2

One, Two, Three, . . . , Infinity

Ten little Indian boys went out to dine;
One choked his little self and then there were Nine.
Nine little Indian boys sat up very late;
One overslept himself and then there were Eight.
. . . . . .
Two little Indian boys playing with a gun;
One shot the other one and then there was One.
One little Indian boy left all alone;
He went out and hanged himself and then there were none.

—Agatha Christie (1939), *Ten Little Niggers (And Then There Were None)*

1 Introduction

In this class we deal with the deterministic optimization problems, i.e. uncertainty is not yet considered here.

An economic problem is often featured by the agent's optimal decision problem, maximizing her utility with the constraints of the resources she can use. And if the problem is made dynamic, one has to consider a bigger problem concerning the decisions in multiple periods. As a result, new techniques as well as new pitfalls emerge.

The beginning sections are more like warming-up sessions. In Section 2 we start with the static optimization problem with equality constraints, using our old friend, the Theorem of Lagrange. However, we then formally explore the implication of Lagrange multiplier and introduce the optimization problem with inequality constraints (the Kuhn-Tucker Theorem). Next by manipulating the tools introduced before, we arrive at the method of optimal control.

Everything works fine in the finite time horizon. However in Section 3 we will see the danger if we transplant the approaches in the infinite time horizon, and discuss how to rescue the infinitely living world.

2 The Deterministic Finite Horizon Optimization Problem

A problem with a finite horizon means that there is a terminal point in the decision process. As we've seen before, it doesn't matter (or, doesn't matter that much) whether you construct the problem in a discrete-time or continuous-time manner (in fact we will frequently switch in between in the coming weeks). The issue that matters more is how the constraints are imposed, in equalities or inequalities.

2.1 Basic Tools

2.1.1 Problems with Equality Constraints: The General Case

Readers may have already practiced the static optimization problems with equality constraints for thousands of times in their undergraduate studies, and the problems won't change much if we simply introduce a finite time dimension, i.e. some constraints must hold for each of the periods $t \in \{0, \dots, T\}$ — In a static problem people do maximization with respect to n variables $(x_1, \dots, x_n)$, and in a dynamic context with finite periods $t \in \{0, \dots, T\}$ we just solve basically the same problem with $n(T+1)$ variables $(x_1, \dots, x_{n(T+1)})$. As we know the Theorem of Lagrange, as Theorem A.1 states, provides a powerful characterization of local optima of equality constrained optimization problems in terms of the behavior of the objective function and the constraint functions at these points. Generally such problems have the form as following

$$\max\ f(\mathbf{x})$$
$$s.t.\ \mathbf{x} \in \mathcal{D} = U \cap \{\mathbf{x}|\mathbf{g}(\mathbf{x}) = 0\}\,,$$

in which object function $f : \mathbb{R}^{n(T+1)} \to \mathbb{R}$ and constraints $g_i : \mathbb{R}^{n(T+1)} \to \mathbb{R}^{k(T+1)}, \forall i \in \{1, \ldots, k(T+1)\}$ be continuously differentiable functions, and $U \subseteq \mathbb{R}^{n(T+1)}$ is open. To solve it we set up a function called *Lagrangian* $\mathscr{L} : \mathcal{D} \times \mathbb{R}^{k(T+1)} \to \mathbb{R}$

$$\mathscr{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{k(T+1)} \lambda_i g_i(\mathbf{x})$$

in which the vector $\lambda = (\lambda_1, \ldots, \lambda_{k(T+1)}) \in \mathbb{R}^{k(T+1)}$ is called *Lagrange multiplier*.

Then by Theorem A.1 [1] we find the set of all critical points of $\mathscr{L}(\mathbf{x}, \lambda)$ for $\mathbf{x} \in U$, i.e. the *first order conditions*

$$\frac{\partial \mathscr{L}}{\partial x_j} = 0,\ \forall j \in \{1, \ldots, n(T+1)\},$$
$$\frac{\partial \mathscr{L}}{\partial \lambda_i} = 0,\ \forall i \in \{1, \ldots, k(T+1)\},$$

which simply say that these conditions should hold for each x and λ in every period.

Now we continue to explore the interpretation for the Lagrange multiplier λ. We relax the equality constraints by adding a sufficiently small constant to each of them, i.e.

$$\mathbf{g}(\mathbf{x}, \mathbf{c}) = \mathbf{g}(\mathbf{x}) + \mathbf{c}$$

in which $\mathbf{c} = (c_1, \ldots, c_k)$ is a vector of constants. Now the set of constraints becomes

$$\mathcal{D} = U \cap \{\mathbf{x}|\mathbf{g}(\mathbf{x}, \mathbf{c}) = 0\}\,.$$

Then by Theorem A.1 at the optimum $\mathbf{x}^*(\mathbf{c})$ there exists $\lambda^*(\mathbf{c}) \in \mathbb{R}^{k(T+1)}$ such that

$$Df\,(\mathbf{x}^*(\mathbf{c})) + \sum_{i=1}^{k(T+1)} \lambda_i^*(\mathbf{c}) Dg_i(\mathbf{x}^*(\mathbf{c})) = 0. \tag{1}$$

[1] Please note that as a tradition people denote the derivative of a multi-variate function $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ by $Df(\mathbf{x})$, which is an n dimensional vector $Df(\mathbf{x}) := \left[\frac{\partial f(x_1,\ldots,x_n)}{\partial x_1}, \ldots, \frac{\partial f(x_1,\ldots,x_n)}{\partial x_n}\right]$.

Define a new function of $\mathbf{c}$, $F(\mathbf{c}) = f(\mathbf{x}^*(\mathbf{c}))$. Then by chain rule,

$$DF(\mathbf{c}) = Df\,(\mathbf{x}^*(\mathbf{c}))\,D\mathbf{x}^*(\mathbf{c}).$$

Insert (1) into the equation above, one can get

$$DF(\mathbf{c}) = -\left(\sum_{i=1}^{k(T+1)} \lambda_i^*(\mathbf{c}) Dg_i(\mathbf{x}^*(\mathbf{c}))\right) D\mathbf{x}^*(\mathbf{c}),$$

and this is equivalent to

$$DF(\mathbf{c}) = -\sum_{i=1}^{k(T+1)} \lambda_i^*(\mathbf{c}) Dg_i(\mathbf{x}^*(\mathbf{c})) D\mathbf{x}^*(\mathbf{c}). \tag{2}$$

Define another new function of $\mathbf{c}$, $G_i(\mathbf{c}) = g_i(\mathbf{x}^*(\mathbf{c}))$. Then again by chain rule,

$$DG_i(\mathbf{c}) = Dg_i\,(\mathbf{x}^*(\mathbf{c}))\,D\mathbf{x}^*(\mathbf{c}).$$

Insert this into (2), and one can get

$$DF(\mathbf{c}) = -\sum_{i=1}^{k(T+1)} \lambda_i^*(\mathbf{c}) DG_i(\mathbf{c}). \tag{3}$$

By the equality constraint $\mathbf{g}(\mathbf{x}) + \mathbf{c} = 0$ one can easily see that

$$DG_i(\mathbf{c}) = -\mathbf{e}_i$$

in which $\mathbf{e}_i$ is the i-th unit vector in $\mathbb{R}^{k(T+1)}$, i.e. the vector that has a 1 in the i-th place and zeros elsewhere. Therefore (3) turns out to be

$$\begin{aligned} DF(\mathbf{c}) &= -\sum_{i=1}^{k(T+1)} \lambda_i^*(\mathbf{c})\,(-\mathbf{e}_i) \\ &= \lambda^*(\mathbf{c}). \end{aligned}$$

From the equation above one can clearly see that the Lagrange multiplier λ_i measures the sensitivity of the value of the objective function at it maxima $\mathbf{x}^*$ to a small relaxation of the constraint g_i. Therefore λ_i has a very straightforward economic interpretation, that λ_i represents the maximum amount the decision maker would be willing to pay for a marginal relaxation of constraint i — this is sometimes called the *shadow price* of constraint i at the optima.

2.1.2 *Problems with Equality Constraints: A Simplified Version*

The general case may be a little bit messy to go through, now we deal with the same problem in a much simplified version, i.e. the univariate case which we are quite familar with. Suppose that an agent maximizes her neoclassical utility function with respect to a single good x, and x must follow an equality constraint,

$$\begin{aligned} \max_{x}\ & u(x), \\ s.t.\ & g(x) = 0. \end{aligned}$$

Then the problem can be easily solved by setting up Lagrangian

$$\mathscr{L} = u(x) + \lambda g(x),$$

and the optimal x, denoted by x^*, can be derived from the first order conditions

$$\frac{\partial \mathscr{L}}{\partial x} = 0, \frac{\partial \mathscr{L}}{\partial \lambda} = 0.$$

Now relax the constraint a little bit by ϵ around x^*, and rewrite the optimization problem at x^* as

$$\begin{aligned} \max_{\epsilon}\ & u(x^*, \epsilon), \\ s.t.\ & g(x^*) = \epsilon. \end{aligned}$$

By Theorem A.1 the optimal value of ϵ can be solved from the first order conditions of the Lagrangian

$$\mathscr{L}' = u(x^*, \epsilon) + \lambda \left[g(x^*) - \epsilon \right].$$

However, since we already know that x^* is the optimal solution of the original problem, and the optimal value of ϵ must be achieved when $\epsilon \to 0$, i.e.

$$\begin{aligned} \left. \frac{\partial \mathscr{L}'}{\partial \epsilon} \right|_{\epsilon \to 0} = \frac{\partial u(x^*, \epsilon)}{\partial \epsilon} - \lambda = 0, \\ \lambda = \frac{\partial u(x^*, \epsilon)}{\partial \epsilon}. \end{aligned}$$

The last step clearly shows that the Lagrange multiplier λ measures how much the utility changes when the constraint is relaxed a little bit at the optimum, i.e. the shadow price at the optimum.

2.1.3 Problems with Inequality Constraints: The General Case

For problems with inequality constraints, the solution procedure is pretty similar. The only differences are the following: First, of course, the prototype problem is different in the contraints, which are now

$$\mathbf{x} \in \mathcal{D} = U \cap \{\mathbf{x}|\mathbf{h}(\mathbf{x}) \geq 0\}.$$

Second, besides the first order conditions, there is an additional *complementary slackness* condition saying that at optimum

$$\begin{aligned} \lambda^* &\geq \mathbf{0}, \\ \lambda^* \mathbf{h}^* &= \mathbf{0}. \end{aligned}$$

The economic intuition behind the condition is pretty clear: If any resource i has a positive value at the optima, i.e. $\lambda_i^* > 0$, then it must be exhausted to maximize the object function, i.e. $h_i^* = 0$; and if any resource j is left at a positive value at the optima, i.e. $h_j^* > 0$, then it must be worthless at all, i.e. $\lambda_j^* = 0$. To see how one can arrive at such results, an example is exposed in the next section.

2.1.4 Problems with Inequality Constraints: An Example

(Adapted from Heer and Mauβner (2005), Chapter 1) Consider the following two-period Ramsey-Cass-Koopmans problem [2] of a farmer. Suppose that

- Time is divided into two intervals of unit length indexed by $t = 0, 1$;
- K_t and N_t denote the amounts of seeds and labor available in period t;
- Seeds and labor input produce an amount Y_t of corn according to the neoclassical production function $Y_t = F(K_t, L_t)$;
- For each period t the farmer must decide
  - how much corn to produce,
  - how much corn to eat, and
  - how much corn to put aside for future production;
- Next period's seed is next period's stock of capital K_{t+1};
- Choice of consumption C_t and investment
  - is constrained by current production

    $$C_t + K_{t+1} \leq Y_t,$$

  - aims at maximizing the utility function (assume that $U(\cdot)$ satisfies Inada condition)

[2] We will interpret all the settings and results later. Readers are only asked to understand the techniques here.

$U(C_0, C_1) = u(C_0) + \beta u(C_1)$;

- Leisure does not appear in the utility function; assume that the farmer works a given number of hours N each period.

Then the maximization problem turns out to be

$$\begin{aligned}
\max_{C_0,C_1} \ & U(C_0, C_1) = u(C_0) + \beta u(C_1), \\
s.t. \ & C_0 + K_1 \leq F(K_0), \\
& C_1 + K_2 \leq F(K_1), \\
& 0 \leq C_0, \\
& 0 \leq C_1, \\
& 0 \leq K_1, \\
& 0 \leq K_2.
\end{aligned}$$

Comparing with the prototype problem presented in Theorem A.2 we may define that

$$\begin{aligned}
\mathbf{x} &= (C_0, C_1, K_1, K_2), \\
f(C_0, C_1, K_1, K_2) &= U(C_0, C_1), \\
n &= 4
\end{aligned}$$

as well as the constraints

$$\begin{aligned}
&h_1 = F(K_0) - C_0 - K_1 \geq 0, \\
&h_2 = F(K_1) - C_1 - K_2 \geq 0, \\
&h_3 = C_0 \geq 0, \\
&h_4 = C_1 \geq 0, \\
&h_5 = K_1 \geq 0, \\
&h_6 = K_2 \geq 0.
\end{aligned}$$

By Theorem A.2 the first order conditions are

$$0 = \frac{\partial U}{\partial C_0} + \lambda_1 \frac{\partial h_1}{\partial C_0} + \ldots + \lambda_6 \frac{\partial h_6}{\partial C_0} = \frac{\partial U}{\partial C_0} - \lambda_1 + \lambda_3, \tag{4}$$

$$0 = \frac{\partial U}{\partial C_1} + \lambda_1 \frac{\partial h_1}{\partial C_1} + \ldots + \lambda_6 \frac{\partial h_6}{\partial C_1} = \frac{\partial U}{\partial C_1} - \lambda_2 + \lambda_4, \tag{5}$$

$$0 = \frac{\partial U}{\partial K_1} + \lambda_1 \frac{\partial h_1}{\partial K_1} + \ldots + \lambda_6 \frac{\partial h_6}{\partial K_1} = -\lambda_1 + \lambda_2 F'(K_1) + \lambda_5, \tag{6}$$

$$0 = \frac{\partial U}{\partial K_2} + \lambda_1 \frac{\partial h_1}{\partial K_2} + \ldots + \lambda_6 \frac{\partial h_6}{\partial K_2} = -\lambda_2 + \lambda_6, \tag{7}$$

as well as $\lambda_i \geq 0$, $\forall i \in \{1,\ldots,6\}$. And complementary slackness gives $\lambda_i h_i = 0$, $\forall i \in \{1,\ldots,6\}$.

Now let's try to simplify all the statements above. Knowing by Inada condition that

$$\lim_{c_i \to 0} \frac{\partial U}{\partial C_i} = +\infty$$

we infer that $C_0 > 0$ and $C_1 > 0$. From complementary slackness one can directly see that $\lambda_3 = \lambda_4 = 0$. Then by the strict concavity of $U(\cdot)$, $\frac{\partial U}{\partial C_i} > 0$. Therefore (4) (5) simply imply that $\lambda_1 = \frac{\partial U}{\partial C_0} > 0$ and $\lambda_2 = \frac{\partial U}{\partial C_1} > 0$, as well as $\lambda_6 > 0$ from (7) — this further implies that $K_2 = 0$ by complementary slackness. And from h_2 one can see that $F(K_1) \geq C_1 > 0$, implying that $K_1 > 0$ as well as $\lambda_5 = 0$. From (6) one can see that

$$F'(K_1) = \frac{\lambda_1}{\lambda_2} = \frac{\frac{\partial U}{\partial C_0}}{\frac{\partial U}{\partial C_1}}.$$

This is just the Euler condition. With two other conditions $h_1 = 0$ and $h_2 = 0$ one can easily solve for (C_0, C_1, K_1).

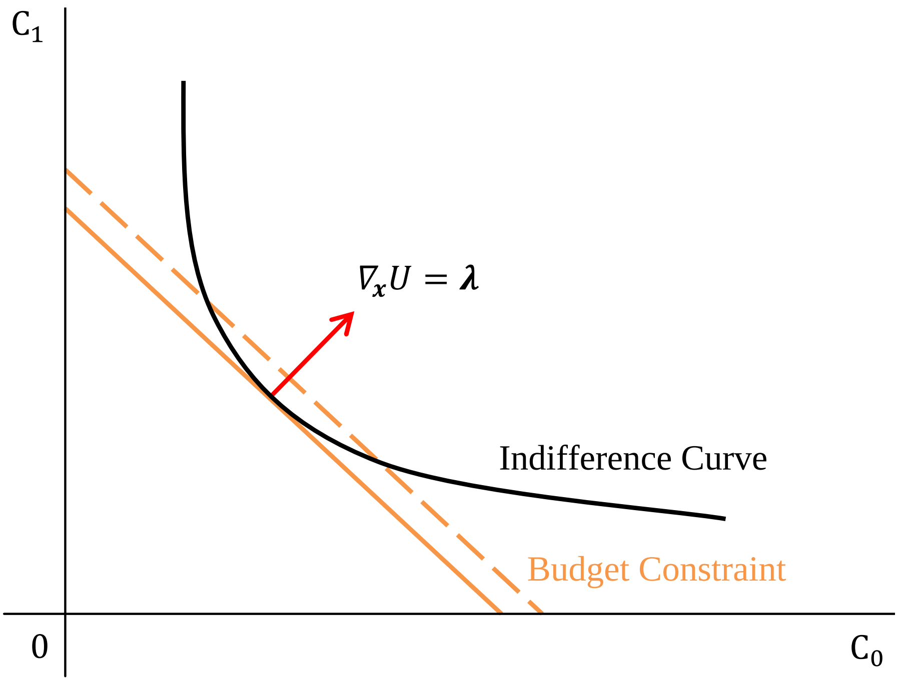

Fig. 1. Shadow Price

Figure 1 gives a graphical interpretation to this inequality constrained optimization problem (one can find more theoretical arguments in Mas-Colell *et al.* (1995), Chapter 2). The agent maximizes her life-time utility by choosing the consumption level for each of the two periods, on the basis of her intertemporal budget constraints. The optimum is achieved where the indifferent curve is exactly tangent to the frontier of the budget constraint. Suppose that we relax the budget constraint by adding a little bit to it, the vector λ just describes by how much the indifferent curve responds to the relaxation — in mathematical term, exactly the gradient $\nabla_{\mathbf{x}} U$ as the graph shows.

2.2 The General Deterministic Finite Horizon Optimization Problems: From Lagrangian to Hamiltonian

Let's take a closer look at the structure of the problem in the example. What makes it interesting is that the variables from the different periods are linked through the constraints (otherwise we can solve the problem by simply repeating dealing with the insulated $T + 1$ static problems), therefore one variable's change in one period may have pervasive effects into the other periods. So one may wonder whether there exists a solution method by exploiting such linkage — this is just the widely applied *optimal control* method.

As a general exposure, the prototype problem can be described as following. Think about the simplest case with only two variables k_t, c_t in each period $t \in \{0, 1, \dots, T\}$, $T < +\infty$. The problem is to maximize the object function $U : \mathbb{R}^{2(T+1)} \to \mathbb{R}$ which is the summation of the function $u : \mathbb{R}^2 \to \mathbb{R}$ for each period, constrained by the intertemporal relations of k and c as well as the boundary values

$$
\begin{aligned}
\max_{\{c_t\}} U &= \sum_{t=0}^{T} \frac{1}{(1+\rho)^t} u\left(k_t, c_t, t\right), \\
s.t.\ k_{t+1} - k_t &= g\left(k_t, c_t\right), \\
k_{t=0} &= k_0, \\
k_{T+1} &\geq \overline{k}_{T+1}.
\end{aligned}
$$

k and c represent two kinds of variables. Variable k_t is the one with which each period starts and on which the decision is based, therefore it's usually called *state variable*. And variable c_t is the one the decision maker can change in each period and what is left over is fed back into the next period state variable, therefore it's usually called *control variable*. The constraint linking these variables across periods is called the *law of motion*.

If we express everything in continuous time, we only need to rewrite the summation by integration and the intertemporal change by the derivative with respect to time. However solving the continuous time problems with the Lagrangian would be a bit tricky. And in order to give the readers more exposures to the continuous time models, in the section that follows we start with building up the foundations of finite horizon optimization problems in continuous time. Readers may extend the same idea into the discrete time problems as an exercise.

2.2.1 Continuous Time

Suppose that time is continous such that $t \in [0, T]$, $T \leq +\infty$. A typical deterministic continuous time optimization problem can be written as (often people simply set $\bar{k}(T)$ to be zero)

$$\begin{aligned}
\max_{\{c(t)\}} \; U &= \int_0^T e^{-\rho t} u\left(k(t), c(t), t\right) dt, \\
s.t. \; \dot{k}(t) &= g\left(k(t), c(t), t\right), \\
k(0) &= k_0, \\
k(T) &\geq \bar{k}(T).
\end{aligned}$$

Set up Lagrangian for this problem

$$\mathscr{L} = \int_0^T e^{-\rho t} u\left(k(t), c(t), t\right) dt + \int_0^T \mu(t)\left(g\left(k(t), c(t), t\right) - \dot{k}(t)\right) dt + \nu\left[k(T) - \bar{k}(T)\right], \tag{8}$$

and we are supposed to find the first order conditions with respect to $k(t)$ and $c(t)$. However the second term in $\mathscr{L}$ involves $\dot{k}(t)$, and this make it difficult to derive it with respect to $k(t)$. Therefore we rewrite this term with integration by parts

$$\begin{aligned}
\int_0^T \mu(t)\dot{k}(t) dt &= \mu(t)k(t)|_0^T - \int_0^T k(t)\dot{\mu}(t) dt \\
&= \mu(T)k(T) - \mu(0)k_0 - \int_0^T k(t)\dot{\mu}(t) dt.
\end{aligned}$$

Insert it back into Lagrangian, we get

$$\begin{aligned}
\mathscr{L} = &\int_0^T \left[e^{-\rho t} u\left(k(t), c(t), t\right) + \mu(t) g\left(k(t), c(t), t\right)\right] dt \\
&- \left(\mu(T)k(T) - \mu(0)k_0 - \int_0^T k(t)\dot{\mu}(t) dt\right) + \nu\left[k(T) - \bar{k}(T)\right].
\end{aligned}$$

Define Hamiltonian function as

$$\mathcal{H}(k, c, \mu, t) = e^{-\rho t} u\left(k(t), c(t), t\right) + \mu(t) g\left(k(t), c(t), t\right), \tag{9}$$

then Lagrangian turns out to be

$$\mathscr{L} = \int_0^T \left[\mathcal{H}(k, c, \mu, t) + k(t)\dot{\mu}(t)\right] dt - \mu(T)k(T) + \mu(0)k_0 + \nu\left[k(T) - \bar{k}(T)\right].$$

Now let $k^*(t), c^*(t)$ be the optimal path for state and control variable. Define $p_1(t)$ as an arbitrary perturbation for $c^*(t)$, then a neighbouring path around $c^*(t)$ can be defined as

$$c(t) = c^*(t) + \epsilon p_1(t).$$

Similarly define $p_2(t)$ as an arbitrary perturbation for $k^*(t)$, then a neighbouring path around $k^*(t)$ can be defined as

$$k(t) = k^*(t) + \epsilon p_2(t)$$

as well as the end-period state variable

$$k(T) = k^*(T) + \epsilon dk(T).$$

Rewrite $\mathscr{L}$ in terms of ϵ

$$\mathscr{L}^*(\cdot, \epsilon) = \int_0^T \left[\mathcal{H}(k(t, \epsilon), c(t, \epsilon), t) + k(t, \epsilon)\dot{\mu}(t)\right] dt - \mu(T)k(T, \epsilon) + \mu(0)k_0 + \nu\left[k(T, \epsilon) - \bar{k}(T)\right]$$

and the first order condition must hold

$$\begin{aligned}\left.\frac{\partial \mathscr{L}^*(\cdot, \epsilon)}{\partial \epsilon}\right|_{\epsilon \to 0} &= 0 \\ &= \int_0^T \left[\frac{\partial \mathcal{H}}{\partial \epsilon} + \dot{\mu}(t)\frac{\partial k}{\partial \epsilon}\right] dt + (\nu - \mu(T))\,\frac{\partial k(T)}{\partial \epsilon}.\end{aligned}$$

By the chain rule

$$\begin{aligned}\frac{\partial \mathcal{H}}{\partial \epsilon} &= \frac{\partial \mathcal{H}}{\partial k}\frac{\partial k}{\partial \epsilon} + \frac{\partial \mathcal{H}}{\partial c}\frac{\partial c}{\partial \epsilon} \\ &= \frac{\partial \mathcal{H}}{\partial k}p_2(t) + \frac{\partial \mathcal{H}}{\partial c}p_1(t),\end{aligned}$$

and insert it into the first order condition

$$
\begin{aligned}
\frac{\partial \mathscr{L}^*(\cdot,\epsilon)}{\partial \epsilon}\bigg|_{\epsilon\to 0} &= \int_0^T \left[\frac{\partial \mathcal{H}}{\partial k} p_2(t) + \frac{\partial \mathcal{H}}{\partial c} p_1(t) + \dot{\mu}(t) p_2(t)\right] dt + (\nu - \mu(T))\, dk(T) \\
&= \int_0^T \left[\left(\frac{\partial \mathcal{H}}{\partial k} + \dot{\mu}(t)\right) p_2(t) + \frac{\partial \mathcal{H}}{\partial c} p_1(t)\right] dt + (\nu - \mu(T))\, dk(T) \\
&= 0.
\end{aligned}
$$

Therefore the first order condition is equivalent to the following equations

$$
\frac{\partial \mathcal{H}}{\partial c} = 0, \tag{10}
$$
$$
\frac{\partial \mathcal{H}}{\partial k} = -\dot{\mu}(t), \tag{11}
$$
$$
\mu(T) = \nu. \tag{12}
$$

Since we assume that $k^*(t), c^*(t)$ be the optimal path, then these conditions must hold. Condition (10) is called the *Euler equation*, and condition (11) is the *Maximum Principle*. Condition (12) requires that the terminal date costate variable, $\mu(T)$, equal the terminal date static Lagrange multiplier ν.

There is still something missing — Go back to the Lagrangian (8), we also have to address the concern on complementary slackness regarding the terminal time capital constriant, i.e.

$$
\nu \left[k(T) - \overline{k}(T)\right] = 0 \text{ with } \nu \geq 0.
$$

Combining with condition (12) the complementary slackness is simply equivalent to

$$
\mu(T) \left[k(T) - \overline{k}(T)\right] = 0, \tag{13}
$$

which is often called *transversality condition*. The intuition behind it is pretty clear: If there is strictly positive amount of more capital is left at the end date T than required, i.e. $k(T) - \overline{k}(T) > 0$, then its price must be zero, i.e. $\mu(T) = 0$, because it is worthless at all. On the other hand if the capital stock at the end date has a strictly positive value, i.e. $\mu(T) > 0$, then the agent must leave no excessive capital at all, i.e. $k(T) - \overline{k}(T) = 0$.

Now the lengthy procedure which we went through simply tells us that one can actually start from the Hamiltonian and directly arrive at the first order conditions. As a summary, to solve the deterministic multi-period optimization problem the whole procedure can be simplified into the following steps:

- Formulate the optimization problem as we did in the beginning of this section, and write down its Hamiltonian as (9);

- Derive the first order conditions regarding control and state variables respectively, such as (10) and (11);
- Add the transversality condition such as (13);
- Make further treatments on these equations to get whatever you are interested in.

In addition, please note that the menu also works for the problems with more than one state and / or control variables. The first order conditions are in the same forms as equations (10) and (11), for control and state variables respectively.

2.2.2 Discrete Time

Since discrete time problems have the same nature as the ones for continuous time, therefore here we simply present the results without going into the details of proofs.

A typical deterministic discrete time optimization problem can be written as

$$\begin{aligned}
\max_{\{c_t\}} \; & U = \sum_{t=0}^{T} \frac{1}{(1+\rho)^t} u\left(k_t, c_t, t\right), \\
s.t. \; & k_{t+1} - k_t = g\left(k_t, c_t\right), \\
& k_{t=0} = k_0, \\
& k_{T+1} \geq \overline{k}_{T+1}.
\end{aligned}$$

Construct the present value Hamiltonian $\mathcal{H}_t = u\left(k_t, c_t, t\right) + \lambda_t g\left(k_t, c_t\right)$, and the first order conditions are $\forall t \in \{0, 1, \ldots, T\}$

$$\begin{aligned}
\frac{\partial \mathcal{H}_t}{\partial c_t} &= 0, \\
\frac{\partial \mathcal{H}_t}{\partial k_t} &= -\left(\lambda_t - \lambda_{t-1}\right), \\
\frac{\partial \mathcal{H}_t}{\partial \lambda_t} &= k_{t+1} - k_t,
\end{aligned}$$

as well as the complementary slackness such that $\lambda_T \geq 0$ and $\lambda_T \left(k_{T+1} - \overline{k}_{T+1}\right) = 0$.

2.2.3 Present versus Current Value Hamiltonian

Often what we consider in economics is the optimization problem regarding a discounted object function (in contrast to the prototype model by Ramsey), such as

$$
\begin{aligned}
\max_{\{c(t)\}} U &= \int_0^T e^{-\rho t} u\left(k(t), c(t), t\right) dt, \\
s.t.\ \dot{k}(t) &= g\left(k(t), c(t), t\right), \\
k(0) &= k_0, \\
k(T) - \bar{k}(T) &\geq 0
\end{aligned}
$$

in which ρ is the discount rate. As we did in Section 2.2.1 the *present value* Hamiltonian can be expressed as

$$
\mathcal{H} = e^{-\rho t} u\left(k(t), c(t), t\right) + \mu(t) g\left(k(t), c(t), t\right)
$$

— notice that $\mu(t)$ is the present value shadow price, for it correspondents to the discounted object function. Same as before, the first order conditions can be derived as equations (10) and (11), plus the transversality condition (13).

Sometimes it's convenient to study a problem in the current time terms, and people set up the *current value* Hamiltonian as

$$
\hat{\mathcal{H}} = u\left(k(t), c(t), t\right) + q(t) g\left(k(t), c(t), t\right)
$$

in which $q(t) = \mu(t)e^{\rho t}$ is the current value shadow price, for it correspondents to the non-discounted object function. Now the first order conditions are slightly different in $\frac{\partial \hat{\mathcal{H}}}{\partial k}$

$$
\frac{\partial \hat{\mathcal{H}}}{\partial c} = 0, \tag{14}
$$

$$
\frac{\partial \hat{\mathcal{H}}}{\partial k} = \rho q(t) - \dot{q}(t), \tag{15}
$$

as well as the transversality condition

$$
q(T) e^{-\rho t} \left[k(T) - \bar{k}(T) \right] = 0. \tag{16}
$$

Although equation (15) is a little more complicated, it is very intuitive. Notice that $\frac{\partial \hat{\mathcal{H}}}{\partial k}$ is just the marginal contribution of the capital to utility, i.e. the divident received by the agent, the equation reflects the idea of asset pricing: given that $\dot{q}(t)$ is the capital gain (the change in the price of the asset), and ρ is the rate of return on an alternative asset, i.e. consumption, equation (15) says that at the optimum the agent is indifferent between the two types of the investment, for the overall rate of return to the capital,

$$\frac{\frac{\partial \hat{\mathcal{H}}}{\partial k} + \dot{q}(t)}{q(t)},$$

equals the return to consumption, ρ. For this reason, equation (15) is also called *non-arbitrage condition*.

3 Going Infinite: Something Rotten in Denmark

[3] As a general principle we know that whenever one jumps into infinity there must be a problem waiting for him or her. So what can go wrong, when we extend the results of finite horizon optimization problems into the ones with infinite horizon?

The optimization itself is only a little different — $T = +\infty$ in the object function

$$U = \int_0^{+\infty} e^{-\rho t} u\left(k(t), c(t), t\right) dt,$$

and there will be no terminal time condition any more, because the time doesn't terminate at all. But this makes a big change of the problem: Now the optimal time path looks like a kite — we hold the thread at hand, but we don't know where it ends.

Note that the principles behind the finite time optimization problem are that following the optimal time path nothing valuable is left over in the end of the world (such that $\mu(T)\left[k(T) - \bar{k}(T)\right]$ is non-positive) and the agent doesn't exit the world with debt (such that $\mu(T)\left[k(T) - \bar{k}(T)\right]$ is non-negative), which are captured in the transversality condition. To maintain the same principles in the infinite time horizon, we may assume that there is an end of the world, but after a nearly infinitely long time. Therefore we may impose a similar transversality condition for the problems of infinite time horizon

$$\lim_{T\to+\infty} \mu(T)\left[k(T) - \bar{k}(T)\right] = 0,$$

[3] William Shakespeare (1602): *Hamlet*, Act 1, Scene 4
...
HORATIO
Have after. To what issue will this come?
MARCELLUS
Something is rotten in the state of Denmark.
HORATIO
Heaven will direct it.
...

i.e. the value of the state variable must be *asymptotically* zero: If the quantity of $k(T)$ remains different from the constraint asymptotically, then its price, $\mu(T)$, must approach 0 asymptotically; If $k(T) - \overline{k}(T)$ grows forever at a positive rate, then the price $\mu(T)$ must approach 0 at a faster rate so that the product, $\mu(T)\left[k(T) - \overline{k}(T)\right]$, goes to 0.

However, such asymptotical transversality condition is not free from controversies (See Barro and Sala-í-Martin (2004), Appendix of Mathematical Methods A.2 and A.3). But for the problems with discounted object functions (the form we take throughout this course), it is sufficient for one to arrive at the correct solutions.

4 Example: Working with the Hamiltonian

In this section we rewrite the problem presented in Chapter 2.1 – 2.2, Romer (2006) in the framework of optimal control. Basically Romer attempts to solve the problem in a least demanding approach such that readers can follow the reasoning with just the most elementary knowledge of calculus. But this treatment comes with a price: The arguments there are pretty tedious, sometimes confusing, and lack of strict proofs. Now by introducing the method of optimal control, readers will see that everything becomes simple and straight forward.

You will learn later that this problem is equivalent to the following setup [4]

$$\max_{\{c(t),k(t)\}_{t=0}^{+\infty}} U = \int_0^{+\infty} e^{-\rho t} \frac{[A(0)e^{gt}c(t)]^{1-\theta}}{1-\theta} \frac{L(0)e^{nt}}{H} dt,$$
$$s.t.\ \dot{k}(t) = f(k(t)) - (n+g)k(t) - c(t),$$

in which the equation

$$\dot{k}(t) = f(k(t)) - (n+g)k(t) - c(t) \tag{17}$$

is the law of motion, and the parameters ρ, $A(0)$, g, θ, $L(0)$, n and H are constant.

Set up the present value Hamiltonian for this problem [5]

$$\mathcal{H} = e^{-\rho t} \frac{\left[A(0)e^{gt}c\right]^{1-\theta}}{1-\theta} \frac{L(0)e^{nt}}{H} + \mu\left(f(k) - (n+g)k - c\right),$$

the first order conditions are

[4] Again here you don't have to understand the economics behind. You may prove the equivalence when you arrive at the Ramsey-Cass-Koopmans model in the lecture.

[5] From now on we drop off the time variable t when there is no confusion.

$$\frac{\partial \mathcal{H}}{\partial c} = e^{-\rho t}\frac{L(0)e^{nt}}{H}\left[A(0)e^{gt}c\right]^{-\theta}A(0)e^{gt} - \mu = 0, \tag{18}$$

$$\frac{\partial \mathcal{H}}{\partial k} = \mu\left(f'(k) - (n+g)\right) = -\dot{\mu} \tag{19}$$

as well as the transversality condition

$$\lim_{T\to+\infty} \mu(T)k(T) = 0. \tag{20}$$

Take logarithm on (18),

$$\begin{aligned} \ln\left\{e^{-\rho t}\frac{L(0)e^{nt}}{H}\left[A(0)e^{gt}c\right]^{-\theta}A(0)e^{gt}\right\} &= \ln\mu, \\ -\rho t + \ln L(0) - \ln H + nt - \theta\left[\ln A(0) + gt + \ln c\right] + \ln A(0) + gt &= \ln\mu, \end{aligned}$$

then take derivatives with respect to t on both sides (note that c, k and μ are functions of t)

$$-\rho + n - \theta\left[g + \frac{\dot{c}}{c}\right] + g = \frac{\dot{\mu}}{\mu}. \tag{21}$$

Notice that (19) implies that

$$\frac{\dot{\mu}}{\mu} = -\left(f'(k) - (n+g)\right), \tag{22}$$

combine (21) and (22) to get

$$-\rho + n - \theta\left[g + \frac{\dot{c}}{c}\right] + g = -f'(k) + (n+g),$$

rearrange to get

$$\frac{\dot{c}}{c} = \frac{f'(k) - \rho - \theta g}{\theta}, \tag{23}$$

which is just the Euler equation.

Further more, one can solve for the time path of μ from the ordinary differential equation (22)

$$\mu(t) = \mu(0)\exp\left[-\left(f'(k) - (n+g)\right)t\right]. \tag{24}$$

Equation (24) determines the steady state value of k

$$k^* = (f')^{-1}(\rho + \theta g). \tag{25}$$

Insert (24) and (25) into the transversality condition (20)

$$\begin{aligned}
&\lim_{T\to+\infty} \mu(0)\exp\left[-\left(f'(k(T)) - (n+g)\right)t\right]k(T) \\
=&\lim_{T\to+\infty} \mu(0)\exp\left[-\left(\rho + \theta g - (n+g)\right)t\right](f')^{-1}(\rho + \theta g).
\end{aligned}$$

The equation above is asymptotically zero only if

$$\rho + \theta g - (n+g) > 0. \tag{26}$$

Given (26) satisfied, the dynamics of the system are fully characterized by the equations (17) and (23), which we'll explore further in the next class.

5 Readings

Barro and Sala-í-Martin (2004), Appendix of Mathematical Methods A.2 and A.3.

6 Bibliographic Notes

Among the huge literature on intertemporal optimization problems, Chiang and Wainwright (2005) and Dixit (1990) are distinguished by providing junior researchers a good balance between both theoretical sufficiency and practical accessibility. Sundaram (1996) is based on sound analytical foundations and perfect for the first course of mathematical economics. Also it is much self-contained so that readers can work it throughout by timely referring to Kolmogorov and Fomin (1970), or Rudin (1976). Chow (1997) is an interesting reading, by showing that Lagrange method has a much wider range of applications than people usually think and is even suitable for many occasions in which the popular approach of recursive method fails to work. Kamien and Schwartz (1991) is the classic from the golden age of dynamic control in economics and management science.

7 Exercises

7.1 Dynamic Optimization in Continuous Time

An individual receives a steady stream of income over time $y(t)$. She maximizes her discounted utility from consumption. Her intertemporal utility function is given by

$$\int_0^{+\infty} e^{-\rho t} U(t) dt \quad \text{with} \quad U(t) = \frac{1}{\alpha} c(t)^{\alpha}$$

The consumer has access to a perfect capital market at which she can lend or borrow at an interest rate r.

a)[A] Give an interpretation of the parameter ρ. Calculate the elasticity of substitution between consumption of two points in time and the rate of relative risk aversion.

b)[A] What is the transition equation for consumer's wealth?

c)[A] Formulate the Hamiltonian of this problem and derive first order conditions.

d)[B] Derive the Euler equation and show how consumption changes over time. Distinguish two cases: a rate of time preference being lower / exceeding the rate of interest.

e)[B] Let $r = 0.1$ and $\rho = 0.2$. Determine the optimal consumption path, if the present value of the income stream is $y_0 = 100$. Discuss the relation between the transversality condition and the household's intertemporal budget constraint.

7.2 Dynamic Optimization in Discrete Time

A representative consumer maximizes

$$u = \sum_{t=0}^{+\infty} U(C_t) \frac{1}{(1+\rho)^t}$$

subject to the per period budget constraint $B_{t+1} - B_t = Y_t + rB_t - C_t$ with B_0 and $\{Y_t\}_{t=0}^{+\infty}$ given.

a)[A] Derive the first order conditions and characterize the optimal consumption paths.

b)[A] Explain the relation of this discrete time approach to the continuous time approach in the previous exercise.

c)[B] Using the No-Ponzi-Game condition, formulate the consumer's intertemporal wealth constraint. Discuss the relation between the No-Ponzi-Game condition and the transversality condition.

7.3 Euler Equation: A General Proof and Its Application[C]

Consider the problem of finding a path $x(t)$ that solves

$$\max_{x(t)} \int_a^b F(t, x(t), x'(t))dt \quad \text{subject to} \quad x(a) = x_a \quad \text{and} \quad x(b) = x_b.$$

Show that the first order condition to this problem is given by the Euler equation

$$\frac{\partial F}{\partial x} = \frac{d\frac{\partial F}{\partial x'}}{dt}.$$

Using the result, find the Euler equation and the optimal path $x(t)$ for

$$\max_{x(t)} \int_0^T \left((x'(t))^2 + cx(t)\right) dt \quad \text{subject to} \quad x(0) = 0 \quad \text{and} \quad x(T) = B.$$

7.4 Application of Dynamic Optimization in Growth Theory: Ramsey Model

An infinitely lived representative agent has the utility function

$$U_0 = \int_0^{+\infty} e^{-\rho t} (c(t))^\beta \, dt, \quad 0 < \beta < 1.$$

The aggregate production function is $Y = K^\alpha N^{1-\alpha}$ $(0 < \alpha < 1)$, in which K is capital input and N is labor input. The growth rate of labor force is n, the rate depreciation of capital is δ. Both rates are constant over time.

a)$^{\mathbf{A}}$ Show that the production function has constant returns to scale and formulate output per capita (suppose that everyone in this economy provides a unit of labor force) as a function of capital intensity (capital per capita).

b)$^{\mathbf{A}}$ Derive the transition equation for capital intensity.

c)$^{\mathbf{A}}$ Using the Hamiltonian, derive first order conditions of the agents optimization problem.

d)$^{\mathbf{A}}$ Derive the Euler equation for per capita consumption.

e)$^{\mathbf{A}}$ Calculate capital intensity and per capita consumption of the steady state.

f)$^{\mathbf{B}}$ Sketch the phase diagram and explain the optimal growth path from an arbitrary starting value of capital intensity.

g)[C] How should the economy respond to a *foreseeable* change in the growth rate of labor force? To put it clear, suppose the economy is already in the steady state at t_0 with a constant growth rate of labor force n_0, and then for whatever reason it becomes public information at t_0 that from t_1 in the future the growth rate of labor force will be $n_1 > n_0$, $\forall t \in [t_1, +\infty)$. Using phase diagram characterize the response of the economy from t_0 on.

References

Barro, R. J. and X. Sala-í-Martin (2004): *Economic Growth (2nd Ed.)*. Cambridge: MIT Press.

Chiang, A. C. and K. Wainwright (2005): *Fundamental Methods of Mathematical Economics (4th Ed.)*. Boston: McGraw-Hill Irwin.

Chow, G. C. (1997): *Dynamic Economics - Optimization by the Lagrange Method*. New York: Oxford University Press.

Dixit, A. (1990): *Optimization in Economic Theory (2nd Ed.)*. New York: Oxford University Press.

Heer, B. and A. Mauβner (2005): *Dynamic General Equilibrium Modelling*. Heidelberg: Springer Verlag.

Kamien, M. I. and N. L. Schwartz (1991): *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management (2nd Ed.)*. Amsterdam: North-Holland.

Kolmogorov, A. N. and S. V. Fomin (1970): *Introductory Real Analysis*. New Jersey: Prentice-Hall.

Mas-Colell, A., Whinston, M. D. and J. R. Green (1995): *Microeconomic Theory*. New York: Oxford University Press.

Romer, D. (2006): *Advanced Macroeconomics (3rd Ed.)*. Boston: McGraw-Hill Irwin.

Rudin, W. (1976): *Principles of Mathematical Analysis (3rd Ed.)*. New York: McGraw-Hill.

Sundaram, R. K. (1996): *A First Course in Optimization Theory*. New York: Cambridge University Press.

Appendix

A Useful Results of Mathematics

A.1 The Theorem of Lagrange

Theorem A.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}^k$ be continuously differentiable functions, $\forall i \in \{1, \ldots, k\}$. Suppose that $\boldsymbol{x}^*$ is a local maximum or minimum of f on the set*

$$\mathcal{D} = U \cap \{\boldsymbol{x} | g_i(\boldsymbol{x}) = 0, \forall i \in \{1, \dots, k\}\},$$

in which $U \subseteq \mathbb{R}^n$ *is open. Suppose also that* $\operatorname{rank}(D\boldsymbol{g}(\boldsymbol{x}^*)) = k$. *Then, there exists a vector* $\lambda^* = \left(\lambda_1^*, \dots, \lambda_k^*\right) \in \mathbb{R}^k$ *such that*

$$Df(\boldsymbol{x}^*) + \sum_{i=1}^{k} \lambda_i^* Dg_i(\boldsymbol{x}^*) = 0.$$

A.2 Kuhn-Tucker Theorem

Theorem A.2 *Let* f *be a concave, continuously differentiable function mapping* U *into* $\mathbb{R}$, *where* $U \subseteq \mathbb{R}^n$ *is open and convex. For* $i = 1, \dots, l$, *let* $h_i : U \to \mathbb{R}$ *be concave, continuously differentiable functions. Suppose there is some* $\overline{\boldsymbol{x}} \in U$ *such that*

$$h_i(\overline{\boldsymbol{x}}) > 0, i = 1, \dots, l.$$

Then $\boldsymbol{x}^*$ *maximizes* f *over*

$$\mathcal{D} = \{\boldsymbol{x} \in U | h_i(\boldsymbol{x}) \geq 0, i = 1, \dots, l\}$$

if and only if there is $\lambda^* \in \mathbb{R}^l$ *such that the Kuhn-Tucker first-order conditions hold:*

$$\begin{aligned} \frac{\partial f(\boldsymbol{x}^*)}{\partial x_j} + \sum_{i=1}^{j} \lambda_i^* \frac{\partial h_i(\boldsymbol{x}^*)}{\partial x_j} &= 0,\ j = 1, \dots, n, \\ \lambda_i^* &\geq 0,\ i = 1, \dots, l, \\ \lambda_i^* h_i(\boldsymbol{x}^*) &= 0,\ i = 1, \dots, l. \end{aligned}$$

A.3 Miscellaneous

Integration by Parts Suppose that $u(x)$ and $v(x)$ are both functions of x and differentiable for $x \in [a, b]$. Then by the product rule of differentiation

$$d[u(x)v(x)] = v(x)du(x) + u(x)dv(x).$$

Then integrate both sides on $[a, b]$ and get

$$\int_a^b u(x)dv(x) = u(x)v(x)|_a^b - \int_a^b v(x)du(x).$$