Convex Optimization

Part 4: Duality IV

Namhoon Lee

POSTECH

9 Nov 2022

Quadratic program

Primal problem ($P \succ 0$)

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

Dual function

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

Dual problem

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

- 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

Entropy maximization

Recall the entropy maximization problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^\top x = 1 \end{array}$$

Dual problem

$$\begin{array}{ll} \text{maximize} & -b^\top \lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^\top \lambda} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- Slater's condition: the optimal duality gap is zero if there exists an $x \succ 0$ with $Ax \preceq b$ and $\mathbf{1}^\top x = 1$

We can simplify the dual problem by maximizing over the dual variable ν analytically. For fixed λ, the objective function is maximized when the derivative with respect to ν is zero, *i.e.*

$$\nu = \log \sum_{i=1} n e^{-a_i^\top \lambda} - 1$$

Substituting this optimal value of ν into the dual problem gives

$$\begin{array}{ll} \text{maximize} & -b^\top \lambda - \log \left(\sum_{i=1}^{n} e^{-a_i^\top \lambda} \right) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

which is a geometric program (in convex form) with nonnegativity constraints

Geometric interpretation of duality

Consider a set of values taken on by the constraint and objective functions

$$\mathcal{G} = \left\{(f_1(x), \ldots, f_m(x), h_1(x), \ldots, h_p(x), f_0(x)) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid x \in \mathcal{D}\right\}$$

Optimal value $p^\star$ can be expressed in terms of $\mathcal{G}$

$$p^\star = \inf\{t \mid (u, v, t) \in \mathcal{G}, u \preceq 0, v = 0\}$$

To evaluate the dual function, we minimize the affine function

$$(\lambda, \nu, 1)^\top (u, v, t) = \sum_{i=1}^{m} \lambda_i u_i + \sum_{i=1}^{p} \nu_i v_i + t$$

i.e., we have the dual function as

$$g(\lambda, \nu) = \inf\{(\lambda, \nu, 1)^\top (u, v, t) \mid (u, v, t) \in \mathcal{G}\}$$

If the infimum is finite, then the inequality

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

defines a supporting hyperplane to $\mathcal{G}$

Next suppose $\lambda \succeq 0$. Then, obviously, $t \geq (\lambda, \nu, 1)^\top (u, v, t)$ if $u \preceq 0$ and $v = 0$. Therefore

$$\begin{aligned} p^\star &= \inf\{t \mid (u, v, t) \in \mathcal{G}, u \preceq 0, v = 0\} \\ &\geq \inf\{(\lambda, \nu, 1)^\top (u, v, t) \mid (u, v, t) \in \mathcal{G}, u \preceq 0, v = 0\} \\ &\geq \inf\{(\lambda, \nu, 1)^\top (u, v, t) \mid (u, v, t) \in \mathcal{G}\} \\ &= g(\lambda, \nu) \end{aligned}$$

i.e., we have weak duality

Consider a simple problem with one inequality constraint $f_1(x) \leq 0$

Interpretation of dual function

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

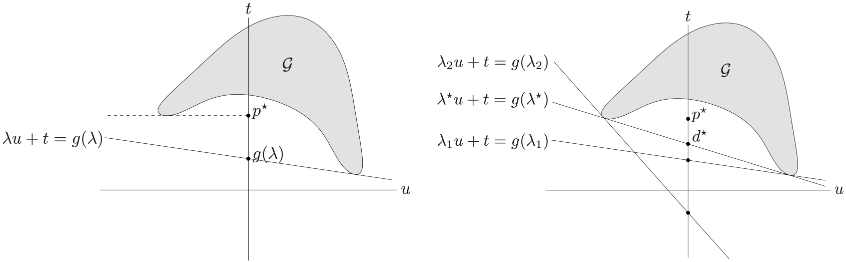

Figure: geometric interpretation of dual function and lower bound $g(\lambda) \leq p^\star$ figure from BV

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 $i = 1, \ldots, p$ (primal feasibility)
2. $\lambda \succeq 0$ (dual feasibility)
3. $f_0(x) = g(\lambda, \nu)$ (strong duality)

conversely, these three conditions imply optimality of $x, (\lambda, \nu)$, 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 $\lambda \succeq 0$

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

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

- 1st 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.*,

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

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. $f_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, \ldots, p$ (primal feasibility)
2. $\lambda \succeq 0$ (dual feasibility)
3. $\lambda_i f_i(x) = 0$ for $i = 1, \ldots, m$ (complementary slackness)
4. x is a minimizer of $L(\cdot, \lambda, \nu)$ (stationarity)

conversely, these four conditions imply optimality of $x, (\lambda, \nu)$, and strong duality

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

4' the gradient of the Lagrangian with respect to x vanishes: (stationarity')

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

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 conditions 1-4 are satisfied
- if functions are differentiable, condition 4 can be replaced with 4'

Summary

- KKT conditions imply zero duality gap, and vice versa (sufficient and necessary), *i.e.*

$$x^\star \text{ and } (\lambda^\star, \nu^\star) \text{ are optimal} \quad \Longleftrightarrow \quad x^\star \text{ and } (\lambda^\star, \nu^\star) \text{ satisfy the KKT conditions}$$

- If Slater's condition holds (for a convex problem) then KKT conditions are met.
- To note further, the KKT conditions are essentially the same as the optimality conditions derived from subgradient

$$0 \in \partial f(x^\star) + \sum_{i=1}^{m} \mathcal{N}_{f_i \le 0}(x^\star) + \sum_{i=1}^{p} \mathcal{N}_{h_i = 0}(x^\star)$$

Example: equality constrained convex quadratic minimization

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^\top P x + q^\top x + r \\ \text{subject to} & Ax = b, \end{array}$$

where $P \in S^n_+$; this problem leads to Newton's method for constrained case (later)

By KKT conditions we have

$$Ax^\star = b, \quad Px^\star + q + A^\top \nu^\star = 0$$

(note that complementary slackness and dual feasibility are vacuous)

which we can write as

$$\begin{bmatrix} P & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

Example: water-filling

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

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

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

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

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

Solution

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

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

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

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

Interpretation

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

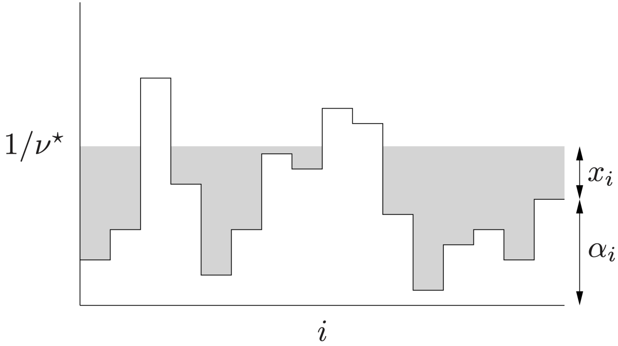

Figure: Illustration of water-filling algorithm; figure from BV

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

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

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:

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

Any questions?