# Chapter 3. Extensions and Generalizations

Xiaoxiao Hu

February 22, 2022

# 3. Extensions and Generalizations

In this Chapter, we will learn several extentions and generalizations to the two-variable, one-constraint *Lagrange's Theorem* we learned in Chapter 2, namely,

(i) Allowing for more variables and constraints;

(ii) Including non-negative variables;

(iii) Adding inequality constraints (*Kuhn-Tucker Theorem*).

# 3.A. More variables and constraints

Recall *Lagrange's Theorem* we learned in Chapter 2:

**Theorem 2.1** (Lagrange's Theorem). *Suppose $x$ is a two-dimensional vector, $c$ is a scalar, and $F$ and $G$ functions taking scalar values. Suppose $x^*$ solves the following maximization problem:*

$$\max_x F(x)$$

$$s.t. \; G(x) = c,$$

*and the constraint qualification holds, that is, if $G_j(x^*) \neq 0$ for at least one $j$.*

## More variables and constraints

**Theorem 2.1 (continued).**

*Define*

$$\mathcal{L}(x, \lambda) = F(x) + \lambda\left[c - G(x)\right]. \tag{2.10}$$

*Then there is a value of* $\lambda$ *such that*

$$\mathcal{L}_j(x^*, \lambda) = 0 \text{ for } j = 1, 2 \qquad \mathcal{L}_\lambda(x^*, \lambda) = 0. \tag{2.11}$$

## More variables and constraints

- In Theorem 2.1, $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, and we only have one constraint: $G(x) = c$.

- In this section, we will extend the theorem to $n$ choice variables $x = (x_1, x_2, ..., x_n)^T$, and $m$ constraints[1]

$$G^i(x) = c_i, \quad i = 1, 2, ..., m.$$

[1] Here, the superscript on $G(x)$ denotes the constraint number. For instance, $G^i(x) = c_i$ denotes the $i^{th}$ constraint. Please remember that we used the subscript on $G(x)$ to denote partial derivatives. For instance, $G_j(x) = \partial G(x)/\partial x_j$.

## More variables and constraints

**Assumption.** $m < n$.

This assumption is to ensure that the maximization problem is solvable and interesting.

- If $m = n$, the variables could be solved solely using constraints, and the maximization problem would become trivial.

- If $m > n$, the constraints themselves could be mutually inconsistent, leading to non-existence of solutions.

## More variables and constraints

To extend the Lagrange's method, we define $\lambda_i$ as the Lagrange multiplier for each constraint, and the we could write the Lagrangian:

$$\begin{aligned}\mathcal{L}(x_1, ..., x_n, \lambda_1, ..., \lambda_m) =& F(x_1, ..., x_n) \\ &+ \sum_{i=1}^{m} \lambda_i \left[c_i - G^i(x_1, ..., x_n)\right]. \quad (3.1)\end{aligned}$$

## More variables and constraints

First-order necessary conditions are

$$\mathcal{L}_j = \partial\mathcal{L}/\partial x_j = F_j(x_1, ..., x_n) - \sum_{i=1}^{m} \lambda_i G_j^i(x_1, ..., x_n) = 0$$

$$\text{for } j = 1, 2, ..., n; \tag{3.2}$$

$$\mathcal{L}_{\lambda_i} = \partial\mathcal{L}/\partial\lambda_i = c_i - G^i(x) = 0 \text{ for } i = 1, 2, ..., m. \tag{3.3}$$

We have $(n + m)$ equations in (3.2) and (3.3) to solve for $(n + m)$ variables $x_1^*, x_2^*, ..., x_n^*, \lambda_1, ..., \lambda_m$.

## Vector-Matrix Form

- There is nothing conceptually new.
- It is only introduced to make the equations look neat.

### Vector-Matrix Form

$$G(x) = \begin{pmatrix} G^1(x) \\ \vdots \\ G^m(x) \end{pmatrix}; \qquad c = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}; \qquad \lambda = \Big(\lambda_1, ..., \lambda_m\Big).$$

With the new notations, (3.1)

$$\mathcal{L}(x_1, ..., x_n, \lambda_1, ..., \lambda_m) = F(x_1, ..., x_n) + \sum_{i=1}^{m} \lambda_i \left[c_i - G^i(x_1, ..., x_n)\right].$$

could be written as

$$\mathcal{L}(x, \lambda) = F(x) + \lambda \left[c - G(x)\right] \tag{3.4}$$

## Vector-Matrix Form

$$F_x(x) = \Big(F_1(x), ..., F_n(x)\Big); \quad G^i_x(x) = \Big(G^i_1(x), ..., G^i_n(x)\Big);$$

$$G_x(x) = \begin{pmatrix} G^1_x(x) \\ G^2_x(x) \\ \vdots \\ G^m_x(x) \end{pmatrix} = \begin{bmatrix} G^1_1(x) & ... & G^1_n(x) \\ G^2_1(x) & ... & G^2_n(x) \\ \vdots & \ddots & \vdots \\ G^m_1(x) & ... & G^m_n(x) \end{bmatrix}.$$

We adopt the convention that when the argument of a function is a column vector, the vector of partial derivatives is a row vector, and vice versa.

## Vector-Matrix Form

First-order necessary conditions

$$\mathcal{L}_j = \partial\mathcal{L}/\partial x_j = F_j(x_1, ..., x_n) - \sum_{i=1}^{m} \lambda_i G^i_j(x_1, ..., x_n) = 0$$

$$\text{for } j = 1, 2, ..., n; \tag{3.2}$$

$$\mathcal{L}_{\lambda_i} = \partial\mathcal{L}/\partial\lambda_i = c_i - G^i(x) = 0 \text{ for } i = 1, 2, ..., m. \tag{3.3}$$

could be written as

$$\mathcal{L}_x(x^*, \lambda) = F_x(x) - \lambda G_x(x) = 0, \tag{3.5}$$

$$\mathcal{L}_\lambda(x^*, \lambda) = c - G(x) = 0. \tag{3.6}$$

## Constraint Qualification

- In Chapter 2, we have learned that for two-variable, one-constraint case, to ensure the validity of the first-order necessary conditions, we need to check *Constraint Qualification.*

- We also learned that the condition is $(G_1(x^*), G_2(x^*))$ being a non-zero vector.

## Constraint Qualification

- For n-variable, m-constraint cases, *Constraint Qualification* is also required.

- The condition is that the matrix $G_x(x^*)$ should not have any singularity.

- That is, $G_x^i(x^*)$'s should be linearly independent, or $G_x(x^*)$ should have rank $m$.[2]

[2] Formal proofs are not required and will not be discussed in this course.

## Constraint Qualification

- Again, in practice, failure of *Constraint Qualification* is rarely a problem.
- However, you should be alerted and check *Constraint Qualification* if standard methods are problematic.
- Failure of *Constraint Qualification* could usually be circumvented by writing the algebriac form of the constraints differently.[3]

[3] See Chapter 2 Section 2.C.

## Lagrange's Theorem for $n$ varaibles and $m$ constraints

**Theorem 3.1** (Lagrange's Theorem). *Suppose $x$ is a $n$-dimensional vector, $c$ an $m$-dimensional vector, $F$ a function taking scalar values, $G$ a function taking $m$-dimensional vector values, with $m < n$. Suppose $x^*$ solves the following maximization problem:*

$$\max_x F(x)$$

$$s.t. \; G(x) = c,$$

*and the constraint qualification holds, i.e., rank $G_x(x^*) = m$.*

**Theorem 3.1 (continued).**

*Define*

$$\mathcal{L}(x, \lambda) = F(x) + \lambda \left[ c - G(x) \right], \tag{3.4}$$

*where $\lambda$ is an $m$-dimensional row vector. Then there is a value of $\lambda$ such that*

$$\mathcal{L}_x(x^*, \lambda) = 0, \tag{3.5}$$

$$\mathcal{L}_\lambda(x^*, \lambda) = 0. \tag{3.6}$$

## 3.B. Non-negative variables

- Suppose that $x_j$ must be non-negative to make economic sense.

- If the optimum $x^*$ happens to be $x_j^* > 0$ for all $j$, then what we learned in Section 3.A continues to hold.

- However, if it is not true, say if $x_1^* = 0$, then only one side of the arbitrage argument would apply.

- More specifically, we can only consider infinitesimal changes $\mathrm{d}x$ for which $dx_1 > 0$.

## Non-negative variables

Therefore, when $x_1^* = 0$, Condition (3.2)

$$\mathcal{L}_1(x^*, \lambda) = F_1(x^*) - \sum_{i=1}^{m} \lambda_i G_1^i(x^*) = 0. \tag{3.2}$$

is modified as

$$\mathcal{L}_1(x^*, \lambda) = F_1(x^*) - \sum_{i=1}^{m} \lambda_i G_1^i(x^*) \leq 0. \tag{3.7}$$

## Non-negative variables

Therefore,

i. when $x_j^* > 0$, (3.2) $\mathcal{L}_j(x^*, \lambda) = 0$ holds;

ii. when $x_j^* = 0$, (3.7) $\mathcal{L}_j(x^*, \lambda) \leq 0$ holds.

In other words, for every $j$

$$\mathcal{L}_j(x^*, \lambda) \leq 0 \text{ and } x_j^* \geq 0 \tag{3.8}$$

with at least one holding with equality.[4]

[4] This qualification rules out the case when both expressions hold with inequality.

## Non-negative variables

The requirement that at least one inequalities hold with equality could be equivalently written as

$$x_j^* \mathcal{L}_j(x^*, \lambda) = 0,$$

and is called *complementary slackness*: one inequality complements the slackness in the other.[5]

[5] An inequality is called *binding* if it holds with equality; and *slack* if it holds with strict inequality.

## Vector-matrix Form

- We use vector-matrix form for simple exposition.
- Note that for a vector $x$:

  (i) $x \geq 0$ means that $x_j \geq 0$ for all $j$;

  (ii) $x > 0$ means that $x_j \geq 0$ for all $j$ and at least one of $x_j > 0$;

  (iii) $x \gg 0$ means that $x_j > 0$ for all $j$.

## Vector-matrix Form

Using the new notation, we could rewrite (3.8)

For every $j$, $\mathcal{L}_j(x^*, \lambda) \le 0$ and $x_j^* \ge 0$, with at least one holding with equality.

as follows:

$$\mathcal{L}_x(x^*, \lambda) \le 0 \text{ and } x^* \ge 0, \text{ with complementary slackness}^6 \tag{3.8}$$

[6] Compelementary slackness holds for each component pair.

## Lagrange's Theorem with Non-Negative Variables

**Theorem 3.2.** *Suppose $x$ is a $n$-dimensional vector, $c$ an $m$-dimensional vector, $F$ a function taking scalar values, $G$ a function taking $m$-dimensional vector values, with $m < n$. Suppose $x^*$ solves the following maximization problem:*

$$\max_x F(x)$$

$$s.t.\ G(x) = c \ and\ x \geq 0,$$

*and the constraint qualification holds, i.e., rank $G_x(x^*) = m$.*

# Lagrange's Theorem with Non-Negative Variables

**Theorem 3.2 (continued).**

*Define*

$$\mathcal{L}(x, \lambda) = F(x) + \lambda \left[c - G(x)\right], \tag{3.4}$$

*where $\lambda$ is an $m$-dimensional row vector. Then there is a value of $\lambda$ such that*

$$\mathcal{L}_x(x^*, \lambda) \leq 0, \;\; x^* \geq 0, \textit{ with complementary slackness}, \tag{3.8}$$

$$\mathcal{L}_\lambda(x^*, \lambda) = 0. \tag{3.6}$$

## Lagrange's Theorem with Non-Negative Variables

- Applying Theorem 3.2, one systematic way to search for an optimum is that we assume a particular pattern, say $x_1^* > 0$, $x_2^* = 0$, ..., $x_n^* > 0$.

- Then from (3.8),

$$\mathcal{L}_x(x^*, \lambda) \leq 0,\ x^* \geq 0, \text{with complementary slackness,} \quad (3.8)$$

we get $n$ equations: $\mathcal{L}_1(x^*, \lambda) = 0$, $x_2^* = 0$,..., $\mathcal{L}_n(x^*, \lambda) = 0$.

## Lagrange's Theorem with Non-Negative Variables

- Together with $m$ equations in (3.6),

$$\mathcal{L}_\lambda(x^*, \lambda) = 0, \tag{3.6}$$

  we could solve for the $n + m$ unknowns $x^*$ and $\lambda$.

- If a solution exists, and further it satisfy the other inequality conditions required from the pattern, then it is a candidate for the optimum.

## Lagrange's Theorem with Non-Negative Variables

- There are in total $2^n$ such patterns to consider.
- Therefore, to have a complete list of candidates for the optimum, we need to repeat the above algorithm $2^n$ times.
- The simplex method for solving linear programming problems is one application of the algorithm.

## Lagrange's Theorem with Non-Negative Variables

- However, in general, this algorithm is exhaustive and exhausting.

- In practice, we should use our economic intuition to make good guesses about the pattern, proceed on that basis, and use second-order sufficient conditions to verify our guesses.

## 3.C. Inequality constraints

- In this section, we introduce the inequality constraints.
- This is of considerable economic importance, since it is not always optimal to use up all the resources.

## Inequality constraints

Suppose that the first constraint holds with inequality:

$$G^1(x) \leq c_1.$$

Therefore, the problem is

$$\max_{x_1,...,x_n} F(x_1, ..., x_n)$$

$$\text{s.t. } G^1(x_1, ..., x_n) \leq c_1,$$

$$G^2(x_1, ..., x_n) = c_2, ..., G^m(x_1, ..., x_n) = c_m.$$

## Inequality constraints

- Invoking the "unspent income" argument we introduced in Chapter 1, we could define a new variable $x_{n+1}$ as follows:

$$x_{n+1} = c_1 - G^1(x). \tag{3.9}$$

- Now the constraint becomes

$$G^1(x) + x_{n+1} = c_1,$$

with the additional requirement $x_{n+1} \geq 0$.

## Inequality constraints

Thus, the maximization problem becomes

$$\max_{x_1,...,x_n,x_{n+1}} F(x_1,...,x_n)$$

$$\text{s.t. } G^1(x_1,...,x_n) + x_{n+1} = c_1 \text{ and } x_{n+1} \geq 0;$$

$$G^2(x_1,...,x_n) = c_2, ..., G^m(x_1,...,x_n) = c_m.$$

We have learned how to handle such problems in Section 3.B.

## Inequality constraints

Instead of transforming the problem and invoking Theorem 3.2 each time we saw such a problem, we want to find conditions for the maximization problems with the inequality constraints.

## Inequality constraints

Let $\widehat{\mathcal{L}}$ be the Lagrangian for the new problem with $G^1(x_1, ..., x_n) + x_{n+1} = c_1$ and $x_{n+1} \geq 0$, to distinguish from $\mathcal{L}$ of the old one with $G^1(x_1, ..., x_n) \leq c_1$. Then

$$\widehat{\mathcal{L}}(x_1, ..., x_n, x_{n+1}, \lambda_1, ..., \lambda_m)$$

$$= \mathcal{L}(x_1, ..., x_n, \lambda_1, ..., \lambda_m) - \lambda_1 x_{n+1}.$$

## Inequality constraints

Applying Theorem 3.2, we have

$$\widehat{\mathcal{L}}_j = \mathcal{L}_j = 0 \text{ for } j \neq n+1, \tag{3.10}$$

$$\widehat{\mathcal{L}}_{n+1} = -\lambda_1 \leq 0, \text{ and } x_{n+1} \geq 0, \text{ with CS}, \tag{3.11}$$

$$\widehat{\mathcal{L}}_{\lambda_1} = \mathcal{L}_{\lambda_1} - x_{n+1} = 0, \tag{3.12}$$

$$\widehat{\mathcal{L}}_{\lambda_i} = \mathcal{L}_{\lambda_i} = 0 \text{ for } i \neq 1. \tag{3.13}$$

(3.10) and (3.13) are already expressed with respect to $\mathcal{L}$, so we only need to deal with (3.11) and (3.12).

## Inequality constraints

- By (3.12), $x_{n+1} = \mathcal{L}_{\lambda_1}$.

- Plugging into (3.11), we get

$$\lambda_1 \geq 0 \text{ and } \mathcal{L}_{\lambda_1} \geq 0, \text{with CS} \tag{3.14}$$

## Inequality constraints

Therefore, the solution could be expressed in terms of $\mathcal{L}$:

$$\mathcal{L}_j = 0, \text{ for } j = 1, ..., n, \tag{3.10}$$

$$\mathcal{L}_{\lambda_1} \geq 0 \text{ and } \lambda_1 \geq 0, \text{ with CS}, \tag{3.14}$$

$$\mathcal{L}_{\lambda_i} = 0 \text{ for } i = 2, 3, ..., m. \tag{3.13}$$

## Inequality constraints

- We could extend the above reasoning to allow all constraints to be inequalities.

- Inequality constriants:

$$G^1(x) \leq c_1, G^2(x) \leq c_2, ..., G^m(x) \leq c_m.$$

- Lagrangian

$$\mathcal{L}(x, \lambda) = F(x) + \sum_{i=1}^{m} \lambda_i [c_i - G^i(x)].$$

## Inequality constraints

- If all constraints are inequality constraints, then there is no reason in restricting $m < n$, since any number of inequality constriants can still leave a non-trivial range of variation for $x$.

## Constraint qualification

- The **Constraint qualification** needs to be altered.

- We only require the matrix formed by the binding constraints to have full rank.

## Kuhn-Tucker Theorem

**Theorem 3.3.** *Suppose $x$ is a $n$-dimensional vector, $c$ an $m$-dimensional vector, $F$ a function taking scalar values, $G$ a function taking $m$-dimensional vector values, with $m < n$. Suppose $x^*$ solves the following maximization problem:*

$$\max_x F(x)$$
$$s.t.\ G(x) \leq c \text{ and } x \geq 0,$$

*and the constraint qualification holds.*

**Theorem 3.3 (continued).**

*Define*

$$\mathcal{L}(x,\lambda) = F(x) + \lambda\left[c - G(x)\right], \qquad (3.4)$$

*where* $\lambda$ *is an* $m$*-dimensional row vector. Then there is a value of* $\lambda$ *such that*

$$\mathcal{L}_x(x^*,\lambda) \leq 0,\ x^* \geq 0, \textit{with complementary slackness}, \qquad (3.8)$$

$$\mathcal{L}_\lambda(x^*,\lambda) \geq 0,\ \lambda \geq 0, \textit{with complementary slackness}. \qquad (3.15)$$

## Kuhn-Tucker Theorem

- Once again, an exhaustive procedure for finding a solution involves searching among all $2^{m+n}$ patterns from the $(m+n)$ complementary slackness conditions.

- And in practice, we should use our economic intuition to narrow down the search.

## 3.D. Examples

In this section, we will apply the *Kuhn-Tucker Theorem* in examples.

**Example 3.1: Quasi-linear Preferences.**

Suppose there are two goods $x$ and $y$, whose quantities must be non-negative, and whose prices are $p > 0$ and $q > 0$ respectively. Consider a consumer with income $I$ and the utility function.

$$U(x, y) = y + a \ln(x).$$

What is the consumer's optimal bundle $(x, y)$?

**Solution.**

See Lecture Notes.

### Example 3.2: Technological Unemployment

Suppose an economy has 300 units of labor and 450 units of land. These can be used in the production of wheat and beef. Each unit of wheat requires 2 of labor and 1 of land; each unit of beef requires 1 of labor and 2 of land.

A plan to produce $x$ units of wheat and $y$ units of beef is feasible if

$$2x + y \leq 300, \tag{3.16}$$

$$x + 2y \leq 450. \tag{3.17}$$

## Example 3.2: Technological Unemployment (continued)

Suppose the society has an objective, or social welfare function as follows:

$$W(x, y) = \alpha \ln(x) + \beta \ln(y). \qquad (3.18)$$

where $\alpha + \beta = 1$.

What is the optimal amount of wheat and beef production?

**Solution.**

See Lecture Note.