Optimal Control

Written by Paul Schrimpf and modified by Hiro Kasahara
October 14, 2020
University of British Columbia
Economics 526

1. References

These notes are about optimal control. References from our text books are chapter 10 of Dixit (1990), chapter 20 Chiang and Wainwright (2005), and chapter 12.2 of De la Fuente (2000) (and chapter 13 for more examples). Other readily available useful references invclude Dorfman (1969) who derives the maximum principle in a model of capital accumulation and Intriligator (1975) which has some economic examples. There are two textbooks about optimal control in economics available online through UBC libraries. Sethi and Thompson (2000) focuses on examples. Caputo (2005) also has many examples, but goes into a bit more mathematical detail.

Although more advanced than what these notes cover, Luenberger (1969) is the classic mathematics text on optimal control and is excellent. Clarke (2013) is available online through UBC libraries and covers similar material as Luenberger (1969), but at a more advanced level.

2. Introduction

In the past few lectures we have focused on optimization problems of the form

$$\max_x f(x) \text{ s.t. } h(x) = c$$

where $x \in \mathbb{R}^n$. The variable that we are optimizing over, x, is a finite dimensional vector. There are interesting optimization problems in economics that involve an infinite dimensional choice variable. The most common example are models where something is chosen at every instant of time.

Example 2.1. [Optimal growth] Consider an simple one sector economy. The only input to production is capital. Output can be used for consumption or investment.

The optimal paths of consumption and capital solve

$$\max_{c(t),k(t)} \int_0^\infty e^{-\delta t} u(c(t))dt$$
$$\text{s.t. } \frac{dk}{dt} = f(k) - \phi k - c$$
$$k(0) = k_0$$
$$0 \leq c \leq f(k)$$

Here, δ is the discount rate, ϕ is the depreciation rate, and we will assume that both f and u are increasing, twice differentiable, and concave (have negative second derivatives).

Example 2.2. [Investment costs] A firm that produces output from capital and labor. The price of output is p and the wage is w. The rate of change of capital is equal to investment minus current capital times the depreciation rate, δ. The price of capital goods is q. It is costly for the firm to adjust its capital stock (because, e.g. the firm might have to temporarily shut down to install new equipment). The adjustment cost is given by $c(i(t), k(t))$. It is such that $c(0, k) = 0$, and $i\frac{\partial c}{\partial i} \geq 0$. The firm's problem is

$$\max_{k(t),i(t),l(t)} \int_0^T e^{-rt} \left[pf(k(t), l(t)) - qi(t) - wl(t) - c(i(t), k(t))\right] dt$$
$$\text{s.t. } \frac{dk}{dt} = i(t) - \delta k(t)$$
$$k(0) = k_0$$
$$k(t) \geq 0$$
$$l(t) \geq 0.$$

All of these examples have a common structure. They each have the following form:

$$\max_{x(t),y(t)} \int_0^T F(x(t), y(t), t)dt$$
$$\text{s.t.}$$
$$\frac{dy}{dt} = g(x(t), y(t), t) \forall t \in [0, T]$$
$$y(0) = y_0$$

This is a generic continuous time optimal control problem. x is called a control variable, and y is called a state variable. The choice of x controls the evolution of y through the first constraint. We will now derive some necessary conditions for a maximum.

3. The maximum principle

We will now derive some necessary conditions on the maximizers of an optimal control problem. That is, we will come up with a first order condition. We will do this in two ways. First, we will approximate the continuous problem with a discrete one, solve the discrete problem, and then take a limit to make the solution continuous again. Second, we will set up a Lagrangian for the above problem and then take functional derivatives to get the first order condition. Both approaches will lead to the same solution.

3.1. **Discretization.** Let's begin by approximating our continuous time problem with a discrete time problem. This will be useful because we already know how to solve optimization problems with a finite dimensional choice variable.

The integral can be approximated by a sum. If divide $[0, T]$ into J intervals of length Δ, we have

$$\int_0^T F(x(t), y(t), t)dt \approx \sum_{j=1}^{T} \Delta F(x(\Delta j), y(\Delta j), \Delta j).$$

Similarly, we can approximate $\frac{dy}{dt} \approx \frac{y(\Delta j) - y(\Delta(j-1))}{\Delta}$. Thus, we can approximate the whole problem by:

$$\begin{aligned} \max_{x_1,\ldots,x_J,y_1,\ldots,y_J} \sum_{j=1}^{J} \Delta F(x_j, y_j, \Delta j) & \qquad \text{s.t.} \\ y_j - y_{j-1} = \Delta g(x_j, y_j, \Delta j) \text{ for} j = 1, \ldots, J & \end{aligned}$$

where we are letting $x_j = x(\Delta j)$, $y_j = y(\Delta j)$. This is just a usual optimization problem with the Lagrangian given by

$$\mathcal{L} = \sum_{j=1}^{J} \Delta F(x_j, y_j, \Delta j) - \lambda_j (y_j - y_{j-1} - \Delta g(x_j, y_j, \Delta j)).$$

The first order conditions are

$$\begin{aligned} [x_j]: && \Delta \frac{\partial F}{\partial x} + \lambda_j \Delta \frac{\partial g}{\partial x} &= 0 \\ [y_j]: && \Delta \frac{\partial F}{\partial y} - \lambda_j + \lambda_{j+1} + \lambda_j \Delta \frac{\partial g}{\partial y} &= 0 \\ [\lambda_j]: && y_j - y_{j-1} - \Delta g(x_j, y_j, \Delta j) &= 0. \end{aligned}$$

Each of these equations hold for $j = 1, \ldots, J$. Also, since there is no $J + 1$ constraint, we set $\lambda_{J+1} = 0$. Rearranging these slightly gives

$$\begin{aligned} [x_j]: && \frac{\partial F}{\partial x} + \lambda_j \frac{\partial g}{\partial x} &= 0 \\ [y_j]: && \frac{\partial F}{\partial y} + \lambda_j \frac{\partial g}{\partial y} &= -\frac{\lambda_{j+1} - \lambda_j}{\Delta} \\ [\lambda_j]: && g(x_j, y_j, \Delta j) &= \frac{y_j - y_{j-1}}{\Delta}. \end{aligned}$$

Taking the limit as $\Delta \to 0$ to go back to continuous time, these become,

$$\begin{aligned}
[x_j]: && \frac{\partial F}{\partial x} + \lambda(t)\frac{\partial g}{\partial x} &= 0 \\
[y_j]: && \frac{\partial F}{\partial y} + \lambda(t)\frac{\partial g}{\partial y} &= -\frac{d\lambda}{dt} \\
[\lambda_j]: && g(x_j, y_j, \Delta j) &= \frac{dy}{dt}.
\end{aligned}$$

Any optimal $x(t)$, $y(t)$, and $\lambda(t)$ must satisfy these equations. This result is known as Pontryagin's maximum principle. It is often stated by defining the Hamiltonian,

$$H(x, y, \lambda, t) = F(x(t), y(t), t) + \lambda(t) g(x(t), y(t), t).$$

Using the Hamiltonian, the three equations can be written as

$$\begin{aligned}
[x]: && 0 &= \frac{\partial H}{\partial x}(x^*, y^*, \lambda^*, t) \\
[y]: && -\frac{d\lambda}{dt}(t) &= \frac{\partial H}{\partial y}(x^*, y^*, \lambda^*, t) \\
[\lambda]: && \frac{dy}{dt}(t) &= \frac{\partial H}{\partial \lambda}(x^*, y^*, \lambda^*, t).
\end{aligned}$$

The $[y]$ equation is called the co-state equation, and λ are called co-state variables (λ is still also called the Lagrange multiplier).

Theorem 3.1 (Pontryagin's maximum principle)**.** *Consider*

$$\begin{aligned}
& \max_{x,y} \int_0^T F(x(t), y(t), t) dt \\
& s.t. \frac{dy}{dt} = g(x(t), y(t), t) \quad \forall t \in [0, T] \\
& y(0) = y_0.
\end{aligned} \tag{1}$$

where x and y are functions from $[0, T]$ to $\mathbb{R}$, and $F, g : \mathbb{R}^2 \times [0, T] \to \mathbb{R}$ are continuously differentiable. Define the Hamiltonian as

$$H(x, y, \lambda, t) = F(x(t), y(t), t) + \lambda(t) g(x(t), y(t), t).$$

If x^* and y^* are local constrained maximizers of (1), then there exists $\lambda^*(t)$ such that

$$\begin{aligned}
[x]: && 0 &= \frac{\partial H}{\partial x}(x^*, y^*, \lambda^*, t) \\
[y]: && -\frac{d\lambda}{dt}(t) &= \frac{\partial H}{\partial y}(x^*, y^*, \lambda^*, t) \\
[\lambda]: && \frac{dy}{dt}(t) &= \frac{\partial H}{\partial \lambda}(x^*, y^*, \lambda^*, t)
\end{aligned}$$

1

[1] The condition for x is often stated somewhat more generally as

$$H(x^*, y^*, \lambda^*, t) = \max_x H(x, y^*, \lambda^*, t).$$

3.2. **Lagrangian approach.** Although it leads to a correct maximum principle, the above discretization approach has some downsides. One is that the derivation is somewhat tedious, and it only applies to problems of the form we considered. Similar problems will still have something like a Hamiltonian and maximum principle, but the details can change. Discretizing every problem could be time consuming and error prone. More importantly, it is challenging to prove that discretizing actually gives the correct solution. Specifically, it is difficult to show that the limit as $\Delta \to 0$ of the discretized solution actually solves the continuous problem.

History 3.1. Pontryagin laid out the maximum principle around 1950, a remarkably recent date. Mathematicians have been studying related problems for hundreds (or perhaps thousands) of years. Various physical and geometric problems lead to a maximization of the form

$$\max_{x} \int_a^b F(t, y(t), y'(t))dt \; s.t. \; y(a) = A, \; y(b) = B. \tag{2}$$

Hero of Alexandria studied one such problem in around 0 A.D. Many of the mathematicians associated with the development of calculus in the 17th century works on such problems, including Newton, Leibniz, Bernoulli, and l'Hôpital. In 1744, Euler made the first systematic study of (2) and showed that a necessary condition for y^* to be a maximizer is that

$$\frac{d}{dt}\frac{\partial F}{\partial y'}(t, y(t), y'(t)) = \frac{\partial F}{\partial y}(t, y(t), y'(t)),$$

which is why we still call similar results Euler(-Lagrange) equations. Euler proved this by discretization. In 1755, Lagrange approached (2) using a different technique. Instead of discretization, Lagrange worked with functional directional derivatives. That is, Lagrange considered what happens to the problem when you replace $x(t)$ with $x(t)+v(t)$ for some small function v. He referred to the direction, v, as a variation. Lagrange wrote a letter to Euler describing this new approach. Euler instantly agreed that it was more elegant and began referring to this area of mathematics as the calculus of variations.

Optimal control problems were not systematically studied until 200 years later. Part of this reason was practical. In the 18th and 19th century, there were many practical problems of the form (2), but not very many of the more general optimal control form:

$$\begin{aligned} &\max_{x(t),y(t)} \int_0^T F(x(t), y(t), t)dt \\ &\text{s.t.} \\ &\frac{dy}{dt} = g(x(t), y(t), t) \forall t \in [0, T] \\ &y(0) = y_0. \end{aligned} \tag{3}$$

If Euler and Lagrange had thought (3) was important, they likely could have derived the maximum principle. Optimal control problems were studied in the 1950s because questions like how to launch a satellite using the least amount of fuel, how to land a spaceship on the moon as softly as possible, and how to fire an ICBM from one side of the world to the other as quickly as possible all have this form. Pontryagin was a Soviet mathematician.[a]

One technical difference between classic calculus of variation problems (2) and optimal control problems (3) is that calculus of variation problems typically have solutions that are twice continuously differentiable, but optimal control problems (3) often have y^* only once differentiable, and x^* discontinuous. Euler and Lagrange may have been able to handle this difference. However, one thing that modern mathematics does, but 18th century mathematics could not have done, is give general conditions under which a maximum exists.

[a] At roughly the same time that Pontryagin developed optimal control, Richard Bellman, an American mathematician, developed dynamic programming to solve the same sort of problems.

Discretization can be avoided by considering the similarities between optimal control problems and the finite optimization problems we studied earlier. A powerful technique for solving new problems is to think about the features the new problem shares with old ones. Then, define and study a new abstraction that subsumes both the old and new problem. This process can also give new insights to the old problem, and potentially apply to other problems that you were not even thinking about. We will do some of this here.

To begin, recall how we arrived at the first order condition for standard optimization problems. We considered looking at small changes around a maximizer and approximated how the objective function would change using directional derivatives. We can apply a similar idea here. For $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$, a directional derivative is:

$$df(x;v) = \lim_{\alpha\to 0} \frac{f(x+\alpha v) - f(x)}{\alpha}$$

where α is a scalar and $v \in \mathbb{R}^n$ is a direction. When $x(\cdot)$ is a function, we can apply the same idea. Let Q be a function from some set of functions to $\mathbb{R}$. For example, in the optimal control example, $Q(x,y) = \int_0^T F(x(t),y(t),t)dt$, where $x : [0,T]\to\mathbb{R}$, $y : [0,T]\to\mathbb{R}$, $F : \mathbb{R}^3\to\mathbb{R}$. Let $v : [0,T]\to\mathbb{R}$ and $w : [0,T]\to\mathbb{R}$ be another function that we'll think of as a direction. We'll consider changing x by v and y by w. The directional derivative of Q with respect to in direction v, w is

$$dQ(x,y;v,w) = \lim_{\alpha\to 0} \frac{Q(x+\alpha v, y+\alpha w) - Q(x,y)}{\alpha}$$

For example, when $Q(x, y) = \int_0^T F(x(t), y(t), t)dt$, then

$$
\begin{aligned}
dQ(x, y; v, w) &= \lim_{\alpha \to 0} \frac{\int_0^T F(x(t) + \alpha v(t), y(t) + \alpha w(t), t) - F(x(t), y(t), t)dt}{\alpha} \\
&\text{assuming we can interchange the limit and integral} \\
&= \int_0^T \lim_{\alpha \to 0} \frac{F(x(t) + \alpha v(t), y(t) + \alpha w(t), t) - F(x(t), y(t), t)}{\alpha} dt \\
&= \int_0^T \frac{\partial F}{\partial x}(x(t), y(t), t)v(t) + \frac{\partial F}{\partial y}(x(t), y(t), t)w(t)dt
\end{aligned}
$$

[2] Using directional derivatives the first order condition for (3) can be stated as if x^*, y^* are maximizers, then for all v and w that satisfy the constraints, we have

$$
0 = dQ(x^*, y^*; v, w) = \int_0^T \frac{\partial F}{\partial x}(x(t), y(t), t)v(t) + \frac{\partial F}{\partial y}(x(t), y(t), t)w(t)dt \tag{4}
$$

Just as when maximizing with respect to vectors, the set of v and w that satisfy the constraints is the set of v and w for which the derivative of the constraints in directions v and w is 0. The derivative of the constraints in direction v and w is

$$
0 = \lim_{\alpha \to 0} \frac{g(x^*(t) + \alpha v(t), y^*(t) + \alpha w(t), t) - d(y^* + \alpha w)/dt - [g(x^*(t), y^*(t), t) - dy^*/dt]}{\alpha}
$$

$$
0 = \frac{\partial g}{\partial x}(x^*(t), y^*(t), t)v(t) + \frac{\partial g}{\partial y}(x^*(t), y^*(t), t)w(t) - \frac{dw}{dt} \tag{5}
$$

Thus, the first order condition is that if v and w satisfy (5) then v and w must also satisfy (4). In the previous set of notes, when working with finite dimensional optimization problems, we then asserted that some results from linear algebra imply that this first order condition is equivalent to there existing Lagrange multipliers. These results can be generalized to the current situation by recognizing that both functions and vectors in $\mathbb{R}^n$ share some important properties. Namely, they can be added together and multiplied by scalars. We will return to this idea in the second half of the course. Suppose there exists $\lambda(t)$ such that

$$
\begin{aligned}
-\lambda(t)\frac{\partial g}{\partial x}(x^*(t), y^*(t), t) &= \frac{\partial F}{\partial x}(x^*(t), y^*(t), t) \\
-\lambda(t)\left[\frac{\partial g}{\partial y}(x^*(t), y^*(t), t)w(t) - \frac{dw}{dt}\right] &= \frac{\partial F}{\partial y}(x^*(t), y^*(t), t)w(t)
\end{aligned}
$$

then (5) implies (4). Showing the reverse (i.e. that the first order condition above implies the existence of multipliers) is more difficult. For now, let's take as given that this first order condition can be equivalently expressed using a Lagrangian.

[2]The dominated convergence theorem gives conditions for which $\lim \int something = \int \lim something$. Here, a simple (but not the weakest) sufficient condition is existence of $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial x} \leq M$ for some constant M.

What should the Lagrangian look like here? Recall that our problem is:

$$\max_{x(t),y(t)} \int_0^T F(x(t), y(t), t)dt$$
$$\text{s.t.}$$
$$\frac{dy}{dt} = g(x(t), y(t), t) \forall t \in [0, T]$$
$$y(0) = y_0.$$

The Lagrangian is the objective function minus the sum of the multipliers times the constraints. Here, we have a continuum of constraints on dy/dt, so instead of a sum we should use an integral. The Lagrangian is then

$$L(x, y, \lambda, \mu_0) = \int_0^T F(x(t), y(t), t)dt - \int_0^T \lambda(t) \left(\frac{dy}{dt} - g(x(t), y(t), t) \right) dt - \mu_0(y(0) - y_0)$$

It is somewhat difficult to think about the derivative of L with respect to y because L involves both y and dy/dt. We can get around this by eliminating dy/dt through integration by parts (it would also work okay to differentiate first and then integrate by parts like we did in class).[3] Integrating by parts gives

$$\begin{aligned} L(x, y, \lambda, \mu_0) = & \int_0^T F(x(t), y(t), t)dt + \int_0^T \left(\lambda(t) g(x(t), y(t), t) + y(t)\frac{d\lambda}{dt} \right) dt + \\ & - \lambda(T)y(T) + \lambda(0)y(0) - \mu_0(y(0) - y_0). \end{aligned}$$

We can then differentiate with respect to $x(t)$ and $y(t)$ to get the first order conditions.

$$[x]: \quad \frac{\partial F}{\partial x} + \lambda(t)\frac{\partial g}{\partial x} = 0$$
$$[y]: \quad \frac{\partial F}{\partial y} + \lambda(t)\frac{\partial g}{\partial y} = -\frac{d\lambda}{dt}$$

These, along with the constraint, are once again the Maximum principle.

3.3. **Transversality conditions.** The three equations of the maximum principle are not necessarily enough to determine x, y, and λ. The reason is that they tell us about dy/dt and $d\lambda/dt$ instead of y and λ. For any constant, c, $d(y + c)/dt = dy/dt$, so the three equations only determine y and λ up to a constant. To pin down the constant for y, the problem tells us what $y(0)$ must be. To pin down the constant for λ, there is an extra condition on $\lambda(T)$. This condition is called a transversality condition. Unfortunately, the heuristic derivations we did above obscure the transversality condition somewhat. In the discrete time approach, we set $\lambda_{J+1} = 0$, so we also should impose $\lambda(T) = 0$.

The same condition appears in the Lagrangian approach, if we are more careful about taking derivatives. Taking the derivative of $\int_0^T F(x(t), y(t), t)dt$ with respect to $x(\tau)$ holding $x(\cdot)$ constant for all other periods does not really make sense because changing x at a single point will not change the integral at all. Instead, we need to consider directional

[3] Integration by parts says that $\int_a^b u(x)v'(x)dx = u(b)v(b) - u(a)v(a) - \int_a^b v(x)u'(x)dx$.

derivatives. Let v be another function of t. Then the derivative of L with respect to x in direction v is

$$d_x L(x, y, \lambda, \mu_0; v) = \lim_{\alpha \to 0} \frac{L(x + \alpha v, y, \lambda, \mu_0) - L(x, y, \lambda, \mu_0)}{\alpha} = \frac{d}{d\alpha} L(x + \alpha v, y, \lambda, \mu_0).$$

This is exactly the same as our previous definition of a directional derivative, except now the direction is a function. The first order conditions are that the directional derivatives are zero for all directions (functions) v. Assuming that F is well-behaved so that we can interchange integration and differentiation,[4], the first order conditions are then

$$\begin{aligned}
[x]: \quad 0 &= \int_0^T \frac{\partial F}{\partial x}(x(t), y(t), t) v(t) dt - \int_0^T -\frac{\partial g}{\partial x}(x(t), y(t), t) v(t) \lambda(t) dt \\
&= \int_0^T v(t) \left(\frac{\partial F}{\partial x}(x(t), y(t), t) + \frac{\partial g}{\partial x}(x(t), y(t), t) \lambda(t) \right) dt \\
[y]: \quad 0 &= \int_0^T \frac{\partial F}{\partial y}(x(t), y(t), t) v(t) dt - \int_0^T \left(\frac{\partial v}{\partial t}(t) - \frac{\partial g}{\partial y}(x(t), y(t), t) v(t) \right) \lambda(t) dt - \mu_0 v(0) \\
&= \int_0^T v(t) \left(\frac{\partial F}{\partial y}(x(t), y(t), t) + \frac{\partial g}{\partial y}(x(t), y(t), t) \lambda(t) \right) dt - \int_0^T \frac{dv}{dt}(t) \lambda(t) dt - \mu_0 v(0) \\
&= \int_0^T v(t) \left(\frac{\partial F}{\partial y}(x(t), y(t), t) + \frac{\partial g}{\partial y}(x(t), y(t), t) \lambda(t) \right) dt + \int_0^T \frac{d\lambda}{dt}(t) v(t) dt - \\
&\quad - \lambda(T) v(T) + \lambda(0) v(0) - \mu_0 v(0).
\end{aligned}$$

The last line comes from integration by parts. These conditions must hold for for all functions v. If we take $v(t) = \begin{cases} 0 & \text{if } t < T \\ 1 & \text{if } t = T \end{cases}$, then the second equation requires $\lambda(T) = 0$.

We can also deduce the transversality condition by thinking about the interpretation of the multiplier as the shadow price of the relaxing the constraint. Relaxing the constraint affects the objective function by changing future y. At time T, there is no future y, so the value of relaxing the constraint must be $0 = \lambda(T)$.

It is important to understand where the transversality condition comes from, because the transversality condition can change depending on whether T is finite or infinite and whether or not there are constraints on $y(T)$ or $x(T)$.

Example 3.1. A landlord can rent $y(t)$ units of housing at price $p(t)$. The landlord can adjust her housing at rate $x(t)$ for cost $s(t)x(t) + c(t)x(t)^2$, where $s(t)$ represents the price of buying or selling housing and $c(t)x(t)^2$ is an adjustment cost meant to capture the idea that it can be increasingly costly to buy or sell a large amount at once. The landlord has a finite time horizon and no discounting. The landlord's profit

[4] By Leibniz's rule, F and its partial derivatives being continuous is sufficient.

maximization problem is

$$\begin{aligned}
\max_{x,y} \int_0^T & p(t)y(t) - s(t)x(t) - c(t)x(t)^2 dt \\
\text{s.t.} \frac{dy}{dt} &= x(t) \\
y(0) &= y_0.
\end{aligned}$$

The maximum principle for this problem is that

$$\begin{aligned}
[x]: \quad & 0 = -s(t) - 2c(t)x(t) + \lambda(t) \\
[y]: \quad & -\frac{d\lambda}{dt} = p(t)
\end{aligned}$$

and the tranversality condition is

$$\lambda(T) = 0.$$

Using the transversality condition and the first order condition for $[y]$, we can solve for λ. From the fundamental theorem of calculus, we know that

$$\lambda(t) = \lambda(T) - \int_t^T \frac{d\lambda}{dt}(\tau) d\tau$$

so,

$$\begin{aligned}
\lambda(t) =& 0 - \int_t^T -p(\tau) d\tau \\
\lambda(t) =& \int_t^T p(\tau) d\tau
\end{aligned}$$

We can then substitute this into the first order condition for $[x]$ to solve for $x(t)$,

$$\begin{aligned}
x(t) =& \frac{\lambda(t) - s(t)}{2c(t)} \\
x(t) =& \frac{\int_t^T p(\tau) d\tau - s(t)}{2c(t)}.
\end{aligned}$$

Finally, we can solve for $y(t)$ by integrating the constraint.

$$\begin{aligned}
y(t) =& y(0) + \int_0^t \frac{dy}{dt}(r) dr \\
=& y(0) + \int_0^t \left[\frac{\int_r^T p(\tau) d\tau - s(r)}{2c(r)} \right] dr
\end{aligned}$$

We can evaluate these integrals for some simple $p(t)$, $s(t)$, and $c(t)$. For example, when everything is constant, $p(t) = p$, $s(t) = s$, and $c(t) = c$, then

$$\begin{aligned} x(t) =& \frac{p(T-t)-s}{2c} \\ y(t) =& y(0) + t\frac{pT-s}{2c} - \frac{p}{4c}t^2 \end{aligned}$$

3.4. **Additional constraints .** Many optimal control problems include some additional constraints. In the four examples above, there were often bounds constraints on the control and/or state variables. In general, we might have some constraints of the form $h(x(t), y(t), t) \leq 0$ for all t. It also common for there to be constraints on the initial and/or final values of y. Having additional constraints can change the first order conditions of the maximum principle. You can always figure out the correct first order conditions by starting with the Lagrangian. Either the discrete approach that we took to first arrive at the maximum principle, or the directional derivative approach of the previous section will work.

Let us consider the following problem:

$$\begin{aligned} \max_{x,y} & \int_0^T F(x(t), y(t))dt \\ \text{s.t.} & \frac{dy}{dt} = g(x(t), y(t)) \\ & h(x(t), y(t)) \leq 0 \qquad\qquad y(0) = y_0 \\ & y(T) = y_T \end{aligned}$$

The Lagrangian is

$$\begin{aligned} L(x, y, \lambda, \mu, \psi_0, \psi_T) = & \int_0^T F(x(t), y(t)) - \lambda(t)(\frac{dy}{dt} - g(x(t), y(t))) - \mu(t)h(x(t), y(t))dt + \\ & - \psi_0(y(0) - y_0) - \psi_T(y(T) - y_T). \end{aligned}$$

The first order conditions for x and y are that for any function v,

$$\begin{aligned} [x] : 0 = & \int_0^T v(t) \left(\frac{\partial F}{\partial x}(x(t), y(t)) + \lambda(t)\frac{\partial g}{\partial x}(x(t), y(t)) - \mu(t)\frac{\partial h}{\partial x}(x(t), y(t)) \right) dt \\ [y] : 0 = & \int_0^T v(t) \left(\frac{\partial F}{\partial y}(x(t), y(t)) + \lambda(t)\frac{\partial g}{\partial y}(x(t), y(t)) - \mu(t)\frac{\partial h}{\partial y}(x(t), y(t)) + \frac{d\lambda}{dt} \right) dt + \\ & - \lambda(T)v(T) + \lambda(0)v(0) - \psi_0 v(0) - \psi_T v(T). \end{aligned}$$

Also, the constraints must be met, and $\mu(t) \geq 0$ with complementary slackness to $h(x(t), y(t)) \leq 0$. The first order condition for x implies that

$$\frac{\partial F}{\partial x}(x(t), y(t)) + \lambda(t)\frac{\partial g}{\partial x}(x(t), y(t)) - \mu(t)\frac{\partial h}{\partial x}(x(t), y(t)) = 0.$$

The first order condition for y implies that

$$\frac{\partial F}{\partial y}(x(t), y(t)) + \lambda(t)\frac{\partial g}{\partial y}(x(t), y(t)) - \mu(t)\frac{\partial h}{\partial y}(x(t), y(t)) = -\frac{d\lambda}{dt}.$$

The first order condition for y also implies that $\lambda(T) = -\psi_T$, and $\lambda(0) = \psi_0$. Each of the ψ could be anything, so there is no restriction on either $\lambda(T)$ or $\lambda(0)$. This is okay because the constraints on both the initial and final value of y will be enough to fully determine the solution.

We still have not covered all possible forms of constraint. For example, the optimal taxation problem includes a constraint of the form $\int_0^T h(x(t), y(t))dt \leq 0$, the contracting problem has a constraint involving both the derivative of y and x instead of just y, and the taxation problem has a constraint on $y'(x(t)t)$ instead of just $y'(t)$. Rather than trying to derive a form of the Hamiltonian and maximum principle that encompasses all possible types of constraints, I find it easier to work from the Lagrangian. Nonetheless, theorem A.1 in the appendix gives a version of the maximum principle that encompasses all the examples that we have encountered.

4. Applications

We will begin with some examples where we can explicitly solve for the optimal control and state variables. We will then analyze examples that cannot be explicitly solved, but we can characterize the solution and do some comparative statics. We have three different ways to arrive at the first order conditions for an optimal control problem. (1) We can discretize the problem as in section 3. (2) We can write down the continuous time Lagrangian and take directional (functional) derivatives. (3) We can write down the Hamiltonian and use the maximum principle. If the problem does not have the same form as that considered in theorem (3.1), then the usual Hamiltonian will not give the correct answer. In that case we can either take approach (1) or (2), or perhaps use theorem (A.1).

Example 4.1 (Linear production and savings)**.** Consider an economy that has a linear production function, $y = k$. The model begins at time 0 and lasts until time T. Each instant, output can be either saved to produce capital or consumed. There is no depreciation or discounting. The objective is to maximize consumption. Let $s(t)$ be the portion of output saved at time t. Then the problem can be written as

$$\begin{aligned}
&\max_{s,k} \int_0^T (1 - s(t))k(t)dt \\
&\text{s.t. } \frac{dk}{dt} = s(t)k(t) \\
&\qquad k(0) = k_0 \\
&\qquad k(t) \geq 0 \\
&\qquad 0 \leq s(t) \leq 1
\end{aligned}$$

Notice that this problem has the sorts of constraints considered in section 3.4. The Lagrangian for this problem is

$$L = \int_0^T \left((1-s)k - \lambda(\frac{dk}{dt} - sk) + \mu_k k + \mu_{s0} s - \mu_{s1}(s-1) \right) dt - \psi_0(k(0) - k_0)$$

where s k, λ, μ_k, μ_{s0}, and μ_{s1} are all functions of t.

The first order condition for the control variable, s, is that

$$-k + \lambda k + \mu_{s0} - \mu_{s1} = 0.$$

The first order condition for k is

$$(1-s) + \lambda s + \mu_k = -\frac{d\lambda}{dt}.$$

The transversality condition, which also comes from the first order condition for k, is $\lambda(T) = 0$.

To solve, first notice that if $k(t) \geq 0$ for all t, we must have $k_0 \geq 0$. If $k_0 = 0$, then $dk/dt = 0$ regardless of s, and any choice of $s(t) \in [0,1]$ is a maximizer. Therefore, for the remainder of the solution, we can assume $k_0 > 0$.

Notice that the constraints imply $dk/dt \geq 0$. Therefore if $k_0 > 0$, then for all t, $k(t) \geq k_0 > 0$. At time T, the first order condition for s says that

$$-k(T) + \mu_{s0}(T) - \mu_{s1}(T) = 0.$$

By complementary slackness, if $\mu_{s0}(t) > 0$, then $s(t) = 0$ and $\mu_{s1}(t) = 0$. Conversely, if $\mu_{s1}(t) > 0$, then $s(t) = 1$ and $\mu_{s0}(t) = 0$. Therefore, since $k(T) > 0$, it must be that $\mu_{s0}(T) > 0$ and $s(T) = 0$.

Now, consider other t. The first order condition for s is that

$$k(t)(\lambda(t) - 1) + \mu_{s0}(t) - \mu_{s1}(t) = 0.$$

For the same reason that $s(T) = 0$, we must have $s(t) = 0$ whenever $\lambda(t) < 1$. In that case the first order condition for y is that

$$1 = -\frac{d\lambda}{dt}.$$

Therefore, for t near T, we know that

$$\lambda(t) = \int_T^t -1 d\tau = T - t.$$

This $\lambda(t) < 1$ for $t > T - 1$. At $T - 1$, $\lambda(T-1) = 1$ and

$$(1-s) + s = 1 = -d\lambda/dt$$

regardless of t. Therefore for t just below $T - 1$, we must have $\lambda(t) > 1$. In that case the first order condition requires $s(t) = 1$. Then, $\frac{d\lambda}{dt} = -\lambda(t) < 0$, so for even smaller t, $\lambda(t)$ will be even bigger and $s(t)$ must still be 1.

We can conclude that $s^*(t) = \begin{cases} 1 & \text{if } t < T-1 \\ 0 & \text{if } t \geq T-1 \end{cases}$. For $t < T-1$, we have $dk/dt = k$ and $k(0) = k_0$. This implies that $k(t) = k_0 e^t$. For $t > T-1$, $dk/dt = 0$. Therefore, $k(t) = \begin{cases} k_0 e^t & \text{if } t < T-1 \\ k_0 e^{T-1} & \text{if } t \geq T-1 \end{cases}$. The maximum is then $k_0 e^{T-1}$ if $T > 1$ and $k_0 T$ if $T < 1$.

Example 4.2 (Inventory)**.** A dairy has an order for y_T units of cheese to be delivered at time T at price p. Currently, the firm has $y_0 = 0$ units available. Producing at a rate of $x(t)$ costs the firm $cx(t)^2$. Storing cheese requires refrigeration, so it is costly. Storing $y(t)$ units costs $sy(t)$. The dairy chooses its production schedule to maximize profits.

$$
\begin{aligned}
\max_{x,y} & p y_T - \int_0^T \left(cx(t)^2 + sy(t)\right) dt \\
\text{s.t. } & \frac{dy}{dt} = x(t) \\
& y(T) = y_T \\
& y(0) = 0 \\
& x(t) \geq 0
\end{aligned}
$$

Let's solve this problem using the maximum principle. This problem includes inequality constraints for all t, so we should use the form of the maximum principle derived in section 3.4.

$$
\begin{aligned}
[x]: & \quad -2cx + \lambda + \mu = 0 \\
[y]: & \quad -s = -\frac{d\lambda}{dt}
\end{aligned}
$$

Since $d\lambda/dt = s$ is constant, $\lambda(t)$ must be linear with slope s, i.e.

$$\lambda(t) = st + \lambda(0).$$

Substituting that into the first order condition we have

$$
\begin{aligned}
-2cx(t) + st + \lambda(0) + \mu(t) &= 0 \\
x(t) &= \frac{st + \mu(t) + \lambda(0)}{2c}.
\end{aligned}
$$

There is also a complementary slackness condition on μ and x. First, suppose that x is always positive, then $\mu(t) = 0$, and we can solve for $\lambda(0)$ using the initial and final y.

$$
\begin{aligned}
y(T) - y(0) &= \int_0^T \frac{dy}{dt} dt \\
y_T - 0 &= \int_0^T x(t) dt \\
&= \int_0^T \frac{st + \lambda(0)}{2c} \\
&= \frac{sT^2 + 2\lambda(0)T}{4c} \\
\lambda(0) &= \frac{1}{T} 2c y_T - sT
\end{aligned}
$$

Substituting,

$$x(t) = \frac{s}{2c}(t - T) + \frac{y_T}{T}.$$

This $x(t) \geq 0$ for all t if $y_T \geq s/2cT^2$.

If this inequality does not hold, then since storage is costly, we should expect that if production, x, is ever 0, it would be for early t. Let $\bar{t}$ be the time at which x is first non-zero. Then,

$$
\begin{aligned}
y_T &= \int_{\bar{t}}^T \frac{st + \lambda(0)}{2c} \\
y_T &= \frac{s(T^2 - \bar{t}^2) + 2\lambda(0)(T - \bar{t})}{4c} \\
\lambda(0) &= y_T \frac{2c}{T - \bar{t}} - s(T + \bar{t}).
\end{aligned}
$$

Finally, we can use $x(\bar{t}) = 0$ to solve for $\bar{t}$.

$$\bar{t} = T - \frac{y_T 2c}{Ts}.$$

Thus,

$$x(t) = \begin{cases} \frac{s}{2c}(t - T) + \frac{y_T}{T} & \text{if } t > T - \frac{y_T 2c}{Ts} \\ 0 & \text{otherwise} \end{cases}$$

4.1. **Optimal growth.** Let's characterize the solution to our optimal growth model in example 2.1. The optimal growth problem is

$$
\begin{aligned}
\max_{c(t), k(t)} & \int_0^\infty e^{-\delta t} u(c(t)) dt \\
\text{s.t. } & \frac{dk}{dt} = f(k(t)) - \phi k(t) - c(t) \\
& k(0) = k_0 \\
& 0 \leq c(t) \leq f(k(t))
\end{aligned}
$$

The Lagrangian is

$$L = \int_0^\infty \left[e^{-\delta t} u(c) - \lambda(dk/dt - f(k) + \phi k + c) + \mu_0 c - \mu_1(c - f(k)) \right] dt - \psi(k(0) - k_0)$$

The first order conditions are:

$$[c]: \qquad e^{-\delta t} u'(c(t)) - \lambda(t) + \mu_0(t) - \mu_1(t) = 0$$

$$[k]: \qquad \lambda(t)\left(f'(k(t)) - \phi\right) + \mu_1(t) f'(k(t)) = -\frac{d\lambda}{dt}.$$

The multipliers on the inequality constraints are somewhat annoying to deal with. Fortunately, under standard conditions, the constraints will not bind and their multipliers are 0. For example, if $\lim_{c \to 0} u'(c) = \infty$, as is the case for many specifications of u, then the optimal $c(t) > 0$ for all t and $\mu_0(t) = 0$. Also, when k_0 is small, it turns out that $c(t) < f(k(t))$ along the optimal path. Setting $\mu_0 = \mu_1 = 0$ and differentiating $[c]$ with respect to time gives

$$-\delta e^{-\delta t} u'(c) + e^{-\delta t} u''(c) \frac{dc}{dt} = \frac{d\lambda}{dt}.$$

Substituting into $[k]$ and rearranging gives

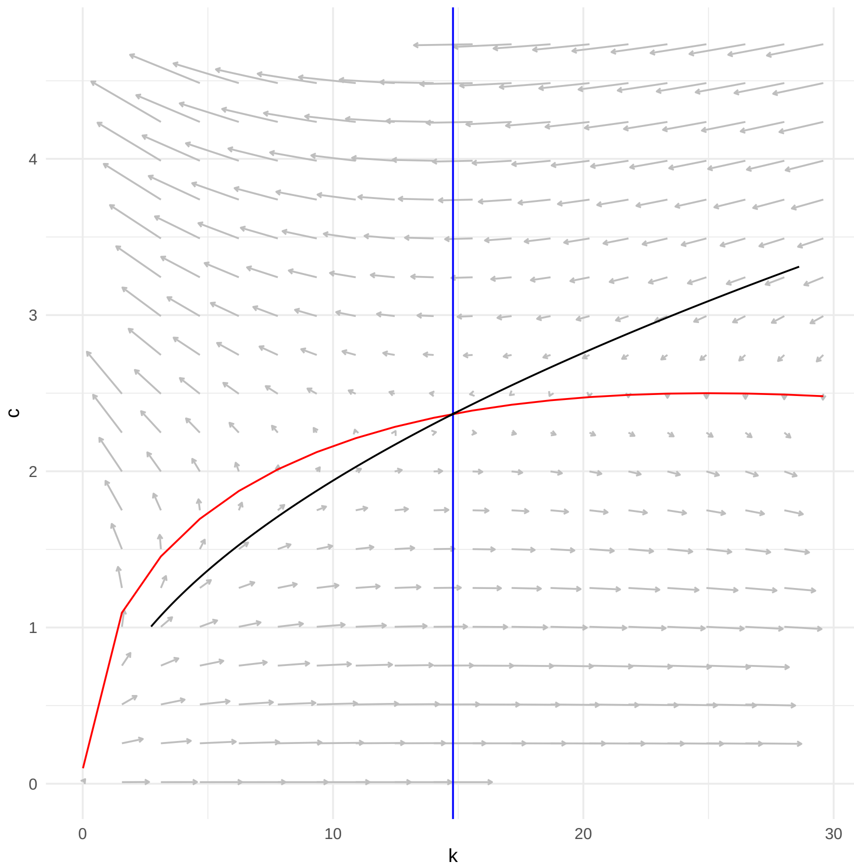
$$\frac{dc}{dt} = -\frac{u'(c)}{u''(c)} \left(f'(k(t)) - \phi - \delta\right)$$

This equation along with the constraint on k give us a pair of differential equations expressing dc/dt and dk/dt as functions of c and k. We can plot dc/dt and dk/dt as functions of c and k in a phase diagram as shown in figure 1. First, we can plot the line where $dk/dt = 0$. This is just $c = f(k) - \phi k$. Above this red line, $dk/dt < 0$, and below $dk/dt > 0$. Then, we can also plot the line where $dc/dt = 0$ in blue. This is just where $f'(k^*) = \phi + \delta$. Usually we assume f is concave, so then to the left of this line, $dc/dt > 0$ and to the right $dc/dt < 0$.

Given any initial $k(0)$ we can use the phase diagram to trace out the path of k and c. Initial capital, $k(0) = k_0$ is fixed. The optimal $c(0)$ will be such that the subsequent path converges to the steady-state. The steady-state is where both capital and consumption are constant. In the phase diagram, the steady-state is the intersection of the blue and red lines. For each possible k_0, there is a unique $c(0)$ (on the black line in the figure) that leads to the stead state. This stable path is shown in black in the figure. If $k_0 < k^*$, then any $c(0)$ above the red cannot be optimal because then capital just decreases further. If $c(0)$ is too low, then capital increases, but consumption initially increases then decreases and fails to reach the steady-state.

The phase diagram is also useful to describing what will happen if some part of the model changes unexpectedly. For example if productivity increases, so that the production function changes from $f(k)$ to $Af(k)$ with $A > 1$, then the blue $dc/dt = 0$ curve will shift to the right and the red $dk/dt = 0$ curve will rotate upward. There will be a new stable path to go with the shifted curves. Suppose we start from the old steady-steady. If $f(k)$ shifts at time T, then $k(T)$ will remain at the old stead-state capital and $c(T)$ will jump immediately to the new stable path. As time progresses k and c will adjust toward the new steady-state.

FIGURE 1. Phase diagram



Code for figure: https://bitbucket.org/paulschrimpf/econ526/src/master/02-optimalControl/phase.R?at=master

In the optimal growth model, we often consider the following transversality condition:

$$\lim_{t\to\infty} k(t)\lambda(t)e^{-\delta t} = 0,$$

which provides the sufficient condition for the optimality. This has an intuitive interpretation that the present value of the shadow value of future capital stock goes to 0 as $t \longrightarrow \infty$. Because the first order condition is given by $\frac{d\ln\lambda(t)}{dt} = \frac{d\lambda(t)/dt}{\lambda(t)} = -(f'(k(t)) - \phi)$, we have $\lambda(T) = \lambda(0)\exp(-\int_0^T (f'(k(s)) - \phi)ds)$. Therefore, assuming that $\lambda(0) < \infty$, we may rewrite this transversality condition as

$$\lim_{t\to\infty}\left[k(t)\exp\left(-\int_0^t (f'(k(s)) - \phi - \delta)ds\right)\right] = 0.$$

4.1.1. *Transversality conditions for infinite horizon problems.* Transversality conditions for infinite horizon problems are somewhat more delicate than for finite horizon problems.

Consider the problem:

$$\max_{x,y} \int_0^\infty F(x(t), y(t), t)dt$$
$$\text{s.t.} \frac{dy}{dt} = g(x(t), y(t))$$
$$y(0) = y_0$$

As with a finite horizon problem, writing down the Lagragian and looking at the first order condition for y yields:

$$0 = \int_0^\infty \left(\frac{\partial F}{\partial y}(x(t), y(t), t) + \lambda(t)\frac{\partial g}{\partial y}(x(t), y(t), t) + \frac{d\lambda}{dt}(t) \right) v(t)dt - \lim_{T\to\infty} \lambda(T)v(T) + \lambda(0)v(0) - \mu v(0).$$

This implies the transversality condition that

$$\lim_{T\to\infty} \lambda(T)v(T) = 0 \tag{6}$$

for all v that are valid perturbations to y. The difficulty here is that the set of v that are allowed affects what (6) implies about λ. For example, if we allow v to be any function, then we would need $\lambda(T) = 0$ for T sufficiently large (if not, set $v(T) = 1/\lambda(T)$ and then $\lim_{T\to\infty} \lambda(T)v(T) \neq 0$). However, in most cases v cannot be allowed to be any function. For the derivative in direction v and the first order condition to make sense,

$$\int_0^\infty F(x(t), y(t) + \alpha v(t), t)dt$$

must exist for small α. Additionally, to ensure the existence of Lagrange multiplies, the set of allowed v must be rich enough to behave like the set of vectors in $\mathbb{R}^n$. Specifically, when you add two allowed directions, the sum must also be allowed. For most finite horizon problems and many infinite horizon problems with discounting, the set of allowed v is the set of bounded functions. In that case, (6) implies $\lim_{T\to\infty} \lambda(T) = 0$. (Unfortunately, this is not true of all problems, see example 4.3).

Determining right set of allowed pertubations for a given problem can be difficult. For a well behaved problem with a unique steady state and an optimal path that leads to the steady state, the transversality condition is not needed to characterize the optimal path. Any infinite horizon problem that I will ask about in this course will be of that sort.

In a class of models with $F(x(t), y(t), t) = e^{-\delta t}u(x(t))$ such as the optimal growth model in section 4.1, we may prove that the transversality condition given by

$$\lim_{t\to\infty} y(t)\lambda(t)e^{-\delta t} = 0$$

together with the concavity of the "maximized" Hamiltonian

$$\tilde{H}(y, \lambda, t) := \max_x F(x(t), y(t), t) + \lambda(t)g(x(t), y(t), t)$$

provide sufficient conditions for the optimal solution (see Theorem 2.4 of De la Fuente (2000)).

A useful transversality condition for infinite horizon problems can be derived by looking at the set of directions such that

$$\int_0^\infty F(x(t) + \alpha v(t), y(t) + \alpha w(t), t)dt$$

is finite and the constraints are satisfied. Michel (1982) takes such an approach and shows that this leads to

$$\lim_{T\to\infty} F(x(T), y(T), T) + \lambda(T)g(x(T), y(T)) = 0.$$

See Acemoglu (2008) or Chiang (2000) for further discussion.

Example 4.3 (Non-standard transversality conditions)**.** The following two examples feature solutions with $\lim_{T\to\infty} \lambda(T) \neq 0.$, but that satisfy the condition

$$\lim_{T\to\infty} F(x(T), y(T), T) + \lambda(T)g(x(T), y(T)) = 0.$$

(1)

$$\begin{aligned} \max_{x,y} & \int_0^\infty (1 - y(t))x(t)dt \\ \text{s.t.} & \frac{dy}{dt} = (1 - y(t))x(t) \\ & y(0) = 0 \\ & 0 \leq x(t) \leq 1 \end{aligned}$$

(2) Let $k^s = (\alpha/\delta)^{1/(1-\alpha)}$ and $c^s = (k^s)^\alpha - \delta k^s$.

$$\begin{aligned} \max_{c,k} & \int_0^\infty \left(\log(c(t)) - \log(c^s)\right) dt \\ \text{s.t.} & \frac{dk}{dt} = k(t)^\alpha - c(t) - \delta k(t) \\ & k(0) = 1 \\ & k(t) \geq 0 \end{aligned}$$

Appendix A. Generalized maximum principle

The theorem below is a generalized version the maximum principle. I find it easier to work from the Lagrangian than to try to remember or apply this theorem. We will suppose we are choosing n instead of just 2 functions of t. We will denote the n functions by $z(t)$ collectively, and $z_j(t)$ individually. We will denote their derivatives by $\frac{dz_j}{dt} = \dot{z}_j$ individually, and simply $\dot{z}(t)$ for the vector of n derivatives. There will some constraints on the derivatives, which we will express as

$$G_m(z(t), \dot{z}, t) = 0$$

where $G_m : \mathbb{R}^{2n} \times [0, T] \to \mathbb{R}$ is continuously differentiable. For example, the canonical optimal control problem has $n = 2$, $z(t) = (x(t), y(t))$, and

$$G_m(z(t), \dot{z}(t), t) = \dot{y}(t) - g(x(t), y(t), t).$$

Theorem A.1 (Generalized maximum principle)**.** *Let* $z : [0,T] \rightarrow \mathbb{R}^n$. *Let* $\dot{z} : [0,T] \rightarrow \mathbb{R}^n$ *denote the derivatives of* z. *Consider*

$$
\begin{aligned}
\max_{z} \int_0^T & F(z(t),t)dt \\
& s.t. \\
& 0 = G_m(z(t),\dot{z}(t),t) \forall t \in [0,T], m \in 1,\ldots,M && (7) \\
& \int_0^T h_k(z(t),t)dt = 0 \; for \; k = 1,\ldots,K && (8) \\
& z(0) = Z_0 \quad , \quad z(T) = Z_T && (9)
\end{aligned}
$$

Assume that F, G, h *are continuously differentiable. If* z^* *is a local constrained maximizer, then there exists functions* $\lambda_m(t)$ *and constants* $\mu_k, \psi_{j,l}$ *such that*

$$[z_j]: \quad 0 = \frac{\partial F}{\partial z_j}(t) - \sum_{m=1}^{M} \lambda_m(t) \left(\frac{\partial G_m}{\partial z_j}(t) - \sum_{i=1}^{n} \frac{\partial^2 G_m}{\partial \dot{z}_j \partial z_i}(t) \dot{z}_i(t) \right) + \sum_{m=1}^{M} \frac{d\lambda_m}{dt}(t) \frac{\partial G_m}{\partial \dot{z}_j}(t) - \sum_{k=1}^{K} \psi_k \frac{\partial h_k}{\partial z_j}(t)$$

and

$$
\begin{aligned}
-\psi_{j,0} &= \sum_{m=1}^{M} \lambda_m(0) \frac{\partial G}{\partial \dot{z}_j}(0) \\
\psi_{j,T} &= \sum_{m=1}^{M} \lambda_m(T) \frac{\partial G}{\partial \dot{z}_j}(T).
\end{aligned}
$$

Exercise A.1. Verify that theorem is the same as the previous maximum principle (3.1), if we set $n = 2$, $z(t) = (x(t), y(t))$, $M = 1$, $K = 0$, and

$$G_1(z(t), \dot{z}(t), t) = \frac{dy}{dt} - g(x(t), y(t), t).$$

Exercise A.2 (Difficult)**.** Prove this theorem from the continuous time Lagrangian or by discretizing.

References

Acemoglu, Daron. 2008. *Introduction to modern economic growth*. Princeton University Press.

Caputo, Michael R. 2005. *Foundations of Dynamic Economic Analysis: Optimal Control Theory and Applications*. Cambridge University Press. URL http://lib.myilibrary.com.ezproxy.library.ubc.ca/Open.aspx?id=16326.

Chiang, AC. 2000. *Elements of dynamic optimization*. Illinois: Waveland Press Inc.

Chiang, AC and K Wainwright. 2005. "Fundamental methods of mathematical economics." *McGraw-Hill, New York* .

Clarke, Francis. 2013. *Functional analysis, calculus of variations and optimal control*, vol. 264. Springer.

De la Fuente, Angel. 2000. *Mathematical methods and models for economists*. Cambridge University Press.

Diamond, Peter A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *The American Economic Review* 88 (1):pp. 83–95. URL http://www.jstor.org/stable/116819.

Dixit, Avinash K. 1990. *Optimization in economic theory*. Oxford University Press Oxford.

Dorfman, Robert. 1969. "An Economic Interpretation of Optimal Control Theory." *The American Economic Review* 59 (5):pp. 817–831. URL http://www.jstor.org/stable/1810679.

Intriligator, Michael D. 1975. "Applications of Optimal Control Theory in Economics." *Synthese* 31 (2):pp. 271–288. URL http://www.jstor.org/stable/20115071.

Luenberger, David G. 1969. *Optimization by vector space methods*. John Wiley & Sons.

Michel, Phillippe. 1982. "On the Transversality Condition in Infinite Horizon Optimal Problems." *Econometrica* 50 (4):975–985. URL http://www.jstor.org/stable/1912772.

Mirrlees, J. A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *The Review of Economic Studies* 38 (2):pp. 175–208. URL http://www.jstor.org/stable/2296779.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *The Review of Economic Studies* 68 (1):205–229. URL http://restud.oxfordjournals.org/content/68/1/205.abstract.

Sethi, Suresh P and Gerald L. Thompson. 2000. *Optimal Control Theory: Applications to Management Science and Economics*. Springer-Verlag. URL http://site.ebrary.com/lib/ubc/detail.action?docID=10133718.