R401: Statistical and Mathematical Foundations

Nonlinear Programming and Concave Optimization

Andrey Vassilev

Lecture Contents

Static optimization with inequality constraints

Basic formulation with inequality constraints

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

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

s.t.

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

In vector notation:

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

s.t.

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

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 $\mathbf{x}^* = (x_1^*, \dots, x_n^*)'$ be a solution to (1)-(2). Suppose that the functions f and g_i, $i = 1, \dots, m$, 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 $\lambda_1^*, \dots, \lambda_m^*$ such that equations (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:

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

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:

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

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 $m-k$.

Remark: While we work with a specific version of the CQ here, bear in mind that the term "constraint qualification" refers to a whole class of different regularity conditions used in optimization problems.

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.

For more complicated problems one approach is to work over the various combinations of constraints (assume all are binding, one is binding etc.) and to study these cases one by one.

Some illustrations

Example 2 (Example 1 revisited)

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

$$g(x,y) = x^2 + y^2 \quad \Rightarrow \quad \nabla g(x,y) = (2x, 2y)'.$$

Since any single vector except the zero vector is linearly independent in itself, the CQ fails only at $(0,0)'$.
Therefore, a strict application of Fact 1 would have required to treat the case $x = 0,\ y = 0$ separately, to establish the other two possibilities $x = y = \frac{1}{\sqrt{2}}$ and $x = y = -\frac{1}{\sqrt{2}}$, and then to compare the three cases. In working through Example 1 we obtained the first case via the K-T algorithm.

Some illustrations

Example 3

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 3 (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.

Sufficient conditions

The following fact provides a sufficient condition for optimality by imposing a concavity requirement on the Lagrangian.

Fact 2

Let $\mathbf{x}^*$ be an admissible point for the problem (1)-(2).
Suppose that:

1. We can find a vector $(\lambda_1^*, \ldots, \lambda_m^*)'$ which, together with $\mathbf{x}^*$, satisfies equations (3) and (4) in the K-T algorithm .
2. The Lagrangian is concave.

Then the point $\mathbf{x}^*$ is optimal.

Sufficient conditions

Remarks

- The requirement of concavity or convexity of a Lagrangian or a similar object is a popular way of obtaining sufficient conditions for optimality.
- Rather than working directly with the definition of concavity, it is often useful to apply special results to show concavity. Examples include:
  - The sum of concave functions is concave.
    - As a special case, if $f(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i)$ and f_i are concave, then f is also concave.
  - If f is concave and $\alpha > 0$, then αf is concave.
  - A twice differentiable function of one variable is concave on an interval I if $f''(x) \leq 0, \ \forall x \in I$.
  - Let $f(\mathbf{x})$ be a twice differentiable function defined on an open convex set $S \subset \mathbb{R}^n$. The function $f(\mathbf{x})$ is concave in S if and only if the Hessian $\mathbf{f}''(\mathbf{x})$ is negative semidefinite for all $\mathbf{x} \in S$. If the Hessian $\mathbf{f}''(\mathbf{x})$ is negative definite for all $\mathbf{x} \in S$, it follows that $f(\mathbf{x})$ is *strictly* concave in S.
- Similar results hold for convex functions. See SHSS, 2.2-2.4, for more information on convex sets and concave/convex functions.

Example: utility maximization

Consider a consumer with income I who plans to buy certain quantities, x_1 and x_2, of two goods at prices p_1 and p_2, respectively . Obviously, we require $p_1, p_2, I > 0$. The consumer's preferences are described via the utility function

$$U(x_1, x_2) = x_1^{\alpha} x_2^{\beta}, \quad \alpha, \beta > 0,\ \alpha + \beta < 1.$$

The consumer's feasible set is characterized by the budget constraint and nonnegativity constraints on the quantities:

$$p_1 x_1 + p_2 x_2 \leq I,$$

$$x_1 \geq 0,$$

$$x_2 \geq 0.$$

The consumer seeks to maximise utility subject to the above constraints.

Example: utility maximization

Notice that the consumer is **not** required to spend all his income!

To cast the problem in canonical form, we rewrite the nonnegativity constraints as

$$-x_1 \leq 0, \qquad -x_2 \leq 0.$$

The Lagrangian is

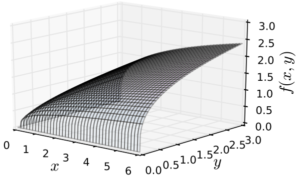
$$\mathcal{L} = x_1^{\alpha} x_2^{\beta} - \lambda_1 (p_1 x_1 + p_2 x_2 - I) - \lambda_2(-x_1) - \lambda_3(-x_2)$$

and is concave in (x_1, x_2).

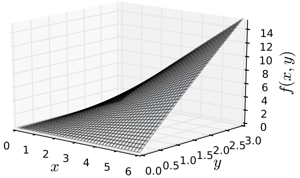
Note: The concavity of the Lagrangian depends on the concavity of $U(x_1, x_2)$, which hinges on the assumption $\alpha + \beta < 1$.

Example: utility maximization

The graph of $x^\alpha y^\beta$ for α = 0.3, β = 0.3



The graph of $x^\alpha y^\beta$ for α = 0.95, β = 0.95



Example: utility maximization

Notice that the Lagrangian is not differentiable at $x_1 = 0$ or $x_2 = 0$. However, we can exclude those cases as it is immediately seen that any feasible point with positive elements yields a positive value of the utility function (compared to the zero utility when one or both x_i is zero).

Then, we can apply the K-T algorithm:

$$
\begin{aligned}
&\frac{\partial \mathcal{L}}{\partial x_1} = \alpha x_1^{\alpha-1} x_2^{\beta} - \lambda_1 p_1 + \lambda_2 = 0, \\
&\frac{\partial \mathcal{L}}{\partial x_2} = \beta x_1^{\alpha} x_2^{\beta-1} - \lambda_1 p_2 + \lambda_3 = 0, \\
&\lambda_1 \geq 0, \quad \lambda_1(p_1 x_1 + p_2 x_2 - I) = 0, \\
&\lambda_2 \geq 0, \quad \lambda_2 x_1 = 0, \\
&\lambda_3 \geq 0, \quad \lambda_3 x_2 = 0.
\end{aligned}
$$

Example: utility maximization

We will work through the various possibilities for the Lagrange multipliers. These are:

Case	λ_1	λ_2	λ_3
1	0	0	0
2	0	0	+
3	0	+	0
4	0	+	+
5	+	+	+
6	+	0	+
7	+	+	0
8	+	0	0

Example: utility maximization

Our task is made easy by the following observation:

If either $\lambda_2 > 0$ or $\lambda_3 > 0$, then the complementary slackness conditions imply $x_1 = 0$ or $x_2 = 0$, which was excluded as a possibility. Thus, we are left to study only Cases 1 and 8.

Consider **Case 1**: $\lambda_1 = \lambda_2 = \lambda_3 = 0$. The complementary slackness conditions now imply $x_1, x_2 > 0$.
The Lagrangian maximization conditions yield

$$\alpha x_1^{\alpha-1} x_2^{\beta} = 0,$$

$$\beta x_1^{\alpha} x_2^{\beta-1} = 0,$$

which requires that at least one of x_1 or x_2 is zero, a contradiction. Therefore, we can disregard this case.

Example: utility maximization

Consider next **Case 8**: $\lambda_1 > 0,\ \lambda_2 = \lambda_3 = 0$. Again the complementary slackness conditions imply $x_1, x_2 > 0$.
The first complementary slackness condition now yields

$$p_1 x_1 + p_2 x_2 = I,$$

i.e. the budget constraint is binding.
The Lagrangian maximization conditions take the form

$$\alpha x_1^{\alpha-1} x_2^{\beta} - \lambda_1 p_1 = 0,$$

$$\beta x_1^{\alpha} x_2^{\beta-1} - \lambda_1 p_2 = 0.$$

Rearranging these to isolate λ_1 and equating the resulting expressions, we get

$$\frac{\alpha x_1^{\alpha-1} x_2^{\beta}}{p_1} = \frac{\beta x_1^{\alpha} x_2^{\beta-1}}{p_2} \quad \Rightarrow \quad \frac{\alpha}{p_1 x_1} = \frac{\beta}{p_2 x_2} \quad \Rightarrow \quad x_1 = \frac{\alpha p_2}{\beta p_1} x_2.$$

Example: utility maximization

Substitute the expression for x_1 in the budget constraint:

$$p_1 \left(\frac{\alpha p_2}{\beta p_1} x_2 \right) + p_2 x_2 = I.$$

Solve for x_2 to obtain

$$x_2 = \frac{\beta}{\alpha + \beta} \frac{I}{p_2}.$$

Next, substitute the solution for x_2 in the expression for x_1 obtained earlier:

$$x_1 = \frac{\alpha p_2}{\beta p_1} \frac{\beta}{\alpha + \beta} \frac{I}{p_2} \quad \Rightarrow \quad x_1 = \frac{\alpha}{\alpha + \beta} \frac{I}{p_1}.$$

Given that the formulas for x_1 and x_2 computed above confirm that $x_1, x_2 > 0$, one can in turn verify that $\lambda_1 > 0$.

Example: utility maximization

Comments

- The above analysis provides

$$x_1 = \frac{\alpha}{\alpha + \beta} \frac{I}{p_1}, \qquad x_2 = \frac{\beta}{\alpha + \beta} \frac{I}{p_2}$$

  as the only K-T candidate. In view of Fact 2, this is the solution to the consumer's problem.
- The solutions for x_1 and x_2 are standard demand functions familiar from microeconomics. The quantities depend negatively on the respective prices and positively on income.
- We obtained the fact that the budget constraint binds in the course of solving the problem, rather than as an assumption. It was a consequence of having $\lambda_1 > 0$ in the CS condition. The economic interpretation is that if a resource is valuable (as captured by the fact that its shadow price is positive), then it should not be wasted or left unutilized.

Static optimization with mixed constraints

The case of mixed inequality and equality constraints

Formulation

Sometimes an optimization problem features a combination of equality and inequality constraints:

$$f(x_1, \dots, x_n) \to \max \tag{5}$$

s.t.

$$\begin{array}{l} g_1(x_1, \dots, x_n) = b_1 \\ g_2(x_1, \dots, x_n) = b_2 \\ \dots \\ g_r(x_1, \dots, x_n) = b_r \\ h_1(x_1, \dots, x_n) \leq c_1 \\ h_2(x_1, \dots, x_n) \leq c_2 \\ \dots \\ h_s(x_1, \dots, x_n) \leq c_s \end{array}, \tag{6}$$

i.e.

$$\max f(\mathbf{x}) \text{ s.t. } \left\{ \begin{array}{l} \mathbf{g}(\mathbf{x}) = \mathbf{b} \\ \mathbf{h}(\mathbf{x}) \leq \mathbf{c} \end{array} \right. .$$

The case of mixed inequality and equality constraints

Constructing the Lagrangian

As before, we will require that $r < n$.

The Lagrangian for problem (5)-(6) is constructed as follows:

$$\mathcal{L} = f(\mathbf{x}) - \sum_{i=1}^{r} \lambda_i (g_i(\mathbf{x}) - b_i) - \sum_{j=1}^{s} \mu_j (h_j(\mathbf{x}) - c_j).$$

The Lagrange multipliers conform to the well-known principle: λ_i are free of sign (since they correspond to equality constraints) and μ_j must fulfil complementary slackness conditions.

Necessary conditions for problems with mixed constraints

Fact 3

Let $\mathbf{x}^*$ be a solution to (5)-(6). Suppose all functions involved are continuously differentiable and $r < n$. Suppose further that (without loss of generality) the first s_0 of the inequality constraints are binding at $\mathbf{x}^*$ and that the Jacobian comprising those constraints and the equality constraints has rank $s_0 + r$. Then there exist unique numbers $\lambda_1^*, \ldots, \lambda_r^*$ and $\mu_1^*, \ldots, \mu_s^*$ for which

$$\nabla \mathcal{L}(\mathbf{x}^*) = \mathbf{0},$$

$$\mu_j^* \geq 0 \text{ and } \mu_j^*(h_j(\mathbf{x}^*) - c_j) = 0,\ j = 1, \ldots, s.$$

Necessary conditions for problems with mixed constraints

In more detail, the Jacobian involving the equality constraints and the active inequality constraints is

$$\begin{pmatrix} \nabla g_1(\mathbf{x}^*) \\ \cdots \\ \nabla g_r(\mathbf{x}^*) \\ \nabla h_1(\mathbf{x}^*) \\ \cdots \\ \nabla h_{s_0}(\mathbf{x}^*) \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial g_1(\mathbf{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_r(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial g_r(\mathbf{x}^*)}{\partial x_n} \\ \frac{\partial h_1(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial h_1(\mathbf{x}^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_{s_0}(\mathbf{x}^*)}{\partial x_1} & \cdots & \frac{\partial g_{s_0}(\mathbf{x}^*)}{\partial x_n} \end{pmatrix}.$$

The requirement on the rank of the Jacobian is the familiar constraint qualification imposing linear independence.

Sufficient conditions for problems with mixed constraints

Analogously to Fact 2, we have the following result for the case of mixed constraints:

Fact 4

Let $\mathbf{x}^*$ be an admissible point satisfying the conditions stated in Fact 3. Suppose that the Lagrangian for problem (5)-(6) is concave in $\mathbf{x}$.

Then the point $\mathbf{x}^*$ is a solution to problem (5)-(6).

Concave programming

Problem description

Consider the familiar maximization problem with inequality constraints:

$$f(\mathbf{x}) \to \max \tag{7}$$

s.t.

$$\mathbf{g}(\mathbf{x}) \leq \mathbf{b}, \tag{8}$$

with the additional assumptions that the objective function f is concave and the functions $g_i,\ i = 1, \ldots, m$, are convex. Then the optimization problem is called a *concave programming problem* or simply a *concave program*.

Note: Recall that convex and concave functions are properly defined on *convex* sets (typically convex subsets of $\mathbb{R}^n$). This is an implicit requirement in the above formulation.

The concave case without constraints

It is useful to start with the following result for *unconstrained* concave optimization problems:

Fact 5

Let U be a convex subset of $\mathbb{R}^n$ and f be a continuously differentiable concave function on U. A point $\mathbf{x}^* \in U$ is a *global maximum* of f if and only if

$$\nabla' f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \leq 0, \quad \forall \mathbf{x} \in U.$$

If U is open or $\mathbf{x}^* \in \operatorname{int} U$, then $\mathbf{x}^*$ is a global maximum of f on U if and only if

$$\nabla f(\mathbf{x}^*) = 0.$$

Quasiconcavity and quasiconvexity

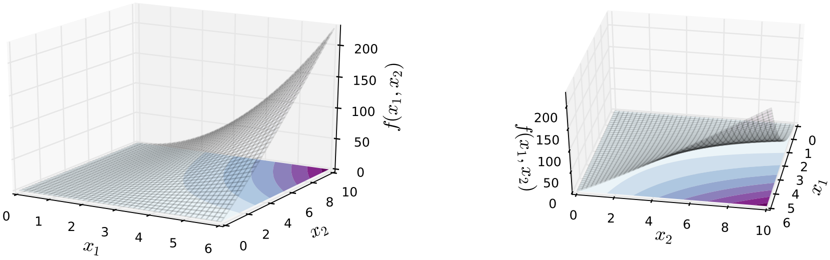
- Note that for a concave program as defined above the concavity condition on the Lagrangian stated in Fact 2 is fulfilled and we can apply Fact 2 directly.
- However, in economic applications it is sometimes useful to generalize the concepts of concave and convex functions and apply them to the concave programming problem as described below.
- A function f defined on a convex subset U of $\mathbb{R}^n$ is called *quasiconcave* if, for any $a \in \mathbb{R}$, the set
$$C_a^+ = \{\mathbf{x} \in U \mid f(\mathbf{x}) \geq a\}$$
is convex. A set C_a^+ is called an *upper level set* for f.
- A function f defined on a convex subset U of $\mathbb{R}^n$ is called *quasiconvex* if, for any $a \in \mathbb{R}$, the set
$$C_a^- = \{\mathbf{x} \in U \mid f(\mathbf{x}) \leq a\}$$
is convex. A set C_a^- is called a *lower level set* for f.

Quasiconcavity and quasiconvexity

- If f is quasiconcave, then $-f$ is quasiconvex.
- A concave function is quasiconcave and a convex function is quasiconvex.
- Increasing transformations of quasiconcave (quasiconvex) functions preserve quasiconcavity (quasiconvexity).
- Decreasing transformations of quasiconcave (quasiconvex) functions reverse the property (i.e. quasiconcavity turns into quasiconvexity and vice versa).
- Quasiconcavity is equivalent to the statements that for all $\mathbf{x}, \mathbf{y} \in U$ and $t \in [0,1]$:

  1. $f(\mathbf{x}) \geq f(\mathbf{y}) \quad \Rightarrow \quad f(t\mathbf{x} + (1-t)\mathbf{y}) \geq f(\mathbf{y})$,
  2. $f(t\mathbf{x} + (1-t)\mathbf{y}) \geq \min\{f(\mathbf{x}), f(\mathbf{y})\}$.

Example: The Cobb-Douglas function $Ax_1^{\alpha}x_2^{\beta}$ for $A = 0.5,\ \alpha = 2,\ \beta = 1.1$ is quasiconcave but not concave



Quasiconcavity and quasiconvexity

Caveats

- A function can be both quasiconcave and quasiconvex. (Consider $f(x) = x$.)
- As a consequence of the previous property, e.g. a quasiconvex function can be concave (consider $\ln x$ for $x \in \mathbb{R}_{++}$).
- The sum of quasiconcave functions is not necessarily quasiconcave.
- A quasiconcave (quasiconvex) function can be discontinuous.

Sufficient conditions for optimality in concave programs

By virtue of the structure of concave programs, obtaining sufficiency results for them is more direct:

Fact 6

Let the functions f and $g_i,\ i = 1, \ldots, m$, be defined and continuously differentiable on an open convex set $U \subseteq \mathbb{R}^n$. Let f be quasiconcave with non-vanishing gradient and g_i be quasiconvex on U. Suppose further that the following constraint qualification, known as *Slater's condition*, is true:
There exists a point $\mathbf{z} \in U$ such that $g_i(\mathbf{z}) < b_i,\ i = 1, \ldots, m$.
Consider the problem (7)-(8) and form the Lagrangian in the usual manner as $\mathcal{L}(\mathbf{x}) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i (g_i(\mathbf{x}) - b_i)$. If there exist a feasible point $\mathbf{x}^*$ and multipliers $\lambda_1^*, \ldots, \lambda_m^*$ such that

$$\nabla \mathcal{L}(\mathbf{x}^*) = \mathbf{0},$$

$$\lambda_i^* \geq 0 \text{ and } \lambda_i^* (g_i(\mathbf{x}^*) - b_i) = 0,\ i = 1, \ldots, m,$$

then $\mathbf{x}^*$ is a global maximum for the problem (7)-(8).

Comments

- If the functions g_i in Fact 6 are linear, then the result is valid in the absence of Slater's condition.
- The statement of Fact 6 can be weakened slightly by requiring that only the *active* constraints be quasiconvex.

Readings

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

Additional readings:
Simon and Blume. *Mathematics for economists*. Chapters 18 and 19.
Chiang and Wainwright. *Fundamental methods of mathematical economics*. Chapter 13.