MAT 3007 – Optimization

KKT Conditions

Lecture 13 *July 7th*

Andre Milzarek SDS / CUHK-SZ

Repetition

First-Order Conditions for Unconstrained Problems:

- Points with $\nabla f(x) = 0$ are called stationary points.
- FONC: every local minimizer (maximizer) is a stationary point.

Second-Order Conditions for Unconstrained Problems:

- If x^* is a local minimizer (maximizer), then $\nabla^2 f(x^*)$ is positive (negative) semidefinite.
- If x^* is a stationary point and $\nabla^2 f(x^*)$ is positive (negative) definite, then x^* is a strict local minimizer (maximizer).
- If x^* is a stationary point and $\nabla^2 f(x^*)$ is indefinite, then x^* is a saddle point.

Feasible Directions and Descent Directions:

- We call d a feasible direction at $x \in \Omega$ if there exists $\bar{t} > 0$ such that $x + td \in \Omega$ for all $0 \leq t \leq \bar{t}$.
- d is called a descent direction at x if and only if $\nabla f(x)^\top d < 0$.

First-Order Conditions for Constrained Problems:

- If x^* is a local minimum of $\min_{x \in \Omega} f(x)$, then $\nabla f(x^*)^\top d \geq 0$ for any feasible direction d at x^*.

$\rightsquigarrow$ There are no feasible descent directions.

Existence of Global Solutions:

- Weierstraß Theorem: If Ω is compact, then the optimization problem $\min_{x \in \Omega} f(x)$ attains a global min. (and max.).
- If f is coercive ($f(x) \to \infty$ if $\|x\| \to \infty$), then the problem $\min_{x \in \mathbb{R}^n} f(x)$ attains a global solution.

Optimality Conditions for Constrained Problems

We want to use the notion of feasible and descent directions to obtain optimality conditions for nonlinear programs of the form:

General Nonlinear Optimization Problem:

$$\begin{aligned} \text{minimize}_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \quad \forall\ i = 1, ..., m, \\ & h_j(x) = 0, \quad \forall\ j = 1, ..., p. \end{aligned}$$

- The feasible set is $\Omega = \{x \in \mathbb{R}^n : g(x) \le 0,\ h(x) = 0\}$.

Definition: Active and Inactive Set

At a point $x \in \Omega$, the set $\mathcal{A}(x) := \{i : g_i(x) = 0\}$ denotes the set of active constraints. The set of inactive constraints is given by $\mathcal{I}(x) := \{i : g_i(x) < 0\}$.

$\rightsquigarrow$ We first consider linear constraints as special case.

Optimality Conditions: Linear Constraints

We now first consider an inequality constrained problem:

$$\text{minimize}_x \; f(x) \quad \text{s.t.} \quad Ax \geq b. \tag{1}$$

How can we express the necessary optimality conditions?

Theorem: FONC for Linearly Constrained Problems

If x^* is a local minimum of (1), then there exists some $y \geq 0$ with

$$\begin{aligned} \nabla f(x^*) - A^\top y &= 0 \\ y_i \cdot (a_i^\top x^* - b_i) &= 0 \quad \forall \; i, \end{aligned}$$

where $a_i^\top$ is the ith row of A.

As a consequence, the first-order conditions for the problem

$$\text{minimize}_x \; f(x) \quad \text{s.t.} \quad Ax = b \tag{2}$$

are given by:

Theorem: Linear Equality Constraints

If x^* is a local minimum of (2), then there is some $y \in \mathbb{R}^m$ with

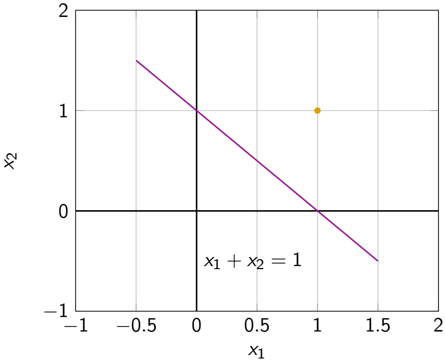
$$\nabla f(x^*) = A^\top y.$$

Example

Consider the problem:

$$\begin{array}{ll} \text{minimize} & (x_1 - 1)^2 + (x_2 - 1)^2 \\ \text{s.t.} & x_1 + x_2 = 1 \end{array}$$

- This problem finds the nearest point on the line $x_1 + x_2 = 1$ to the point $(1, 1)$



Example: Continued

By the FONC, if $x = (x_1, x_2)$ is a local minimizer, then there exists y such that

$$A^\top y = \nabla f(x).$$

Here, we have $A = (1, 1)$ and $\nabla f(x) = (2x_1 - 2, 2x_2 - 2)^\top$.

Thus, there exists y with

$$2x_1 - 2 = y \quad \text{and} \quad 2x_2 - 2 = y.$$

Combined with the constraint $x_1 + x_2 = 1$, we can infer:

$$x_1 = x_2 = \frac{1}{2}, \quad y = -1.$$

This is the only candidate for a local solution.

$\rightsquigarrow$ Indeed, it is a local minimizer (also a global minimizer)!

We have discussed cases with linear equality constraints or linear inequality constraints and derived first optimality conditions:

- We want to extend them to more general cases $\rightsquigarrow$ KKT conditions.
- The first-order necessary conditions for general nonlinear programs are called KKT conditions.
- The KKT conditions were originally named after H. Kuhn and A. Tucker, who first published the conditions in 1951. Later scholars discovered that the conditions had been stated by W. Karush in his master's thesis in 1939.

General Inequality Constraints

We consider the nonlinear program:

$$\begin{aligned} \text{minimize}_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \quad \forall\ i = 1, ..., m. \end{aligned} \tag{3}$$

- f and g_i, $i = 1, ..., m$, are assumed to be cont. differentiable.

Lemma: No Feasible Descent for Inequality Constraints

Let x^* be a local minimum of (3). Then, there does not exist a vector $d \in \mathbb{R}^n$ such that:

$$\nabla f(x^*)^\top d < 0 \quad \text{and} \quad \nabla g_i(x^*)^\top d < 0 \quad \forall\ i \in \mathcal{A}(x^*).$$

Fritz-John Conditions

Theorem: Fritz-John Conditions

Let x^* be a local minimum of (3). Then, there exists λ_0, λ_1, ..., $\lambda_m \geq 0$, which are not all zeros, such that

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) = 0,$$
$$\lambda_i \cdot g_i(x) = 0, \quad \forall\ i = 1, ..., m.$$

- Major Drawback: The choice $\lambda_0 = 0$ is allowed.
- In this case, the Fritz-John conditions just impose linear dependence of the vectors $\{\nabla g_i(x^*)\}_{i \in \mathcal{A}(x^*)}$.

$\rightsquigarrow$ No information about f! Many points might satisfy this condition without being minimizer!

Theorem: KKT-Conditions for Inequality Constraints

Let x^* be a local minimum of (3) and suppose that the vectors

$$\{\nabla g_i(x^*) : i \in \mathcal{A}(x^*)\}$$

are linearly independent. Then, there exist $\lambda_1, ..., \lambda_m \geq 0$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) = 0,$$

$$\lambda_i \cdot g_i(x) = 0, \quad \forall\ i = 1, ..., m.$$

General Inequality and Equality Constraints: KKT Conditions

KKT Conditions: The General Setting

General Nonlinear Optimization Problem:

$$\begin{aligned} \text{minimize}_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad \forall\, i = 1, ..., m, \\ & h_j(x) = 0, \quad \forall\, j = 1, ..., p. \end{aligned}$$

Summary & Agenda:

- We have already derived the KKT conditions for inequality constrained problems!
- Main Tools: Feasible and descent directions, strong duality for linear programs.

$\rightsquigarrow$ Let's slow down a bit and develop the general KKT conditions step by step.

The Lagrangian

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

1. We associate each constraint with a Lagrangian multiplier:

$$\begin{aligned} g_i(x) \le 0 \quad & \cdots \quad \lambda_i \\ h_j(x) = 0 \quad & \cdots \quad \mu_j \end{aligned}$$

2. We define the Lagrangian of this problem by:

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

Lagrangian Multipliers:

- We require the multipliers associated with $g_i(x) \leq 0$ be non-negative: $\lambda_i \geq 0$.
- The multipliers for the equality constraints are free: $\mu_j \in \mathbb{R}$.
- The conditions "$\lambda_i \geq 0$, μ_i free" are called dual feasibility conditions.

Meaning of the Lagrangian:

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

$\rightsquigarrow$ We move all the constraints to the objective and interpret them as reward/penalty functions.

- We want $g_i(x) \leq 0$: setting $\lambda_i \geq 0$, we reduce the Lagrangian when the constraint holds and penalize/increase it when the constraint does not hold.

KKT from Lagrangian

Now we derive the KKT conditions from the Lagrangian.

We first take derivative with respect to x and set it to zero.

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

Finally we have a set of complementarity conditions:

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

$$\begin{aligned} \lambda_i \cdot g_i(x) &= 0, \quad \forall\, i \\ \mu_j \cdot h_j(x) &= 0, \quad \forall\, j. \end{aligned}$$

The second condition is usually omitted since the feasibility of x already guarantees it.

Theorem: KKT Conditions

If x is a local minimizer and if a regularity condition $(\star)$ holds, then there exist λ and μ such that:

1. Main Condition

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{j=1}^{p} \mu_j \nabla h_j(x) = 0.$$

2. Dual Feasibility

$$\lambda_i \geq 0 \quad i = 1, ..., m.$$

3. Complementarity

$$\lambda_i \cdot g_i(x) = 0 \quad \forall\ i = 1, ..., m.$$

We often add primal feasibility as part of the KKT conditions:

4. Primal Feasibility

$$g_i(x) \leq 0, \quad h_j(x) = 0 \quad \forall\ i, \quad \forall\ j.$$

Constraint Qualifications

As before, we require the collection of gradients

$$\{\nabla g_i(x^*) : i \in \mathcal{A}(x^*)\} \cup \{\nabla h_j(x^*) : j = 1, ..., p\} \qquad (\star)$$

to be linearly independent or to have full rank.

- This condition is a constraint qualification (CQ) and is called Linear Independence Constraint Qualification (LICQ).
- A feasible point x^* satisfying the LICQ is called regular.
- There are more CQs: ACQ, GCQ, MFCQ, PLICQ, Slater's condition ($\rightsquigarrow$ Optimization II (?)).

KKT Points and Remarks

Further Remarks:

- A (feasible) point satisfying the KKT conditions is called a KKT point.
- KKT points are candidates for local optimal solutions – just like stationary points.

The main condition can be compactly represented as:

$$\nabla_x L(x, \lambda, \mu) = \nabla f(x) + \nabla g(x)\lambda + \nabla h(x)\mu = 0,$$

where we set $\nabla g(x) := (\nabla g_1(x), ..., \nabla g_m(x)) = Dg(x)^\top \in \mathbb{R}^{m \times n}$ and $\nabla h(x) := (\nabla h_1(x), ..., \nabla h_p(x)) = Dh(x)^\top \in \mathbb{R}^{p \times n}$.

Examples: Formulating and Using KKT Conditions

Consider the problem:

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

⇝ Formulate the KKT conditions!

First we associate the constraints with Lagrange multipliers λ_1 and λ_2 and construct the Lagrangian for this problem:

$$\begin{aligned} L(x, \lambda) = {} & 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ & + \lambda_1(x_1^2 + x_2^2 - 5) + \lambda_2(3 - 3x_1 - x_2) \end{aligned}$$

with $\lambda_1 \ge 0$, $\lambda_2 \ge 0$.

The gradients are given by:

$$\nabla f(x) = \begin{pmatrix} 4x_1 + 2x_2 - 10 \\ 2x_1 + 2_x 2 - 10 \end{pmatrix}, \nabla g_1(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}, \nabla g_2(x) = \begin{pmatrix} -3 \\ -1 \end{pmatrix}$$

The main conditions are:

$$\begin{aligned} 4x_1 + 2x_2 - 10 + 2\lambda_1 x_1 - 3\lambda_2 &= 0 \\ 2x_1 + 2x_2 - 10 + 2\lambda_1 x_2 - \lambda_2 &= 0 \end{aligned}$$

The complementarity conditions are:

$$\begin{aligned} \lambda_1 \cdot (x_1^2 + x_2^2 - 5) &= 0 \\ \lambda_2 \cdot (3 - 3x_1 - x_2) &= 0 \end{aligned}$$

Therefore, the KKT conditions are given by:

1. Main Conditions:

$$4x_1 + 2x_2 - 10 + 2\lambda_1 x_1 - 3\lambda_2 = 0$$
$$2x_1 + 2x_2 - 10 + 2\lambda_1 x_2 - \lambda_2 = 0$$

2. Primal Feasibility

$$x_1^2 + x_2^2 \leq 5, \quad -3x_1 - x_2 \leq -3$$

3. Dual Feasibility

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

4. Complementarity Conditions

$$\lambda_1 \cdot (x_1^2 + x_2^2 - 5) = 0$$
$$\lambda_2 \cdot (3 - 3x_1 - x_2) = 0$$

Formulate the KKT conditions for the problem:

$$\begin{aligned} \text{minimize}_x \quad & \tfrac{1}{2}x^\top Qx - c^\top x \\ \text{s.t.} \quad & Ax = b \\ & Cx \le d \\ & x \ge 0 \end{aligned}$$

- We associate the linear constraints with the multiplier μ.
- We introduce the multipliers λ and η for the constraints.

The Lagrangian for this problem is:

$$L(x, \lambda, \mu, \eta) = \frac{1}{2}x^\top Qx - c^\top x + \mu^\top(Ax - b) + \lambda^\top(Cx - d) - \eta^\top x.$$

We have $\lambda, \eta \ge 0$ and μ is a free variable.

Example II - Continued

The main condition is:

$$\nabla_x L(x, \lambda, \mu, \eta) = Qx - c + A^\top \mu + C^\top \lambda - \eta = 0.$$

The complementarity conditions are:

$$\begin{aligned} \lambda_i \cdot (c_i^\top x - d_i) &= 0, \qquad \forall\ i \\ -\eta_i \cdot x_i &= 0, \qquad \forall\ i, \end{aligned}$$

where $c_i^\top$ is the ith row of $C^\top$.

Therefore, the KKT conditions are:

1. Main condition:

$$\nabla_x L(x, \lambda, \mu, \eta) = Qx - c + A^\top \mu + C^\top \lambda - \eta = 0.$$

2. Primal feasibility

$$Ax = b, \quad Cx \leq d, \quad x \geq 0.$$

3. Dual feasibility

$$\lambda \geq 0, \quad \eta \geq 0.$$

4. Complementarity conditions

$$\begin{aligned} \lambda_i \cdot (c_i^\top x - d_i) &= 0, \qquad \forall\ i \\ -\eta_i \cdot x_i &= 0, \qquad \forall\ i. \end{aligned}$$

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:

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

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

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

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

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 $\lambda \ge 0$

Example III - 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:

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

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.

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

$$2\pi r^2 + 2\pi rh = C \tag{4}$$

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

$$-2\pi rh + 4\pi r\lambda + 2\pi h\lambda = 0 \tag{5}$$

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

$$-\pi r^2 + 2\pi r\lambda = 0 \tag{6}$$

From (6), we have $r = 2\lambda$, then back to (5), we get $h = 4\lambda$. And back to (4), 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.

Questions?