Discussion 9

1. An optimization problem

Consider the primal optimization problem,

$$p^\star = \min_{x \in \mathbb{R}^2} \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 \tag{1}$$
$$\text{s.t. } x_1 \geq 0 \tag{2}$$
$$x_1 + x_2 \geq 2. \tag{3}$$

First we solve the primal problem directly.

(a) Sketch the feasible region and argue that $x^\star = (1, 1)$ and $p^\star = 1$.

Solution: The feasible region is shown in Figure 1 as the green shaded region.

The level curves for the function are concentric circles centered around the origin and the objective value increases as we go away from the origin. Thus the optimal point is the point in the feasible region that is closest to the origin and this point is $(1, 1)$ which gives us $p^\star = 1$.

(b) The *critical points* of an optimization problem are points where the gradient is 0 or undefined, and also points which are on the boundary of the constraint set.

Compute the value of the objective function at its critical points and find $p^\star$ and $x^\star$.

Solution: The critical points are as follows:

i. Point where gradient of objective function is 0. i.e $x_1 = 0, x_2 = 0$. This point is infeasible.

ii. Points at infinity. As we go to infinity along any direction the objective value goes to infinity as well so this cannot be optimal.

iii. Points along the boundary $x_1 = 0, x_1 + x_2 \geq 2$. Here the optimal value is given by,

$$\min_{x_2 \in \mathbb{R}} \frac{1}{2}x_2^2$$
$$x_2 \geq 2$$

The minimum is achieved by $x_2 = 2$ and has value 2.

iv. Points along the boundary $x_1 + x_2 = 2, x_1 \geq 0$. Here the optimal value is given by,

$$\min_{x_1 \in \mathbb{R}} \frac{1}{2}x_1^2 + \frac{1}{2}(2 - x_1)^2$$
$$x_1 \geq 0$$

The minimum is achieved at $x_1 = 1$ and has value 1. The corresponding value for $x_2 = 1$ and by comparing this to values at other critical points we conclude that $x^\star = (1, 1)$ and $p^\star = 1$.

(c) Next we solve the problem with the help of the dual. First, find the Lagrangian $\mathcal{L}(x, \lambda)$.

Solution: The Lagrangian $\mathcal{L}$ is:

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + \lambda_1(-x_1) + \lambda_2(-x_1 - x_2 + 2).$$

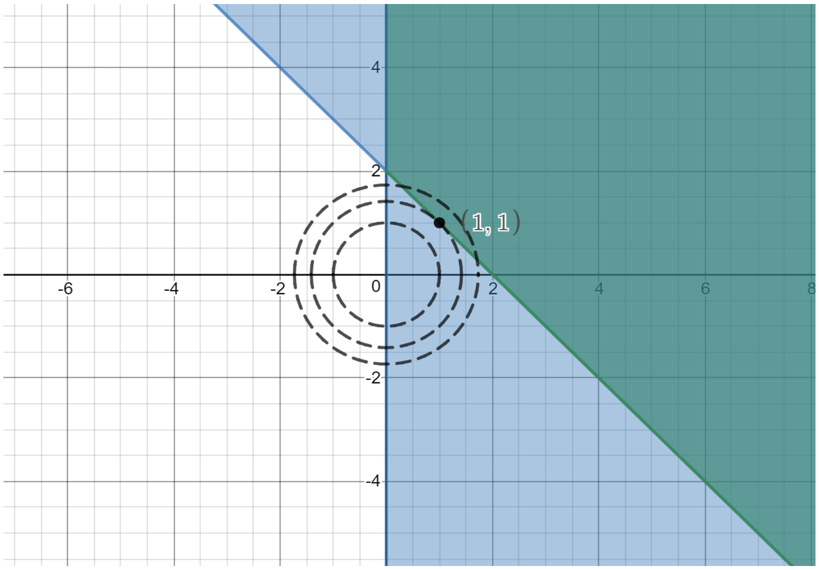

Figure 1

(d) Formulate the dual problem.

Solution: The dual problem writes:

$$d^\star = \max_{\vec{\lambda} \geq 0} \min_{x} \mathcal{L}(x, \vec{\lambda})$$

.

Specifically, we can define $g(\vec{\lambda}) := \min_{\vec{x} \in \mathbb{R}^2} \mathcal{L}(\vec{x}, \vec{\lambda})$. To find the minimizer $\vec{x}^\star(\vec{\lambda})$, we can set the gradient to zero:

$$\nabla \mathcal{L}(\vec{x}^\star(\vec{\lambda}), \vec{\lambda}) = \begin{bmatrix} x_1^\star(\vec{\lambda}) - \lambda_1 - \lambda_2 \\ x_2^\star(\vec{\lambda}) - \lambda_2 \end{bmatrix} = 0.$$

This gives us $\vec{x}^\star(\vec{\lambda}) = (\lambda_1 + \lambda_2, \lambda_2)$. Then, we have:

$$\begin{aligned} g(\vec{\lambda}) &= \mathcal{L}(x^\star(\vec{\lambda}), \vec{\lambda}) \\ &= \frac{1}{2}(\lambda_1 + \lambda_2)^2 + \frac{1}{2}\lambda_2^2 + \lambda_1(-\lambda_1 - \lambda_2) + \lambda_2(-\lambda_1 - 2\lambda_2 + 2) \\ &= -\frac{1}{2}\lambda_1^2 - \lambda_1\lambda_2 - \lambda_2^2 + 2\lambda_2. \end{aligned}$$

so the dual problem is:

$$d^\star = \max_{\vec{\lambda} \geq 0} \quad -\frac{1}{2}\lambda_1^2 - \lambda_1\lambda_2 - \lambda_2^2 + 2\lambda_2$$

(e) Solve the dual problem to find $d^\star$ and $\vec{\lambda}^\star$.

Solution: Our dual problem is:

$$d^\star = \max_{\vec{\lambda} \geq 0} \quad -\frac{1}{2}\lambda_1^2 - \lambda_1\lambda_2 - \lambda_2^2 + 2\lambda_2$$

Note that we can bound it as follows, since $\vec{\lambda} \geq 0$:

$$\begin{aligned} -\frac{1}{2}\lambda_1^2 - \lambda_1\lambda_2 - \lambda_2^2 + 2\lambda_2 &\leq -\lambda_2^2 + 2\lambda_2 \\ &= -(\lambda_2 - 1)^2 + 1 \\ &\leq 1. \end{aligned}$$

This bound is also achievable: we have equality if and only if $\vec{\lambda} = (0, 1)$. So, $d^\star = 1$ and $\vec{\lambda}^\star = (0, 1)$.

(f) Does strong duality hold?

Solution: Yes, it does hold, since the primal is convex and Slater's condition holds (we can find a point in the interior of the feasible set, such as $(2, 2)$).

(g) Find $p^\star$ and $x^\star$.

Solution: By strong duality $p^\star = 1$ and $x^\star = (1, 1)$.

(h) Finally, we use KKT conditions to find $x^\star$, $\lambda^\star$. First, Write down the KKT conditions and find $\tilde{x}$ and $\tilde{\lambda}$ that satisfy it.

Solution: The KKT conditions are

i. (Lagrangian Stationarity) $\nabla_x = 0 \Rightarrow x + (-\lambda_1, 0)^\top + (-\lambda_2, -\lambda_2)^\top = 0$
ii. (Dual Feasibility) $\lambda_1, \lambda_2 \geq 0$
iii. (Primal Feasibility) $x_1 \geq 0$ and $x_1 + x_2 \geq 2$
iv. (Complementary Slackness) $\lambda_1 x_1 = 0$ and $\lambda_2(2 - x_1 - x_2) = 0$

From the first condition we have that

$$\tilde{x_1} = \tilde{\lambda_1} + \tilde{\lambda_2}, \quad \tilde{x_2} = \tilde{\lambda_2}$$

This implies from the fourth condition that $\tilde{\lambda_1}(\tilde{\lambda_1} + \tilde{\lambda_2}) = 0$ which implies $\tilde{\lambda_1} = 0$. It follows that $\tilde{x_1} = \tilde{x_2} = \tilde{\lambda_2} = 1$ since $\tilde{x_1} = \tilde{x_2} = \tilde{\lambda_2} = 0$ does not satisfy the third condition.

(i) Argue why the optimal primal and dual solutions are given by $x^\star = \tilde{x}$ and $\lambda^\star = \tilde{\lambda}$.

Solution: The primal problem is convex and has strong duality so the KKT conditions are both necessary and sufficient for optimality.

Aside: An aside connecting the necessity/sufficiency of the KKT conditions, Slater's condition, and strong duality.

i. Necessary and Sufficient conditions. A *necessary* condition for a point to be a local optimum is a condition such that any point that is a local optimum must satisfy this condition. A *sufficient* condition for a point to be a local optimum means that if a point satisfies this condition, then the point must be an optimal point.
While subtle, notice how different necessity and sufficiency are: if the KKT conditions are necessary, an optimal point must satisfy them but that is not to say there aren't points that satisfy the KKT conditions but are not local optima. On the other hand, sufficiency means that if you satisfy the KKT conditions then you must be a local optima. In this sense, if you have necessity and sufficiency of the KKT conditions for an optimization problem, you're in great shape to solve your problem.
ii. Given an arbitrary optimization problem, the KKT conditions **DO NOT** have to be sufficient or necessary; that is the optimal solution of a problem need not satisfy the KKT conditions.

iii. For an optimization problem, if $x^\star$ is a local optima and the problem satisfies strong duality, then the KKT conditions are necessary conditions. This means that there must exist exist $(\lambda^\star, \nu^\star)$ dual variables satisfying such that $x^\star, \lambda^\star, \nu^\star$ all satisfy the KKT conditions.

iv. There are a few weird/funky regularity conditions for some problems that also imply strong duality and make the KKT conditions necessary for optimality (we never discussed these in class). When the problem is nice and convex, the sledgehammer of constraint qualifications is what we all know and love – Slater's Condition or Slater's Constraint Qualification (SCQ). That is not to say there are not others; just that SCQ is the usually the easiest to verify and happens to work in most practical settings.

v. SCQ is also incredible because in addition to providing us a constraint qualification making KKT necessary, it also gives us strong duality. However, we'll see below that KKT is sufficient for convex problems and that strong duality implies $\iff$ KKT conditions for convex problems as well, i.e. if you have strong duality and convexity the the KKT condition are both necessary and sufficient for optimality of a point.

vi. It is an amazing fact that for convex problems, the KKT conditions are **always sufficient**. Note this is not a statement about SCQ. To prove this, assume we have the convex program

$$\begin{aligned} \min_x \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, \quad \forall i = 1, ..., m \\ & Ax = b \end{aligned}$$

Then given $(x^\star, \lambda^\star, \nu^\star)$ that satisfy the KKT conditions, we have

$$\nabla f_0(x^\star) + \sum \lambda_i^\star \nabla f_i(x^\star) + A^\top \nu^\star = 0$$

and $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, ..., m$. Since the f_i are convex, then

$$\mathcal{L}(x, \lambda^\star, \nu^\star) = f_0(x) + \sum \lambda_i^\star f_i(x) + {\nu^\star}^\top (Ax - b)$$

is convex in x and thus $x^\star$ is a minimizer of $\mathcal{L}(x, \lambda^\star, \nu^\star)$ (note we need convexity here because if it were non-convex then an $x^\star$ satisfying lagrangian stationarity might not be the global minimizer of the lagrangian). Thus if x is feasible for the primal problem then

$$\begin{aligned} f_0(x) &\ge \mathcal{L}(x, \lambda^\star, \nu^\star) \\ &\ge \mathcal{L}(x^\star, \lambda^\star, \nu^\star) \\ &= f_0(x^\star) \end{aligned}$$

and hence $x^\star$ is a minimizer (using the fact that $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, ..., m$) to get the equality in the last line.

vii. Slater's theorem is a statement about when we have strong duality. Having strong duality implies the KKT conditions always hold. To see this if you have strong duality you have the chain

$$\begin{aligned} d^\star &= g(\lambda^\star, \nu^\star) \\ &= \min_x f_0(x) + \sum \lambda_i^\star f_i(x) + \sum \nu_i^\star h_i(x) \\ &\le f_0(x^\star) + \sum \lambda_i^\star f_i(x^\star) + \sum \nu_i^\star h_i(x^\star) \\ &\le f_0(x^\star) \end{aligned}$$

$$= p^\star$$

where f_i are convex and h_i are affine. Since $d^\star = p^\star$ we can infer complementary slackness and infer that $x^\star$ minimizes $\mathcal{L}(x, \lambda^\star, \nu^\star)$ by the third line. These combined with primal and dual feasibility are the KKT conditions.

viii. We have

A. Optimality and Strong Duality $\Rightarrow$ KKT (for all problems)

B. KKT and convexity $\Rightarrow$ optimality and Strong Duality (for convex problems)

C. SCQ $\Rightarrow$ Strong duality for convex problems

D. If convexity and strong duality hold then we have: KKT $\iff$ optimality.

ix. For more general theory and more discussion on the topic, see the more general Fritz John conditions. The theory concerning these topics is involved enough and can be combined with duality theory to make an entire course worth of material.

2. Lagrangian Dual of a QP

Consider the standard form of a convex quadratic program, with $Q \succ 0$:

$$\begin{aligned} \min_{\vec{x}} \quad & \frac{1}{2}\vec{x}^\top Q\vec{x} && (4) \\ \text{s.t.} \quad & A\vec{x} \le \vec{b} && (5) \end{aligned}$$

(a) Write the Lagrangian function $\mathcal{L}(\vec{x}, \vec{\lambda})$.

Solution:

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = \frac{1}{2}\vec{x}^\top Q\vec{x} + \vec{\lambda}^\top (A\vec{x} - \vec{b}) \tag{6}$$

(b) Write the Lagrangian dual function, $g(\vec{\lambda})$.

Solution:

$$g(\vec{\lambda}) = \inf_{\vec{x}} \ \mathcal{L}(\vec{x}, \vec{\lambda}) \tag{7}$$

We can find this infimum by setting $\nabla_{\vec{x}}\mathcal{L}(\vec{x}^\star, \vec{\lambda}) = 0$:

$$Q\vec{x}^\star + A^\top\vec{\lambda} = 0 \implies \vec{x}^\star = -Q^{-1}A^\top\vec{\lambda} \tag{8}$$

Substituting, we get

$$\begin{aligned} g(\vec{\lambda}) &= \mathcal{L}(\vec{x}^\star, \vec{\lambda}) && (9) \\ &= \frac{1}{2}\vec{\lambda}^\top AQ^{-\top}A^\top\vec{\lambda} - \vec{\lambda}^\top AQ^{-1}A^\top\vec{\lambda} - \vec{\lambda}^\top\vec{b} && (10) \\ &= -\frac{1}{2}\vec{\lambda}^\top AQ^{-1}A^\top\vec{\lambda} - \vec{\lambda}^\top\vec{b} && (11) \end{aligned}$$

(c) Show that the Lagrangian dual problem is convex by writing it in "standard QP form" – that is, in a form similar to the original problem. Is the Lagrangian dual problem convex in general?

Solution: The Lagrangian dual problem writes

$$\max_{\vec{\lambda}\ge 0} \ g(\vec{\lambda}) = \max_{\vec{\lambda}\ge 0} \ -\frac{1}{2}\vec{\lambda}^\top AQ^{-1}A^\top\vec{\lambda} - \vec{\lambda}^\top\vec{b}, \tag{12}$$

the maximization of a concave function of $\vec{\lambda}$ over the convex region given by the non-negative orthant $\vec{\lambda} \geq 0$. The dual problem is therefore convex.

While in this problem, the primal problem was convex, it turns out that the Lagrangian dual problem is a convex problem even when the primal is not. To see this, examine its general form:

$$\max_{\vec{\lambda} \geq 0} \min_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda}) = \max_{\vec{\lambda} \geq 0} \min_{\vec{x}} \left[f_0(\vec{x}) + \sum_{i=1}^{n} \lambda_i f_i(\vec{x}) \right] \tag{13}$$

This represents the pointwise minimum of affine functions of $\vec{\lambda}$, which we know to be concave. The resulting maximization problem of a concave objective in $\vec{\lambda}$ over the convex region $\vec{\lambda} \geq 0$ is then a convex optimization problem!

