p^*$. Therefore, strong duality holds and dual optimum is attained when $\mu > 0$.
Assume $\mu = 0$. From (3),

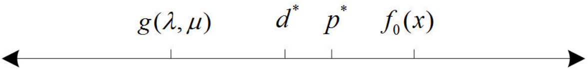
$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T(Ax - b) \geq 0, \ \forall x \in \mathcal{D}. \quad (4)$$

Therefore, for $\tilde{x}$ satisfying Slater's condition, we have $\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) \geq 0$. But, $f_i(\tilde{x}) < 0$, $\tilde{\lambda}_i \geq 0$ and we conclude $\tilde{\lambda} = 0$. Therefore, from $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$, we should have $\nu \neq 0$.

From (4),

$$\nu^T (Ax - b) \geq 0, \ \forall x \in \mathcal{D}. \ (5)$$

But, $\tilde{\nu}^T (A\tilde{x} - b) = 0$, and since $\tilde{x} \in \text{int}\mathcal{D}$ there exists points in $\mathcal{D}$ with $\tilde{\nu}^T (Ax - b) < 0$ unless $A^T \nu = 0$ which is impossible as rank $A = p$. A contradiction to (5). □



- A dual feasible (λ, ν) is a certificate that $p^* \geq g(\lambda, \nu)$. A primal feasible x is a certificate that $d^* \leq f_0(x)$.
- If x and (λ, ν) are a feasible pair, then

$$d^* - g(\lambda, \nu), f_0(x) - p^* \leq f_0(x) - g(\lambda, \nu).$$

  Thus x and (λ, ν) are ϵ-optimal solutions, where $\epsilon = f_0(x) - g(\lambda, \nu)$.
- If duality gap is zero, i.e. $f_0(x) = g(\lambda, \nu)$, then x and (λ, ν) are an optimal pair. Therefore, (λ, ν) is a certificate that x is optimal, and vice versa.

- Suppose an algorithm produces a sequence of primal feasible $x^{(k)}$ and dual feasible $(\lambda^{(k)}, \nu^{(k)})$ for $k = 1, 2, \cdots$, and $\epsilon_{abs} > 0$ is required absolute accuracy. Then, the stopping criterion

$$f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \leq \epsilon_{abs}$$

provides ϵ_{abs}-optimal solution $x^{(k)}$, and $(\lambda^{(k)}, \nu^{(k)})$ is a certificate.

- Similarly, for a relative accuracy $\epsilon_{rel} > 0$, the following conditions can work as a proper stopping criterion:

$$g(\lambda^{(k)}, \nu^{(k)}) > 0, \quad \frac{f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)})}{g(\lambda^{(k)}, \nu^{(k)}))} < \epsilon_{rel},$$

$$f_0(x^{(k)}) < 0, \quad \frac{f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)})}{-f_0(x^{(k)})} < \epsilon_{rel}.$$

Assume strong duality holds, x^* and (λ^*, ν^*) are primal-dual pair. Then,

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

From the above, we can derive the following:

- x^* minimizes $L(x, \lambda^*, \nu^*)$ over $\mathcal{D}$ (which we assume open) and hence gradient $\nabla_x L(x, \lambda^*, \nu^*)$ vanishes at $x = x^*$.
- $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \cdots, m$, or

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

known as *complementary slackness*.

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

1. Primal feasibility: $f_i(x) \leq 0$, $i = 1, \cdots, m$; $h_j(x) = 0$, $j = 1, \cdots, p$,
2. Dual feasibility: $\lambda \geq 0$,
3. Complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \cdots, m$,
4. Gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{j=1}^{p} \nu_j \nabla h_j(x) = 0.$$

We have seen if strong duality holds and x and (λ, ν) are optimal, then they satisfy KKT conditions.

Proposition

Suppose primal optimization is convex. If $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ satisfy KKT conditions, then they are optimal.

Proof From complementary slackness, $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ From the 4th condition and convexity, $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$. Hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$. □

Corollary

If Slater's condition is satisfied, x is optimal iff there exist λ, ν satisfying KKT conditions.

Examples

Consider equality constrained convex quadratic minimization

$$\begin{array}{ll} \min & (1/2)x^T P x + q^T x + r \\ \text{s.t.} & Ax = b, \end{array}$$

where $P \in \mathbb{S}^n_+$. KKT conditions are $Ax^* = b$, $Px^* + q + A^T \nu^* = 0$, or

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

Examples(*cont'd*)

Consider following optimization:

$$\begin{array}{ll} \min & -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \text{s.t.} & x \geq 0,\ \mathbf{1}^T x = 1, \end{array}$$

where $\alpha_i > 0$. KKT conditions for this problem are

$$x^* \geq 0,\ \mathbf{1}^T x^* = 1,\ \lambda^* \geq 0,\ \lambda_i^* x_i^* = 0,\ i = 1, \cdots, n,$$
$$-1/(\alpha_i + x_i^*) - \lambda_i^* + \nu^* = 0,\ i = 1, \cdots, n.$$

Solving the equations, we have

$$x_i^* = \begin{cases} 1/\nu^* - \alpha_i, & \nu^* \leq 1/\alpha_i \\ 0 & \nu^* \geq 1/\alpha_i \end{cases},\ \text{or } x_i^* = \max\{0, 1/\nu^* - \alpha_i\}.$$

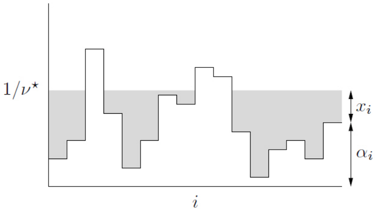
Since $\mathbf{1}^T x^* = 1$, we can obtain

$$\sum_{i=1}^{n} \max\{0, 1/\nu^* - \alpha_i\} = 1.$$

Examples(*cont'd*)

This solution method is called *water-filling* for the following reason:

- α_i is ground level above patch i.
- $1/\nu^*$ is target depth for flood.
- Total amount of water used is $\sum_i \max\{0, 1/\nu^* - \alpha_i\}$.
- We increase flood level until we have used total amount of water equal to one. Then, final depth of water above patch i is x_i^*.



Suppose we have strong duality and an optimal (λ^*, μ^*) is known, and minimizer of $L(x, \lambda^*, \nu^*)$ is unique (e.g. due to strict convexity). Then,

- if the minimizer is primal feasible, then it must be primal optimal.
- otherwise, primal optimum can not be attained.

This idea can be helpful when dual is easier to solve than primal.

Example: Entropy maximization

$$\begin{array}{lll} \min & f_0(x) = \sum_{i=1}^n x_i \log x_i & \xrightarrow{\text{dual}} \\ \text{s.t.} & Ax \le b,\ \mathbf{1}^T x = 1 & \end{array} \quad \begin{array}{ll} \max & -b^T\lambda - \nu - e^{-\nu-1}\sum_{i=1}^n e^{-a_i^T\lambda} \\ \text{s.t.} & \lambda \ge 0 \end{array}$$

Assume weak form of Slater's condition: $\exists$ an $x > 0$ with $Ax \le b$ and $\mathbf{1}^T x = 1$, so strong duality holds and an optimal solution (λ^*, ν^*) exists. Then,

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

is strictly convex on $\mathcal{D}$ and bounded below. Hence it has unique minimizer

$$x_i^* = 1/\exp(a_i^T\lambda^* + \nu^* + 1), i = 1, \cdots, n,$$

where a_i are columns of A. If it is primal feasible, then optimal. Otherwise, primal optimum is not attained.

Example: Equal'ty-const'ned separable function minimization

Objective is called *separable* when it is sum of functions of individual variables $x_1, \cdots, x_n$:

$$\begin{aligned} \min \quad & f_0(x) = \sum_{i=1}^{n} f_i(x_i) \\ \text{s.t.} \quad & a^T x = b. \end{aligned}$$

Lagrangian is

$$L(x, \nu) = \sum_{i=1}^{n} f_i(x_i) + \nu(a^T x - b) = -b\nu + \sum_{i=1}^{n} (f_i(x_i) + \nu a_i x_i),$$

which is also separable.

Example: Equal'ty-const'ned separable function minimization(*cont'd*)

Therefore, dual function is

$$\begin{aligned} g(\nu) &= -b\nu + \inf_x \left(\sum_{i=1}^{n} (f_i(x_i) + \nu a_i x_i) \right) \\ &= -b\nu + \sum_{i=1}^{n} \inf_x \left((f_i(x_i) + \nu a_i x_i) \right) \\ &= -b\nu - \sum_{i=1}^{n} f_i^*(-\nu a_i). \end{aligned}$$

Dual problem is then

$$\max \ -b\nu - \sum_{i=1}^{n} f_i^*(-\nu a_i).$$

Consider a system of inequalities and equalities,

$$\begin{aligned} f_i(x) \leq 0,\ i = 1, \ldots, m, \\ h_j(x) = 0,\ j = 1, \ldots, p. \end{aligned} \tag{1}$$

Assume domain of system (1) is nonempty. Consider the following problem:

$$\begin{array}{lll} \min & 0 & \\ \text{s.t.} & f_i(x) \leq 0, & i = 1, \cdots, m \\ & h_j(x) = 0, & j = 1, \cdots, p, \end{array} \tag{2}$$

Its optimal value is $p^* = \begin{cases} 0, & \text{if (1) is feasible} \\ \infty, & \text{otherwise.} \end{cases}$.

So solving (1) is the same as solving (2).

Dual function of (2) is

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \left(\sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x) \right).$$

Since $f_0 = 0$, dual is positively homogeneous in (λ, ν) and its optimal value is

$$d^* = \begin{cases} \infty & \lambda \geq 0, \text{if } g(\lambda, \nu) > 0 \text{ is feasible,} \\ 0 & \lambda \geq 0, \text{ otherwise.} \end{cases}$$

Combining this with weak duality, feasibility of (3)

$$\lambda \geq 0, \; g(\lambda, \nu) > 0 \tag{3}$$

implies infeasibility of (1). Hence such (λ, ν) is a certificate of infeasibility of (1).
Conversely, if (1) is feasible, then (3) must be infeasible. Hence feasible x of (1) is a certificate of infeasibility of (3).

Thus, (1) and (3) are *weak alternatives*, in sense that at most one of two is feasible.

Similarly, following systems are weak alternatives:

$$f_i(x) < 0, i = 1, \cdots, m, \; h_j(x) = 0, j = 1, \cdots, p. \tag{4}$$

$$\lambda \succeq 0, \; \lambda \neq 0, \; g(\lambda, \nu) \geq 0. \tag{5}$$

For, suppose $\exists \; \tilde{x}$ that satisfies (4). Then, for any $\lambda \succeq 0$, $\lambda \neq 0$, and ν,

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{j=1}^{p} \nu_j h_j(\tilde{x}) < 0.$$

It follows that

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \left(\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{j=1}^{p} \nu_j h_j(\tilde{x}) \right) \leq \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{j=1}^{p} \nu_j h_j(\tilde{x}) < 0.$$

Therefore, feasibility of (4) leads to infeasibility of (5).

When a convex system $\{f_i(x) \leq 0,\ i = 1, \ldots, m,\ Ax = b\}$ satisfies a constraint qualification, we get *strong alternatives*: exactly one of two systems is feasible.

- Strict inequalities. If $\exists\ x \in \text{relint}\mathcal{D}$ satisfying $Ax = b$, then the followings are strong alternatives.

$$f_i(x) < 0,\ i = 1, \cdots, m,\ Ax = b \tag{6}$$

$$\lambda \geq 0,\ \lambda \neq 0,\ g(\lambda, \nu) \geq 0 \tag{7}$$

  Proof: Consider optimization $\min\{s | f_i(x) \leq s,\ i = 1, \ldots, m,\ Ax = b\}$ and its dual. □

- Nonstrict inequalities. Similarly, if $\exists\ x \in \text{relint}\mathcal{D}$ satisfying $Ax = b$ and p^* is attained, the followings are strong alternatives.

$$f_i(x) \leq 0,\ i = 1, \cdots, m,\ Ax = b$$
$$\lambda \geq 0, g(\lambda, \nu) > 0.$$

- Linear inequalities. Consider system $Ax \leq b$ whose dual function is

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

  Thus, the alternative inequality system is

$$\lambda \geq 0,\ A^T\lambda = 0,\ b^T\lambda < 0,$$

  and moreover, they are strong alternatives. This is a version of Farkas lemma.

- Similarly, the followings are strong alternative:

$$\begin{gathered} Ax < b, \\ \lambda \geq 0,\ \lambda \neq 0,\ A^T\lambda = 0,\ b^T\lambda \leq 0. \end{gathered}$$

Farkas lemma

Theorem

The followings are strong alternatives:

$$Ax \le 0,\ c^T x < 0, \tag{8}$$

$$A^T y + c = 0,\ y \ge 0. \tag{9}$$

Proof Similarly with the previous proofs or by LP duality. □

LP duality and no-arbitrage bounds on price