MS-C2105 - Introduction to Optimization
Lecture 10

Fabricio Oliveira (with modifications by Harri Hakula)

Systems Analysis Laboratory
Department of Mathematics and Systems Analysis

Aalto University
School of Science

March 22, 2022

Outline of this lecture

Optimality conditions for constrained problems

Karush-Kuhn-Tucker (KKT) conditions

Reading: Taha: Chapter 20; Winston: Chapter 11

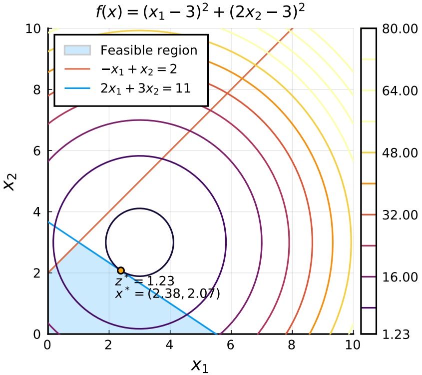
Optimality for constrained problems

In the presence of constraints, first-order conditions for unconstrained problems might never be achieved.

For example, consider the problem

$$\text{min. } \{f(x) : g(x) \leq 0\} \, .$$

- Notice that $\nabla f(x) = 0$ does not belong to the feasible region.
- In this case, the optimal on the frontier, but is not a vertex.



Optimality for constrained problems

To consider a more general setting, we rely on an alternative framework for stating optimality conditions.

- The key underlying concept is to represent constraint violations by means of penalties in the objective function.
- Coordinates feasibility and optimality simultaneously.
- Lagrangian duality provides the theoretical support for this approach.

Consider an equality constrained problem of the form:

$$\begin{aligned} \text{min. } & z = f(x) \\ \text{s.t.: } & h_i(x) = 0, \ i = 1, \dots, l. \end{aligned}$$

with $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$, all differentiable.

Optimality for constrained problems

We associate with each constraint a (Lagrangian) multiplier $\mu \in \mathbb{R}^l$, and define the Lagrangian function

$$L(x, \mu) = f(x) + \sum_{i=1}^{l} \mu_i h_i(x)$$

and then proceed to optimise the Lagrangian function $L(x, \mu)$.

- Notice that the problem becomes "unconstrained".
- $h_i(x)$ is a measure of infeasibility.

First-order (unconstrained) optimality conditions require that:

$$\frac{\partial L(x, \mu)}{\partial x} = 0 \Rightarrow \nabla f(x) + \sum_{i=1}^{l} \mu_i \nabla h_i(x) = 0$$

$$\frac{\partial L(x, \mu)}{\partial \mu_i} = 0 \Rightarrow h_i(x) = 0, i = 1, \dots, l.$$

Optimality for constrained problems

Theorem 1 (Necessary condition - equality const. problems)

Let P be min. $\{f(x) : h(x) = 0\}$ with differentiable $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$. If $\overline{x}$ is optimal for P, then $(\overline{x}, \overline{\mu})$ satisfies

$$\frac{\partial L(x,\mu)}{\partial x} = 0 \Rightarrow \nabla f(x) + \sum_{i=1}^{l} \mu_i \nabla h_i(x) = 0 \tag{1}$$

$$\frac{\partial L(x,\mu)}{\partial \mu} = 0 \Rightarrow h(x) = 0. \tag{2}$$

Proof.

Take any feasible point x^0. Since (1) and (2) are optimality conditions for $L(x, \mu)$, for any μ^0 we have

$$L(\overline{x}, \overline{\mu}) \leq L(x^0, \mu^0)$$
$$f(\overline{x}) + \overline{\mu}^\top h(\overline{x}) \leq f(x^0) + \mu^{0\top} h(x^0)$$
$$f(\overline{x}) \leq f(x^0). \quad \square$$

Optimality for constrained problems

Remark: these are necessary conditions for local optimality.

For these to become sufficient conditions for global optimality, we need stronger assumptions on f and g.

Theorem 2

Consider the problem P : min. $\{f(x) : h(x) = 0\}$ with $f : \mathbb{R}^n \to \mathbb{R}$ convex and $h : \mathbb{R}^n \to \mathbb{R}^l$ affine. Then, $\overline{x}$ is optimal for P if and only if $(\overline{x}, \overline{\mu})$ satisfies

$$\frac{\partial L(x,\mu)}{\partial x} = 0 \Rightarrow \nabla f(x) + \sum_{i=1}^{l} \mu_i \nabla h_i(x) = 0$$

$$\frac{\partial L(x,\mu)}{\partial \mu} = 0 \Rightarrow h(x) = 0.$$

Optimality for constrained problems

Example:

max. $z = -2x_1^2 - x_2^2 + x_1x_2 + 8x_1 + 3x_2 : 3x_1 + x_2 = 10.$

The Lagrangian function is given by:

$L(x_1, x_2, \mu) = -2x_1^2 - x_2^2 + x_1x_2 + 8x_1 + 3x_2 + \mu(3x_1 + x_2 - 10)$

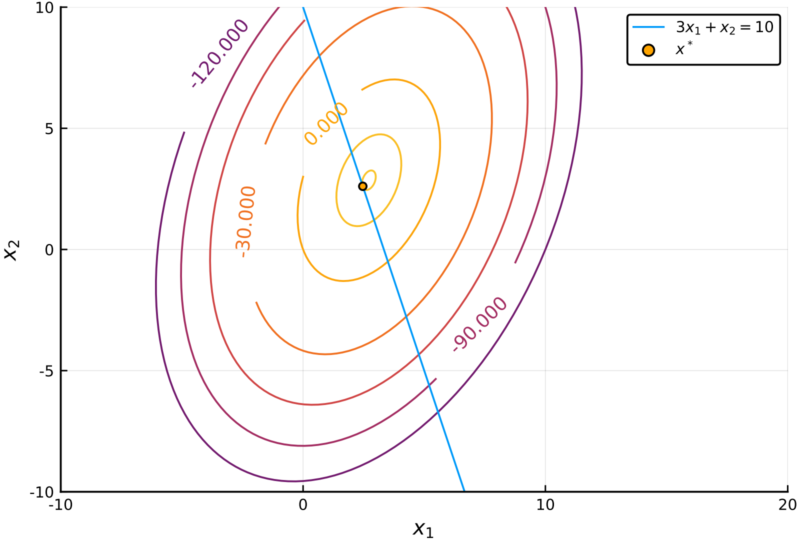
Optimality conditions are:

$$\partial \frac{L(x_1, x_2, \mu)}{\partial x_1} = -4x_1 + x_2 + 8 + 3\mu = 0$$
$$\partial \frac{L(x_1, x_2, \mu)}{\partial x_2} = -2x_2 + x_1 + 3 + \mu = 0$$
$$\partial \frac{L(x_1, x_2, \mu)}{\partial \mu} = 3x_1 + x_2 - 10 = 0$$

Solving this system, we obtain $\overline{x} = (2.46, 2.60)$ and $\overline{\mu} = -0.25$.

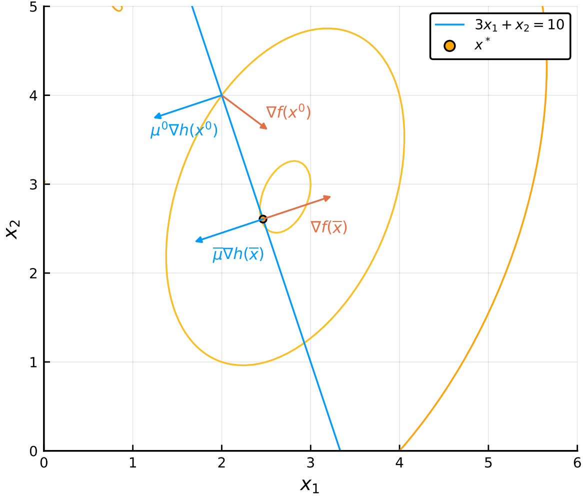
Since z is concave and constraints are affine, optimality conditions are necessary and sufficient for global optimality.

Optimality for constrained problems



Geometry of optimality conditions

The condition $\nabla f(x) = -\sum_{i=1}^{m} \mu_i \nabla h_i(x)$ can be interpret as a "force equilibrium".



- $\overline{x}$ will be optimal if $\nabla f(\overline{x})$ can be written as a linear combination of $\nabla h(\overline{x})$'s.
- If $\nabla f(x^0)$ components cannot be cancelled by $\nabla h(x^0)$'s, there is a possibility for improvement.

Optimality for constrained problems - inequalities

We now consider the most general case:

$$(P): \text{min. } z = f(x)$$
$$\text{s.t.: } g_i(x) \leq 0, \ i = 1, \dots, m$$

For now, we assume the following:

- $g_i(x)$'s satisfy regularity conditions (constraint qualification); we assume that $\nabla g_i(x)$'s are linearly independent (LICQ).

The Karush-Kuhn-Tucker conditions represent the necessary conditions for optimality in the inequality-constrained case.

- Can be derived similarly to the equality case, using the Lagrangian function
$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x).$$
- $\lambda_i \geq 0$ and $\lambda_i g_i(x) = 0$ are imposed for $i = 1, \dots, m$ since penalties are only needed for $g(x) > 0$.

The KKT conditions

Theorem 3 (Necessary condition - inequality const. problems)

Let P be min. $\{f(x) : g(x) \leq 0\}$ with differentiable $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$. If $\overline{x}$ is optimal for P, then $(\overline{x}, \overline{\lambda})$ satisfies

$$
\begin{aligned}
&\nabla f(\overline{x}) + \sum_{i=1}^{m} \overline{\lambda}_i \nabla g_i(\overline{x}) = 0 \\
&g(\overline{x}) \leq 0 \\
&\overline{\lambda}_i g_i(\overline{x}) = 0, \ i = 1, \dots, m \\
&\overline{\lambda}_i \geq 0, \ i = 1, \dots, m.
\end{aligned}
$$

Remarks:

- There is a strong connection between KKT conditions and Lagrangian duality.
- In particular, $\overline{\lambda}$ are the optimal values of the dual variables, as seen in the LP case.

The KKT conditions

Example:

min. $_x \left\{(x_1-3)^2+(x_2-3)^2 : -x_1+x_2 \leq 4;\ 2x_1+3x_2 \leq 11\right\}$

The Lagrangian function is given by: $L(x_1,x_2,\lambda_1,\lambda_2) = (x_1-3)^2+(x_2-3)^2+\lambda_1(-x_1+x_2-4)+\lambda_2(2x_1+3x_2-11)$

KKT conditions are:

$$\begin{bmatrix}2x_1-6\\2x_2-6\end{bmatrix}+\lambda_1\begin{bmatrix}-1\\1\end{bmatrix}+\lambda_2\begin{bmatrix}2\\3\end{bmatrix}=0$$

$$x_1+x_2-2 \leq 0$$

$$2x_1+3x_2-11 \leq 0$$

$$\lambda_1(x_1+x_2-2)=0$$

$$\lambda_2(2x_1+3x_2-11)=0$$

$$\lambda_1,\lambda_2 \geq 0$$

The KKT conditions

In theory, KKT conditions can be solved analytically.

For example, for two constraints, complementarity conditions $\lambda_i g_i(x) = 0, i = 1, \ldots, m$ imply that one of the following holds:

1. both $\lambda_1 = 0$ and $\lambda_2 = 0$; thus $g_1(x) < 0$ and $g_2(x) < 0$;
2. $\lambda_1 > 0$ and $\lambda_2 = 0$; thus $g_1(x) = 0$ and $g_2(x) < 0$;
3. $\lambda_1 = 0$ and $\lambda_2 > 0$; thus $g_1(x) < 0$ and $g_2(x) = 0$;
4. both $\lambda_1 > 0$ and $\lambda_2 > 0$; thus $g_1(x) = 0$ and $g_2(x) = 0$;

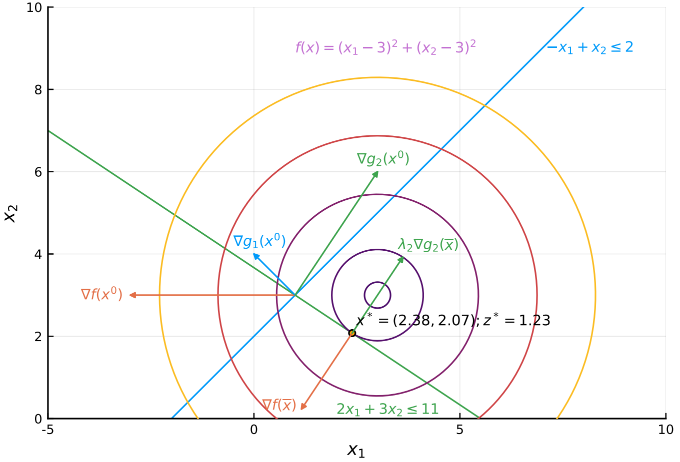
One might need to test all cases to find solutions satisfying the KKT conditions, unless sufficiency can be established.

In the previous example: $\lambda_1 = 0, \lambda_2 > 0$ leads to a (unique optimal) solution satisfying KKT conditions:

$$(\overline{x}_1, \overline{x}_2, \overline{\lambda}_1, \overline{\lambda}_2) = (2.38, 2.07, 0, 0.61)$$

Geometry of optimality conditions II

Similarly to the equality case, the "force equilibrium" also holds, but only for active constraints ($\lambda_i = 0$ for $g_i(x) < 0$).



The complete KKT conditions

For the sake of completeness, we state the KKT conditions for general problems.

Theorem 4 (KKT general conditions)

Let P be min. $\{f(x) : g(x) \leq 0, h(x) = 0\}$ with differentiable $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^l$. If $\overline{x}$ is optimal for P, then $(\overline{x}, \overline{\lambda}, \overline{\mu})$ satisfies

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} \overline{\lambda}_i \nabla g_i(\overline{x}) + \sum_{i=1}^{l} \overline{\mu}_i \nabla h_i(\overline{x}) = 0$$

$$g_i(\overline{x}) \leq 0, \ i = 1, \dots, m$$

$$h_i(\overline{x}) = 0, \ i = 1, \dots, l$$

$$\overline{\lambda}_i g_i(\overline{x}) = 0, \ i = 1, \dots, m$$

$$\overline{\lambda}_i \geq 0, \ i = 1, \dots, m.$$

Sufficiency of optimality conditions

If Slater's constraint qualification (CQ) holds, the KKT conditions become necessaru and sufficient for global optimality. Slater's CQ conditions are

1. f convex (concave for max.) function
2. g convex functions with strict interior (i.e. , exists x such that $g(x) < 0$)
3. h affine functions.

Theorem 5 (Necessary and sufficient optimality conditions)

Consider the problem P : min. $\{f(x) : g(x) \leq 0, h(x) = 0\}$ with $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ such that Slater's CQ are met. Then $\overline{x}$ is globally optimal for P if and only if $(\overline{x}, \overline{\lambda})$ satisfies the KKT conditions.

The complete KKT conditions

Example:

$$\begin{aligned} \text{min. } & z = (x_1 - 1)^2 + (x_2 - 2)^2 \\ \text{s.t.: } & -x_1 + x_2 = 1 \\ & x_1 + x_2 \leq 2 \end{aligned}$$

KKT conditions are:

$$\begin{bmatrix} 2(x_1 - 1) \\ 2(x_2 - 2) \end{bmatrix} + \mu \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-x_1 + x_2 - 1 = 0$$

$$x_1 + x_2 - 2 \leq 0$$

$$\lambda(x_1 + x_2 - 2) = 0$$

$$\lambda \geq 0$$

1. For $\lambda = 0$: $x = (1, 2)$, which violates $g(x) < 0$, and $\mu = 0$.
2. For $\lambda > 0$: $\overline{x} = (0.5, 1.5)$, $\overline{\mu} = 0$, and $\overline{\lambda} = 1$.

As Slater's CQ hold, KKT conditions are also sufficient for global optimality. Thus, $\overline{x} = (0.5, 1.5)$ is a global optimum.