Norwegian University of Science and Technology
Department of Engineering Cybernetics

Exercise 1
Solution

Problem 1 (25 %)

Optimization problem:

$$\min x_1 + 2x_2 \qquad \text{s.t.} \qquad 2 - x_1^2 - x_2^2 \geq 0, \quad x_2 \geq 0 \tag{1}$$

a) The optimal point is $x^* = (-\sqrt{2}, 0)^\top$, this can be found by inspection.

b) The Lagrangean for the problem is

$$\mathcal{L}(x, \lambda) = x_1 + 2x_2 - \lambda_1 c_1(x) - \lambda_2 c_2(x) \tag{2}$$

with

$$c_1(x) = 2 - x_1^2 - x_2^2 \tag{3a}$$
$$c_2(x) = x_2 \tag{3b}$$
$$\mathcal{I} = \{1, 2\} \tag{3c}$$

Hence,

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} 1 + 2\lambda_1^* x_1^* \\ 2 + 2\lambda_1^* x_2^* - \lambda_2 \end{bmatrix} = 0 \Rightarrow \lambda^* = \begin{bmatrix} \frac{1}{2\sqrt{2}} \\ 2 \end{bmatrix} \tag{4}$$

and all KKT conditions are satisfied.

c) The gradients of the active constraints and the objective function at the solution are illustrated in Figure 1.

d) This problem has two constraints, both of which are inequality constraints. The KKT conditions tells that for a point to be optimal, each of the multipliers corresponding to the constraints have to be positive; if we had a negative multiplier for an inequality constraint the KKT conditions would not be satisfied and our point could not be optimal. For this problem, the first KKT condition can be written

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \lambda_1^* \nabla c_1(x^*) - \lambda_2^* \nabla c_2(x^*) = 0 \tag{5}$$

which we can rewrite as

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*) + \lambda_2^* \nabla c_2(x^*) \tag{6}$$

This equation can be interpreted to mean that the vector $\nabla f(x^*)$ can be expressed as a linear combination of the two vectors $\nabla c_1(x^*)$ and $\nabla c_2(x^*)$ (a weighted sum

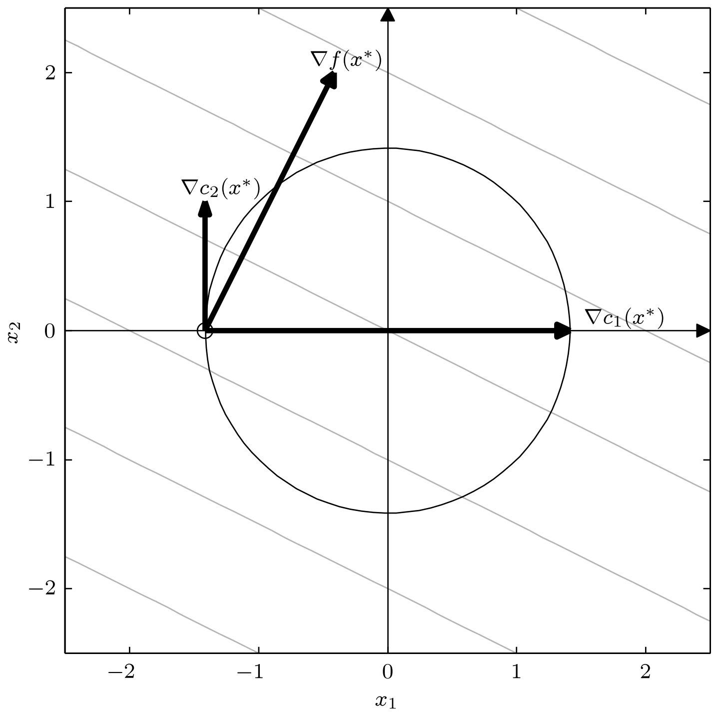

Figure 1: Gradients at the optimal point in Problem 1.

of the two vectors, where λ_1^* and λ_2^* are the weights). Looking at Figure 1, we see that this is possible if and only if λ_1^* and λ_2^* are both positive. Just based on that illustration alone, we can make a good guess of what the multiplier values must be. If we first look at the vector $\nabla c_1(x^*)$, we see that is it "too long", meaning it goes further to the right (the x_1 direction) than $\nabla f(x^*)$ does. This means that we need to make it shorter if this vector's length in the x_1 direction is to match $\nabla f(x^*)$'s length in the x_1 direction. How much shorter must it be? Judging from the figure it looks like 1/3 of the length would be about right. This means that we guess $\lambda_1^* = 1/3$.

Similarly, we see that the vector $\nabla c_2(x^*)$ is shorter in the x_2 direction than $\nabla f(x^*)$ is. This means we have to make it longer if $\nabla c_2(x^*)$ is to match the length of $\nabla f(x^*)$ in the x_2 direction. Again, judging from the figure, it looks like making $\nabla c_2(x^*)$ twice as long would make it reach as far in the x_2 direction as $\nabla f(x^*)$ does. This means we guess that $\lambda_2^* = 2$. With respect to the first KKT condition, we can now say that we believe

$$\nabla f(x^*) \approx \frac{1}{3}\nabla c_1(x^*) + 2\nabla c_2(x^*) \tag{7}$$

If this does not make sense, try to sketch a small illustration and convince yourself that the above equation describes $\nabla f(x^*)$ as a linear combination of the two vectors $\nabla c_1(x^*)$ and $\nabla c_2(x^*)$ with our guesses of the multipliers λ_1^* and λ_2^*.

Once you are convinced that $\lambda_1^* \approx 1/3$ and $\lambda_2^* \approx 2$, you can also see that it would be impossible to express $\nabla f(x^*)$ as a linear combination of $\nabla c_1(x^*)$ and $\nabla c_2(x^*)$ if one or both of the multipliers were negative — a negative multiplier would flip the corresponding gradient around at it would point in the "opposite" direction of

$\nabla f(x^*)$. Hence, we can see from drawing all the gradients at the optimal point that multipliers *corresponding to inequality constraints* have to be positive if the first KKT condition is to be satisfied.

Note that we were able to guess fairly accurate values of the multipliers just by looking at Figure 1; compare 1/3 and 2 to the values we found in part b.

We can think of this in another way: the constraint normals $\nabla c_1(x)$ and $\nabla c_2(x)$ point toward the inside of the feasible region ($\nabla c_1(x)$ points in the direction of the center of the circle, and $\nabla c_2(x)$ points in the direction of the top half plane. These directions can be seen as the directions which we are allowed to move in. Since this is a minimization problem, and $\nabla f(x)$ always points in the direction where the objective function f increases the most, we need all inequality-constraint gradients to point in directions that are not "opposite" of $\nabla f(x)$. If an inequality-constraint gradient pointed away from $\nabla f(x)$ by more than 90°, it means there is a feasible direction we can move in where f decreases. If this is the case, our point cannot be optimal! A way of saying that all constraint gradients must have an angle with $\nabla f(x)$ that is less than 90° is

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* \nabla c_i(x^*) = 0 \quad \text{with all } \lambda_i^*, i \in \mathcal{I} \tag{8}$$

which is nothing more than the first and fourth KKT conditions.

Both of these ways of thinking about gradients and multipliers are useful when we later learn about sensitivity analysis, which amounts to analyzing how "important" the different constraints are if we want to improve the solution by changing a constraint. Using sensitivity analysis, we can say a positive number λ_i means any small change of x results in an increase in the objective function. The bigger the value of the multiplier, the more improvement in $f(x^*)$ we can expect.

e) This is a convex problem, since the objective function is convex (all linear functions are convex) and the feasible set is convex. For a problem with only inequality constraints, the feasible set is convex if all inequalities are concave functions. Here, both $c_1(x)$ (a paraboloid with a unique maximizer) and $c_2(x)$ (a linear function) are concave functions. (Note that a linear function is both concave and convex).

Problem 2 (30 %)

Optimization problem:

$$\min 2x_1 + x_2 \qquad \text{s.t.} \qquad x_1^2 + x_2^2 - 2 = 0 \tag{9}$$

a) The extreme points are

$$x = \left(-2\sqrt{\frac{2}{5}}, -\sqrt{\frac{2}{5}}\right)^\top \tag{10a}$$

and

$$x = \left(2\sqrt{\frac{2}{5}}, \sqrt{\frac{2}{5}}\right)^\top \tag{10b}$$

(Use, for instance, $\nabla f = \lambda_1 \nabla c_1$ and $x_1^2 = 2 - x_2^2$ to find these.)

b) The Lagrangian for the problem is

$$\mathcal{L}(x, \lambda) = 2x_1 + x_2 - \lambda_1 c_1(x) \tag{11}$$

with

$$c_1(x) = x_1^2 + x_2^2 - 2,\ \mathcal{E} = \{1\} \tag{12}$$

Hence,

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} 2 - 2\lambda_1^* x_1^* \\ 1 - 2\lambda_1^* x_2^* \end{bmatrix} = 0 \Rightarrow \lambda^* = \pm\frac{\sqrt{10}}{4} \tag{13}$$

c) The gradients of the active constraint and the objective function at the optimal point are illustrated in Figure 2.

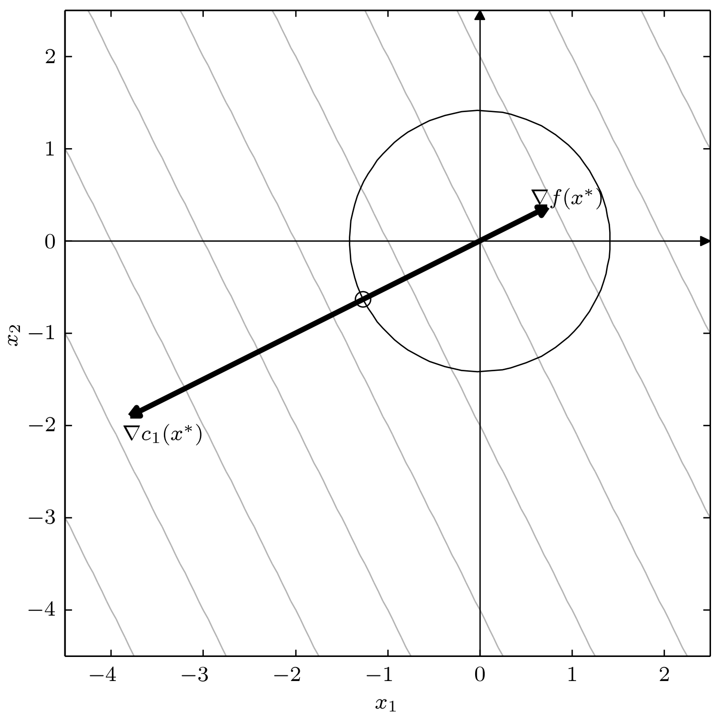

Figure 2: Gradients at the optimal point in Problem 2.

d) The values of the multiplier above mean that the KKT conditions are satisfied at both extreme points.

e) We have that

$$\nabla^2_{xx}\mathcal{L}(x^*,\lambda^*) = \begin{bmatrix} -2\lambda^* & 0 \\ 0 & -2\lambda^* \end{bmatrix} > 0 \tag{14}$$

holds for $\lambda^* = -\frac{\sqrt{10}}{4}$. Then,

$$w^\top \nabla^2_{xx}\mathcal{L}(x^*,\lambda^*)w > 0, \qquad \text{for all } w \in \mathcal{C}(x^*,\lambda^*),\ w \neq 0 \tag{15}$$

which means that both the necessary and sufficient second-order conditions in Chapter 12.5 hold (Theorems 12.5 and 12.6, respectively).

f) The problem is nonconvex due to the nonlinear equality constraint.

Problem 3 (20 %)

Optimization problem:

$$\min_{x\in\mathbb{R}^2} f(x) = -2x_1 + x_2 \qquad \text{s.t.} \qquad \begin{cases} c_1(x) = (1-x_1)^3 - x_2 \geq 0 \\ c_2(x) = x_2 + 0.25x_1^2 - 1 \geq 0 \end{cases} \tag{16}$$

a) The set of active constraint gradients at the solution,

$$\nabla c_1(x^*) = \begin{bmatrix} -3 \\ -1 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{17}$$

is linearly independent. Hence, LICQ hold.

b) The Lagrangian for the problem is

$$\mathcal{L}(x,\lambda) = -2x_1 + x_2 - \lambda_1 c_1(x) - \lambda_2 c_2(x) \tag{18}$$

with

$$c_1(x) = (1-x_1)^3 - x_2 \tag{19a}$$
$$c_2(x) = x_2 + 0.25x_1^2 - 1 \tag{19b}$$
$$\mathcal{I} = \{1,2\} \tag{19c}$$

Hence,

$$\nabla_x\mathcal{L}(x^*,\lambda^*) = \begin{bmatrix} -2 + 3\lambda_1^*(1-x_1^*)^2 - 0.5\lambda_2^* x_1^* \\ 1 + \lambda_1 - \lambda_2 \end{bmatrix} = 0 \Rightarrow \lambda^* = \begin{bmatrix} 2/3 \\ 5/3 \end{bmatrix} \tag{20}$$

and all KKT conditions are satisfied.

d) The Hessian of the Lagrangian at the solution,

$$\nabla^2_{xx}\mathcal{L}(x^*,\lambda^*) = \begin{bmatrix} -6\lambda_1^*(1-x_1^*) - 0.5\lambda_2^* & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -29/6 & 0 \\ 0 & 0 \end{bmatrix} \tag{21}$$

is negative semidefinite. However, the critical cone $\mathcal{C}(x^*,\lambda^*)$ contains only the vector $w^\top = [0,\ 0]$. Hence, the second-order necessary conditions are satisfied, whereas the second-order sufficient conditions are not.

Problem 4 (25 %)

Finding the maximizer for $f(x) = x_1 x_2$ is equivalent to finding the minimizer for the function $\bar{f} = -x_1 x_2$. We therefore state the optimization problem as

$$\min -x_1 x_2 \qquad \text{s.t.} \qquad 1 - x_1^2 - x_2^2 \geq 0 \tag{22}$$

Note that the objective function represents a saddle, and it is therefore clear that the minimizer(s) exist(s) on the boundary of the unit disk. The Lagrangian is given by

$$\mathcal{L}(x, \lambda) = -x_1 x_2 - \lambda_1 (1 - x_1^2 - x_2^2) \tag{23}$$

From the KKT conditions,

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} -x_2^* + 2\lambda_1^* x_1^* \\ -x_1^* + 2\lambda_1^* x_2^* \end{bmatrix} = 0 \Rightarrow \lambda_1^* = \pm\frac{1}{2} \tag{24}$$

Since λ_1^* has to be nonnegative, we have

$$x_1^* = x_2^* = \pm\frac{1}{\sqrt{2}} \tag{25}$$

Hence, $f(x) = x_1 x_2$ has two maximizers, $x^* = \left[-\frac{1}{\sqrt{2}},\ -\frac{1}{\sqrt{2}}\right]$ and $x^* = \left[\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}\right]$, both with $\lambda_1^* = +\frac{1}{2}$. The gradients at the optimal point $x^* = \left[-\frac{1}{\sqrt{2}},\ -\frac{1}{\sqrt{2}}\right]$ are illustrated in Figure 3, while the gradients at the optimal point $x^* = \left[\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}\right]$ are illustrated in Figure 4.

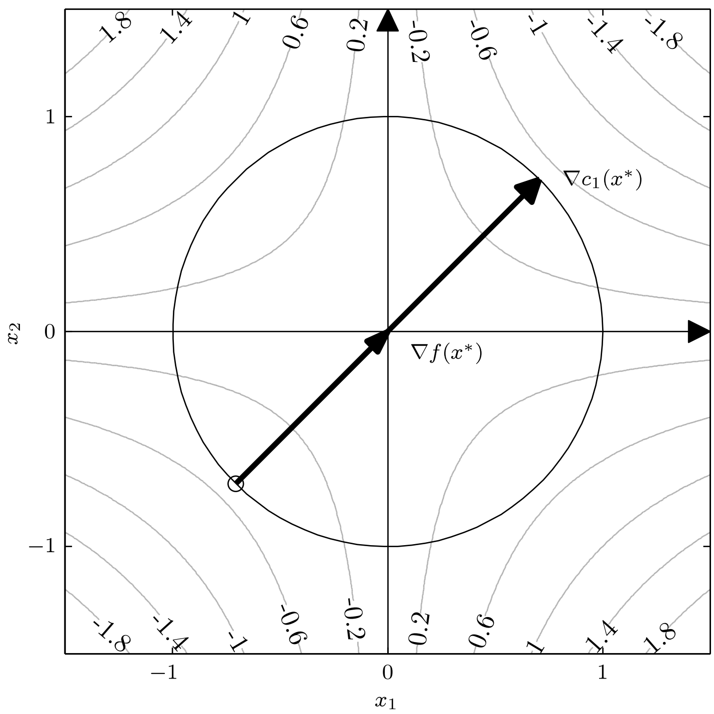

Figure 3: Gradients at the solution $x^* = \left[-\frac{1}{\sqrt{2}},\ -\frac{1}{\sqrt{2}}\right]$ of Problem 4.

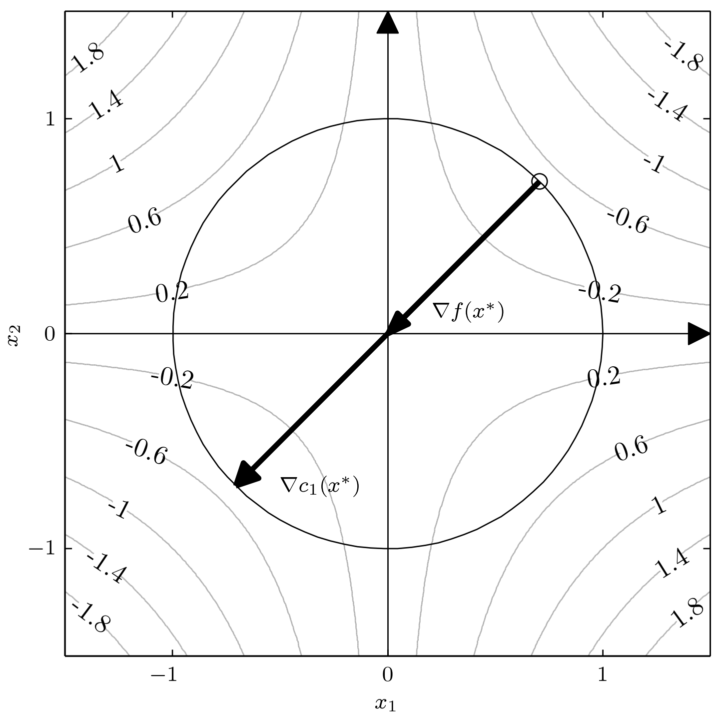

Figure 4: Gradients at the solution $x^* = \left[\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}\right]$ of Problem 4.