Chapter 2. Constrained Optimization*

We in this chapter study the first order necessary conditions for an optimization problem with equality and/or inequality constraints. The former is often called the **Lagrange problem** and the latter is called the **Kuhn-Tucker problem**. We will not discuss the unconstrained optimization problem separately but treat it as a special case of the constrained problem because the unconstrained problem is rare in economics. Related materials of this chapter can be found in Chapter 17-19 of Simon and Blume (1994) and Chapter 4-6 of Sundaram (1996). Other useful references include Peressini et al. (1988), Bazaraa et al. (2006), Bertsekas (2016) and Luenberger and Ye (2016).

For an optimization problem, we first define what maximum/minimum and maximizer/minimizer mean.

![Figure 1: Local and Global Maxima and Minima for \( \cos(3\pi x)/x \), 0.1 \( \leq x \leq 1.1 \)](image)

**Definition 1** A function \( f : X \to \mathbb{R} \) has a **global (or absolute) maximum point (or maximizer)** at \( x^* \) if \( f(x^*) \geq f(x) \) for all \( x \in X \) and \( x \neq x^* \). Similarly, the function has a **global minimum point** at \( x^* \) if \( f(x^*) \leq f(x) \) for all \( x \in X \) and \( x \neq x^* \). If the domain \( X \) is a metric space, usually a subset of \( \mathbb{R}^n \), then \( f \) is said to have a **local (or relative) maximum point (or maximizer)**

*Email: pingyu@hku.hk*
at the point $x^*$ if there exists $r > 0$ such that $f(x^*) \geq f(x)$ for all $x \in B_r(x^*) \cap X \setminus \{x^*\}$. Similarly, the function has a **local minimum point** at $x^*$ if $f(x^*) \leq f(x)$ for all $x \in B_r(x^*) \cap X \setminus \{x^*\}$.

**Definition 2** In both the global and local cases, the value of the function at a maximum point is called the **maximum (value)** of the function and the value of the function at a minimum point is called the **minimum (value)** of the function.

**Remark 1** The maxima and minima (the respective plurals of maximum and minimum) are called **optima** (the plural of **optimum**) or **extrema** (the plural of **extremum**), and the maximizer and minimizer are called the **optimizer**. Figure [1] shows local and global extrema. Usually, the optimizer and optimum without any qualifier means the global ones.

**Remark 2** Note that a global optimizer is always a local optimizer but the converse is not correct.

**Remark 3** In both the global and local cases, the concept of a **strict optimum** and a **strict optimizer** can be defined by replacing weak inequalities by strict inequalities. The global strict optimizer and optimum, if exist, are unique.

The problem of maximization is usually stated as

$$\max_x f(x) \quad \text{s.t. } x \in X,$$

where "s.t." is a short for "subject to" and $X$ is called the **constraint set** or **feasible set**. The maximizer is denoted as

$$\arg \max \{f(x) | x \in X\} \quad \text{or} \quad \arg \max_{x \in X} f(x),$$

where "arg" is a short for "arguments". The difference between the Lagrange problem and Kuhn-Tucker problem lies in the definition of $X$.

## 1 Equality-Constrained Optimization

### 1.1 Lagrange Multipliers

Consider the problem of a consumer who seeks to distribute his income across the purchase of the two goods that he consumes, subject to the constraint that he spends no more than his total income. Let us denote the amount of the first good that he buys $x_1$ and the amount of the second good $x_2$, the prices of the two goods $p_1$ and $p_2$, and the consumer’s income $y$. The utility that the consumer obtains from consuming $x_1$ units of good 1 and $x_2$ of good two is denoted $u(x_1, x_2)$. Thus the consumer’s problem is to maximize $u(x_1, x_2)$ subject to the constraint that $p_1 x_1 + p_2 x_2 \leq y$. (We shall soon write $p_1 x_1 + p_2 x_2 = y$, i.e., we shall assume that the consumer must spend all of his income.

\footnote{"s.t." is also a short for "such that" in some books.}
income.) Before discussing the solution of this problem let us write it in a more “mathematical” way:

\[
\max_{x_1, x_2} u(x_1, x_2) \\
\text{s.t. } p_1 x_1 + p_2 x_2 = y.
\]  

We read this “Choose \(x_1\) and \(x_2\) to maximize \(u(x_1, x_2)\) subject to the constraint that \(p_1 x_1 + p_2 x_2 = y\).”

Let us assume, as usual, that the indifference curves (i.e., the sets of points \((x_1, x_2)\) for which \(u(x_1, x_2)\) is a constant) are convex to the origin. Let us also assume that the indifference curves are nice and smooth. Then the point \((x_1^*, x_2^*)\) that solves the maximization problem \((1)\) is the point at which the indifference curve is tangent to the budget line as given in Figure 2.

One thing we can say about the solution is that at the point \((x_1^*, x_2^*)\) it must be true that the marginal utility with respect to good 1 divided by the price of good 1 must equal the marginal utility with respect to good 2 divided by the price of good 2. For if this were not true then the consumer could, by decreasing the consumption of the good for which this ratio was lower and increasing the consumption of the other good, increase his utility. Marginal utilities are, of course, just the partial derivatives of the utility function. Thus we have

\[
\frac{\partial u}{\partial x_1} (x_1^*, x_2^*) = \frac{\partial u}{\partial x_2} (x_1^*, x_2^*) \frac{p_1}{p_2}.
\]

The argument we have just made seems very “economic.” It is easy to give an alternate argument that does not explicitly refer to the economic intuition. Let \(x_2^p\) be the function that defines the
indifference curve through the point \((x_1^*, x_2^*)\), i.e.,

\[ u(x_1, x_2^u(x_1)) \equiv \bar{u} \equiv u(x_1^*, x_2^*). \]

Now, totally differentiating this identity gives

\[
\frac{\partial u}{\partial x_1}(x_1, x_2^u(x_1)) + \frac{\partial u}{\partial x_2}(x_1, x_2^u(x_1)) \frac{dx_2^u}{dx_1}(x_1) = 0.
\]

That is,

\[
\frac{dx_2^u}{dx_1}(x_1) = -\frac{\frac{\partial u}{\partial x_1}(x_1, x_2^u(x_1))}{\frac{\partial u}{\partial x_2}(x_1, x_2^u(x_1))}.
\]

Now \(x_2^u(x_1^*) = x_2^*.\) Thus the slope of the indifference curve at the point \((x_1^*, x_2^*)\)

\[
\frac{dx_2^u}{dx_1}(x_1^*) = -\frac{\frac{\partial u}{\partial x_1}(x_1^*, x_2^*)}{\frac{\partial u}{\partial x_2}(x_1^*, x_2^*)}.
\]

Also, the slope of the budget line is \(-\frac{p_1}{p_2}.\) Combining these two results again gives result (2).

Since we also have another equation that \((x_1^*, x_2^*)\) must satisfy, viz.,

\[ p_1 x_1^* + p_2 x_2^* = y, \quad (3) \]

we have two equations in two unknowns and we can (if we know what the utility function is and what \(p_1, p_2,\) and \(y\) are) go happily away and solve the problem. (This isn’t quite true but we shall not go into that at this point.) What we shall develop is a systemic and useful way to obtain the conditions (2) and (3). Let us first denote the common value of the ratios in (2) by \(\lambda.\) That is,

\[
\frac{\frac{\partial u}{\partial x_1}(x_1^*, x_2^*)}{p_1} = \lambda = \frac{\frac{\partial u}{\partial x_2}(x_1^*, x_2^*)}{p_2}
\]

and we can rewrite this and (3) as

\[
\frac{\partial u}{\partial x_1}(x_1^*, x_2^*) - \lambda p_1 = 0, \\
\frac{\partial u}{\partial x_2}(x_1^*, x_2^*) - \lambda p_2 = 0, \\
y - p_1 x_1^* - p_2 x_2^* = 0. \tag{4}
\]

Now we have three equations in \(x_1^*, x_2^*,\) and the new artificial or auxiliary variable \(\lambda.\) Again we can, perhaps, solve these equations for \(x_1^*, x_2^*,\) and \(\lambda.\) Consider the following function

\[ L(x_1, x_2, \lambda) = u(x_1, x_2) + \lambda(y - p_1 x_1 - p_2 x_2) \]

This function is known as the Lagrangian. Now, if we calculate \(\frac{\partial L}{\partial x_1}, \frac{\partial L}{\partial x_2},\) and \(\frac{\partial L}{\partial \lambda},\) and set the results equal to zero we obtain exactly the equations given in (4). We now describe this technique in a somewhat more general way.
Suppose that we have the following maximization problem

\[
\begin{align*}
\max_{x_1, \ldots, x_n} & \quad f(x_1, \ldots, x_n) \\
\text{s.t.} & \quad g(x_1, \ldots, x_n) = c,
\end{align*}
\]  

(5)

and we let

\[
\mathcal{L}(x_1, \ldots, x_n, \lambda) = f(x_1, \ldots, x_n) + \lambda(c - g(x_1, \ldots, x_n)),
\]

then if \((x_1^*, \ldots, x_n^*)\) solves (5) there is a value of \(\lambda\), say \(\lambda^*\) such that

\[
\frac{\partial \mathcal{L}}{\partial x_i}(x_1^*, \ldots, x_n^*, \lambda^*) = 0, \quad i = 1, \ldots, n,
\]

(6)

\[
\frac{\partial \mathcal{L}}{\partial \lambda}(x_1^*, \ldots, x_n^*, \lambda^*) = 0.
\]

(7)

Notice that the conditions (6) are precisely the first order conditions for choosing \(x_1, \ldots, x_n\) to maximize \(\mathcal{L}\), once \(\lambda^*\) has been chosen. This provides an intuition into this method of solving the constrained maximization problem. In the constrained problem we have told the decision maker that he must satisfy \(g(x_1, \ldots, x_n) = c\) and that he should choose among all points that satisfy this constraint the point at which \(f(x_1, \ldots, x_n)\) is greatest. We arrive at the same answer if we tell the decision maker to choose any point he wishes but that for each unit by which he violates the constraint \(g(x_1, \ldots, x_n) = c\) we shall take away \(\lambda\) units from his payoff. Of course we must be careful to choose \(\lambda\) to be the correct value. If we choose \(\lambda\) too small the decision maker may choose to violate his constraint, e.g., if we made the penalty for spending more than the consumer’s income very small the consumer would choose to consume more goods than he could afford and to pay the penalty in utility terms. On the other hand if we choose \(\lambda\) too large the decision maker may violate his constraint in the other direction, e.g., the consumer would choose not to spend any of his income and just receive \(\lambda\) units of utility for each unit of his income.

It is possible to give a more general statement of this technique, allowing for multiple constraints. Consider the problem

\[
\begin{align*}
\max_{x_1, \ldots, x_n} & \quad f(x_1, \ldots, x_n) \\
\text{s.t.} & \quad g_1(x_1, \ldots, x_n) = c_1, \\
& \quad \vdots \\
& \quad g_m(x_1, \ldots, x_n) = c_m,
\end{align*}
\]  

(8)

where \(m \leq n\), i.e., we have fewer constraints than we have variables. Again we construct the Lagrangian

\[
\mathcal{L}(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m) = f(x_1, \ldots, x_n) + \lambda_1(c_1 - g_1(x_1, \ldots, x_n)) + \cdots + \lambda_m(c_m - g_m(x_1, \ldots, x_n))
\]
and again if \( \mathbf{x}^* \equiv (x_1^*, \ldots, x_n^*)' \) solves \([8]\) there are values of \( \lambda, \) say \( \lambda_1^*, \ldots, \lambda_m^* \), such that

\[
\frac{\partial L}{\partial x_i}(x_1^*, \ldots, x_n^*, \lambda_1^*, \ldots, \lambda_m^*) = 0, \quad i = 1, \ldots, n, \\
\frac{\partial L}{\partial \lambda_j}(x_1^*, \ldots, x_n^*, \lambda_1^*, \ldots, \lambda_m^*) = 0, \quad j = 1, \ldots, m.
\]

These conditions are often labeled as "first order conditions" or "FOCs" for the corresponding maximization problem.

### 1.2 Caveats and Extensions

Notice that we have been referring to the set of conditions which a solution to the maximization problem must satisfy. (We call such conditions necessary conditions, so the FOCs usually mean the first order necessary conditions) So far we have not even claimed that there necessarily is a solution to the maximization problem. There are many examples of maximization problems which have no solution. One example of an unconstrained problem with no solution is

\[
\max_x 2x,
\]

maximizing over the choice of \( x \) the function \( 2x \). Clearly the greater we make \( x \) the greater is \( 2x \), and so, since there is no upper bound on \( x \) there is no maximum. Thus we might want to restrict maximization problems to those in which we choose \( x \) from some bounded set. Again, this is not enough. Consider the problem

\[
\max_{0 \leq x \leq 1} \frac{1}{x}.
\]

The smaller we make \( x \) the greater is \( 1/x \) and yet at zero \( 1/x \) is not even defined. We could define the function to take on some value at zero, say 7. But then the function would not be continuous. Or we could leave zero out of the feasible set for \( x \), say \( 0 < x \leq 1 \). Then the set of feasible \( x \) is not closed. Since there would obviously still be no solution to the maximization problem in these cases we shall want to restrict maximization problems to those in which we choose \( x \) to maximize some continuous function from some closed (and because of the previous example) bounded set. (Recall from the Heine-Borel Theorem that a set of numbers, or more generally a set of vectors, that is both closed and bounded is a compact set.) Is there anything else that could go wrong? No. The following result says that if the function to be maximized is continuous and the set over which we are choosing is compact, then there is a solution to the maximization problem.

**Theorem 1 (The Weierstrass Theorem)** Let \( S \) be a compact set and \( f : S \to \mathbb{R} \) be continuous. Then there is some \( x^* \) in \( S \) at which the function is maximized. More precisely, there is some \( x^* \) in \( S \) such that \( f(x^*) \geq f(x) \) for any \( x \in S \).

Notice that in defining such compact sets we typically use inequalities, such as \( x \geq 0 \). However in Section 1 we did not consider such constraints, but rather considered only equality constraints.
However, even in the example of utility maximization at the beginning of Section 1.1, there were implicitly constraints on \( x_1 \) and \( x_2 \) of the form

\[
x_1 \geq 0, \; x_2 \geq 0.
\]

A truly satisfactory treatment would make such constraints explicit. It is possible to explicitly treat the maximization problem with inequality constraints, at the price of a little additional complexity. We shall return to this question later in this chapter.

Also, notice that had we wished to solve a minimization problem we could have transformed the problem into a maximization problem by simply multiplying the objective function by \(-1\). That is, if we wish to minimize \( f(x) \) we could do so by maximizing \(-f(x)\). From the following exercise, we know that if \( x_1^*, \; x_2^* \), and \( \lambda^* \) satisfy the FOCs for \( \max_{x_1, x_2} u(x_1, x_2) \) s.t. \( p_1 x_1 + p_2 x_2 = y \), then \( x_1^*, \; x_2^* \), and \(-\lambda^* \) satisfy the FOCs for \( \min_{x_1, x_2} u(x_1, x_2) \) s.t. \( p_1 x_1 + p_2 x_2 = y \). Thus we cannot tell whether there is a maximum at \((x_1^*, \; x_2^*)\) or a minimum. This corresponds to the fact that in the case of a function of a single variable over an unconstrained domain at a maximum we require the first derivative to be zero, but that to know for sure that we have a maximum we must look at the second derivative. We shall not develop the analogous conditions for the constrained problem with many variables here. However, again, we shall return to it in the next chapter.

Finally, the unconstrained problem is a special case of the constrained problem where \( \lambda^* \) is set at zero. In other words, since no constraints exist, no penalty is imposed on constraints.

**Exercise 1** Write out the FOCs for \( \min_{x_1, x_2} u(x_1, x_2) \) s.t. \( p_1 x_1 + p_2 x_2 = y \). Show that if \( x_1^*, \; x_2^* \), and \( \lambda^* \) satisfy \( \text{(3)} \), then \( x_1^*, \; x_2^* \), and \(-\lambda^* \) satisfy the new FOCs.

## 2 Inequality-Constrained Optimization

Up until now we have been concerned with solving optimization problems where the variable could be any real number and where the constraints were of the form that some function should take on a particular value. However, for many optimization problems, and particularly in economics, the more natural formulation of the problem has some inequality restriction on the variables, such as the requirement that the amount consumed of any good should be non-negative. And it is often the case that the constraints on the problem are more naturally thought of as inequality constraints. For example, rather than thinking of the budget constraint as requiring that a consumer consume a bundle exactly equal to what he can afford the requirement perhaps should be that he consume no more than he can afford. The theory of such optimization problems is called **nonlinear programming**. We give here an introduction to the most basic elements of this theory.

### 2.1 Kuhn-Tucker Conditions

We start by considering one of the very simplest optimization problems, namely the maximization of a continuously differentiable function of a single real variable:
\[ \max_x f(x). \] (9)

We read this as “Choose \( x \) to maximize \( f(x) \).” As mentioned at the end of Section 1.2, the first order necessary condition for \( x^* \) to be a solution to this maximization problem is that
\[ \frac{df}{dx}(x^*) = 0. \]

Suppose now we add the constraint that \( x \geq 0 \). How does this change the problem and its solution? We can write the problem as follows:
\[ \max_x f(x) \quad \text{s.t. } x \geq 0. \] (10)

We read this as “Choose \( x \) to maximize \( f(x) \) subject to the constraint that \( x \geq 0 \).”

How does it change the conditions for \( x^* \) to be a solution? Well, if there is a solution to the maximization problem, there are two possibilities. Either the solution could occur when \( x^* > 0 \) or it could occur when \( x^* = 0 \). In the first case then \( x^* \) will also be (at least a local) maximum of the unconstrained problem and a necessary condition for this is that \( \frac{df}{dx}(x^*) = 0 \). In the second case it is not necessary that \( \frac{df}{dx}(x^*) = 0 \) since we are in any case not permitted to decrease \( x^* \). (It’s already 0 which is as low as it can go.) However, it should not be the case that we increase the value of \( f(x) \) when we increase \( x \) from \( x^* \). A necessary condition for this is that
\[ \frac{df}{dx}(x^*) \leq 0. \]

Figure 3 illustrates the intuition why \( \frac{df}{dx}(x^*) \leq 0 \) when \( x^* = 0 \). The solution \( x^* = 0 \) is called a corner solution.

So, our FOCs for a solution to this maximization problem are that \( \frac{df}{dx}(x^*) \leq 0 \), that \( x^* \geq 0 \) and that either \( x^* = 0 \) or that \( \frac{df}{dx}(x^*) = 0 \). We can express this last by the equivalent requirement that \( x^* \frac{df}{dx}(x^*) = 0 \), that is that the product of the two is equal to zero. We state this as a (very small and trivial) theorem.

**Theorem 2** Suppose that \( f : \mathbb{R} \to \mathbb{R} \) is continuously differentiable (or a \( C^1 \) function). Then, if \( x^* \) maximizes \( f(x) \) over all \( x \geq 0 \), \( x^* \) satisfies
\[ \frac{df}{dx}(x^*) \leq 0, \]
\[ x^* \frac{df}{dx}(x^*) = 0, \]
\[ x^* \geq 0. \] (11)

**Exercise 2** Give a version of Theorem 2 for minimization problems.
Figure 3: Illustration of Why $\frac{df}{dx}(x^*) \leq 0$ When $x^* = 0$

How to understand these FOCs? If we form the Lagrangian
\[ \mathcal{L}(x, \lambda) = f(x) + \lambda x, \]
then we can express these FOCs as
\[ \frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda^*) = \frac{df}{dx}(x^*) + \lambda^* = 0, \]
\[ \frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*) = x^* \geq 0, \]
\[ \lambda^* \geq 0, \lambda^* \frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*) = \lambda^* x^* = 0. \]

For the general inequality-constrained problem,
\[ \max_{x_1, \ldots, x_n} f(x_1, \cdots, x_n) \]
\[ \text{s.t. } g_j(x_1, \cdots, x_n) \geq 0, \text{ } j = 1, \cdots, J, \]
or more compactly,
\[ \max_x f(x) \]
\[ \text{s.t. } g(x) \geq 0, \]
we can form the Lagrangian
\[ \mathcal{L}(x, \lambda) = f(x) + \lambda \cdot g(x). \]
and express the FOCs as
\[
\frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda^*) = \frac{\partial f}{\partial x}(x^*) + \frac{\partial g(x^*)}{\partial x} \lambda^* = 0,
\]
\[
\frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*) = g(x^*) \geq 0,
\]
\[
\lambda^* \geq 0, \lambda^* \odot \frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*) = \lambda^* \odot g(x^*) = 0,
\]

where \(\odot\) is the element-by-element product. These FOCs are called the **Kuhn-Tucker conditions** due to Kuhn and Tucker.\(^2\)\(^3\)

We summarize the discussions on the equality-constrained and inequality-constrained problem in the following theorem. First, define the **mixed constrained problem** as follows,

\[
\max_{x_1, \ldots, x_n} f(x_1, \ldots, x_n)
\begin{align*}
\text{s.t. } & g_j(x_1, \ldots, x_n) \geq 0, \; j = 1, \ldots, J, \\
& h_k(x_1, \ldots, x_n) = 0, \; k = 1, \ldots, K,
\end{align*}
\]

or more compactly,

\[
\max_x f(x) \\
\text{s.t. } g(x) \geq 0, \\
\quad \quad \quad \quad \quad \quad h(x) = 0,
\]

where \(K \leq n\). The term "mixed constrained problem" is only for convenience because any equality constraint can be transformed to two inequality constraints, e.g., \(h_k(x) = 0\) is equivalent to \(h_k(x) \geq 0\) and \(h_k(x) \leq 0\). Form the Lagrangian

\[
\mathcal{L}(x, \lambda, \mu) = f(x) + \lambda \cdot g(x) + \mu \cdot h(x).
\]

**Theorem 3 (Theorem of Kuhn-Tucker)** Suppose that \(f : \mathbb{R}^n \to \mathbb{R}, \ g : \mathbb{R}^n \to \mathbb{R}^J\) and \(h : \mathbb{R}^n \to \mathbb{R}^K\) are \(C^1\) functions. Then, if \(x^*\) maximizes \(f(x)\) over all \(x\) satisfying the constraints \(g(x) \geq 0\) and \(h(x) = 0\), and if \(x^*\) satisfies the **nondegenerate constraint qualification** (NDCQ) as will be specified in the next subsection, then there exists a vector \((\lambda^*, \mu^*)\) such that \((x^*, \lambda^*, \mu^*)\) satisfies the Kuhn-Tucker conditions given as follows:

\[
\begin{align*}
\frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda^*, \mu^*) &= 0, \quad \frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*, \mu^*) = h(x^*) = 0, \\
\frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*, \mu^*) &= g(x^*) \geq 0, \quad \lambda^* \geq 0, \\
\lambda^* \odot \frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*, \mu^*) &= \lambda^* \odot g(x^*) = 0.
\end{align*}
\]

**Remark 4** Rigorously speaking, the Kuhn-Tucker conditions are necessary conditions for "local"

---

\(^2\) Albert W. Tucker (1905-1995) is the supervisor of John Nash, the Nobel Prize winner in Economics in 1994, and Lloyd Shapley, the Nobel Prize winner in Economics in 2012.

\(^3\) The Kuhn-Tucker conditions were originally named after Harold W. Kuhn, and Albert W. Tucker. Later scholars discovered that the necessary conditions for this problem had been stated by William Karush in his master’s thesis in 1939, so this group of conditions is also labeled as the Karush-Kuhn-Tucker (KKT) conditions.
optima, and of course are also necessary conditions for global optima.

The \( x^* \)'s that satisfy the Kuhn-Tucker conditions are called the \textbf{critical points} of \( \mathcal{L} \). Usually, critical points mean the points that satisfy the FOCs; the Kuhn-Tucker conditions are a special group of FOCs. Parallel to Lagrange multipliers in the Lagrange problem, \((\lambda^*, \mu^*)\) are called \textbf{Kuhn-Tucker multipliers}.

**Exercise 3** Write out the Kuhn-Tucker conditions given in (12) in the long form without using vector and matrix notation.

**Exercise 4** Give a version of the Kuhn-Tucker conditions such as in (12) for a constrained minimization problem.

**Exercise 5** Show that the Kuhn-Tucker conditions for the maximization problem, \( \max_x f(x) \text{ s.t. } x \geq 0 \text{ and } g(x) \leq 0 \), can be expressed as

\[
\begin{align*}
\frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda^*) &\leq 0, \\
x^* \odot \frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda^*) &= 0, \\
x^* &\geq 0, \\
\frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*) &\geq 0, \\
\lambda^* \odot \frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*) &= 0, \\
\lambda^* &\geq 0,
\end{align*}
\]

where the Lagrangian is defined as

\[
\mathcal{L}(x, \lambda) = f(x) - \lambda \cdot g(x).
\]

2.2 The Constraint Qualification

We in this subsection describe the NDCQ in the Theorem of Kuhn-Tucker. A constraint \( g_j(x) \geq 0 \) is \textbf{binding} (or \textbf{effective}) at \( x^* \) if \( g_j(x^*) = 0 \). Suppose the first \( J_0 \) inequality constraints are binding at \( x^* \); then the NDCQ states that the rank at \( x^* \) of the Jacobian matrix of the equality constraints and the binding inequality constraints

\[
\begin{pmatrix}
\frac{\partial g_{j1}}{\partial x_1}(x^*) & \cdots & \frac{\partial g_{j1}}{\partial x_n}(x^*) \\
\vdots & \ddots & \vdots \\
\frac{\partial g_{j0}}{\partial x_1}(x^*) & \cdots & \frac{\partial g_{j0}}{\partial x_n}(x^*) \\
\frac{\partial h_{11}}{\partial x_1}(x^*) & \cdots & \frac{\partial h_{11}}{\partial x_n}(x^*) \\
\vdots & \ddots & \vdots \\
\frac{\partial h_{K1}}{\partial x_1}(x^*) & \cdots & \frac{\partial h_{K1}}{\partial x_n}(x^*) \\
\end{pmatrix}
\]
is \( J_0 + K \) - as large as it can be. When for some \( x \)'s the NQCD does not hold, compare the values of \( f(\cdot) \) at critical points and also these \( x \)'s to determine the ultimate maximizer.

**Exercise 6** Write out the Kuhn-Tucker conditions and the NDCQ in the unconstrained problem and the equality-constrained problem.

The following example provides a case where the constraint qualification fails.

**Example 1** This example follows Example 19.9 of Simon and Blume (1994). We want to maximize 
\[
f(x_1, x_2) = x_1 \text{ s.t. } g(x_1, x_2) = x_1^3 + x_2^2 \leq 0.
\]
From Figure 4, the constraint set is a cusp and it is easy to see that \((x_1^*, x_2^*) = (0, 0)\). However, at \((x_1^*, x_2^*)\), there is no \( \lambda^* \) satisfying the Kuhn-Tucker conditions. To see why, set the Lagrangian
\[
L(x, \lambda) = x_1 - \lambda (x_1^3 + x_2^2),
\]
and then the Kuhn-Tucker conditions are
\[
\begin{align*}
1 - 3\lambda x_1^2 &= 0, 2\lambda x_2 = 0, \\
x_1^3 + x_2^2 &\leq 0, \lambda \geq 0, \lambda (x_1^3 + x_2^2) = 0.
\end{align*}
\]
It is not hard to see that there is no \( \lambda^* \) satisfying these conditions when \((x_1^*, x_2^*) = (0, 0)\).

What can we learn from this example? Note that \( g(x_1, x_2) \) is binding at \((0, 0)\), while \((0, 0)\) is the critical point of \( g(x_1, x_2) \) (i.e., \( \frac{\partial g}{\partial x_1}(0, 0) = \frac{\partial g}{\partial x_2}(0, 0) = 0 \)), so the constraint qualification fails. If we compare \( f(\cdot) \) at the critical values of \( L \) (which is empty) and \((0, 0)\), we indeed get the correct maximizer \((0, 0)\). □

(***) There are alternative constraint qualifications (CQs). We list some of them for a taste. To ease our statement, consider the inequality-constrained problem,
\[
\begin{align*}
\max_x f(x) \\
\text{s.t. } g(x) &\geq 0
\end{align*}
\]
(13)

because as mentioned above, an equality constraint can be expressed as two inequality constraints. Suppose the first \( J_0 \) constraints are binding.

(a) **(Karush-Kuhn-Tucker CQ)** For any \( v \in \mathbb{R}^n \) with \( Dg_j(x^*)v \geq 0, j = 1, \ldots, J_0 \), there exists \( \epsilon > 0 \) and a \( C^1 \) curve \( \alpha : [0, \epsilon) \to \mathbb{R}^n \) such that (i) \( \alpha(0) = x^* \), (ii) \( \alpha'(0) = v \), and (iii) \( g_j(\alpha(t)) \geq 0 \) for \( j = 1, \ldots, J \) and \( t \in [0, \epsilon) \).

(b) **(Slater CQ)** There is a ball \( B \) about \( x^* \) in \( \mathbb{R}^n \) such that \( g_1, \ldots, g_{J_0} \) are concave functions on \( B \) and there exists some \( x \) such that \( g_j(x) > 0, j = 1, \ldots, J \).

(c) \( g_1, \ldots, g_{J_0} \) are concave functions.
(d) $g_1, \cdots, g_J$ are linear functions.

The first CQ is designed to rule out cusps in the constraint set, like the one in the above example. The condition that there exists some $\mathbf{x}$ such that $g_j(\mathbf{x}) > 0$, $j = 1, \cdots, J$, in Slater CQ is called **Slater’s condition** due to Slater (1950), which requires that the constraint set have a nonempty interior (which excludes equality constraints). The Slater CQ also requires that the binding constraint functions be concave functions which will be discussed in the next chapter. The last three CQs are particularly helpful in economic applications, e.g., the budget set in the utility maximization problem of consumer theory $\{(x_1, x_1) \mid p_1x_1 + p_2x_2 \leq y, x_1 \geq 0, x_2 \geq 0\}$ satisfies any of these CQs.

If we want to avoid the CQ completely, then we need to introduce one more multiplier - a multiplier for the objective function itself. This result is known as the **Fritz John Theorem**.

**Theorem 4 (Fritz John Theorem)** Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^J$ are $C^1$ functions in $\{13\}$. Form the Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_0, \lambda) = \lambda_0 f(\mathbf{x}) + \lambda \cdot g(\mathbf{x}).$$

If $\mathbf{x}^*$ is a local maximizer of $\{13\}$, then there exist multipliers $(\lambda_0^*, \lambda^*)$ such that

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}^*, \lambda_0^*, \lambda^*) = 0, \quad \frac{\partial \mathcal{L}}{\partial \lambda_0}(\mathbf{x}^*, \lambda_0^*, \lambda^*) = g(\mathbf{x}^*) \geq 0, \quad \lambda^* \geq 0, \quad \lambda^* \circ \frac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}^*, \lambda_0^*, \lambda^*) = \lambda_0^* \circ g(\mathbf{x}^*) = 0, \quad \lambda_0^* = 0 \text{ or } 1, \text{ and } (\lambda_0^*, \lambda^*) \neq 0.$$

CQs guarantee that $\lambda_0^* = 1$. (***)
We use an example to illustrate how to find the maximizer in practice. A caution here is that "Never blindly apply the Kuhn-Tucker conditions".

**Example 2** \[
\text{max } x_1^2 + (x_2 - 5)^2 \quad \text{s.t. } x_1 \geq 0, \ x_2 \geq 0, \text{ and } 2x_1 + x_2 \leq 4.
\]

**Solution:** First, since the objective function is continuous and the constraint set is compact (why?), by the Weierstrass theorem, the maximizer exists. We then check the NDCQ. \( g_1(x) = x_1, \ g_2(x) = x_2 \) and \( g_3(x) = 4 - 2x_1 - x_2 \), so the Jacobian of the constraint functions is

\[
\begin{pmatrix}
1 & 0 \\
0 & 1 \\
-2 & -1
\end{pmatrix},
\]

whose any one or two rows are linearly independent. Since at most two of the three constraints can be binding at any one time, the NDCQ holds at any solution candidate.

The Lagrangian is

\[
\mathcal{L}(x, \lambda, \mu) = x_1^2 + (x_2 - 5)^2 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 (4 - 2x_1 - x_2),
\]

and the Kuhn-Tucker conditions are

\[
2x_1 + \lambda_1 - 2\lambda_3 = 0, \ 2(x_2 - 5) + \lambda_2 - \lambda_3 = 0, \\
x_1 \geq 0, \ x_2 \geq 0, \ 4 - 2x_1 - x_2 \geq 0, \ \lambda_1 \geq 0, \ \lambda_2 \geq 0, \ \lambda_3 \geq 0, \\
\lambda_1 x_1 = 0, \ \lambda_2 x_2 = 0, \ \lambda_3 (4 - 2x_1 - x_2) = 0.
\]

Totally eight possibilities depending whether \( \lambda_j = 0 \) or not, \( j = 1, 2, 3 \).

(i) \( \lambda_1 > 0, \lambda_2 > 0 \) and \( \lambda_3 > 0 \) \( \implies \) \( x_1 = 0, x_2 = 0, \) and \( 2x_1 + x_2 = 4 \). Impossible.

(ii) \( \lambda_1 = 0, \lambda_2 > 0 \) and \( \lambda_3 > 0 \) \( \implies \) \( x_1 \geq 0, x_2 = 0, \) and \( 2x_1 + x_2 = 4 \). So \( (x_1, x_2) = (2, 0) \). From \( 4 - 2\lambda_3 = 0 \) and \( -10 + \lambda_2 - \lambda_3 = 0 \), we have \( (\lambda_1, \lambda_2, \lambda_3) = (0, 12, 2) \).

(iii) \( \lambda_1 > 0, \lambda_2 = 0 \) and \( \lambda_3 > 0 \) \( \implies \) \( x_1 = 0, x_2 \geq 0, \) and \( 2x_1 + x_2 = 4 \). So \( (x_1, x_2) = (0, 4) \). From \( \lambda_1 - 2\lambda_3 = 0 \) and \( -2 - \lambda_3 = 0 \), we have \( (\lambda_1, \lambda_2, \lambda_3) = (-2, 0, -4) \). Impossible.

(iv) \( \lambda_1 = \lambda_2 = 0 \) and \( \lambda_3 > 0 \) \( \implies \) \( x_1 \geq 0, x_2 \geq 0, \) and \( 2x_1 + x_2 = 4 \). So from \( 2x_1 - 2\lambda_3 = 0, 2(x_2 - 5) - \lambda_3 = 0, \) and \( 2x_1 + x_2 = 4 \), we have \( (x_1, x_2) = (-2/5, 24/5) \). Impossible.

(v) \( \lambda_1 > 0, \lambda_2 > 0 \) and \( \lambda_3 = 0 \) \( \implies \) \( x_1 = 0, x_2 = 0, \) and \( 2x_1 + x_2 \leq 4 \). So \( (x_1, x_2) = (0, 0) \). From \( \lambda_1 = 0 \) and \( -10 + \lambda_2 = 0 \), we have \( (\lambda_1, \lambda_2, \lambda_3) = (0, 10, 0) \). Impossible.

(vi) \( \lambda_1 = \lambda_3 = 0, \lambda_2 > 0 \) \( \implies \) \( x_1 \geq 0, x_2 = 0, \) and \( 2x_1 + x_2 \leq 4 \). From \( 2x_1 = 0 \) and \( -10 + \lambda_2 = 0 \), we have \( (x_1, x_2) = (0, 0) \) and \( (\lambda_1, \lambda_2, \lambda_3) = (0, 10, 0) \).

(vii) \( \lambda_1 > 0, \lambda_2 = \lambda_3 = 0 \) \( \implies \) \( x_1 = 0, x_2 \geq 0, \) and \( 2x_1 + x_2 \leq 4 \). So from \( \lambda_1 = 0 \) and \( 2(x_2 - 5) = 0 \), we have \( (x_1, x_2) = (0, 5) \) and \( (\lambda_1, \lambda_2, \lambda_3) = (0, 0, 0) \). Impossible.

(viii) \( \lambda_1 = \lambda_2 = \lambda_3 = 0 \) \( \implies \) \( x_1 \geq 0, x_2 \geq 0, \) and \( 2x_1 + x_2 \leq 4 \). So from \( 2x_1 = 0 \) and \( 2(x_2 - 5) = 0 \), we have \( (x_1, x_2) = (0, 5) \). Impossible.
Candidate maximizers are \((2, 0)\) and \((0, 0)\). The objective function values at these two candidates are 29 and 25, so \((2, 0)\) is the maximizer and the associated Lagrange multipliers are \((0, 12, 2)\).

From this above example and discussions in this chapter, we summarize a "cookbook" procedure for a constrained optimization problem. First, define the feasible set of the general mixed constrained maximization problem as

\[
G = \{ x \in \mathbb{R}^n | g(x) \geq 0, h(x) = 0 \}.
\]

**Step 1:** Apply the Weierstrass theorem to show that the maximum exists. If the feasible set \(G\) is compact, this is usually straightforward; if \(G\) is not compact, truncate \(G\) to a compact set, say \(G_o\), such that there is a point \(x_o \in G_o\) and \(f(x_o) > f(x)\) for all \(x \in G \setminus G_o\).

**Step 2:** Check whether the constraint qualification is satisfied. If not, denote the set of possible violation points as \(Q\).

**Step 3:** Set up the Lagrangian and find the critical points. Denote the set of critical points as \(R\).

**Step 4:** Check the value of \(f\) on \(Q \cup R\) to determine the maximizer or maximizers.

It is quite often for practitioners to apply Step 3 directly to find the maximizer. Although this may work in most cases, it is possible to fail in some cases which do not seem bizarre at all. First, the Lagrangian may fail to have any critical points due to nonexistence of maximizers or failure of constraint qualification. Second, even if the Lagrangian does have one or more critical points, this set of critical points need not contain the solution still due to these two reasons. Let us repeat our caveat, "Never blindly apply the Kuhn-Tucker conditions"!
This cookbook procedure works well in most cases, especially when the set $Q \cup R$ is small, e.g., $Q \cup R$ includes only a few points. If this set is large, it is better to employ more necessary conditions (e.g., the second order conditions (SOCs)) to screen the points in $Q \cup R$. Another solution is to employ sufficient conditions, i.e., as long as $x^*$ satisfies these conditions, it must be the maximizer. Sufficient conditions are very powerful especially combined with the uniqueness result because as long as we find one solution, it is the solution and we can stop. These topics are the main theme of the next chapter.

3 Applications to Macroeconomic Theory (*)

In this section we examine one part of one of the central models used in macroeconomics, the overlapping generations model. Related materials can be found in Chapters 2-3 of Auerbach and Kotlikoff (1995).

3.1 Overview of The Basic Overlapping Generations Model of the Economy

The model we consider in this section is one of the basic ones used in modern macroeconomics. The idea of using such a framework was introduced by Maurice Allais\(^4\) (1947) and Paul Samuelson\(^5\) (1958). It was developed in more or less the form we shall give by Peter Diamond\(^6\) (1965). This model forms an important part of Kydland and Prescott’s 1988 paper called “Time to Build and Aggregated Fluctuations,” one of the seminal works in the real business cycle theories.\(^7\)

The basic idea of the model is that in each period a new generation of economic agents is born. Each individual lives for two periods, inelastically\(^8\) supplying one unit of labor in his first period of life and consuming in both periods of his life.

Each agent from generation $t$ supplies one unit of labor to the production process at the beginning of his first period of life in order to produce period $t$’s output of the single good of this economy. For this he receives whatever the wage rate is in period $t$. Part of this he consumes and part he saves. The part he saves becomes the capital used in the production process in period $t + 1$. He provides the part of his wage that he didn’t consume in period $t$ at the beginning of period $t + 1$ in return for a greater amount of the good for him to consume at the end of period $t + 1$. The extra bit that they get we call the interest rate for period $t + 1$.

---

\(^4\) Maurice F.C. Allais (1911-2010)) received the Nobel Prize in Economics in 1988.

\(^5\) Paul A. Samuelson (1915-2009) received the Nobel Prize in Economics in 1970.

\(^6\) Peter A. Diamond (1940-) received the Nobel prize in Economics in 2010.

\(^7\) Finn E. Kydland (1943-) and Edward C. Prescott (1940-) received the Nobel Prize in Economics in 2004.

\(^8\) This means that his supply of labor is fixed; it does not depend on the price system, and in particular does not depend on the price of labor, that is, the wage rate.
3.2 The Consumption-Saving Decision

3.2.1 The Intertemporal Budget Constraint

Recall from your microeconomics the idea of a budget constraint. This is the set of consumption bundles that are affordable for the consumer. This set typically depends on the consumer’s wealth (sometimes we say income, but wealth is more accurate) and the prices of the various consumption goods. A similar situation arises here.

Though it isn’t really very complicated we’ll develop the intertemporal budget constraint a step at a time. Let us consider the generation born in period \( t \). We have said that when the individuals in this generation are young they supply one unit of labor for which they receive the wage of that period, \( w_t \). This amount they can either consume in period \( t \) or take over into period \( t + 1 \) as assets. We’ll call the consumption in period \( t \) of such an individual \( c_{1t} \) (1 for the first period of this individual’s life and \( t \) for the time period) and the amount of assets this individual takes over into period \( t + 1 \) we’ll call \( s_t \). Both \( c_{1t} \) and \( s_t \) are measured in units of the one good produced in period \( t \). We’ll assume that the wages are also measured in units of this good so we have

\[
s_t = w_t - c_{1t}. \tag{14}
\]

In period \( t + 1 \) this individual will be old and will not supply any labor. He will supply his assets, \( s_t \) to the production process and will receive, in return, at the end of the period \( s_t(1 + r_{t+1}) \) which he will consume. Thus there is no depreciation of capital in this model. For every unit of the (capital) good individual supplies the production process, he is paid \( r_{t+1} \), and also receives back the full unit of the good after the production process is completed. Sadly, this individual will then die. We’ll call the amount he consumes in period \( t + 1 \) as an old person \( c_{2t+1} \) (2 for the second period of this individual’s life and \( t + 1 \) for the period in which he consumes). Again, we can write this as a budget constraint, namely

\[
c_{2t+1} = s_t(1 + r_{t+1}). \tag{15}
\]

Of course, what the consumer actually cares about is the amount he consumes is each period. He doesn’t directly care about the amount of assets he takes over from period \( t \) to period \( 1 + 1 \). If we combine equations (14) and (15) by substituting equation (14)’s formula for \( s_t \) into equation (15) we obtain a budget constraint that relates the two quantities that the individual does directly care about, his consumption in each period, namely

\[
c_{1t} + \frac{c_{2t+1}}{1 + r_{t+1}} = w_t. \tag{16}
\]

The term \( 1/(1 + r_{t+1}) \) is the price of period \( t + 1 \) consumption in terms of period \( t \) consumption, it is the number of units of period \( t \) consumption that the individual would have to give up in order to obtain an additional unit of period \( t + 1 \) consumption, just as a price of a good in dollars represents the number of dollars one would give up in order to obtain an additional unit of the good.
3.2.2 Intertemporal Preferences

So we have one half of the information we need to determine how the individuals of this generation will behave. We have said what choices they have available to them. We also need to say what they like or what they want, that is, we have to describe their preferences. As usual we do this by specifying a utility function. In particular we assume that the individuals of generation $t$ have the utility function

$$U_t(c_{1t}, c_{2t+1}) = u(c_{1t}) + \left(\frac{1}{1 + \theta}\right) u(c_{2t+1}),$$

(17)

with $\theta > 0$. The parameter $\theta$ is called the discount factor and indicates how the individual trades off consumption for present consumption. If $\theta = 0$ the individual values future consumption as much as present consumption. As $\theta$ increases the individual values future less. Notice that we have labeled the utility function with the subscript $t$ to indicate that it is generation $t$ preferences we are describing. However, we have assumed that each generation has the same preferences - there are no time subscripts on the single period utility functions $u$ on the right hand side of the equation.

The individuals of generation $t$ may have a different level of utility than another generation, but only because their levels of consumption may differ. We assume that the functional relationship describing how the level of the individual’s utility depends on $c_{1t}$ and $c_{2t+1}$ is the same for each generation, that is we assume that each generation has the same intertemporal preferences.

In fact, we shall make a much stronger assumption about the individuals’ preferences. In this course we shall always assume that the individual’s utility functions come from the family of CES (constant elasticity of substitution) utility functions in which $u(c) = c^{1-\sigma}/(1 - \sigma)$. That is, we assume that

$$U_t(c_{1t}, c_{2t+1}) = \frac{c_{1t}^{1-\sigma}}{1 - \sigma} + \left(\frac{1}{1 + \theta}\right) \frac{c_{2t+1}^{1-\sigma}}{1 - \sigma}.$$  

(18)

A special case of these preferences occurs when $\sigma = 1$. In fact as $\sigma$ approaches 1 this utility function diverges to infinity. However, adding a constant to the utility that does not depend of $c_{1t}$ or $c_{2t+1}$ does not change the preferences. So consider the utility function

$$U_t(c_{1t}, c_{2t+1}) = \frac{c_{1t}^{1-\sigma} - 1}{1 - \sigma} + \left(\frac{1}{1 + \theta}\right) \frac{c_{2t+1}^{1-\sigma} - 1}{1 - \sigma}.$$  

(19)

This represents the same preferences as the utility function given in equation (18). If we consider the behaviour of this function as $\sigma \to 1$ we can show by L’Hôital’s Rule that

$$\lim_{\sigma \to 1} \frac{c_{1t}^{1-\sigma} - 1}{1 - \sigma} + \left(\frac{1}{1 + \theta}\right) \frac{c_{2t+1}^{1-\sigma} - 1}{1 - \sigma} = \log c_{1t} + \left(\frac{1}{1 + \theta}\right) \log c_{2t+1}.$$  

(20)

It is also worth noting that this represents the same preferences as the utility function

$$U_t(c_{1t}, c_{2t+1}) = c_{1t}^{\frac{1-\sigma}{1+\theta}} c_{2t+1}^{\frac{1}{1+\theta}}.$$  

(21)

\footnote{Sometimes, $\frac{1}{1+\theta}$ is called the discount factor.}
If we take the log of this utility function we obtain a constant times the utility function given in equation (20). We shall refer to an individual whose utility function is of the form of either equation (21) or (20) as having Cobb-Douglas preferences. Such an assumption on preferences is very strong and for many purposes we would not want to make it. However, as we shall see, it does make our model much easier to solve, and so, we shall, at least as a first cut, be happy to make this assumption.

3.2.3 The Solution to the Intertemporal Decision Problem

We can state the decision problem of an individual in the generation born in period \( t \), as the following constrained maximization problem:

\[
\max_{c_{1t}, c_{2t+1}} \log c_{1t} + \left( \frac{1}{1 + \theta} \right) \log c_{2t+1} \\
\text{s.t. } c_{1t} + \frac{c_{2t+1}}{1 + r_{t+1}} = w_t.
\]

In the next subsection we shall examine in detail how to solve this maximization problem. Here we simply note a general useful fact about Cobb-Douglas preferences, namely that in a standard utility maximization subject to a linear budget constraint problem if the preferences are Cobb-Douglas, then, whatever the relative prices, the consumer will spend a fraction of his income on a particular good equal to the relative size of the exponent on that good in the Cobb-Douglas utility function. More precisely, the fraction is the exponent divided by the sum of the exponents. In our case this means that the individual will spend a fraction \((1 + \theta)/(2 + \theta)\) of \( w_t \) on his consumption in his first period of life, that is, since \( w_t \) and \( c_{1t} \) are measured in the same units (or equivalently the price of \( c_{1t} \) is 1)

\[
c_{1t} = \left( \frac{1 + \theta}{2 + \theta} \right) w_t.
\]

Similarly he will spend the fraction \(1/(2 + \theta)\) of \( w_t \) on \( c_{2t+1} \). Now, if we look at the budget constraint we see that the price of \( c_{2t+1} \) is not 1, but \(1/(1 + r_{t+1})\). Thus

\[
\frac{c_{2t+1}}{1 + r_{t+1}} = \left( \frac{1}{2 + \theta} \right) w_t
\]

or

\[
c_{2t+1} = \left( \frac{1 + r_{t+1}}{2 + \theta} \right) w_t.
\]

We can also think what this means in terms of the two step constraints we originally used to describe the individual’s maximization problem. In the first period the individual supplies 1 unit of labor and receives \( w_t \). Of this he consumes in that period \( c_{1t} = (1 + \theta)/(2 + \theta)w_t \), and saves the rest, that amount becoming the assets he takes into period \( t + 1 \),

\[
s_t = \left( \frac{1}{2 + \theta} \right) w_t. \quad (22)
\]
He provides this amount to the production process in period $t+1$ and receives in return $(1+r_{t+1})s_t = (1 + r_{t+1})w_t/(2 + \theta)$, which he consumes at the end of period $t+1$.

### 3.2.4 A Detailed Look at the Maximization Problem

The Lagrangian function is

$$\mathcal{L}(c_{1t}, c_{2t+1}, \lambda) = \log c_{1t} + \left(\frac{1}{1 + \theta}\right) \log c_{2t+1} + \lambda \left(w_t - c_{1t} - \frac{c_{2t+1}}{1 + r_{t+1}}\right).$$

and the FOCs are

$$\frac{\partial \mathcal{L}}{\partial c_{1t}}(c_{1t}, c_{2t+1}, \lambda) = \frac{1}{c_{1t}} - \lambda = 0, \quad (23)$$

$$\frac{\partial \mathcal{L}}{\partial c_{2t+1}}(c_{1t}, c_{2t+1}, \lambda) = \left(\frac{1}{1 + \theta}\right) \frac{1}{c_{2t+1}} - \frac{\lambda}{1 + r_{t+1}} = 0, \quad (24)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda}(c_{1t}, c_{2t+1}, \lambda) = w_t - c_{1t} - \frac{c_{2t+1}}{1 + r_{t+1}} = 0. \quad (25)$$

If we solve equation (23) for $c_{1t}$ we obtain

$$c_{1t} = \frac{1}{\lambda}. \quad (26)$$

And if we solve equation (24) for $c_{2t+1}$ we obtain

$$c_{2t+1} = \frac{1 + r_{t+1}}{\lambda(1 + \theta)}. \quad (27)$$

And substituting (26) and (27) into equation (25) (and simplifying a bit) gives

$$\frac{1}{\lambda} = \left(\frac{1 + \theta}{2 + \theta}\right) w_t, \quad (28)$$

and substituting this back into equations (26) and (27) gives

$$c_{1t} = \left(\frac{1 + \theta}{2 + \theta}\right) w_t, \quad (29)$$

$$c_{2t+1} = \left(\frac{1 + r_{t+1}}{2 + \theta}\right) w_t. \quad (30)$$

**Exercise 7** Another way of solving this problem is to solve the budget constraint it gives $c_{2t+1}$ as a function of $c_{1t}$ (and $w_t$ and $r_{t+1}$) and to substitute this into the utility function giving a function that depends on $c_{1t}$, but not on $c_{2t+1}$. We can then find the solution by finding the unconstrained maximum of this function. Solve the problem in this way and confirm that the solution is the same as the one found above.
Exercise 8  Solve the maximization problem using the utility function

\[ \tilde{u}_t(c_{1t}, c_{2t+1}) = c_{1t} \left( \frac{1+\theta}{1+\theta} \right) c_{2t+1} \left( \frac{1}{1+\theta} \right) \]

first by using the Lagrangian method, and then by substituting the formula for \( c_{2t+1} \) into the utility function. Confirm that the solutions for \( c_{1t} \) and \( c_{2t+1} \) are the same as those found above.

Exercise 9  Solve the maximization problem using the general CES utility function we discussed above

\[ U_t(c_{1t}, c_{2t+1}) = \frac{c_{1t}^{1-\sigma}}{1-\sigma} + \left( \frac{1}{1+\theta} \right) c_{2t+1}^{1-\sigma} \]

first by using the Lagrangian method, and then by substituting the formula for \( c_{2t+1} \) into the utility function. Comment on the difference between the solutions for \( c_{1t} \) and \( c_{2t+1} \) here and those we derived above (and you found in your answer to the previous exercise), noting, in particular, on which of the parameters of the problem the solutions depend. This does not change in a fundamental way the solution to this problem, but does make the solution of the complete model substantially harder when we do not assume Cobb-Douglas preferences.