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)

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

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 $\operatorname{dom} L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$,

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

- weighted sum of objective and constraint functions
- 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: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

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

- a concave function of λ, ν
- can be $-\infty$ for some λ, ν; this defines the domain of g

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

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

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

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

Least norm solution of linear equations

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

- Lagrangian is

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

- to minimize L over x, set gradient equal to zero:

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

- plug in in L to obtain g:

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

a concave function of ν

Lower bound property: $p^\star \geq -\frac{1}{4} \nu^T A A^T \nu - b^T \nu$ for all ν

Standard form LP

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

- 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 $\mathbf{dom}\, g = \{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$, hence concave

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

Equality constrained norm minimization

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

- $\|\cdot\|$ is any norm; dual norm is defined as

$$\|v\|_* = \sup_{\|u\| \leq 1} u^T v$$

- define Lagrangian $L(x, \nu) = \|x\| + \nu^T(b - Ax)$
- dual function (proof on next page):

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

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

proof of expression for g: follows from

$$\inf_x \left(\|x\| - y^T x\right) = \begin{cases} 0 & \|y\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases} \qquad (1)$$

Case $\|y\|_* \le 1$:

$$\inf_x \left(\|x\| - y^T x\right) = 0$$

- $y^T x \le \|x\| \|y\|_* \le \|x\|$ for all x (by definition of dual norm)
- $y^T x = \|x\|$ for $x = 0$

Case $\|y\|_* > 1$:

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

- there exists an $\tilde{x}$ with $\|\tilde{x}\| \le 1$ and $y^T \tilde{x} = \|y\|_* > 1$; hence $\|\tilde{x}\| - \|y\|_* < 0$
- consider $x = t\tilde{x}$ with $t > 0$:

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

Two-way partitioning

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

- a nonconvex problem; feasible set $\{-1, 1\}^n$ contains 2^n discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets, $x_i \in \{-1, 1\}$ is assignment for i
- cost function is

$$\begin{aligned} x^T W x &= \sum_{i=1}^n W_{ii} + 2 \sum_{i>j} W_{ij} x_i x_j \\ &= \mathbf{1}^T W \mathbf{1} + 2 \sum_{i>j} W_{ij}(x_i x_j - 1) \end{aligned}$$

cost of assigning i, j to different sets is $-4W_{ij}$

Lagrange dual of two-way partitioning problem

Dual function

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

Lower bound property

$$
p^\star \geq -\mathbf{1}^T \nu \quad \text{if } W + \mathbf{diag}(\nu) \succeq 0
$$

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

Lagrange dual and conjugate function

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

Dual function

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

- recall definition of conjugate $f^*(y) = \sup_x \left(y^T x - f(x) \right)$
- simplifies derivation of dual if conjugate of f_0 is known

Example: entropy maximization

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

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 $p^\star$, obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted by $d^\star$
- often simplified by making implicit constraint $(\lambda, \nu) \in \operatorname{dom} g$ explicit
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \operatorname{dom} g$
- $d^\star = -\infty$ if problem is infeasible; $d^\star = +\infty$ if unbounded above

Example: standard form LP and its dual (page 5.5)

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

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 5.8

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

- does not hold in general
- (usually) holds for convex problems
- sufficient conditions that guarantee strong duality in convex problems are called *constraint qualifications*

Slater's constraint qualification

Convex problem

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

Slater's constraint qualification: the problem is strictly feasible, *i.e.*,

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

- guarantees strong duality: $p^\star = d^\star$
- also guarantees that the dual optimum is attained if $p^\star > -\infty$
- can be sharpened: *e.g.*, can replace $\operatorname{int} \mathcal{D}$ with $\operatorname{relint} \mathcal{D}$ (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

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

Dual problem

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

- 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 ($p^\star = \infty$, $d^\star = -\infty$)

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

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

- 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

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

we allow $A \not\succeq 0$, hence problem may be nonconvex

Dual function (derivation on next page)

$$\begin{array}{rcl} g(\lambda) & = & \inf_x (x^T(A+\lambda I)x + 2b^T x - \lambda) \\ & = & \begin{cases} -b^T(A+\lambda I)^\dagger b - \lambda & A+\lambda I \succeq 0 \text{ and } b \in \mathcal{R}(A+\lambda I) \\ -\infty & \text{otherwise} \end{cases} \end{array}$$

Dual problem and equivalent SDP:

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

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

proof of expression for g: unconstrained minimum of $f(x) = x^T P x + 2q^T x + r$ is

$$\inf_x f(x) = \begin{cases} -q^T P^{-1} q + r & P \succ 0 \\ -q^T P^{\dagger} q + r & P \not\succ 0,\ P \succeq 0,\ q \in \mathcal{R}(P) \\ -\infty & P \succeq 0,\ q \notin \mathcal{R}(P) \\ -\infty & P \not\succeq 0 \end{cases}$$

- if $P \not\succeq 0$, function f is unbounded below: choose y with $y^T P y < 0$ and $x = ty$

$$f(x) = t^2(y^T P y) + 2t(q^T y) + r \to -\infty \qquad \text{if } t \to \pm\infty$$

- if $P \succeq 0$, decompose q as $q = Pu + v$ with $u = P^{\dagger} q$ and $v = (I - PP^{\dagger})q$

  Pu is projection of q on $\mathcal{R}(P)$, v is projection on nullspace of P

- if $v \neq 0$ (*i.e.*, $q \notin \mathcal{R}(P)$), the function f is unbounded below: for $x = -tv$,

$$f(x) = t^2(v^T P v) - 2t(q^T v) + r = -2t\|v\|^2 + r \to -\infty \qquad \text{if } t \to \infty$$

- if $v = 0$, $x^\star = -u$ is optimal since f is convex and $\nabla f(x^\star) = 2Px^\star + 2q = 0$;

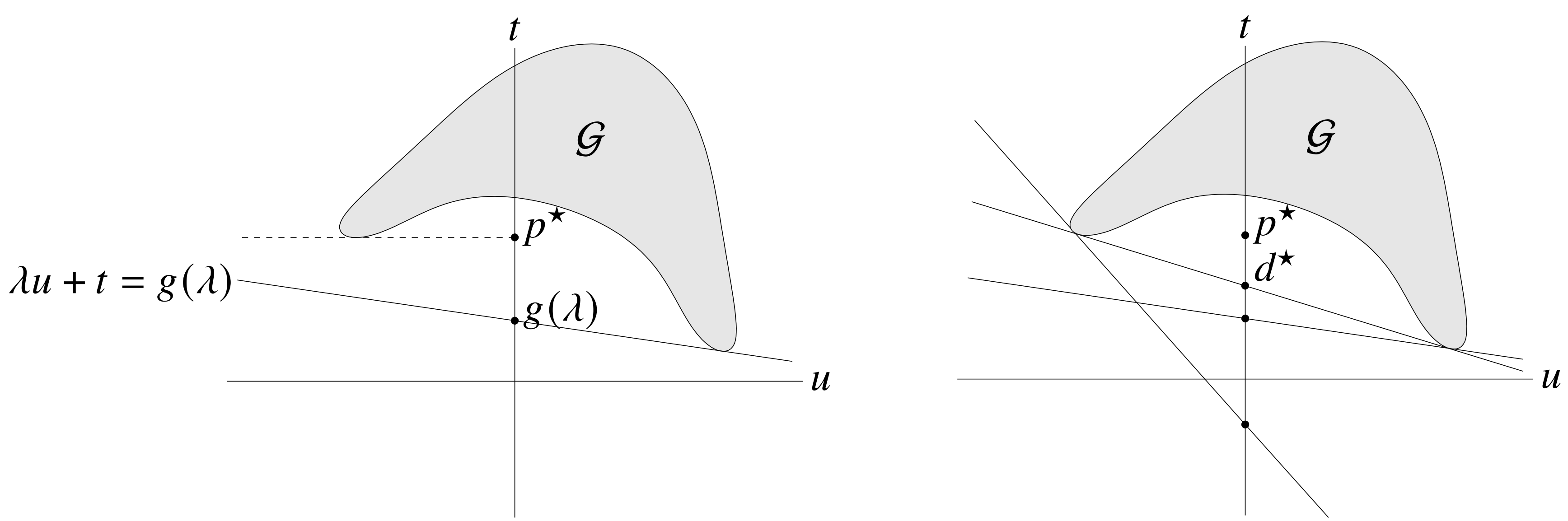
$$f(x^\star) = -q^T P^{\dagger} q + r$$

Geometric interpretation of duality

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

Interpretation of dual function

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

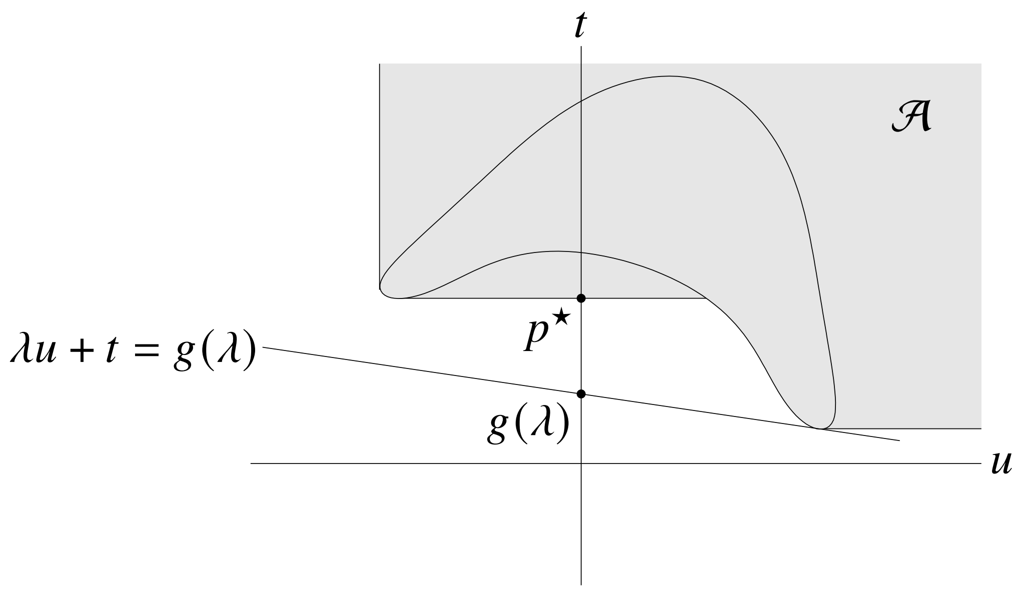


- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t-axis at $t = g(\lambda)$

Geometric interpretation of duality

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 $(0, p^\star)$
- for convex problem, $\mathcal{A}$ is convex, hence has supporting 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

Optimality conditions

if strong duality holds, then x is primal optimal and (λ, ν) is dual optimal if:

1. $f_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, \ldots, p$
2. $\lambda \succeq 0$
3. $f_0(x) = g(\lambda, \nu)$

conversely, these three conditions imply optimality of x, (λ, ν), and strong duality

next, we replace condition 3 with two equivalent conditions that are easier to use

Complementary slackness

assume x satisfies the primal constraints and $\lambda \succeq 0$

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

equality $f_0(x) = g(\lambda, \nu)$ holds if and only if the two inequalities hold with equality:

- first inequality: x minimizes $L(\tilde{x}, \lambda, \nu)$ over $\tilde{x} \in \mathcal{D}$
- 2nd inequality: $\lambda_i f_i(x) = 0$ for $i = 1, \ldots, m$, *i.e.*,

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

this is known as *complementary slackness*

Optimality conditions

if strong duality holds, then x is primal optimal and (λ, ν) is dual optimal if:

1. $f_i(x) \le 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, \ldots, p$
2. $\lambda \succeq 0$
3. $\lambda_i f_i(x) = 0$ for $i = 1, \ldots, m$
4. x is a minimizer of $L(\cdot, \lambda, \nu)$

conversely, these four conditions imply optimality of x, (λ, ν), and strong duality

if problem is convex and the functions f_i, h_i are differentiable, #4 can written as

4'. the gradient of the Lagrangian with respect to x vanishes:

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

conditions 1,2,3,4' are known as *Karush–Kuhn–Tucker* (KKT) conditions

Convex problem with Slater constraint qualification

recall the two implications of Slater's condition for a convex problem

- strong duality: $p^\star = d^\star$
- if optimal value is finite, dual optimum is attained: there exist dual optimal λ, ν

hence, if problem is convex and Slater's constraint qualification holds:

- x is optimal if and only if there exist λ, ν such that 1–4 on p. 5.22 are satisfied
- if functions are differentiable, condition 4 can be replaced with 4'

Example: water-filling

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

- we assume that $\alpha_i > 0$
- Lagrangian is $L(\tilde{x}, \lambda, \nu) = -\sum_i \log(\tilde{x}_i + \alpha_i) - \lambda^T \tilde{x} + \nu(\mathbf{1}^T \tilde{x} - 1)$

Optimality conditions: x is optimal iff there exist $\lambda \in \mathbf{R}^n$, $\nu \in \mathbf{R}$ such that

1. $x \succeq 0$, $\mathbf{1}^T x = 1$
2. $\lambda \succeq 0$
3. $\lambda_i x_i = 0$ for $i = 1, \ldots, n$
4. x minimizes Lagrangian:

$$\frac{1}{x_i + \alpha_i} + \lambda_i = \nu, \quad i = 1, \ldots, n$$

Example: water-filling

Solution

- if $\nu \le 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \ge 1/\alpha_i$: $x_i = 0$ and $\lambda_i = \nu - 1/\alpha_i$
- two cases may be combined as

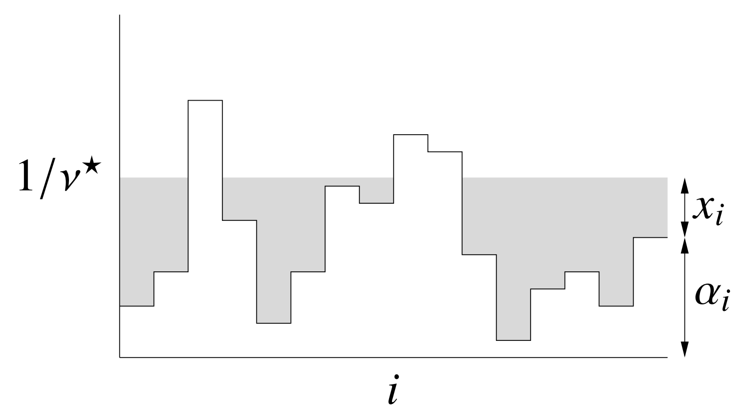
$$x_i = \max\{0, \frac{1}{\nu} - \alpha_i\}, \qquad \lambda_i = \max\{0, \nu - \frac{1}{\alpha_i}\}$$

- determine ν from condition $\mathbf{1}^T x = 1$:

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

Interpretation

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



Example: projection on 1-norm ball

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|x - a\|_2^2 \\ \text{subject to} & \|x\|_1 \leq 1 \end{array}$$

Optimality conditions

1. $\|x\|_1 \leq 1$
2. $\lambda \geq 0$
3. $\lambda(1 - \|x\|_1) = 0$
4. x minimizes Lagrangian

$$\begin{aligned} L(\tilde{x}, \lambda) &= \frac{1}{2}\|\tilde{x} - a\|_2^2 + \lambda(\|\tilde{x}\|_1 - 1) \\ &= \sum_{k=1}^{n} \left(\frac{1}{2}(\tilde{x}_k - a_k)^2 + \lambda|\tilde{x}_k|\right) - \lambda \end{aligned}$$

Example: projection on 1-norm ball

Solution

- optimization problem in condition 4 is separable; solution for $\lambda \geq 0$ is

$$x_k = \begin{cases} a_k - \lambda & a_k \geq \lambda \\ 0 & -\lambda \leq a_k \leq \lambda \\ a_k + \lambda & a_k \leq -\lambda \end{cases}$$

- therefore $\|x\|_1 = \sum_k |x_k| = \sum_k \max\{0, |a_k| - \lambda\}$
- if $\|a\|_1 \leq 1$, solution is $\lambda = 0$, $x = a$
- otherwise, solve piecewise-linear equation in λ:

$$\sum_{k=1}^{n} \max\{0, |a_k| - \lambda\} = 1$$

Perturbation and sensitivity analysis

(Unperturbed) optimization problem and its dual

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

Perturbed problem and its dual

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq u_i, \quad i = 1, \ldots, m \\
 & h_i(x) = v_i, \quad i = 1, \ldots, p
\end{array}
\qquad\qquad
\begin{array}{ll}
\text{maximize} & g(\lambda, \nu) - u^T \lambda - v^T \nu \\
\text{subject to} & \lambda \succeq 0
\end{array}
$$

- 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:

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

Sensitivity interpretation

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

Local sensitivity result

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

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

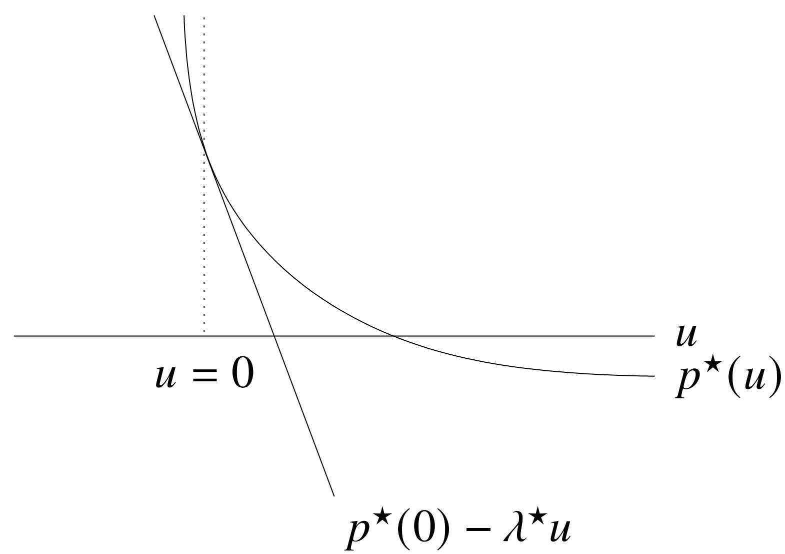
proof (for $\lambda_i^\star$): from global sensitivity result,

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

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

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

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

- 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

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

Example: norm approximation

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

Dual function

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

(last step follows from (1))

Dual of norm approximation problem

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

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 $\mathbf{R}^{k_i}$

Lagrangian and dual function: 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

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

Lagrange dual of problems with generalized inequalities

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

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

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

minimizing over all feasible 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

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

matrices $F_1, \ldots, F_n, G$ are symmetric $k \times k$

Lagrangian and dual function

- Lagrange multiplier is matrix $Z \in \mathbf{S}^k$; Lagrangian is

$$\begin{aligned} L(x, Z) &= c^T x + \operatorname{tr}\left(Z(x_1 F_1 + \cdots + x_n F_n - G)\right) \\ &= \sum_{i=1}^{n} (\operatorname{tr}(F_i Z) + c_i) x_i - \operatorname{tr}(GZ) \end{aligned}$$

- dual function

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

Dual semidefinite program

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

Weak duality: $p^\star \geq d^\star$ always

proof: for primal feasible x, dual feasible Z,

$$
\begin{aligned}
c^T x &= -\sum_{i=1}^n \operatorname{tr}(F_i Z) x_i \\
&= -\operatorname{tr}(GZ) + \operatorname{tr}(Z(G - \sum_{i=1}^n x_i F_i)) \\
&\geq -\operatorname{tr}(GZ)
\end{aligned}
$$

inequality follows from $\operatorname{tr}(XZ) \geq 0$ for $X \succeq 0$, $Z \succeq 0$

Strong duality: $p^\star = d^\star$ if primal SDP or dual SDP is strictly feasible

Complementary slackness

(P) minimize $c^T x$
subject to $\sum_{i=1}^n x_i F_i \preceq G$

(D) maximize $-\operatorname{tr}(GZ)$
subject to $\operatorname{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n$
$Z \succeq 0$

the primal and dual objective values at feasible x, Z are equal if

$$\begin{aligned} 0 &= c^T x + \operatorname{tr}(GZ) \\ &= -\sum_{i=1}^n x_i \operatorname{tr}(F_i Z) + \operatorname{tr}(GZ) \\ &= \operatorname{tr}(XZ) \qquad \text{where } X = G - x_1 F_1 - \cdots - x_n F_n \end{aligned}$$

for $X \succeq 0$, $Z \succeq 0$, each of the following statements is equivalent to $\operatorname{tr}(XZ) = 0$:

- $ZX = 0$: columns of X are in the nullspace of Z
- $XZ = 0$: columns of Z are in the nullspace of X

(see next page)

proof: factorize X, Z as

$$X = UU^T, \qquad Z = VV^T$$

- columns of U span the range of X, columns of V span the range of Z
- $\mathrm{tr}(XZ)$ can be expressed as

$$\mathrm{tr}(XZ) = \mathrm{tr}(UU^TVV^T) = \mathrm{tr}((U^TV)(V^TU)) = \|U^TV\|_F^2$$

- hence, $\mathrm{tr}(XZ) = 0$ if and only if

$$U^TV = 0$$

the range of X and the range of Z are orthogonal subspaces

Example: two-way partitioning

recall the two-way partitioning problem and its dual (page 5.8)

$$
\text{(P)} \quad \begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}
\qquad\qquad
\text{(D)} \quad \begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}
$$

- by weak duality, $p^\star \geq d^\star$
- the dual problem (D) is an SDP; we derive the dual SDP and compare it with (P)
- to derive the dual of (D), we first write (D) as a minimization problem:

$$
\begin{array}{ll} \text{minimize} & \mathbf{1}^T y \\ \text{subject to} & W + \mathbf{diag}(y) \succeq 0 \end{array} \tag{2}
$$

  the optimal value of (2) is $-d^\star$

Example: two-way partitioning

Lagrangian

$$\begin{aligned} L(y,Z) &= \mathbf{1}^T y - \operatorname{tr}(Z(W + \mathbf{diag}(y))) \\ &= -\operatorname{tr}(WZ) + \sum_{i=1}^{n} y_i(1 - Z_{ii}) \end{aligned}$$

Dual function

$$g(Z) = \inf_y L(y,Z) = \begin{cases} -\operatorname{tr}(WZ) & Z_{ii} = 1,\ i = 1,\ldots,n \\ -\infty & \text{otherwise} \end{cases}$$

Dual problem: the dual of (2) is

$$\begin{array}{ll} \text{maximize} & -\operatorname{tr}(WZ) \\ \text{subject to} & Z_{ii} = 1, \quad i = 1,\ldots,n \\ & Z \succeq 0 \end{array}$$

by strong duality with (2), optimal value is equal to $-d^\star$

Example: two-way partitioning

replace (D) on page 5.41 by its dual

$$
\begin{array}{lll}
\text{(P)} & \text{minimize} & x^T W x \\
 & \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n
\end{array}
\qquad\qquad
\begin{array}{lll}
\text{(P')} & \text{minimize} & \mathrm{tr}(WZ) \\
 & \text{subject to} & \mathbf{diag}(Z) = \mathbf{1} \\
 & & Z \succeq 0
\end{array}
$$

optimal value of (P') is equal to optimal value $d^\star$ of (D)

Interpretation as relaxation

- reformulate (P) by introducing a new variable $Z = xx^T$:

$$
\begin{array}{ll}
\text{minimize} & \mathrm{tr}(WZ) \\
\text{subject to} & \mathbf{diag}(Z) = \mathbf{1} \\
 & Z = xx^T
\end{array}
$$

- replace the constraint $Z = xx^T$ with a weaker convex constraint $Z \succeq 0$