EECS 127/227AT **Fall 2020**
Solutions for Discussion 6

1 The flexibility of duality

In class and in our two textbooks the emphasis when Lagrange duality is introduced is on dualizing the indicator functions $I_{\mathbb{R}_-}$ and $I_{\{0\}}$ (each of which is a convex and lower semicontinuous function). However, the general scheme behind Lagrange duality is much broader and one can dualize by expressing any lower semicontinuous convex function in terms of its Fenchel conjugate. This flexibility is very important in creating more insightful and useful duals.

In this question we will illustrate this flexibility of duality in a toy problem. In the next question in this discussion set we will see a more interesting example of the use of this flexibility. Recall from Question 3 of Discussion 5 that we know that that $I^{**}_{\mathbb{R}_-} = I_{\mathbb{R}_-}$ and $I^{**}_{\{0\}} = I_{\{0\}}$, because $\mathbb{R}_- := \{x \in \mathbb{R} : x \leq 0\}$ and $\{0\}$ are closed convex subsets of the real line.

(a) Verify that the Fenchel conjugate of $I_{\mathbb{R}_-}$ is $I_{\mathbb{R}_+}$. Conclude that we can write

$$I_{\mathbb{R}_-}(x) = \sup_{z \geq 0} zx. \tag{1}$$

(It is of course directly obvious, without going through Fenchel conjugates, that the equality claimed in equation (1) is true. However, we aim to make a much broader point, for which it is useful to take a less direct route to this equality.)

Solution:
We have

$$\begin{aligned} I^{*}_{\mathbb{R}_-}(z) &:= \sup_{x \in \mathbb{R}} zx - I_{\mathbb{R}_-}(x) \\ &= \sup_{x \leq 0} zx \\ &= \begin{cases} \infty & \text{if } z < 0, \\ 0 & \text{if } z \geq 0 \end{cases} \\ &= I_{\mathbb{R}_+}(z), \end{aligned}$$

which establishes the claimed form of the Fenchel conjugate of $I_{\mathbb{R}_-}$. After this, we have

$$\begin{aligned} I_{\mathbb{R}_-}(x) &= I^{**}_{\mathbb{R}_-}(x) \\ &:= \sup_{z \in \mathbb{R}} xz - I^{*}_{\mathbb{R}_-}(z) \\ &= \sup_{z \geq 0} xz, \end{aligned}$$

which establishes equation (1)

(b) Verify that the Fenchel conjugate of $I_{\{0\}}$ is 0. Conclude that we can write

$$I_{\{0\}}(x) = \sup_{z \in \mathbb{R}} zx. \tag{2}$$

(Again, it is directly obvious, without going through Fenchel conjugates, that the equality claimed in equation (2) is true, but we are deliberately showing this in a more complicated way.)

Solution:
We have

$$\begin{aligned} I^*_{\{0\}}(z) &:= \sup_{x\in\mathbb{R}} zx - I_{\{0\}}(x) \\ &= \sup_{x=0} zx \\ &= 0, \end{aligned}$$

which establishes the claimed form of the Fenchel conjugate of $I_{\{0\}}$. After this, we have

$$\begin{aligned} I_{\{0\}}(x) &= I^{**}_{\{0\}}(x) \\ &:= \sup_{z\in\mathbb{R}} xz - I^*_{\{0\}}(z) \\ &= \sup_{z\in\mathbb{R}} xz, \end{aligned}$$

which establishes equation (2)

What is going on in traditional Lagrange duality is that we recognize that the constrained convex optimization problem:

$$\begin{array}{ll} \min\limits_x & f_0(x) \\ \text{subject to:} & f_i(x) \leq 0, i = 1, \ldots, m, \\ & h_i(x) = 0, i = 1, \ldots, p, \end{array}$$

is the same as the unconstrained optimization problem

$$\min_x f_0(x) + \sum_{i=1}^m I_{\{f_i(x)\leq 0\}} + \sum_{j=1}^p I_{\{h_j(x)=0\}},$$

namely

$$\min_x f_0(x) + \sum_{i=1}^m I_{\mathbb{R}_-}(f_i(x)) + \sum_{j=1}^p I_{\{0\}}(h_j(x)),$$

and then we use equation (1) to replace each term of the form $I_{\mathbb{R}_-}(f_i(x))$ (the corresponding dual variable is denoted λ_i and must satisfy $\lambda_i \geq 0$), and equation (2) to replace each term of the form $I_{\{0\}}(h_j(x))$ (the corresponding dual variable is denoted ν_j and is unconstrained). This gives us a min-max problem (min over the primal variables and max over the dual variables) which we dualize to a max-min problem, as discussed in class and in the two textbooks. In the max-min problem, after we take the min (over the primal variables) we get a function of the dual variables (which is what we call the dual objective function) and the problem of maximizing the dual objective function subject to the constraints on the dual variables is what is called the dual problem (as we saw, strictly speaking we have to consider the problem of minimizing the negative of the dual objective function to get a convex optimization problem)

However, we can play the same game in a more sophisticated way to create many other dual problems for a given primal problem, as we will now illustrate in a toy problem (see also the next question in this discussion set).

First, let us note that in both the general convex optimization problem and its unconstrained version above, it is implicitly assumed that the minimization is over x in the intersection of the domains of all the functions $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$. This is important to keep in mind for what follows.

(c) Consider the following convex optimization problem

$$\min_x \frac{1}{2}x^2 - \log x.$$

Here the logarithm is to the natural base, the function $\frac{1}{2}x^2$ has domain $\mathbb{R}$ and the function $-\log x$ has domain $\mathbb{R}_{++}$. Also note that it is implicitly assumed that the minimization is over the set of x that lie in the domains of all the functions involved, which in this case is $\mathbb{R}_{++}$.

Find the optimal value of the problem and an optimal point.

Solution:
We can solve the problem by calculus. As the sum of two convex functions, the objective is convex. Its derivative (on $\mathbb{R}_{++}$) is $x - \frac{1}{x}$, which equals zero at $x = 1$, so the optimal point is $x^* = 1$ and the optimal value is $p^* = \frac{1}{2} - \log(1) = \frac{1}{2}$.

(d) Dualize $-\log x$ to write the convex optimization problem of the preceding part of the question as a min-max problem. For this, recall that in Question 2(a) of Discussion 5 you have shown that the Fenchel conjugate of $\phi(x) := -\log x$, with $\mathrm{dom}(\phi) = \mathbb{R}_{++}$, is given by

$$\phi^*(z) = \begin{cases} -1 - \log|z| & \text{if } z < 0, \\ \infty & \text{otherwise.} \end{cases}$$

Further, note that ϕ is a lower semicontinuous convex function, so it equals the Fenchel conjugate of its Fenchel conjugate.

Solution:
Since ϕ is a lower semicontinuous convex function, we have $\phi(x) = \phi^{**}(x)$. Hence we have

$$\phi(x) = \sup_{z\in\mathbb{R}}\ xz - \phi^*(z) = \sup_{z<0}\ xz + 1 + \log|z|.$$

This allows us to write [1]

$$\min_x \frac{1}{2}x^2 - \log x = \min_x \max_{z<0} \frac{1}{2}x^2 + xz + 1 + \log|z|.$$

An important point to make here is that on the left hand side it is implicitly assumed that x is restricted to the domain where we can evaluate the functions involved, namely $\mathbb{R}_{++}$, while on the right hand side x can range over all of $\mathbb{R}$, because we have dualized away the part of the function that was causing the restriction in the domain. Another way to understand this is by recognizing that our convention is that the convex function $-\log x$ equals ∞ outside its domain, so the equation above is actually valid for all $x \in \mathbb{R}$ with this convention.

(e) Dualize the min-max problem found in the preceding part of the question to a max-min problem, find the corresponding dual objective and state the corresponding dual problem.

[1] Please note that in the course we use $\sup$ and $\max$ interchangeably and we use $\min$ and $\inf$ interchangeably. This is because in convex optimization when one writes $\min$ one means "minimize", without necessarily implying that the minimum exists and, similarly, when one write $\max$ one means "maximize", without necessarily implying that the maximum exists. In other mathematical subjects people care about the distinction between $\sup$ and $\max$ and the one between $\inf$ and $\min$.

Solution:
The max-min version of the min-max problem found in the preceding part of this question is

$$\max_{z<0} \min_{x} \frac{1}{2}x^2 + xz + 1 + \log|z|.$$

The dual objective, which we denote by $g(z)$, is defined for $z < 0$ via

$$g(z) = \min_{x} \frac{1}{2}x^2 + xz + 1 + \log|z|.$$

By calculus we see that the minimum occurs at $x = |z|$ and so

$$g(z) = -\frac{1}{2}z^2 + 1 + \log|z|,$$

(defined for $z < 0$).

The dual optimization problem is therefore

$$\max_{z<0} -\frac{1}{2}z^2 + 1 + \log|z|.$$

(f) Find the optimal value of the dual problem found in the preceding part of this question and also an optimal point for this optimization problem. Verify that strong duality holds.

Solution:
The dual objective function is concave over its domain, which is the convex set $\mathbb{R}_{--}$. The derivative of $-\frac{1}{2}z^2 + 1 + \log|z|$ in z, for $z < 0$, is $|z| - \frac{1}{|z|}$. This equals zero at $z = -1$, so this is the dual optimal point. The dual optimal value is then $d^* = -\frac{1}{2} + 1 + \log(1) = \frac{1}{2}$. Since we have $d^* = p^*$, we see that strong duality holds.

(g) **(Optional)**

Returning to the original primal optimization problem in part (c) of this question, dualize the component $\frac{1}{2}x^2$ in the objective to get a different min-max formulation of the primal problem, find the corresponding max-min problem, the corresponding dual objective function, and the corresponding dual optimal point and dual optimal value. Show that strong duality once again holds.

Solution:
We observe that $\frac{1}{2}x^2$ with domain $\mathbb{R}$ is a lower semicontinuous convex function and we know, e.g. from Question 3(b) of Homework 5, that its Fenchel conjugate is $\frac{1}{2}z^2$, with domain $\mathbb{R}$. This means we can write

$$\min_{x} \frac{1}{2}x^2 - \log x = \min_{x} \max_{z} xz - \frac{1}{2}z^2 - \log x.$$

Here, both on the left hand side and on the right hand side, it is implicitly assumed that x is restricted to the domain of the functions involved, namely $\mathbb{R}_{++}$.

The corresponding max-min problem is then

$$\max_{z} \min_{x>0} xz - \frac{1}{2}z^2 - \log x.$$

The dual objective $g(z)$ is now defined for all $z \in \mathbb{R}$ and is given by

$$g(z) := \min_{x>0} \; xz - \frac{1}{2}z^2 - \log x = \begin{cases} 1 - \frac{1}{2}z^2 + \log z & \text{if } z > 0, \\ -\infty & \text{if } z \leq 0. \end{cases}$$

The dual optimization problem can therefore be written as

$$\max_{z>0} \; 1 - \frac{1}{2}z^2 + \log z.$$

The dual optimal point is $z^* = 1$ and the dual optimal value is $d^* = \frac{1}{2}$. Since $d^* = p^*$, strong duality holds.

2 Least absolute deviations with ℓ_∞ regularization

Let $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$ and $\mu > 0$. Consider the problem

$$p^* = \min_x \|Ax - y\|_1 + \mu\|x\|_\infty.$$

In the *least absolute deviations* problem we aim to find the vector of regression coefficients to minimize the sum of the absolute values of the residuals, i.e. given A and y, we seek to find x to minimize $\|Ax - y\|_1$. This is considered a more robust criterion than the least squares criterion because it pays less attention to outliers in the data (the least squares criterion, by squaring large residuals, tends to overemphasize outliers). Here we are considering an ℓ_∞ regularization of the least absolute deviations problem, by introducing a penalty on the ℓ_∞ norm of the vector of regression coefficient, scaled by the hyperparameter μ. Such ℓ_∞ regularization is quite unusual, compared to ℓ_1 or ℓ_2 regularization. It does not promote sparsity, but tends to make all the regression coefficients of roughly the same size, so that many of them will be nonzero but they will all be small.

For $j \in \{1, \ldots, n\}$, denote by $a_j \in \mathbb{R}^m$ the j-th column of A, so that $A = [a_1, \ldots, a_n]$, and define

$$\|A\|_1 := \sum_{j=1}^n \|a_j\|_1.$$

You should check for yourself that this notation makes sense, in that $\|A\|_1$, as defined above, is the induced ℓ_1 norm of the matrix A, namely

$$\|A\|_1 = \max_{x:\|x\|_1=1} \|Ax\|_1$$

(a) Express the problem as an LP.

Solution:
The problem can be written as

$$\begin{aligned} \min_{x,z,t} \quad & z^T\mathbb{1} + \mu t \\ \text{subject to:} \quad & t \geq \|x\|_\infty, \\ & z_i \geq |(Ax - y)_i|, \quad i = 1, \ldots, m, \end{aligned}$$

where $\mathbb{1}$ denotes the all-ones column vector. This is an LP, as can be seen by writing it in more detail as

$$\begin{aligned} \min_{x,z,t} \quad & z^T\mathbb{1} + \mu t \\ \text{subject to:} \quad & t \geq x_j, \quad t \geq -x_j, j = 1, \ldots, n, \\ & z_i \geq (Ax - y)_i, \quad z_i \geq -(Ax - y)_i, \quad i = 1, \ldots, m. \end{aligned} \tag{3}$$

(b) Show that a dual to the problem can be written as

$$\begin{aligned} d^* = \max_u \quad & -u^T y \\ \text{subject to:} \quad & \|u\|_\infty \leq 1, \\ & \|A^T u\|_1 \leq \mu. \end{aligned}$$

Hint: Use the fact that for any vector z we have

$$\max_{u\,:\,\|u\|_1\le 1} u^T z = \|z\|_\infty, \quad \max_{u\,:\,\|u\|_\infty\le 1} u^T z = \|z\|_1.$$

You should check for yourself that these formulas are correct. They correspond to the fact that the ℓ_1 and the ℓ_∞ norms are duals of each other.

Remark: When there are many conditions, as in this problem, it may be more convenient to write optimization problems with the conditions written in line. Thus here we aim to show that the dual problem can be written as

$$d^* := \max_u \; -u^T y \;:\; \|u\|_\infty \le 1, \; \|A^T u\|_1 \le \mu.$$

Solution:

Based on the hint, we see that

$$\|Ax - y\|_1 = \max_{u:\|u\|_\infty\le 1} u^T(Ax - y),$$

and

$$\begin{aligned} \mu\|x\|_\infty &= \max_{v:\|v\|_1\le 1} \mu v^T x \\ &= \max_{v:\|v\|_1\le \mu} v^T x. \end{aligned}$$

These are Fenchel conjugate formulas which, along the lines of the first question of this discussion set, lead us to consider the Lagrangian

$$\mathcal{L}(x, u, v) := u^T(Ax - y) + v^T x,$$

which is such that

$$p^* = \min_x \max_{\|u\|_\infty\le 1, \|v\|_1\le\mu} \mathcal{L}(x, u, v). \tag{4}$$

Exchanging the min and max leads us to consider the dual objective, defined for

$$\{(u, v) : \|u\|_\infty \le 1, \|v\| \le \mu\},$$

and given by

$$g(u, v) := \min_x \; \mathcal{L}(x, u, v) = \begin{cases} -u^T y & \text{if } A^T u + v = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

The dual optimization problem can now be written as

$$d^* = \max_{u,v} \; -u^T y \;:\; A^T u + v = 0, \;\; \|u\|_\infty \le 1, \;\; \|v\|_1 \le \mu.$$

We see that we can eliminate v to get the following simpler form of the dual optimization problem

$$d^* = \max_u \; -u^T y \;:\; \|u\|_\infty \le 1, \;\; \|A^T u\|_1 \le \mu.$$

This is exactly what we were asked to show.

Note that we have weak duality, i.e.

$$p^* \ge d^*,$$

as is always the case.

(c) It can be shown that strong duality holds, i.e. $d^* = p^*$. Assuming this, show that the condition "$\|A^T u\|_1 < \mu$ for every u with $\|u\|_\infty \leq 1$" ensures that $x = 0$ is optimal.

Solution:

If

$$\|A^T u\|_1 < \mu \text{ for every } u \text{ such that } \|u\|_\infty \leq 1, \tag{5}$$

then the constraint $\|A^T u\|_1 \leq \mu$ cannot be active in the dual. Therefore, the dual problem has the same value as one where that constraint is removed, meaning that

$$p^* = d^* = \max_{u\,:\,\|u\|_\infty \leq 1} -u^T y = \|y\|_1.$$

We observe that the value $p^* = \|y\|_1$ is attained for $x = 0$, which shows that $x = 0$ is optimal.

(d) Assuming once again that strong duality holds, show that the condition in the previous part of this question holds if $\mu > \|A\|_1$.

Remark: The point being made in this result is that if the hyperparameter is chosen to be too big, i.e. if we penalize the ℓ_∞ norm of the vector of regression parameters too much, then the solution to the ℓ_∞ regularized least absolute deviations problem will end up simply ignoring the data.

Solution:

The condition (5) is satisfied iff

$$\mu > \max_{u\,:\,\|u\|_\infty \leq 1} \sum_{j=1}^n |a_j^T u|.$$

Since

$$\max_{u\,:\,\|u\|_\infty \leq 1} \sum_{j=1}^n |a_j^T u| \leq \sum_{j=1}^n \max_{u\,:\,\|u\|_\infty \leq 1} |a_j^T u| = \|A\|_1,$$

the desired result follows.

3 KKT conditions

Consider the problem

$$\begin{aligned} \min_{x\in\mathbb{R}^2} \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & (x_1-1)^2 + (x_2-1)^2 \le 2, \\ & (x_1-1)^2 + (x_2+1)^2 \le 2, \end{aligned}$$

where $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top \in \mathbb{R}^2$.

(a) Sketch the feasible region and the level sets of the objective function.

Solution:
See Figure 1.

(b) Is this a convex optimization problem?

Solution:
The objective function is convex and the inequality constraints correspond to sublevel sets of convex functions. This results in a convex optimization problem.

(c) Write the Lagrangian for this problem in the traditional framework of convex duality.

Solution:
The Lagrangian is given by

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

Here λ_1 and λ_2 are the dual variables.

(d) Using Slater's condition, verify that strong duality holds in this problem in the context of the traditional Lagrangian duality.

Solution:
The simple version of Slater's condition tells us that in a convex optimization problem if there is a point in the relative interior of the intersection of the domains of all the functions defining the problem in standard form (i.e. the convex objective function f_0, the convex functions $f_1, \dots, f_m$ defining the inequality constraints, and the affine functions $h_1, \dots, h_p$ defining the equality constraints) such that all the inequality constraints are *strictly* satisfied at this point, then strong duality will hold in the context of the traditional Lagrange duality. Further, if the dual value d^* satisfies $d^* > -\infty$, then a dual optimal pair (λ^*, ν^*) will exist, where $\lambda_1, \dots, \lambda_m$ are the dual variables corresponding to the inequality constraints and $\nu_1, \dots, \nu_p$ are the dual variables corresponding to the equality constraints.

Here the problem is convex; the objective function and the two convex functions defining the constraints have domain $\mathbb{R}^2$ so the intersection of their domains in $\mathbb{R}^2$, whose interior is of course $\mathbb{R}^2$; and there are points in $\mathbb{R}^2$ where the inequality constraints are strictly satisfied (e.g., $x = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$), as can be seen from Figure 1. Hence the simple version of Slater's condition is satisfied and thus strong duality holds.

(e) Write the KKT conditions for this optimization problem.

Solution:
Having established that strong duality holds, and since the objective function and the constraint

functions are differentiable, we can seek to find a primal optimal point x^* and a dual optimal point λ^* (in general this would be (λ^*, ν^*)) by writing the KKT conditions. Note that from the theory we know that any primal optimal point and dual optimal point must satisfy these equations.

We can write the KKT conditions as follows:

(a) Lagrangian stationarity, which are the equations that in general read $\nabla_x \mathcal{L}(x, \lambda^*, \nu^*)|_{x=x^*} = 0$ at a proposed dual optimal pair (λ^*, ν^*) and a proposed primal optimal point x^*. In this question (where, of course, there are no equality constraints) these read:

$$x_1^* + (\lambda_1^* + \lambda_2^*)(x_1^* - 1) = 0,$$

$$x_2^* + \lambda_1^*(x_2^* - 1) + \lambda_2^*(x_2^* + 1) = 0.$$

(b) Primal feasibility, which are the equations exhibiting the feasibility of the proposed primal optimal point x^*. In this question, these read:

$$(x_1^* - 1)^2 + (x_2^* - 1)^2 - 2 \leq 0,$$

$$(x_1^* - 1)^2 + (x_2^* + 1)^2 - 2 \leq 0.$$

(c) Dual feasibility, which are the equations exhibiting the feasibility of the proposed dual optimal pair (λ^*, ν^*), i.e. basically that $\lambda^* \geq 0$, this equation being interpreted coordinatewise. In this question, these read:

$$\lambda_1^* \geq 0, \ \lambda_2^* \geq 0.$$

(d) Complementary slackness, which are the equations that tell us that for any proposed dual optimal pair (λ^*, ν^*) and any proposed primal optimal point x^*, for each inequality constraint that is not active at the proposed primal optimal point the corresponding dual variable in the proposed dual optimal pair must equal zero. In this question, these read:

$$\lambda_1^*[(x_1^* - 1)^2 + (x_2^* - 1)^2 - 2] = 0,$$

$$\lambda_2^*[(x_1^* - 1)^2 + (x_2^* + 1)^2 - 2] = 0.$$

(f) Prove from the KKT conditions that $(x_1^*, x_2^*) = (0, 0)$ is a primal optimal point and $(\lambda_1^*, \lambda_2^*) = (0, 0)$ is a dual optimal point. Use this to find the primal optimal value p^*.

Solution:

It suffices to check that the choices $(x_1^*, x_2^*) = (0, 0)$ and $(\lambda_1^*, \lambda_2^*) = (0, 0)$ satisfy all the KKT equations. This is straightforward (check all the equations one by one).

The value of the objective function at $(0, 0)$ is 0, hence $p^* = 0$.

(g) Find all the solutions of the KKT equations.

Solution:

These KKT conditions have a unique solution, given by $x_1^* = x_2^* = 0$ and $\lambda_1^* = \lambda_2^* = 0$.

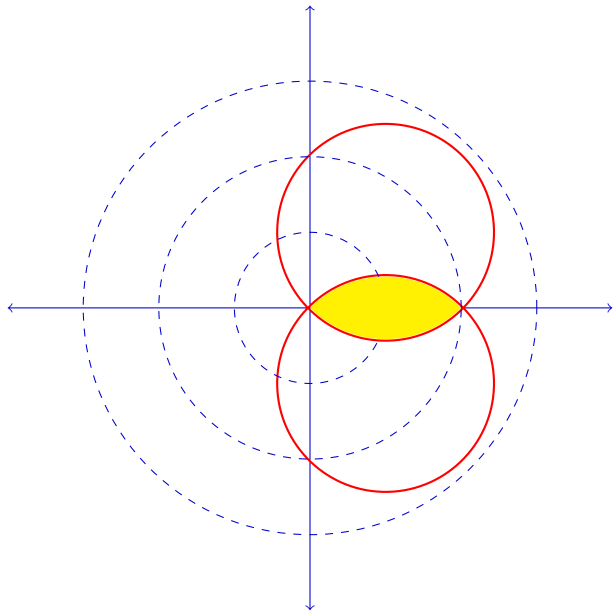

Figure 1: The feasible region of the problem in Question 3 is the closed convex region in yellow given by the intersection of the two disks whose boundaries are indicated by red circles. Some level sets of the objective function are also indicated with dashed blue lines. Since the origin is in the feasible region we can see directly that $x^* = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$ is an optimal point and the optimal value is$p^* = 0$.

To see this, note that the complementary slackness conditions, together with the primal feasibility and the dual feasibility conditions, tell us that there are four cases we need to consider.

Case 1: $\lambda_1^* = 0$ and $\lambda_2^* = 0$.

Case 2: $\lambda_1^* = 0$ and $(x_1^* - 1)^2 + (x_2^* + 1)^2 = 2$.

Case 3: $(x_1^* - 1)^2 + (x_2^* - 1)^2 = 2$ and $\lambda_2^* = 0$.

Case 4: $(x_1^* - 1)^2 + (x_2^* - 1)^2 = 2$ and $(x_1^* - 1)^2 + (x_2^* + 1)^2 = 2$.

These four cases are not mutually exclusive, i.e. more than one of these cases might hold. However, it suffices to analyze each case separately.

In case 1, we see directly from the Lagrangian stationarity conditions that $x_1^* = x_2^* = 0$. We also already have $\lambda_1^* = \lambda_2^* = 0$, by assumption.

In case 2, since $\lambda_1^* = 0$ the Lagrangian stationary conditions become:

$$x_1^* + \lambda_2^*(x_1^* - 1) = 0, \tag{6}$$

$$x_2^* + \lambda_2^*(x_2^* + 1) = 0.$$

Adding these together gives

$$(x_1^* + x_2^*)(1 + \lambda_2^*) = 0.$$

Sincd $\lambda_2^* \geq 0$ is one of the dual feasibility conditions, this implies that

$$x_1^* + x_2^* = 0.$$

Together with the assumption $(x_1^* - 1)^2 + (x_2^* + 1)^2 = 2$, this implies that (x_1^*, x_2^*) is either $(0, 0)$ or $(2, -2)$. The latter case can be excluded by using the primal feasibility condition $(x_1^* - 1)^2 + (x_2^* - 1)^2 \leq 2$, which it violates. We therefore conclude that $x_1^* = x_2^* = 0$.

Substituting this back into, say, equation (6), this gives $\lambda_2^* = 0$. We also already have $\lambda_1^* = 0$ as part of the assumption in case 2.

The analysis of case 3 is similar to that of case 2. The Lagrangian stationary conditions become:

$$x_1^* + \lambda_1^*(x_1^* - 1) = 0, \tag{7}$$

$$x_2^* + \lambda_1^*(x_2^* - 1) = 0.$$

Subtracting the second of these from the first gives

$$(x_1^* - x_2^*)(1 + \lambda_1^*) = 0.$$

Since $\lambda_1^* \geq 0$ is one of the dual feasibility conditions, this implies that

$$x_1^* - x_2^* = 0.$$

Together with the assumption that $(x_1^* - 1)^2 + (x_2^* - 1)^2 = 2$, this implies that (x_1^*, x_2^*) is either $(0, 0)$ or $(2, 2)$. The latter case can be excluded by using the primal feasibility condition $(x_1^* - 1)^2 + (x_2^* + 1)^2 \leq 2$, which it violates. We therefore conclude that $x_1^* = x_2^* = 0$.

Substituting this back into, say, equation (7), this gives $\lambda_1^* = 0$. We also already have $\lambda_2^* = 0$ as part of the assumption in case 3.

In case 4, a look at the figure shows that the only possibilities are that (x_1^*, x_2^*) is either $(0, 0)$ or $(2, 0)$. In the latter case the first of the Lagrangian stationarity conditions would become:

$$2 + \lambda_1^* + \lambda_2^* = 0.$$

However, since the dual feasibility conditions imply that $\lambda_1^* \geq 0$ and $\lambda_2^* \geq 0$, this is impossible. We therefore conclude that $x_1^* = x_2^* = 0$. Substituting this into the first of the Lagrangian stationarity conditions now gives

$$\lambda_1^* + \lambda_2^* = 0,$$

from which we conclude that $\lambda_1^* = \lambda_2^* = 0$. because we have $\lambda_1^* \geq 0$ and $\lambda_2^* \geq 0$ from the dual feasibility conditions.