R401: Statistical and Mathematical Foundations

Lecture 15: Nonlinear Programming and Concave Optimization

Andrey Vassilev

2016/2017

Lecture Contents

Basic formulation with inequality constraints

We now look at a problem which is very similar to the case of optimization with equality constraints:

$$f(x_1, \dots, x_n) \to \max \tag{1}$$

s.t.

$$\begin{array}{l} g_1(x_1, \dots, x_n) \le b_1 \\ g_2(x_1, \dots, x_n) \le b_2 \\ \dots \\ g_m(x_1, \dots, x_n) \le b_m \end{array} . \tag{2}$$

In vector notation:

$$f(\mathbf{x}) \to \max$$

s.t.

$$\mathbf{g}(\mathbf{x}) \le \mathbf{b}.$$

Basic formulation with inequality constraints

- A vector $\mathbf{x}$ satisfying the constraints (2) is called *admissible* or *feasible*.
- In some alternative (but essentially equivalent) formulations the constraints take the form $g_i(x_1, \ldots, x_n) \leq 0$ or $g_i(x_1, \ldots, x_n) \geq 0$ for $i = 1, \ldots, m$.
- The set of admissible vectors is called the *admissible (feasible) set*.
- With inequality constraints the requirement $m < n$ is not necessary. Intuitively, this is because an inequality constraint is much more forgiving: think of a line vs. a half-plane.
- We focus on maximization problems here. Notice that minimizing a function $f(x)$ is equivalent to maximizing $-f(x)$, so there is no loss of generality in our choice.

Basic formulation with inequality constraints

The Lagrangian

We again approach problem (1)-(2) by defining a *Lagrangian*:

$$\mathcal{L}(\mathbf{x}) = f(\mathbf{x}) - \lambda_1(g_1(\mathbf{x}) - b_1) - \cdots - \lambda_m(g_m(\mathbf{x}) - b_m).$$

The Lagrangian takes the familiar form from the case of equality constraints!

The differences arise in the algorithm used to obtain candidates for optimality.

Solution recipe for the case of inequality constraints

When trying to find solutions to (1)-(2), the following procedure is often applied:

Algorithm (Kuhn-Tucker conditions)

1. Form the Lagrangian
2. Differentiate it w.r.t. the elements of $\mathbf{x}$ and set the resulting derivatives equal to zero:

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial f(\mathbf{x})}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(\mathbf{x})}{\partial x_i} = 0, \ i = 1, \ldots, n. \tag{3}$$

3. Check the *complementary slackness* conditions

$$\lambda_j \geq 0 \text{ and } \lambda_j (g_j(\mathbf{x}) - b_j) = 0, \ j = 1, \ldots, m. \tag{4}$$

4. The points satisfying 2) and 3) above are the candidates for optimality

Condition 3) above implies that

$$\lambda_j = 0 \text{ if } g_j(\mathbf{x}) < b_j, \ j = 1, \ldots, m.$$

Comments on the Kuhn-Tucker conditions

- The term *complementary slackness* derives from the fact that according to (4) one of the conditions $\lambda_j \geq 0$ and $g_j(x_1, \ldots, x_n) \leq b_j$ may be *slack* (i.e. be a strict inequality), while the other must bind (i.e. be fulfilled as an equality). Thus, they *complement* each other.
- Let $\mathbf{x}^*$ be an admissible point. If it is true that $g_j(\mathbf{x}^*) = b_j$, the respective constraint is called *active* or *binding*.
- It is possible to have simultaneously $\lambda_j = 0$ and $g_j(\mathbf{x}^*) = b_j$ for some js.

A simple illustration of the Kuhn-Tucker procedure

Example 1

$$\max_{x,y} f(x,y) = xy$$

s.t.

$$x^2 + y^2 \leq 1.$$

The Lagrangian is

$$\mathcal{L} = xy - \lambda(x^2 + y^2 - 1).$$

$$\frac{\partial \mathcal{L}}{\partial x} = y - 2\lambda x = 0 \quad \Rightarrow \quad y = 2\lambda x,$$

$$\frac{\partial \mathcal{L}}{\partial y} = x - 2\lambda y = 0 \quad \Rightarrow \quad x = 2\lambda y.$$

The complementary slackness condition reads

$$\lambda(x^2 + y^2 - 1) = 0.$$

A simple illustration of the Kuhn-Tucker procedure

Example 1 (cont.)

Assume $\lambda = 0$. Then the conditions $\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial \mathcal{L}}{\partial y} = 0$ imply $x = y = 0$, which provides one candidate.

Assume $\lambda \neq 0$. We have $x^2 + y^2 = 1$. Note that $x = 0$ would imply $y = 0$ and vice versa, which would violate the condition $x^2 + y^2 = 1$. Thus, $x, y \neq 0$. Then we obtain

$$\frac{x}{y} = \frac{y}{x} \quad \Rightarrow \quad x^2 = y^2.$$

A simple illustration of the Kuhn-Tucker procedure

Example 1 (cont.)

The result $x^2 = y^2$, combined with $x^2 + y^2 = 1$, yields four possibilities:

$$\begin{aligned}
(x, y) &= \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) &\Rightarrow \lambda &= \frac{1}{2}, \\
(x, y) &= \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) &\Rightarrow \lambda &= -\frac{1}{2}, \\
(x, y) &= \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) &\Rightarrow \lambda &= -\frac{1}{2}, \\
(x, y) &= \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) &\Rightarrow \lambda &= \frac{1}{2}.
\end{aligned}$$

The second and third case can be excluded, as they are associated with $\lambda < 0$. Thus, we are left with the candidates $x = y = \frac{1}{\sqrt{2}}$ and $x = y = -\frac{1}{\sqrt{2}}$ in addition to the candidate $x = y = 0$ obtained above. Note, however, that the last point yields a smaller value of the objective function and can be excluded.

From the Kuhn-Tucker algorithm to necessary conditions

While the Kuhn-Tucker algorithm in its present form provides us with candidates, we need to strengthen it further to obtain proper necessary conditions for optimality.

From the Kuhn-Tucker algorithm to necessary conditions

While the Kuhn-Tucker algorithm in its present form provides us with candidates, we need to strengthen it further to obtain proper necessary conditions for optimality.

Fact 1 (Kuhn-Tucker necessary conditions)

Let $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)'$ be a solution to (1)-(2). Suppose that the functions f and g_i, $i = 1, \ldots, m$, have continuous partial derivatives on a set $S \subseteq \mathbb{R}^n$, and $\mathbf{x}^* \in \operatorname{int} S$. Suppose also that the gradients of the constraints which are binding at $\mathbf{x}^*$, are linearly independent, i.e. if $K := \{k \mid g_k(\mathbf{x}^*) = 0\}$, it is true that

$$\nabla g_k(\mathbf{x}^*),\ k \in K, \text{ are linearly independent.}$$

Then there exist unique numbers $\lambda_1, \ldots, \lambda_m$ such that conditions (3) and (4) hold at $\mathbf{x}^*$.

The linear independence condition in Fact 1 is called the *constraint qualification (CQ)*.

An equivalent formulation of the constraint qualification

Take the *Jacobian* of the constraint function $\mathbf{g}(\mathbf{x})$ and remove the rows that correspond to the inactive constraints:

$$\begin{pmatrix} \dfrac{\partial g_1(\mathbf{x}^*)}{\partial x_1} & \dots & \dfrac{\partial g_1(\mathbf{x}^*)}{\partial x_n} \\ \dfrac{\partial g_2(\mathbf{x}^*)}{\partial x_1} & \dots & \dfrac{\partial g_2(\mathbf{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial g_{m-1}(\mathbf{x}^*)}{\partial x_1} & \dots & \dfrac{\partial g_{m-1}(\mathbf{x}^*)}{\partial x_n} \\ \dfrac{\partial g_m(\mathbf{x}^*)}{\partial x_1} & \dots & \dfrac{\partial g_m(\mathbf{x}^*)}{\partial x_n} \end{pmatrix}$$

An equivalent formulation of the constraint qualification

Take the *Jacobian* of the constraint function $\mathbf{g}(\mathbf{x})$ and remove the rows that correspond to the inactive constraints:

$$
\begin{pmatrix}
\frac{\partial g_1(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial g_1(\mathbf{x}^*)}{\partial x_n} \\
\frac{\partial g_2(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial g_2(\mathbf{x}^*)}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial g_{m-1}(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial g_{m-1}(\mathbf{x}^*)}{\partial x_n} \\
\frac{\partial g_m(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial g_m(\mathbf{x}^*)}{\partial x_n}
\end{pmatrix}
$$

An equivalent formulation of the constraint qualification

Take the *Jacobian* of the constraint function $\mathbf{g}(\mathbf{x})$ and remove the rows that correspond to the inactive constraints:

$$
\begin{pmatrix}
\dfrac{\partial g_1(\mathbf{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial g_1(\mathbf{x}^*)}{\partial x_n} \\[2ex]
\dfrac{\partial g_2(\mathbf{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial g_2(\mathbf{x}^*)}{\partial x_n} \\[2ex]
\vdots & \ddots & \vdots \\[1ex]
\dfrac{\partial g_{m-1}(\mathbf{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial g_{m-1}(\mathbf{x}^*)}{\partial x_n} \\[2ex]
\dfrac{\partial g_m(\mathbf{x}^*)}{\partial x_1} & \cdots & \dfrac{\partial g_m(\mathbf{x}^*)}{\partial x_n}
\end{pmatrix}
$$

Here the first and the $(m-1)$-th constraints are **not** binding.

An equivalent formulation of the constraint qualification

Thus, if k out of m constraints are not binding, we are left with a $(m-k) \times n$ matrix of partial derivatives, evaluated at $\mathbf{x}^*$.

The constraint qualification requires that the rank of this matrix should be equal to the number of rows $m-k$.

Applying the necessary conditions

The CQ is a potential source of problems when applying the Kuhn-Tucker necessary conditions: it is possible to have an optimal point where the CQ (and hence the NCs) fail.

Therefore the general procedure is as follows:

1. Use Fact 1 to find a set of candidates.
2. Find the feasible points where the CQ fails. These are also candidates.
3. Search for the optimum over the union of the preceding two sets.

Some illustrations

Example 2

Consider the problem

$$f(x_1, x_2) = -x_1^3 + x_2 \to \max$$

s.t.

$$x_2 \leq 0.$$

We have

$$\mathcal{L} = -x_1^3 + x_2 - \lambda x_2$$

$$\frac{\partial \mathcal{L}}{\partial x_1} = -3x_1^2 = 0 \quad \Rightarrow \quad x_1 = 0.$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 1 - \lambda = 0 \quad \Rightarrow \quad \lambda = 1.$$

Complementary slackness requires that $\lambda x_2 = 0$, hence $x_2 = 0$.

Some illustrations

Example 2 (cont.)

The gradient of the constraint function is $\nabla g(x_1, x_2) = (g'_x(x_1, x_2), g'_y(x_1, x_2))' = (0, 1)'$, so the CQ is satisfied everywhere.

To sum up, the K-T NCs are satisfied and $x_1 = 0,\ x_2 = 0$ is our only candidate.

However, it is easily seen that any point of the type $(x_1, 0)$ with $x_1 < 0$ is admissible, and for $x_1 \to -\infty$ the objective function grows unboundedly. This again illustrates the need to use NCs with caution.

Some illustrations

Example 3 (Example 1 revisited)

The solution of Example 1 did not make use of the CQ condition. Let us check it now:

$$g(x, y) = x^2 + y^2.$$

Readings

Main references:
Sydsæter et al. *Further mathematics for economic analysis*. Chapter 3.

Additional readings: