EE 127/227A Final Review Problems

EE 127/227AT: Optimization Models in Engineering

Instructor: Profs. Gireeja Ranade, Venkat Anantharam
Authors: Yigit Efe Erginbas,
Kshitij Kulkarni, Chinmay Maheshwari,
Aditya Ramabadran, Chih-Yuan Chiu

December 11, 2023

1 Convex and Non-convex Optimization Problems

Fix non-zero vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$, with $n \geq 2$, fix $\alpha > 0$, $\beta \in \mathbb{R}$, and let $\vec{0} \in \mathbb{R}^n$ denote the n-dimensional zero vector.

1. Is the following optimization problem convex or non-convex? If it is convex, under what conditions do Slater's condition hold? Justify.

$$\begin{aligned} \min_{X \in \mathbb{R}^{n \times n}} \quad & \vec{u}^\top X \vec{v}, \\ \text{s.t.} \quad & \|X\|_F^2 \leq \alpha, \\ & X\vec{w} = \beta \vec{w}. \end{aligned}$$

2. Is the following optimization problem convex or non-convex? If it is convex, under what conditions do Slater's condition hold? Justify.

$$\begin{aligned} \min_{X \in \mathbb{R}^{n \times n}} \quad & \vec{u}^\top X \vec{v}, \\ \text{s.t.} \quad & \|X\|_F^2 = \alpha, \\ & X\vec{w} = \beta \vec{w}. \end{aligned}$$

Solution:

1. The given optimization problem is convex. First, the objective is a linear function in the components of X. One way to see this is to observe that $\vec{u}^\top X \vec{v} = \sum_{i=1}^n \sum_{j=1}^n u_i v_j X_{ij}$ is linear in $\{X_{ij} | i, j \in \{1, \cdots, n\}\}$. Another way is to observe that for any $X_1, X_2 \in \mathbb{R}^{n \times n}$ and $\alpha_1, \alpha_2 \in \mathbb{R}$, we have $\vec{u}^\top(\alpha_1 X_1 + \alpha_2 X_2)\vec{v} = \alpha_1(\vec{u}^\top X_1 \vec{v}) + \alpha_2(\vec{u}^\top X_2 \vec{v})$.

   Second, we claim that the inequality constraint defines a convex set. One way to see this is to observe that the given inequality constraint is equivalent to $\sum_{i=1}^n \sum_{j=1}^n X_{ij}^2 \leq \alpha$, which is the zero sub-level set of the function $f : \mathbb{R}^{n \times n} \to \mathbb{R}$, given by $f(X) := -\alpha + \sum_{i=1}^n \sum_{j=1}^n X_{ij}^2$ for each $X \in \mathbb{R}^{n \times n}$. Since f is a convex quadratic function in the components of X, we conclude that the inequality constraint characterizes a convex set.

   Third, note that the equality constraint is linear in terms of the components of X. Indeed, these constraints are:

$$\sum_{j=1}^n w_j X_{ij} = \beta w_i, \qquad \text{for each } i \in \{1, \cdots, n\}.$$

   Thus, the equality constraint characterizes a convex set (in fact, it is either the empty set, a single point, or an affine subspace of dimension at least 1).

   Finally, we claim that Slater's condition holds if and only if $\beta^2 < \alpha$. To see this, first note that since we are given that $\vec{w} \neq \vec{0}$, the constraint $X\vec{w} = \beta\vec{w}$ requires β to be an eigenvalue of X with corresponding eigenvector $\vec{w}$. We now claim that, if $|\beta| < \sqrt{\alpha}$, the following choice of X is a strictly feasible point that allows Slater's condition to be satisfied:

$$X = \frac{\beta}{\|\vec{w}\|_2^2} \vec{w}\vec{w}^\top.$$

To see this, note that:

$$X\vec{w} = \frac{\beta}{\|\vec{w}\|_2^2}\vec{w}(\vec{w}^\top \vec{w}) = \beta\vec{w},$$

as desired, and:

$$\|X\|_F^2 = \text{tr}(X^\top X) = \text{tr}\left(\frac{\beta^2}{\|\vec{w}\|_2^4}\vec{w}(\vec{w}^\top \vec{w})\vec{w}^\top\right) = \frac{\beta^2}{\|\vec{w}\|_2^2}\text{tr}(\vec{w}\vec{w}^\top) = \beta^2 < \alpha.$$

as desired, where the fact that $\|X\|_F^2 = \text{tr}(X^\top X)$ follows from Proposition 38 in Section 2.7 of the current version of the course reader, and the final equality follows by applying the cyclic property of the trace.

Conversely, we claim that if $\beta^2 \geq \alpha$, then Slater's condition is not satisfied. This is because, if $X \in \mathbb{R}^{n\times n}$ were to satisfy $X\vec{w} = \beta\vec{w}$, then we would have:

$$\|X\|_F \geq \|X\|_2 \geq \frac{\|X\vec{w}\|_2}{\|\vec{w}\|_2} = |\beta| \geq \sqrt{\alpha},$$

so $\|X\|_F^2 \geq \alpha$. Thus, there exists no matrix $X \in \mathbb{R}^{n\times n}$ satisfying $\|X\|_F^2 < \alpha$ and $X\vec{w} = \beta\vec{w}$ at the same time, i.e., the constraint set does not contain a strictly feasible point. As a result, Slater's condition is not satisfied.

2. The constraint set corresponding to the equality $\|X\|_F^2 = \alpha$ is non-convex. To see this, let $X_1 \in \mathbb{R}^{n\times n}$ the matrix whose $(1,1)$ component equals $\sqrt{\alpha}$ and whose remaining components are all 0, and let $X_2 \in \mathbb{R}^{n\times n}$ the matrix whose $(2,2)$ component equals $\sqrt{\alpha}$ and whose remaining components are all 0. Then $\|X_1\|_F^2 = \|X_2\|_F^2 = \alpha$, but:

$$\left\|\frac{1}{2}X_1 + \frac{1}{2}X_2\right\|_F^2 = \frac{1}{2}\alpha \neq \alpha.$$

In short, X_1 and X_2 are both contained in the set $\{X \in \mathbb{R}^{n\times n} | \|X\|_F^2 = \alpha\}$, but $\frac{1}{2}X_1 + \frac{1}{2}X_2$ is not, despite being a convex combination of X_1 and X_2. Thus, the set $\{X \in \mathbb{R}^{n\times n} | \|X\|_F^2 = \alpha\}$ is not convex.

2 Gradient Descent

1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice-differentiable function that we are attempting to minimize using Newton's method. Suppose that at the k^{th} iterate $\vec{x}_k \in \mathbb{R}^n$ we have $\nabla^2 f(\vec{x}_k) = \alpha_k I_n$, where $\alpha_k > 0$ is some positive constant and $I_n \in \mathbb{R}^{n\times n}$ is the identity matrix. **Write the Newton's method step for $\vec{x}_{k+1}$ in terms of $\vec{x}_k$, α_k, and $\nabla f(\vec{x}_k)$.**

2. Now suppose we are trying to minimize the same function f via gradient descent. **Write the gradient descent step for $\vec{x}_{k+1}$ in terms of $\vec{x}_k$ and $\nabla f(\vec{x}_k)$, with some arbitrary step size $\eta_k > 0$ at time k. For what value of η_k is the gradient descent update equation the same as the Newton's update equation from the last part?**

Solution:

1. Since $[\alpha_k I_n]^{-1} = \frac{1}{\alpha_k} I_n$, the Newton's method step is

$$\vec{x}_{k+1} = \vec{x}_k - \frac{1}{\alpha_k} \nabla f(\vec{x}_k). \tag{2.1}$$

2. The gradient descent step is

$$\vec{x}_{k+1} = \vec{x}_k - \eta_k \nabla f(\vec{x}_k). \tag{2.2}$$

The two descent steps are equivalent when $\eta_k = \frac{1}{\alpha_k}$.

3 Duality (Fall 2022 Final)

Consider a convex function

$$f(\vec{x}) = \frac{1}{2}(x_1 + 1)^2 + x_2^2, \qquad \forall \vec{x} = (x_1, x_2) \in \mathbb{R}^2. \tag{3.1}$$

Suppose that we wish to minimize $f(\vec{x})$ subject to the linear constraint $x_1 = 0$.

1. Find the primal optimum p^*.
2. Find the dual function $g(\lambda)$.
3. Find the dual optimum d^*. Conclude from the relation of p^* and d^* whether strong duality holds or not.

Solution:

1. The optimum occurs at $x_1^* = 0$ and $x_2^* = 0$. We have $p^* = \frac{1}{2}$.
2. The dual function takes the form:

$$g(\nu) = \inf_{\vec{x} \in \mathbb{R}^2} \left\{ \frac{1}{2}(x_1 + 1)^2 + x_2^2 + \nu x_1 \right\}. \tag{3.2}$$

   Since the above is a minimization problem and the objective function is convex in $\vec{x}$, we can find the minimizer $\vec{x}^*(\nu)$ for each fixed $\nu \in \mathbb{R}$ by setting the gradient in $\vec{x}$ of the function $\frac{1}{2}(x_1 + 1)^2 + x_2^2 + \nu x_1$ to 0.
   Setting the gradient w.r.t. x_1 equal to 0, we get $x_1^*(\nu) = -\nu - 1$.
   Setting the gradient w.r.t. x_2 equal to 0, we get $x_2^*(\nu) = 0$.
   Plugging $x_1^*(\nu)$ and $x_2^*(\nu)$ back into the function $\frac{1}{2}(x_1 + 1)^2 + x_2^2 + \nu x_1$, we obtain the desired dual function

$$g(\nu) = -\frac{1}{2}\nu^2 - \nu. \tag{3.3}$$

3. To find d^*, we need to solve the dual problem

$$d^* = \sup_{\nu \in \mathbb{R}} g(\nu) \tag{3.4}$$

$$= \sup_{\nu \in \mathbb{R}} -\frac{1}{2}\nu^2 - \nu. \tag{3.5}$$

   Since the above is a maximization problem and the objective function is concave in ν, we can find ν^* by setting the gradient in ν of $-\frac{1}{2}\nu^2 - \nu$ to 0.
   This gives $\nu^* = -1$. So we know, the dual optimum $d^* = g(\nu^*) = \frac{1}{2}$. Since $p^* = d^* = \frac{1}{2}$, strong duality holds.

4 Support Vector Machines (Fall 2020 Final)

Recall that the maximum margin support vector machine problem is to find $\vec{w}$, b that solve the following problem:

$$\begin{aligned} \min_{\vec{w},b} \quad & \frac{1}{2}\|\vec{w}\|_2^2 \\ \text{subject to} \quad & y_i(\vec{w}^\top \vec{x}_i + b) \geq 1, \quad i = 1, \ldots, n. \end{aligned}$$

Here, the data points $(\vec{x}_i, y_i)$ with $\vec{x}_i \in \mathbb{R}^k$, $y_i \in \{+1, -1\}$ for $i = 1, \ldots, n$ are given. Throughout the problem, assume that $(\vec{w}^*, b^*)$ is an optimal primal pair and λ_i^* for $i = 1, \ldots, n$ are optimal dual variables, respectively. Answer true or false for the following questions, with justification:

1. Assume that $(\vec{w}^*, b^*)$ is an optimal primal pair and λ_i^* for $i = 1, \ldots, n$ are optimal dual variables, respectively. Suppose you are told that $\lambda_i^* > 0$ for some $i \in \{1, \ldots, n\}$. Then, $y_i(\vec{w}^{*\top} \vec{x}_i + b^*) = 1$.
   Solution: True. This follows from complementary slackness, which tells us that $(y_i(\vec{w}^{*\top} \vec{x}_i + b^*) - 1)\lambda_i^* = 0$ for all $i = 1, \ldots, n$.

2. The optimal $\vec{w}^*$ is always dependent on the location of every one of the data points, i.e. it is always the case that changing the location of any one of the data points will change the optimal $\vec{w}^*$.
   Solution: False. There are examples where there is a data point $(\vec{x}_i, y_i)$ such that $y_i(\vec{w}^{*\top} \vec{x}_i + b^*) > 1$. In this case, for all sufficiently small changes of the location $\vec{x}_i$ of the data point $(\vec{x}_i, y_i)$ to, say, $\tilde{\vec{x}}_i$, leaving $\tilde{y}_i = y_i$, we will still have $\tilde{y}_i(\vec{w}^{*\top} \tilde{\vec{x}}_i + b^*) \geq 1$, and so the KKT conditions for the maximum margin support vector machine problem will continue to be solved by the same choice of λ_i^* and $(\vec{w}^*, b^*)$ even when the data point $(\vec{x}_i, y_i)$ is replaced by $(\tilde{\vec{x}}_i, \tilde{y}_i)$. Thus, in such cases the optimal $\vec{w}^*$ does not depend on the location of the data point $(\vec{x}_i, y_i)$.

3. Assume that k is very large compared to n. Furthermore, assume that you have a black box that can easily compute the inner product between feature vectors, i.e. that computing $\vec{x}^\top \tilde{\vec{x}}$ for two feature vectors $\vec{x}$ and $\tilde{\vec{x}}$ incurs a very small cost even though k is large. In order to classify a new feature vector $\vec{x}^\dagger \in \mathbb{R}^k$, it is more efficient to directly solve the primal problem and obtain $\vec{w}^*, b^*$ than to solve the dual problem.
   Solution: False. Notice that if we solve the dual problem we will get a dual optimal λ_i^* for $i = 1, \ldots, n$, of which there are very few compared to k. This means that when we classify $\vec{x}^\dagger$, instead of computing $\vec{w}^\top \vec{x}^\dagger + b$, which may require cost require a cost that is determined by k, we can compute $\sum_{i=1}^n y_i \lambda_i^* \vec{x}_i^\top \vec{x}^\dagger + b$, which, since computing $\vec{x}_i^\top \vec{x}^\dagger$ incurs only a fixed cost, will result in an overall cost that is determined by n.

4. If the training data is not linearly separable, then strong duality does not hold.
   Solution: False. While it is true that we cannot find a separating hyperplane, it is not true that strong duality does not hold. In fact, the primal optimal value will be ∞ and the dual optimal value will be ∞. The primal optimal value in this cased will be ∞ because the primal problem is not feasible. The dual optimal value will be ∞ because, if the data is not linearly separable, we

can find a convex combination of the $\vec{x}_i$ that are classified as 1 which equals a convex combination of the $\vec{x}_i$ which are classified as -1. This means we can find nonnegative λ_i, $i = 1, \ldots, n$, not all zero, such that we have both $\sum_{i=1}^{n} \lambda_i y_i = 0$ and $\sum_{i=1}^{n} \lambda_i y_i \vec{x}_i = 0$. Now we can scale the λ_i's arbitrarily large to push the objective of the dual to ∞.

5 Reformulating Convex Optimization Problems (Spring 2020, Spring 2023 Final)

1. Reformulate the following problem as an SOCP:

$$\min_{\vec{x}} \max_{i=1,2,\ldots,m} \|A\vec{x} - B\vec{y}_i\|_2.$$

Solution: Introducing a slack variable $t \in \mathbb{R}$, we can rewrite the problem as

$$\begin{aligned} \min_{\vec{x},t} \quad & t \\ \text{s.t.} \quad & \|A\vec{x} - B\vec{y}_i\|_2 \leq t, \quad i = 1, 2, \ldots, m. \end{aligned}$$

This is an SOCP in standard form.

2. Reformulate the following problem as a QP:

$$\begin{aligned} \min_{\vec{x} \in \mathbb{R}^2} \quad & \vec{x}^\top A\vec{x} \\ \text{s.t.} \quad & \vec{c}^\top \vec{x} \geq 1 \end{aligned}$$

where $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ and $\vec{c} \in \mathbb{R}^2$.

Note: A is not a positive semidefinite symmetric matrix.

Solution: We can equivalently write the objective function as $\vec{x}^\top A\vec{x} = \vec{x}^\top B\vec{x}$ for

$$B = \frac{1}{2}(A + A^\top) = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}.$$

Because the eigenvalues of B are $\{3/2, 1/2\}$, B is PD. Therefore the following problem, which is equivalent to the given problem, is a QP:

$$\begin{aligned} \min_{\vec{x} \in \mathbb{R}^2} \quad & \vec{x}^\top B\vec{x} \\ \text{s.t.} \quad & \vec{c}^\top \vec{x} \geq 1. \end{aligned}$$

3. Reformulate the following problem as a QP:

$$\begin{aligned} \min_{\vec{x} \in \mathbb{R}^d} \quad & \frac{1}{2}\|\vec{x}\|^2 \\ \text{s.t.} \quad & \|A\vec{x} - \vec{y}\|_\infty \leq \epsilon, \end{aligned}$$

where $A \in \mathbb{R}^{n \times d}$, $\vec{y} \in \mathbb{R}^n$, and $\epsilon > 0$.

Solution: This problem can be written as the QP:

$$\begin{aligned} \min_{\vec{x} \in \mathbb{R}^d} \quad & \frac{1}{2}\|\vec{x}\|^2 \\ \text{s.t.} \quad & A\vec{x} - \vec{y} \leq \epsilon\vec{1}, \\ & A\vec{x} - \vec{y} \geq -\epsilon\vec{1}, \end{aligned}$$

where the vector inequalities are meant to hold coordinatewise.

4. Reformulate the following problem as a QP:

$$\min_{\vec{x}\in\mathbb{R}^d} \quad \left\{\frac{1}{2}\|\vec{x}\|^2 + \lambda\sum_{i=1}^{n} \max\{0, |\vec{a}_i^\top \vec{x} - y_i| - \epsilon\}\right\},$$

where $A \in \mathbb{R}^{n\times d}$, $\vec{y} \in \mathbb{R}^n$, $\epsilon > 0$, and $\lambda > 0$.

Hint: Introduce a new variable $\vec{z}$.

Solution:

Note that

$$\max\{0, |\vec{a}_i^\top \vec{x} - y_i| - \epsilon\} = \max\{0, \vec{a}_i^\top \vec{x} - y_i - \epsilon, -\vec{a}_i^\top \vec{x} + y_i - \epsilon\}.$$

Therefore, the optimization problem can be equivalently written as the QP:

$$\begin{aligned} \min_{\vec{x}\in\mathbb{R}^d} \quad & \frac{1}{2}\|\vec{x}\|^2 + \lambda \vec{z}^\top \vec{1} \\ \text{s.t.} \quad & \vec{z} \geq \vec{0}, \\ & \vec{z} \geq A\vec{x} - \vec{y} - \epsilon\vec{1}, \\ & \vec{z} \geq -A\vec{x} + \vec{y} - \epsilon\vec{1}, \end{aligned}$$

where the vector inequalities are meant to hold coordinatewise.

6 Properties of a Linear Program

Consider the following linear program

$$\begin{aligned} \min_{\vec{x}\in\mathbb{R}^2} \quad & c_1x_1 + c_2x_2 \\ \text{s.t.} \quad & x_1 + x_2 \leq 1, \\ & x_1 - x_2 \leq 1, \\ & x_2 - x_1 \leq 1. \end{aligned}$$

(a) Draw the constraint set of this optimization problem.

(b) Draw the level sets of the function $c_1x_1 + c_2x_2 = \{1, 0, -1\}$ corresponding to the following values of c_1 and c_2:

- $c_1 = 1, c_2 = 1$;
- $c_1 = -1, c_2 = 1$;
- $c_1 = 0, c_2 = -1$.

(c) Suppose $c_1 = 1, c_2 = 1$. Does the optimal solution exist? If so, is it unique?

(d) Suppose $c_1 = -1, c_2 = 1$. Does the optimal solution exist? If so, is it unique?

(e) Suppose $c_1 = 0, c_2 = -1$. Does the optimal solution exist? If so, is it unique?

Solution:

(a) The feasible region is shown in Figure 1.

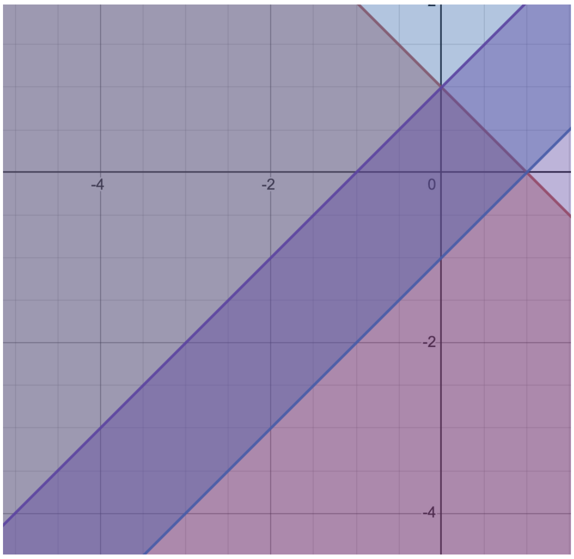

Figure 1: The feasible region is the darkest region in the figure above

(b) (1) The level sets of the function $x_1 + x_2 \in \{-1, 0, 1\}$ is shown in Figure 2.

(2) The level sets of the function $x_1 - x_2 \in \{-1, 0, 1\}$ is shown in Figure 3.

(3) The level sets of the function $-x_2 \in \{-1, 0, 1\}$ is shown in Figure 4.

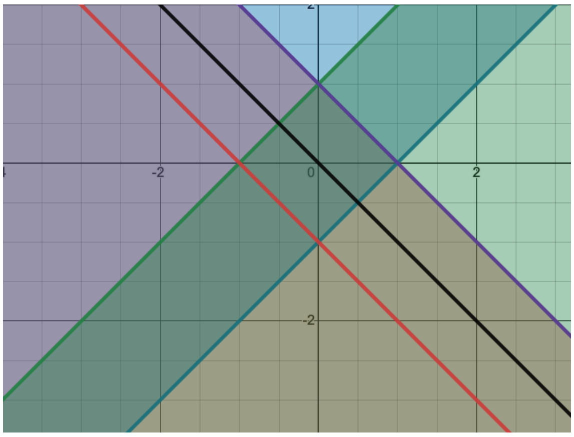

Figure 2: The level sets of the function $x_1 + x_2 \in \{-1, 0, 1\}$. The purple line denotes the set $x_1 + x_2 = 1$. The black line denotes the set $x_1 + x_2 = 0$. The red line denotes the set $x_1 + x_2 = -1$.

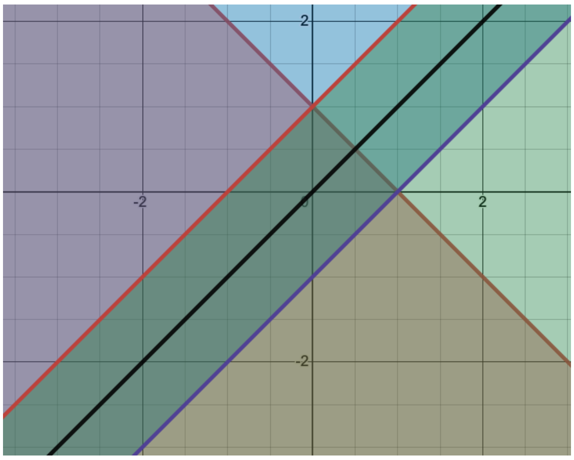

Figure 3: The level sets of the function $x_1 - x_2 \in \{-1, 0, 1\}$. The purple line denotes the set $x_1 - x_2 = 1$. The black line denotes the set $x_1 - x_2 = 0$. The red line denotes the set $x_1 - x_2 = -1$.

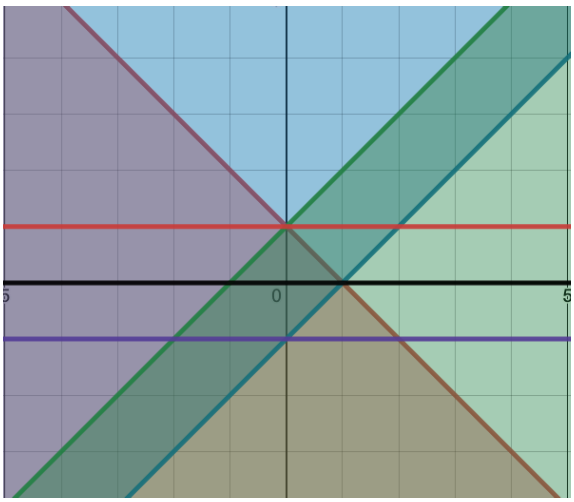

Figure 4: The level sets of the function $-x_2 \in \{-1, 0, 1\}$. The purple line denotes the set $-x_2 = 1$. The black line denotes the set $x_2 = 0$. The red line denotes the set $x_2 = 1$.

(c) If $c_1 = c_2 = 1$ then the optimal solution does not exist. This is because for any $\epsilon > 0$, $(-\epsilon, -\epsilon)$ is a feasible solution. The objective value at this point is -2ϵ. One can keep on increasing ϵ to get lower values of the objective.

(d) If $c_1 = -c_2 = -1$ then an optimal solution exists. From the constraints we see that $-1 \leq c_1x_1 + c_2x_2 \leq 1$. For any $\epsilon > 0$, the solution $(1 - \epsilon, -\epsilon)$ is an optimal solution where the objective takes on value equal to the lower bound of -1. Thus, there are infinitely many optimal solutions to this problem.

(e) If $c_1 = 0, c_2 = -1$ then an optimal solution exists. The constraints $x_1 + x_2 \leq 1$ and $x_1 - x_2 \geq -1$ guarantee that the objective is greater than -1. The only feasible solution that achieves the objective value of -1 is $(0, 1)$. This can be easily observed from Figure 4.

7 KKT Conditions

Consider the problem:

$$\begin{aligned} \min_{x,y\in\mathbb{R}} \quad & 2x + y \\ \text{s.t.} \quad & x^2 + y^2 \leq 4, \\ & x \geq 0, \\ & y \geq \frac{x}{2} - 1. \end{aligned}$$

1. Is the above problem a convex optimization problem?
2. Write the Lagrangian $L(x, y, \lambda_1, \lambda_2, \lambda_3)$ associated to this problem.
3. Write the KKT conditions for this problem.
4. Does strong duality hold for this problem?

Solution:

1. Yes. The objective function is linear in $(x, y) \in \mathbb{R}^2$, and the constraint can be written as $f_1(x, y) \leq 0, f_2(x, y) \leq 0, f_3(x, y) \leq 0$, where $f_1, f_2, f_3 : \mathbb{R}^2 \to \mathbb{R}$ are convex functions defined by:

$$\begin{aligned} f_1(x, y) &:= x^2 + y^2 - 4, \\ f_2(x, y) &:= -x, \\ f_3(x, y) &:= \frac{x}{2} - y - 1. \end{aligned}$$

2. The Lagrangian $L : \mathbb{R}^5 \to \mathbb{R}$ for this problem is given by:

$$L(x, y, \lambda_1, \lambda_2, \lambda_3) = 2x + y + \lambda_1(x^2 + y^2 - 4) + \lambda_2(-x) + \lambda_3 \left(\frac{1}{2}x - y - 1 \right).$$

for each $x, y, \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$.

3. The KKT conditions for this optimization problem are:
   - (**Primal feasibility**) $x^2 + y^2 \leq 4, x \geq 0, y \geq \frac{x}{2} - 1$.
   - (**Dual feasibility**) $\lambda_1, \lambda_2, \lambda_3 \geq 0$.
   - (**Complementary Slackness**) $\lambda_1(x^2 + y^2 - 4) = 0$, $\lambda_2(-x) = 0$, $\lambda_3\big(\frac{x}{2} - y - 1\big) = 0$.
   - (**Stationarity**)

$$\vec{0} = \nabla_{(x,y)} L(x, y, \lambda_1, \lambda_2, \lambda_3) = \begin{bmatrix} 2 + 2\lambda_1 x - \lambda_2 + \frac{1}{2}\lambda_3 \\ 1 + 2\lambda_1 y - \lambda_3 \end{bmatrix}.$$

4. Yes. There exists a strictly feasible point given by $(x, y) = (1, 0)$, so Slater's condition implies that strong duality holds.