Lecture 18: KKT Conditions

Zizhuo Wang

Institute of Data and Decision Analytics (iDDA)
Chinese University of Hong Kong, Shenzhen

Nov 9, 2018

Exam Summary

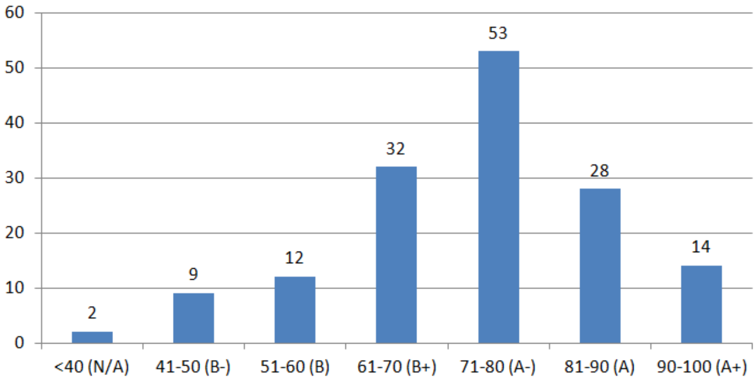
Summary

- Overall average/median: 73
- Overall standard deviation: 13

Solutions have been posted on Blackboard.

- Detailed grading criterion is in the solution. Please consult it before questioning the grading
- If you have grading questions, please come to my office hours in the next two weeks. The remaining exams can also be retrieved during my office hour.

Grade Distributions



Announcements

- Homework 6 posted, due next Wednesday (11/14)

KKT Conditions

We consider the general nonlinear optimization problem:

$$
\begin{array}{rll}
\text{minimize}_{\mathbf{x}} & f(\mathbf{x}) & \\
\text{s.t.} & g_i(\mathbf{x}) \geq 0 & i = 1, ..., m \\
& h_i(\mathbf{x}) = 0 & i = 1, ..., p \\
& \ell_i(\mathbf{x}) \leq 0 & i = 1, ..., r \\
& x_i \geq 0 & i \in M \\
& x_i \leq 0 & i \in N \\
& x_i \text{ free} & i \notin M \cup N
\end{array}
$$

We want to derive the first-order necessary condition for local minimizer — the KKT conditions.

One can use the feasible/descent directions arguments to find the KKT conditions. But it is not very convenient.

- Next we present a direct approach

Lagrangian

The first step is to construct the *Lagrangian* of this problem defined as follows:

1. We associate each constraint (not including the sign constraints) with a *Lagrangian multiplier*

$$
\begin{aligned}
g_i(\mathbf{x}) \geq 0 & \quad \cdots \quad \lambda_i \\
h_i(\mathbf{x}) = 0 & \quad \cdots \quad \nu_i \\
\ell_i(\mathbf{x}) \leq 0 & \quad \cdots \quad \eta_i
\end{aligned}
$$

2. We define the *Lagrangian* of this problem by

$$L(\mathbf{x}, \lambda, \nu, \eta) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) + \sum_{i=1}^{r} \eta_i \ell_i(\mathbf{x})$$

Lagrangian Multipliers

We require that the multipliers associated with ≥ 0 constraints (λ_i's) be non-positive, the multipliers associated with equality constraints (ν_i's) be free variables and the multipliers associated with ≤ 0 constraints (η_i's) be non-negative.

- $\lambda_i \leq 0$, ν_i free and $\eta_i \geq 0$ form one part of the KKT conditions.
- We call them dual feasibility conditions (will explain later)

What is the Meaning of the Lagrangian?

$$L(\mathbf{x}, \lambda, \nu, \eta) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) + \sum_{i=1}^{r} \eta_i \ell_i(\mathbf{x})$$

We move all the constraints to the objective and make them the reward/penalty functions. For example

1. Since we want $g_i(\mathbf{x}) \geq 0$, therefore, by setting $\lambda_i \leq 0$ we reward the objective function when the constraint holds and penalize the objective function when the constraint does not hold.
2. Similar ideas for $\ell_i(\mathbf{x})$

KKT from Lagrangian

Now we derive the KKT conditions from the Lagrangian. We first take derivative with respect to each x_i. We set the derivative to be

1. ≥ 0 if $x_i \geq 0$
2. ≤ 0 if $x_i \leq 0$
3. $= 0$ if x_i is free

This forms one part of the KKT conditions, which we call the main conditions.

KKT from Lagrangian

Finally we have a set of complementarity conditions:

- Each multiplier is complementary to the constraints it is associated to, i.e.,

$$\begin{aligned} \lambda_i \cdot g_i(\mathbf{x}) &= 0, \quad \forall i \\ \nu_i \cdot h_i(\mathbf{x}) &= 0, \quad \forall i \\ \eta_i \cdot \ell_i(\mathbf{x}) &= 0, \quad \forall i \end{aligned}$$

  The second condition is usually omitted since the feasibility of $\mathbf{x}$ already guarantees it.

- Each x_i is complementary to $\nabla_{x_i} L(\mathbf{x}, \lambda, \nu, \eta)$, i.e.,

$$x_i \cdot \nabla_{x_i} L(\mathbf{x}, \lambda, \nu, \eta) = 0$$

  This condition is usually omitted for free variables since in that case the main condition already guarantees it.

KKT Conditions

If $\mathbf{x}$ is a local minimizer, then there exists λ, ν and η such that the following conditions hold:

1. Main Condition

$$\nabla_j f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla_j g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i \nabla_j h_i(\mathbf{x}) + \sum_{i=1}^{r} \eta_i \nabla_j \ell_i(\mathbf{x}) \begin{array}{c} \geq \\ \leq \\ = \end{array} 0, \quad \begin{array}{c} \text{if } x_j \geq 0 \\ \text{if } x_j \leq 0 \\ \text{if } x_j \text{ free} \end{array}$$

2. Dual Feasibility

$$\lambda_i \leq 0, i = 1, ..., m; \quad \eta_i \geq 0, i = 1, ..., r$$

3. Complementarity

$$\lambda_i \cdot g_i(\mathbf{x}) = 0, \forall i; \quad \eta_i \cdot \ell_i(\mathbf{x}) = 0, \forall i; \quad x_i \cdot \nabla_{x_i} L(\mathbf{x}, \lambda, \nu, \eta) = 0, \forall i$$

We often add primal feasibility as part of the KKT conditions

4. Primal Feasibility

$$g_i(\mathbf{x}) \geq 0, h_i(\mathbf{x}) = 0, \ell_i(\mathbf{x}) \leq 0, \forall i$$
$$x_i \geq 0, \forall i \in M; \quad x_i \leq 0, \forall i \in N$$

Constraint Qualifications

The above result requires some constraint qualifications. It requires that:

- The matrix formed by

  $$(\nabla g_1(\mathbf{x}^*); ...; \nabla g_m(\mathbf{x}^*); \nabla h_1(\mathbf{x}^*); ...; \nabla h_p(\mathbf{x}^*); \nabla \ell_1(\mathbf{x}^*); ...; \nabla \ell_r(\mathbf{x}^*))$$

  has full rank.

In most problems, this condition is satisfied. And you don't need to worry about it in this class (you can always assume that this is true, even without being verified)

Example

Consider the problem:

$$\begin{array}{rc} \text{minimize} & 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ \text{subject to} & x_1^2 + x_2^2 \leq 5 \\ & 3x_1 + x_2 \geq 3 \end{array}$$

Find the KKT conditions.

First we associate the constraints with Lagrange multipliers μ_1 and μ_2 and construct the Lagrangian for this problem:

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\mu}) = \quad & 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ & + \mu_1(x_1^2 + x_2^2 - 5) + \mu_2(3x_1 + x_2 - 3) \end{aligned}$$

with $\mu_1 \geq 0$, $\mu_2 \leq 0$

Example Continued

The main conditions are:

$$
\begin{aligned}
4x_1 + 2x_2 - 10 + 2\mu_1 x_1 + 3\mu_2 &= 0 \\
2x_1 + 2x_2 - 10 + 2\mu_1 x_2 + \mu_2 &= 0
\end{aligned}
$$

The complementarity conditions are

$$
\begin{aligned}
\mu_1 \cdot (x_1^2 + x_2^2 - 5) &= 0 \\
\mu_2 \cdot (3x_1 + x_2 - 3) &= 0
\end{aligned}
$$

Example Continued

Therefore the KKT conditions are:

1. Main Conditions:

$$
\begin{aligned}
4x_1 + 2x_2 - 10 + 2\mu_1 x_1 + 3\mu_2 &= 0 \\
2x_1 + 2x_2 - 10 + 2\mu_1 x_2 + \mu_2 &= 0
\end{aligned}
$$

2. Primal Feasibility

$$x_1^2 + x_2^2 \le 5, \quad 3x_1 + x_2 \ge 3$$

3. Dual Feasibility

$$\mu_1 \ge 0, \quad \mu_2 \le 0$$

4. Complementarity Conditions

$$
\begin{aligned}
\mu_1 \cdot (x_1^2 + x_2^2 - 5) &= 0 \\
\mu_2 \cdot (3x_1 + x_2 - 3) &= 0
\end{aligned}
$$

Example

Find the KKT conditions for

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \mathbf{x}^T Q\mathbf{x} - \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b} \\ & C\mathbf{x} \geq \mathbf{d} \\ & \mathbf{x} \geq 0 \end{aligned}$$

We associate the linear constraints with Lagrangian multipliers λ, and the inequality constraints with multipliers η. The Lagrangian for this problem is

$$L(\mathbf{x}, \lambda, \eta) = \mathbf{x}^T Q\mathbf{x} - \mathbf{c}^T\mathbf{x} + \lambda^T(A\mathbf{x} - \mathbf{b}) + \eta^T(C\mathbf{x} - \mathbf{d})$$

And we have λ are free variables and $\eta \leq 0$

Example Continued

The main condition is (since $\mathbf{x} \geq 0$)

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \eta) = 2Q\mathbf{x} - \mathbf{c} + A^T\lambda + C^T\eta \geq 0$$

The complementarity conditions are

$$\begin{aligned} \eta_i \cdot (\mathbf{c}_i^T \mathbf{x} - d_i) &= 0, \forall i \\ x_i \cdot (2\mathbf{q}_i^T \mathbf{x} - c_i + \mathbf{a}_i^T \lambda + \mathbf{c}_i^T \eta) &= 0, \forall i \end{aligned}$$

where $\mathbf{c}_i^T$ is the ith row of C^T, $\mathbf{q}_i^T$ is the ith row of Q and $\mathbf{a}_i^T$ is the ith row of A^T.

Example Continued

Therefore the KKT conditions are:

1. Main condition:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \eta) = 2Q\mathbf{x} - \mathbf{c} + A^T \lambda + C^T \eta \geq 0$$

2. Primal feasibility

$$A\mathbf{x} = \mathbf{b}, C\mathbf{x} \geq \mathbf{d}, \mathbf{x} \geq 0$$

3. Dual feasibility

$$\eta \leq 0$$

4. Complementarity conditions

$$\eta_i \cdot (\mathbf{c}_i^T \mathbf{x} - \mathbf{d}_i) = 0, \forall i$$
$$x_i \cdot (2\mathbf{q}_i \mathbf{x} - \mathbf{c}_i + \mathbf{a}_i^T \lambda + \mathbf{c}_i^T \eta) = 0, \forall i$$

Examples of KKT Conditions

In the following, we are going to show some examples in which KKT conditions can help solve the problem or identify important structure for a problem

Example 1: Cylinder Volume

We want to build a cylinder with the maximum volume, with its surface area no larger than C.

- Decision variables: r (the radius of the base) and h (height).
- Then the optimization problem is:

$$\begin{array}{rc} \text{maximize}_{r,h} & \pi r^2 h \\ \text{subject to} & 2\pi r^2 + 2\pi rh \le C \\ & r, h \ge 0 \end{array}$$

- The optimal solution must satisfy the KKT condition. Therefore, it suffices to search among all KKT conditions to find the optimal solution.

Example 1 Continued

We construct the KKT conditions. We first convert it to a minimization problem.

$$\begin{aligned} \text{minimize}_{r,h} \quad & -\pi r^2 h \\ \text{subject to} \quad & 2\pi r^2 + 2\pi rh \leq C \\ & r, h \geq 0 \end{aligned}$$

We associate the inequality constraint with a Lagrangian multiplier λ. Then the Lagrangian is:

$$L(r, h, \lambda) = -\pi r^2 h + \lambda \cdot (2\pi r^2 + 2\pi rh - C)$$

with $\lambda \geq 0$

KKT Conditions

1. Main condition:

$$-2\pi rh + 4\pi r\lambda + 2\pi h\lambda \geq 0, \quad -\pi r^2 + 2\pi r\lambda \geq 0$$

2. Dual feasibility: $\lambda \geq 0$
3. Complementarity conditions:

$$\begin{aligned} \lambda \cdot (2\pi r^2 + 2\pi rh - C) &= 0 \\ r \cdot (-2\pi rh + 4\pi r\lambda + 2\pi h\lambda) &= 0 \\ h \cdot (-\pi r^2 + 2\pi r\lambda) &= 0 \end{aligned}$$

4. Primal feasibility:

$$2\pi r^2 + 2\pi rh \leq C, \quad r \geq 0, \quad h \geq 0$$

Example 1 Continued

We start by analyzing from the complementarity condition

$$\lambda \cdot (2\pi r^2 + 2\pi rh - C) = 0$$

Either λ or $(2\pi r^2 + 2\pi rh - C)$ has to be 0

First, if $\lambda = 0$, then due to the main condition that $-\pi r^2 + 2\pi r\lambda \geq 0$, we must have $r = 0$.

Indeed, $r = \lambda = 0$ and $h \geq 0$ satisfy all the KKT conditions, and the objective value is 0.

Example 1 Continued

If $\lambda \neq 0$, then by complementarity condition, we have

$$2\pi r^2 + 2\pi rh = C \tag{1}$$

Therefore, $r > 0$, and again by complementarity condition:

$$-2\pi rh + 4\pi r\lambda + 2\pi h\lambda = 0 \tag{2}$$

This means $h \neq 0$ (therefore $h > 0$) and by the last complementarity condition, we have

$$-\pi r^2 + 2\pi r\lambda = 0 \tag{3}$$

From (3), we have $r = 2\lambda$, then back to (2), we get $h = 4\lambda$. And back to (1), we get $\lambda = \sqrt{C/24\pi}$. And therefore the optimal r^* and h^* are:

$$r^* = 2\sqrt{C/24\pi}, \quad h^* = 4\sqrt{C/24\pi}$$

This is the optimal solution to the problem.

Example 2

Suppose we want to build a box with a given volume of at least 64 cubic inches. We want to minimize the total amount of material used.

We can formulate the optimization problem as:

$$\begin{array}{rc} \text{minimize} & 2xy + 2yz + 2xz \\ \text{s.t.} & xyz \geq 64 \end{array}$$

We let λ be the dual multiplier for this problem.

Example 2 Continued

The KKT conditions say:

$$2\begin{pmatrix} y+z \\ x+z \\ x+y \end{pmatrix} = \lambda \begin{pmatrix} yz \\ xz \\ xy \end{pmatrix}, \quad \lambda \geq 0, \quad \lambda \cdot (xyz - 64) = 0$$

To solve these conditions, one first finds that if $\lambda = 0$, then we must have $x = y = z = 0$, however, this doesn't satisfy the constraint.

Therefore $\lambda > 0$ and $xyz = 64$. By the first equality, we have

$$\lambda = 2\left(\frac{1}{x} + \frac{1}{y}\right) = 2\left(\frac{1}{y} + \frac{1}{z}\right) = 2\left(\frac{1}{x} + \frac{1}{z}\right)$$

Thus $x = y = z = 4$ is the only solution to the KKT conditions. Since this problem must have a finite optimal solution, it must be the optimal solution.

Example 3: Power Allocation

We have a collection of n communication channels and we need to decide how much power to allocate to each of them

- The capacity (communication rate) of channel i is $\log(\alpha_i + x_i)$ with a given $\alpha_i > 0$ and when x_i is allocated to it, and we have a budget constraint $\mathbf{e}^T\mathbf{x} = 1$, $\mathbf{x} \geq 0$

The optimization problem is:

$$\begin{array}{rc} \text{maximize}_{\mathbf{x}} & \sum_{i=1}^n \log(\alpha_i + x_i) \\ \text{subject to} & \sum_{i=1}^n x_i = 1 \\ & \mathbf{x} \geq 0 \end{array}$$

- Let's see what the KKT conditions tell us about the structure of this problem

Example 3 Continued

We convert it to a minimization problem, associate a Lagrangian multiplier λ to the equality constraint. The KKT conditions are (no dual feasibility constraints):

1. Main condition:
$$-\frac{1}{\alpha_i + x_i} + \lambda \geq 0, \forall i$$
2. Complementarity:
$$x_i \cdot \left(-\frac{1}{\alpha_i + x_i} + \lambda\right) = 0$$
3. Primal feasible:
$$\sum_{i=1}^{n} x_i = 1, \mathbf{x} \geq 0$$

Example 3 Continued

$$-\frac{1}{\alpha_i + x_i} + \lambda \geq 0,\ x_i \cdot \left(-\frac{1}{\alpha_i + x_i} + \lambda\right) = 0, \forall i,\ \sum_{i=1}^{n} x_i = 1, \quad \mathbf{x} \geq 0$$

- If for a certain i, $\lambda < 1/\alpha_i$, then in order for the first condition to hold, we must have $x_i > 0$. And therefore, by the second condition, we must have $x_i = 1/\lambda - \alpha_i$.
- If for a certain i, $\lambda \geq 1/\alpha_i$, then we must have $x_i = 0$

Therefore, $\mathbf{x}$ and λ must satisfy:

$$x_i = \begin{cases} 1/\lambda - \alpha_i & \lambda < 1/\alpha_i \\ 0 & \lambda \geq 1/\alpha_i \end{cases}$$

This is equivalent to $x_i = \max\{0, 1/\lambda - \alpha_i\}$.

Example 3 Continued

Now we have one last condition to satisfy, that is $\mathbf{e}^T\mathbf{x} = 1$. That means, we must have

$$\sum_{i=1}^{n} \max\{0, 1/\lambda - \alpha_i\} = 1$$

Note that the left hand side is a piecewise-linear increasing function of $1/\lambda$, with breakpoints at α_i, therefore the equation can be easily solved (assuming λ^* to be the solution). Then

$$x_i^* = \max\{0, 1/\lambda^* - \alpha_i\}$$

is the optimal solution to the original problem.

Example 3 Continued: Water-Filling Interpretation

The solution to

$$\sum_{i=1}^{n} \max\{0, 1/\lambda - \alpha_i\} = 1$$

can be viewed as pouring water into a field with n patches, each with original height α_i. And the solution is the flood level when we pour 1 unit of water. See the below figure for an illustration.