

Lecture 19: Examples of duality and KKT

Tuesday, November 15, 2022

Lecturer: Laurent Lessard

Scribe: Rushikesh Sankhye

1 Introduction

In this section we will look at an example of duality. Recall the primal-dual pair and the associated weak duality inequality.

$$p^* = \underbrace{\begin{array}{l} \min_x f(x) \\ \text{s.t. } g_i(x) \leq 0 \quad i = 1, \dots, m \\ h_j(x) = 0 \quad j = 1, \dots, p \end{array}}_{\text{Primal Problem}} \geq \underbrace{\begin{array}{l} \max_{\lambda, \mu} F(\lambda, \mu) \\ \text{s.t. } \lambda \geq 0 \end{array}}_{\text{Dual Problem}} = d^* \quad (1)$$

$$\text{Lagrangian:} \quad L(x, \lambda, \mu) := f(x) + \sum_{i=1}^n \lambda_i g_i(x) + \sum_{j=1}^p \mu_j h_j(x).$$

$$\text{Dual function:} \quad F(\lambda, \mu) := \min_x L(x, \lambda, \mu)$$

Weak duality states that $p^* \geq d^*$. If instead we have $p^* = d^*$, then we have strong duality.

1.1 Example: QCQP

$$\begin{array}{l} \min_x (x^2 + 1) \\ \text{s.t. } (x - 2)(x - 4) \leq 0 \end{array}$$

Now compute the Lagrangian and associated dual function.

$$\begin{aligned} L(x, \lambda) &= x^2 + 1 + \lambda(x^2 - 6x + 8) \\ F(\lambda) &= \min_x (\lambda + 1)x^2 - 6\lambda x + (8\lambda + 1) \end{aligned}$$

Differentiating and setting derivative with respect to x to zero we get

$$2(\lambda + 1)x - 6\lambda = 0 \quad \implies \quad x = \frac{3\lambda}{\lambda + 1}. \quad (2)$$

Substituting this x back into $L(x, \lambda)$, we obtain

$$F(\lambda) = \begin{cases} -\frac{9\lambda}{\lambda+1} + 8\lambda + 1 & \text{if } \lambda > -1 \\ -\infty & \text{if } \lambda \leq -1 \end{cases}$$

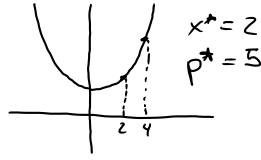


Figure 1: Primal problem. Solution occurs at $x = 2$ with optimal value $p^* = 5$.

Since we always have $\lambda \geq 0$, the dual problem becomes

$$\begin{aligned} \max_{\lambda} \quad & -\frac{9\lambda^2}{\lambda+1} + 8\lambda + 1 \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

The first and second derivatives of $F(\lambda)$ are

$$\begin{aligned} F'(\lambda) &= \frac{9}{(\lambda+1)^2} - 1 \\ F''(\lambda) &= -\frac{18}{(\lambda+1)^3} \end{aligned}$$

We can confirm that $F(\lambda)$ is concave (negative second derivative), which is to be expected because dual functions are *always* concave. The first derivative also tells us that the unique optimal point is at $\lambda^* = 2$ with corresponding $d^* = 5$.

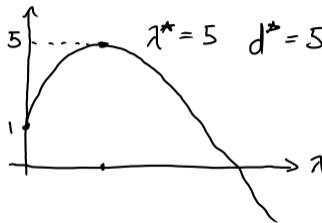
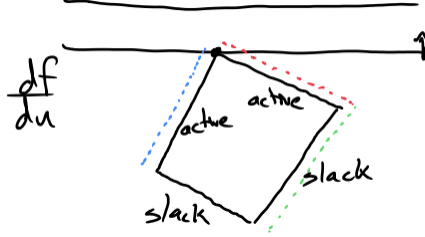


Figure 2: Dual problem. Solution occurs at $\lambda^* = 2$ with optimal value $d^* = 5$.

2 Sensitivity of cost to changes in constraints

We will perturb the constraints with u_i and v_j and see how this affects the cost. Based on the figure below, we can see that perturbing each constraint by the same amount can have a small effect, a larger effect, or no effect (when the constraints are slack).



Here is the perturbed optimization problem and its dual.

$$\begin{aligned}
 p^*(u, v) &= \min_x f(x) \\
 &\text{s.t. } g_i(x) \leq u_i \quad i = 1, \dots, m \\
 &\quad h_j(x) = v_j \quad j = 1, \dots, p \\
 \\
 d^*(u, v) &= \max_{\lambda, \mu} F(\lambda, \mu) - \lambda^\top u - \mu^\top v \\
 &\text{s.t. } \lambda \geq 0
 \end{aligned}$$

This works because we can write the dual function of this perturbed problem in terms of the dual function of the unperturbed problem:

$$\begin{aligned}
 &\min_x \left(f(x) + \sum_{i=1}^n \lambda_i (g_i(x) - u_i) + \sum_{j=1}^p \mu_j (h_j(x) - v_j) \right) \\
 &= \min_x L(x, \lambda, \mu) - \lambda^\top u - \mu^\top v \\
 &= F(\lambda, \mu) - \lambda^\top u - \mu^\top v \quad (\text{last two terms do not depend on } x)
 \end{aligned}$$

Using weak duality, we can write the following inequality relating the primal optimal value of the perturbed problem to that of the unperturbed problem.

$$p^*(u, v) \geq \max_{\lambda, \mu} F(\lambda, \mu) - \lambda^\top u - \mu^\top v \geq \underbrace{F(\lambda^*, \mu^*)}_{p^*(0,0)} - \lambda^{*\top} u - \mu^{*\top} v \quad (3)$$

s.t. $\lambda \geq 0$

The first inequality is due to weak duality, and the second is found by substituting the dual optimal values for the unperturbed problem. Thus, we obtain the inequality:

$$\boxed{p^*(u, v) \geq p^*(0, 0) - \lambda^{*\top} u - \mu^{*\top} v} \quad (4)$$

Now substitute $u = te_i$ and $v = 0$ into (4) and obtain $p^*(e_i t, 0) \geq p^*(0, 0) - \lambda_i^* t$. We have cases.

- If $t > 0$, we can divide by t and obtain $\frac{p^*(e_{it},0) - p^*(0,0)}{t} \geq -\lambda_i^*$.
- If $t < 0$, we can divide by t and obtain $\frac{p^*(e_{it},0) - p^*(0,0)}{t} \leq -\lambda_i^*$.

Taking the limit $t \rightarrow 0$, the left-hand side becomes a partial derivative. We can repeat a similar argument letting $u = 0$ and $v = te_j$ to get an expression involving μ_j^* . The final result is

$$\boxed{\frac{\partial p^*(0,0)}{\partial u_i} = -\lambda_i^* \quad \frac{\partial p^*(0,0)}{\partial v_j} = -\mu_j^*} \quad (5)$$

In other words, the partial derivative of the primal cost with respect to the perturbation of a particular constraint is equal to the negative of the corresponding optimal dual variable.

2.1 Example: LP

$$p^*(u) = \left\{ \begin{array}{l} \max \quad c^\top x \\ \text{s.t.} \quad Ax \leq b + u \\ x \geq 0 \end{array} \right\} = \left\{ \begin{array}{l} \min \quad (b + u)^\top \lambda \\ \text{s.t.} \quad A^\top \lambda \geq c \\ \lambda \geq 0 \end{array} \right\} = d^*(u) \quad (6)$$

This leads to the chain of equalities

$$p^*(u) = c^\top x^* = (b + u)^\top \lambda^* = p^*(0) + \mu^\top \lambda^*.$$

Again, we directly derive the sensitivity of the cost with respect to the constraint

$$\boxed{\frac{\partial p^*(0)}{\partial u_i} = \lambda_i^*}$$

Note that the sign is positive instead of negative as in (5) because our primal problem was a maximization this time rather than a minimization.

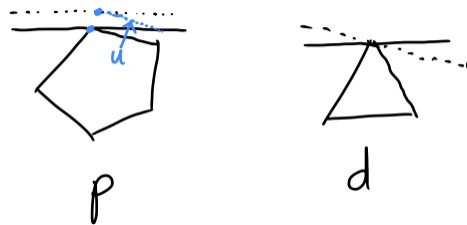


Figure 3: Primal and dual for a perturbed LP. In the left, perturbing a constraint changes the feasible set. If the perturbed constraint was active, this leads to a change in the optimal cost. In the dual problem, the perturbation shows up in the cost. For small perturbations, this does not change the value of the optimal dual variable.

In economics, λ is often referred to as the *shadow price* of the corresponding constraint. When the cost is a dollar amount and the problem is to maximize profit, say, the shadow price is the price you would be willing to pay to increase the constraint. It is what that constraint is *worth* to you.

Note that by complementary slackness, we have $(Ax^* - b)^\top \lambda^* = 0$. So if a constraint is slack, the corresponding dual variable is zero, and so is the shadow price. So slack constraints are worth nothing to us, since small changes in them will not affect the outcome of the optimization problem.

3 Derivation of LQR using KKT conditions

Recall the KKT conditions from Lecture 18. We will use the the Lagrangian stationarity condition in deriving the solution for the LQR problem

$$\begin{aligned} \underset{\substack{u_0, \dots, u_{N-1} \\ x_1, \dots, x_N}}{\text{minimize}} \quad & \frac{1}{2} \sum_{t=0}^{N-1} \left(x_t^\top Q x_t + u_t^\top R u_t \right) + \frac{1}{2} x_N^\top Q_f x_N \\ \text{s.t.} \quad & x_{t+1} = A x_t + B u_t \quad \text{for } t = 0, \dots, N-1 \end{aligned} \quad (7)$$

Here, we will treat this as a constrained optimization problem where the u_t and the x_t are optimization variables. where λ_{t+1} is the dual variable (Lagrange multiplier) associated with the constraint $x_{t+1} = A x_t + B u_t$. The Lagrangian for (7) is

$$L(x, u, \lambda) := \frac{1}{2} \sum_{t=0}^{N-1} \left(x_t^\top Q x_t + u_t^\top R u_t \right) + \frac{1}{2} x_N^\top Q_f x_N + \sum_{t=0}^{N-1} \lambda_{t+1}^\top (A x_t + B u_t - x_{t+1})$$

The stationarity conditions are:

$$\nabla_{u_t} L = 0 : \quad R u_t + B^\top \lambda_{t+1} = 0 \quad \text{for } t = 0, \dots, N-1 \quad (8a)$$

$$\nabla_{x_t} L = 0 : \quad Q x_t + A^\top \lambda_{t+1} - \lambda_t = 0 \quad \text{for } t = 1, \dots, N-1 \quad (8b)$$

$$\nabla_{x_N} L = 0 : \quad Q_f x_N - \lambda_N = 0 \quad (8c)$$

$$\nabla_{\lambda_t} L = 0 : \quad A x_t + B u_t - x_{t+1} = 0 \quad \text{for } t = 1, \dots, N \quad (8d)$$

there are $3N$ variables ($x_{1:N}, u_{0:N-1}, \lambda_{1:N}$) and $3N$ constraints. So under some standard assumptions ($Q \succeq 0$ and $R \succ 0$ suffices), the KKT conditions will have a unique solution.

If we solve (8a) for u_t , we obtain $u_t = -R^{-1} B^\top \lambda_{t+1}$. Substituting this back into (8b)–(8d) to eliminate u_t , we obtain

$$x_{t+1} = A x_t - B R^{-1} B^\top \lambda_{t+1} \quad \text{for } t = 0, \dots, N-1 \quad (\text{state equation})$$

$$\lambda_t = Q x_t + A^\top \lambda_{t+1} \quad \text{for } t = 0, \dots, N-1 \quad (\text{co-state equation})$$

$$\lambda_N = Q_f x_N \quad (\text{boundary condition})$$

From this point, we can use induction to prove that if $\lambda_{t+1} = P_{t+1} x_{t+1}$ (this holds for $t = N-1$ with $P_N = Q_f$, then we will have $\lambda_t = P_t x_t$, where P_t obeys the Riccati recursion. See the supplementary notes on LQR derivations for details.

4 Duality between the Kalman filter and LQR

In this section we will see that the Kalman Filter is in fact the dual of LQR. To do this, we will include a tracking input in the LQR cost function.

$$\begin{aligned} \underset{\substack{u_0, \dots, u_{N-1} \\ x_1, \dots, x_N}}{\text{minimize}} \quad & \frac{1}{2} \sum_{t=0}^{N-1} \left(x_t^\top Q x_t + (u_t - \bar{u}_t)^\top R (u_t - \bar{u}_t) \right) + \frac{1}{2} x_N^\top Q_f x_N \\ \text{s.t.} \quad & x_{t+1} = A x_t + B u_t \quad \text{for } t = 0, \dots, N-1 \end{aligned} \quad (9)$$

The corresponding Lagrangian is:

$$L(x, u, \lambda) = \frac{1}{2} \sum_{t=0}^{N-1} \left(x_t^\top Q x_t + (u_t - \bar{u}_t)^\top R (u_t - \bar{u}_t) \right) + \frac{1}{2} x_N^\top Q_f x_N + \sum_{t=0}^{N-1} \lambda_{t+1}^\top (A x_t + B u_t - x_{t+1})$$

Since our goal is to find the dual, we should minimize the Lagrangian with respect to the primal variables x and u . This amounts to computing gradients with respect to x_t and u_t and setting them equal to zero. This leads to the equations

$$\nabla_{u_t} L = 0 : \quad R(u_t - \bar{u}_t) + B^\top \lambda_{t+1} = 0 \quad \text{for } t = 0, \dots, N-1 \quad (10a)$$

$$\nabla_{x_t} L = 0 : \quad Q x_t + A^\top \lambda_{t+1} - \lambda_t = 0 \quad \text{for } t = 1, \dots, N-1 \quad (10b)$$

$$\nabla_{x_N} L = 0 : \quad Q_f x_N - \lambda_N = 0 \quad (10c)$$

Solving for x_t and u_t (assuming Q and Q_f are invertible), we obtain

$$\begin{aligned} x_N &= Q_f^{-1} \lambda_N \\ x_t &= Q^{-1} (\lambda_t - A^\top \lambda_{t+1}) \quad \text{for } t = 0, \dots, N-1 \\ u_t &= \bar{u}_t - R^{-1} B \lambda_{t+1} \quad \text{for } t = 0, \dots, N-1 \end{aligned} \quad (11)$$

Substituting the expressions (11) back into the formula for the Lagrangian and performing some extensive simplifications, we obtain the expression for the dual function $F(\lambda)$:

$$\begin{aligned} F(\lambda) = -\frac{1}{2} \sum_{t=0}^{N-1} \left((\lambda_t - A^\top \lambda_{t+1})^\top Q^{-1} (\lambda_t - A^\top \lambda_{t+1}) + (B^\top \lambda_{t+1} - R \bar{u}_t)^\top R^{-1} (B^\top \lambda_{t+1} - R \bar{u}_t) \right) \\ - \frac{1}{2} \lambda_N^\top Q_f^{-1} \lambda_N + \frac{1}{2} \sum_{t=0}^{N-1} \bar{u}_t^\top R \bar{u}_t \end{aligned}$$

The dual problem is to maximize $F(\lambda)$, which is the same as minimizing $-F(\lambda)$. Define the new variables $w_t := \lambda_t - A^\top \lambda_{t+1}$ and $v_t := R \bar{u}_t - B^\top \lambda_{t+1}$. Also remove the $\bar{u}_t^\top R \bar{u}_t$ term since it is a constant. We can now rewrite the dual optimization problem in the following compact form.

$$\begin{aligned} \underset{\lambda, w, v}{\text{minimize}} \quad & \frac{1}{2} \lambda_N^\top Q_f^{-1} \lambda_N + \frac{1}{2} \sum_{t=0}^{N-1} \left(w_t^\top Q^{-1} w_t + v_t^\top R^{-1} v_t \right) \\ \text{subject to:} \quad & \lambda_t = A^\top \lambda_{t+1} + w_t \\ & R \bar{u}_t = B^\top \lambda_{t+1} + v_t \end{aligned}$$

This is precisely the MAP optimization formulation for a Kalman filtering problem!

The table below shows how the LQR-related symbols correspond to the associated symbols we used when studying the Kalman filtering problem.

LQR	KF
Q_f	Σ_0
Q	W
R	V
A^\top	A
B^\top	C
$R\bar{u}_t$	y_t
λ_N	x_0
\vdots	\vdots
λ_0	x_N

Therefore, the dual of LQR with input tracking is precisely the MAP formulation of Kalman filtering, where the Lagrange multipliers used to encode the dynamics in LQR (the co-state) becomes the state of the dynamical system we are estimating for the filtering problem, and now time flows in the reverse direction.