COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Math Methods for Political Science

Lecture 18: the Karush–Kuhn–Tucker conditions

Thibault Vatter <tv2233@columbia.edu>
Department of Statistics, Columbia University

November 27, 2017

Reminder: Lagrange's method

Consider the following optimization problem:

$$\begin{aligned} \min_x/\max \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) = b_i,\ i = 1, \dots, m. \end{aligned}$$

Definition 1 (Lagrangian and Lagrange's multipliers)

$L(x, \lambda_1, \cdots, \lambda_m) = f_0(x) - \sum_{i=1}^m \lambda_i (f_i(x) - b_i)$ is the **Lagrangian** and $\lambda_1, \cdots, \lambda_m$ are the **Lagrange's multipliers**.

To solve the problem above, notice that

$$\nabla L(x, \lambda_1, \cdots, \lambda_m) = 0 \iff \begin{cases} \nabla f_0(x) & = \sum_{i=1}^m \lambda_i \nabla f_i(x) \\ f_i(x) & = b_i \ i = 1, \dots, m \end{cases}$$

Theorem 1 (First order conditions)

Under some regularity conditions, if x^* is a local optimum, then there exists λ^*, such that $\nabla L(x^*, \lambda^*) = 0$

Local maximum or minimum?

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Definition 2 (Bordered Hessian)

The **bordered Hessian** is

$$H = \begin{bmatrix} 0 & 0 & \cdots & 0 & \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ 0 & 0 & \cdots & 0 & \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots & \frac{\partial f_m}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \\ \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} & \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 L}{\partial x_1 \partial x_n} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} & \frac{\partial^2 L}{\partial x_2^2} & \cdots & \frac{\partial^2 L}{\partial x_2 \partial x_n} \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_1} & \frac{\partial^2 L}{\partial x_1 \partial x_n} & \frac{\partial^2 L}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 L}{\partial \partial x_n^2} \end{bmatrix}$$

Theorem 2 (Second order conditions)

If the $n - m$ largest principal minors

- *alternate in sign with the smallest one having the sign of* $(-1)^{m+1} \implies$ *local maximum,*
- *have the sign of* $(-1)^m \implies$ *local minimum.*

The Karush–Kuhn–Tucker approach

Consider the following optimization problem:

$$\begin{aligned} \min_x/\max \quad & f_0(x) \\ \text{subject to} \quad & g_i(x) = 0,\ i = 1, \ldots, m \\ & h_j(x) \leq 0,\ j = 1, \ldots, l. \end{aligned}$$

Definition 3 (Lagrangian and KKT multipliers)

The Lagrangian is

$$L(x, \lambda, \mu) = f_0(x_1, \cdots, x_n) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{l} \mu_j h_j(x),$$

and λ and μ are the **Karush–Kuhn–Tucker multipliers**.

The Karush–Kuhn–Tucker conditions

Columbia University in the City of New York

Theorem 3 (First order/Karush–Kuhn–Tucker conditions)

Under some regularity conditions, if x^* is a local optimum, then there exists λ^* and μ^*, such that

- **Stationarity:**
  - *(minimizing)* $-\nabla f_0(x^*) = \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^l \mu_j^* \nabla h_j(x^*)$
  - *(maximizing)* $\nabla f_0(x^*) = \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^l \mu_j^* \nabla h_j(x^*)$
- **Primal feasibility:**
  - *(equalities)* $g_i(x^*) = b_i, \ i = 1, \dots, m$
  - *(inequalities)* $h_j(x^*) \leq c_j, \ j = 1, \dots, l$
- **Dual feasibility:** $\mu_j^* \geq 0, \ j = 1, \dots, l$
- **Complementary slackness:** $\mu_j^*(h_j(x^*) - c_j) = 0, \ j = 1, \dots, l$

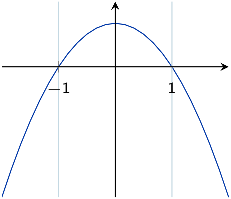
Definition 4 (Binding and non-binding constraints)

Constraint j is **binding** if $h_j(x^*) = c_j$ and **slack** otherwise.

Binding and non-binding constraints

Complementary slackness condition together with constraint j being slack implies $\mu_j^* = 0$. How about $\Longleftarrow$?

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & 1 - x^2 \\ \text{subject to} \quad & x \leq c \end{aligned}$$



Modern portfolio theory

The setup:

- $r = (r_1, \cdots, r_n)$ is the (random) vector of returns on n assets.
- The mean vector for the returns is $\mu = E[r]$.
- The covariance matrix for the returns is $\Sigma = Cov[r]$.
- A portfolio is a set of weights $x = (x_1, \cdots, x_n)$.
- The portfolio return is $r_p = \sum x_i r_i = x^\top r$.

The basic mean-variance optimization problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}\text{Var}[r_p] \\ \text{subject to} & E[r_p] \geq c \\ & x^\top e = 1 \end{array} \iff \begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}x^\top \Sigma x \\ \text{subject to} & x^\top \mu \geq c \\ & x^\top e = 1 \end{array}$$