RANDOM LAGRANGE MULTIPLIERS AND TRANSVERSALITY

1. Introduction

Lagrange multiplier methods are standard fare in elementary calculus courses, and they play a central role in economic applications of calculus because they often turn out to have interpretations as prices or shadow prices. You have seen them generalized to cover dynamic, non-stochastic models as Hamiltonian methods, or as byproducts of using Pontryagin's maximum principle. In static models Lagrangian methods reduce a constrained maximization problem to an equation-solving problem. In dynamic models they result in an ordinary differential equation problem.

In the stochastic models we are about to consider they result in, for discrete time, an integral equation problem or, in continuous time, a partial differential equation problem. Integral equations and partial differential equations are harder to solve than ordinary equations or differential equations — they are both less likely to have an analytical solution and more difficult to handle numerically. The application of Lagrangian methods to stochastic dynamic models therefore appears to be of less help in solving the optimization problem than is their application to non-stochastic problems. Consequently many references on dynamic stochastic optimization give little attention to Lagrange multipliers, instead emphasizing more direct methods for obtaining solutions.

The economic literature has to some extent been guided by this pattern of emphasis. This is unfortunate, because Lagrangian methods are as helpful in economic interpretation of models in stochastic as in non-stochastic models. Also, in general equilibrium models, use of Lagrangian methods turns out sometimes to simplify the computational problem, in comparison to approaches that try to solve by more direct methods all the separate optimizations embedded in the general equilibrium.

2. Statement of the Problem and the Euler Equation First Order Conditions

Since in this course we are more interested in using these results than in proving them, we present them backwards. That is, we begin by writing down the result we are aiming at, then prove that it is part of a set of sufficient conditions for an optimum. The first-order conditions we display are in fact also necessary conditions for an optimum under regularity conditions that often apply in economic models, but we do not in this set of notes prove that. A more complete presentation, that however gives less attention to infinite-horizon problems, is in Kushner (1965b) and (Kushner, 1965a).

Note that in this course you will be responsible for knowing how to use the conditions displayed in these notes to analyze and solve economic models, not for reproducing proofs of necessity or sufficiency.

We consider a problem of the form

$$\max_{C_0^\infty} E\left[\sum_{t=0}^{\infty} \beta^t U_t\left(C_{-\infty}^t, Z_{-\infty}^t\right)\right] \tag{1}$$

subject to

$$g_t\left(C_{-\infty}^t, Z_{-\infty}^t\right) \leq 0,\ t = 0, ..., \infty, \tag{2}$$

where we are using the notation $C_m^n = \{C_s, s = m, \ldots, n\}$.

We assume that the vector Z is an exogenous stochastic process, that is, that it cannot be influenced by the vector of variables that we can choose, C. For a dynamic, stochastic setting, the information structure is an essential aspect of any problem statement. Information is revealed over time, and decisions made at a time t can depend only on the information that has been revealed by time t. Here, we assume that what is known at t is $Z_{-\infty}^t$, i.e. current and past values of the exogenous variables in the model.[1] The class of stochastic processes C that have this property are said to be ***adapted*** to the information structure.

We can generate first order conditions for this problem by first writing down a Lagrangian expression,

$$E\left[\sum_{t=0}^{\infty} \beta^t U_t\left(C_{-\infty}^t, Z_{-\infty}^t\right) - \sum_{t=0}^{\infty} \beta^t \lambda_t g_t\left(C_{-\infty}^t, Z_{-\infty}^t\right)\right], \tag{3}$$

and then differentiating it to form the FOC's:

$$\beta^t E_t\left[\sum_{s=0}^{\infty} \beta^s \frac{\partial U_{t+s}}{\partial C(t)} - \sum_{s=0}^{\infty} \beta^s \frac{\partial g_{t+s}}{\partial C(t)} \lambda_{t+s}\right] = 0,\ t = 0, ..., \infty \tag{4}$$

Notice that:

- In contrast to the deterministic case, the Lagrangian in (3) and the FOC's in (4) involve expectation operators.
- The expectation operator in the FOC is E_t, conditional expectation given the information set available at t, the date of the choice variable vector C with respect to which the FOC is taken.
- Because U and g each depend only on C's dated t and earlier, the infinite sums in (4) involve only U's and g's dated t and later.

[1] It may seem that it would be natural to include also past C's in the information set. But it is our assumption that this would be redundant. Of course a decision maker could make C_t depend on some “extraneous" random element like a coin flip. Our assumption is simply that if this can occur, the coin flip is part of $Z_{-\infty}^t$

3. REVIEW OF FINITE-DIMENSIONAL, NON-STOCHASTIC KUHN-TUCKER CONDITIONS

In finite-dimensional problems, first order conditions are necessary and sufficient conditions for an optimum in a problem with concave objective functions and convex constraint sets. The conditions in (4) are not as powerful, because this is an infinite-horizon problem. First order conditions here, as in simpler problems, are applications of the:

Separating Hyperplane Theorem: If $\bar{x}$ maximizes the continuous, concave function $V(\cdot)$ over a convex constraint set Γ in some linear space, and if there is an (infeasible) x^* with $V(x^*) > V(\bar{x})$, then there is a continuous linear function $f(\cdot)$ such that $f(x) > f(\bar{x})$ implies that x lies outside Γ and $f(x) < f(\bar{x})$ implies $V(x) < V(\bar{x})$.

In a finite-dimensional problem with x $n \times 1$, we can always write any such f as

$$f(x) = \sum_{i=1}^{n} f_i \cdot x_i \tag{5}$$

where the f_i are all real numbers.

If the problem has differentiable V and differentiable constraints of the form $g_i(x) \leq 0$, then it will also be true that we can always pick

$$f_i = \frac{\partial V}{\partial x_i}(\bar{x}) \tag{6}$$

and nearly always write

$$f(x) = \sum_j \lambda_j \frac{\partial g_j(\bar{x})}{\partial x} \cdot x \tag{7}$$

with $\lambda_i \geq 0$, all i. The "nearly" is necessary because of what is known as the "constraint qualification". It is possible that the first-order properties of the constraints near the optimum do not give a good local characterization of the constraint set Γ. However, if we can find an x vector and a set of non-negative λ_i's that satisfy the constraints and (6) and (7), we have found the separating hyperplane and hence the optimum. The standard Lagrange multiplier equations are therefore sufficient conditions for an optimum, and they are "nearly" necessary: We know there will always be a separating hyperplane, and usually we will be able to write it in the form (7), but there are some knife-edge (i.e., rare) special cases in which this will not be true. This justifies the common strategy of trying to solve such problems by looking for solutions to (6) and (7). The sufficiency part of these results can be summarized as:

Kuhn-Tucker Theorem: [2] If

- V is a continuous, concave function on a finite-dimensional linear space,
- V is differentiable at $\bar{x}$,
- g_i, $i = 1, \ldots, k$ are convex functions, each differentiable at $\bar{x}$,

[2] This version of the Kuhn-Tucker theorem is not the most general possible, even for finite-dimensional spaces.

- there is a set of non-negative numbers $\lambda_i, i = 1,\ldots,k$ such that

$$\frac{\partial V(\bar{x})}{\partial x} = \sum_i \lambda_i \frac{\partial g_i(\bar{x})}{\partial x}\text{, and}$$

- $g_i(\bar{x}) \leq 0\,,\ \lambda_i g_i(\bar{x}) = 0\,,\ i = 1,\ldots,k\,,$

then $\bar{x}$ maximizes V over the set of x's satisfying $g_i(x) \leq 0,\ \ i = 1,\ldots,k$.

4. Complications from an Infinite Horizon

But in an infinite dimensional space it may not be true that we can write every continuous linear function as an infinite sum analogous to (5), and the potentially infinite sums in (7) and in (5) with f defined by (6) might not converge. These complications provide additional reasons that there can be models in which the Lagrange multiplier equations are not necessary conditions for an optimum, but more importantly they mean that they are no longer sufficient conditions, even for problems with concave objective functions and convex constraint sets. It is to handle these problems that we impose on infinite horizon problems what are called *transversality conditions*. To apply the Lagrange multiplier idea to our current problem, interpret V as given by the maximand in (1), $\bar{x}$ as being $\bar{C}$, the optimal C sequence, and x as being a generic C sequence. In our stochastic problem, (5)-(7) become

$$E\left[\sum_{t=0}^{\infty}\sum_{s=0}^{t}\beta^t \frac{\partial U_t\left(C_0^t, Z_0^t\right)}{\partial C_s}\cdot C_s\right] = f\left(C_0^{\infty}\right) = E\left[\sum_{t=0}^{\infty}\beta^t\lambda_t\sum_{s=0}^{t}\frac{\partial g_t\left(\bar{C}_0^t, Z_0^t\right)}{\partial C_s}\cdot C_s\right] \tag{8}$$

In order to get from (8) what are given as FOC's in (4) above, we interchange the order of summation in the expressions on the left and right sides of (8), then equate coefficients of correspondingly subscripted C's. The version of (8) with orders of summation interchanged is

$$E\left[\sum_{s=0}^{\infty}\sum_{t=s}^{\infty}\beta^t \frac{\partial U_t\left(\bar{C}_0^t, Z_0^t\right)}{\partial C_s}\cdot C_s\right] = E\left[\sum_{s=0}^{\infty}\sum_{t=s}^{\infty}\beta^t\lambda_t\frac{\partial g_t\left(\bar{C}_0^t, Z_0^t\right)}{\partial C_s}\cdot C_s\right]\,, \tag{9}$$

from which it is easy to see that (4) follows, if we equate the coefficients on C_s terms on the two sides of the equation. But to justify these manipulations, we must be careful about issues of convergence. Dealing with convergence of these sums is checking transversality.

Note that simply "equating coefficients" on the left and right of (9) might seem to imply (4) either without the "E_t" operator or with an unsubscripted "E" operator. To understand why the E_t appears, remember that C_t is a random variable, a rule for choosing a numerical value for C_t as a function of information available at t. Its "coefficient" in (9) is therefore the sum of all the terms that multiply it, over both dates and possible states of the world given information at t. It is the sum over states consistent with information available at t that results in the E_t operator in the FOC's. The need for the E_t is explained more precisely in footnote 3 below, during the formal argument for sufficiency.

5. SUFFICIENT CONDITIONS FOR THE FINITE LAG, STOCHASTIC, INFINITE HORIZON CASE

In most economic models, there are only finitely many lags as arguments to g and U, which makes many of the infinite sums in (8) and (9) become finite. In fact most commonly U has no lags in its arguments. To get versions of transversality that are closer to what is commonly discussed in economic models and allow us to prove results, we now specialize to the case where $U_t = U(C_t, C_{t-1}, Z_t)$ and $g_t = g(C_t, C_{t-1}, Z_t)$. This allows us to write a version of the Kuhn-Tucker theorem for infinite-dimensional spaces as:

Infinite-Dimensional Kuhn-Tucker: Suppose

(i) $V\left(C_{-\infty}^{\infty}, Z_{-\infty}^{\infty}\right) = \liminf_{T\to\infty} E_0\left[\sum_{t=0}^{T} \beta^t U(C_t, C_{t-1}, Z_t)\right]$;

(ii) U is concave and each element of $g(C_t, C_{t-1})$ is convex in C_t and C_{t-1} for each Z_t, and all integer $t \geq 0$;

(iii) there is a sequence of random variables $\bar{C}_0^{\infty}$ such that each $\bar{C}_t$ is a function only of information available at t, $V(\bar{C}_{-\infty}^{\infty}, Z_{-\infty}^{\infty})$ is finite with the partial sums defining it on the right hand side of (5) converging to a limit, and, for each $t \geq 0$, $g(C_t, C_{t-1}, Z_t) \leq 0$;

(iv) U and g are both differentiable in C_t and C_{t-1} for each Z_t and the derivatives have finite expectation;

(v) There is a sequence of non-negative random vectors λ_0^{∞}, with each λ_t in the corresponding information set at t, and satisfying $\lambda_t g\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right) = 0$ with probability one for all t;

(vi)

$$\frac{\partial U\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right)}{\partial C_t} + \beta E_t\left[\frac{\partial U\left(\bar{C}_{t+1}, \bar{C}_t, Z_{t+1}\right)}{\partial C_t}\right] = \lambda_t \frac{\partial g\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right)}{\partial C_t} + \beta E_t\left[\lambda_{t+1}\frac{\partial g\left(\bar{C}_{t+1}, \bar{C}_t, Z_t\right)}{\partial C_t}\right] \quad (10)$$

for all t (i.e., the **Euler equations** hold);

(vii) (**transversality**) for every feasible C sequence $\hat{C}_0^{\infty}$, either $V(\hat{C}_0^{\infty}) < V(\bar{C}_0^{\infty})$ or

$$\limsup_{t\to\infty} \beta^t E\left[\left(\frac{\partial U\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right)}{\partial C_t} - \lambda_t \frac{\partial g\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right)}{\partial C_t}\right)\cdot\left(\hat{C}_t - \bar{C}_t\right)\right] \leq 0 \quad (11)$$

Then $\bar{C}_0^{\infty}$ maximizes V subject to $g(C_t, C_{t-1}, Z_t) \leq 0$ for all $t \geq 0$ and to the given non-random value of C_1.

Proof: Suppose $\hat{C}_0^{\infty}$ is a feasible sequence of consumption choice rules that achieves a higher value of V than does $\bar{C}_0^{\infty}$, despite $\bar{C}_0^{\infty}$'s satisfying the conditions of the theorem. We simplify notation from this point on by using U_t for $U\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right)$ and using g_t for $g(\bar{C}_t, \bar{C}_{t-1}, Z_t))$. By differentiability and by concavity of U and convexity of g, we know that for each t

$$D_1U_t \cdot \left(\hat{C}_t - \bar{C}_t\right) + D_2U_t \cdot \left(\hat{C}_{t-1} - \bar{C}_{t-1}\right) \geq U\left(\hat{C}_t, \hat{C}_{t-1}, Z_t\right) - U_t \tag{12}$$

and similarly

$$D_1g_t \cdot \left(\hat{C}_t - \bar{C}_t\right) + D_2g_t \cdot \left(\hat{C}_{t-1} - \bar{C}_{t-1}\right) \leq g\left(\hat{C}_t, \hat{C}_{t-1}, Z_t\right) - g_t \tag{13}$$

Using (12), the definition of V, and our working hypothesis that $\hat{C}$ gives a higher value of V than does $\bar{C}$, we conclude that

$$\lim_{T\to\infty} E\left[\sum_{t=0}^{T} \beta^t \cdot \left(D_1U_t \cdot \left(\hat{C}_t - \bar{C}_t\right) + D_2U_t \cdot \left(\hat{C}_{t-1} - \bar{C}_{t-1}\right)\right)\right] > 0 \tag{14}$$

where, because C_{-1} is given exogenously, $\hat{C}_{-1} - \bar{C}_{-1} = 0$ for any feasible $\hat{C}$ sequence.

The law of iterated expectations allows us to rewrite (14) as

$$\begin{aligned}
&\lim_{T\to\infty} E\left[\sum_{t=0}^{T} \beta^t \cdot \left(D_1U_t \cdot \left(\hat{C}_t - \bar{C}_t\right) + E_{t-1}D_2U_t \cdot \left(\hat{C}_{t-1} - \bar{C}_{t-1}\right)\right)\right] = \\
&\quad \lim_{T\to\infty} E\left[\sum_{t=0}^{T-1} \left\{\beta^t \cdot \left(D_1U_t + E_t\left[D_2U_{t+1}\right]\right) \cdot \left(\hat{C}_t - \bar{C}_t\right)\right\} + \beta^T D_1U_T \cdot \left(\hat{C}_T - \bar{C}_T\right)\right] \\
&\qquad > 0\,.
\end{aligned} \tag{15}$$

Then the Euler equations as given in (vi) assure us that (15) equates term by term, except for a leftover term on the end, to the expected sum of the gradients of g, weighted by the the λ sequence.[3] In particular, (15) is exactly

$$\lim_{T\to\infty} \left\{ \begin{array}{c} E\left[\sum_{t=0}^{T} \beta^t \lambda_t \cdot \left(D_1g_t \cdot \left(\hat{C}_t - \bar{C}_t\right) + D_2g_t \cdot \left(\hat{C}_{t-1} - \bar{C}_{t-1}\right)\right)\right] \\ +E\left[\beta^T \cdot \left(D_1U_T - \lambda_T D_1g_T\right) \cdot \left(\hat{C}_t - \bar{C}_t\right)\right] \end{array} \right\} \tag{16}$$

Since the $\hat{C}$ sequence is by hypothesis feasible, since $\lambda_t \geq 0$, and since $\lambda_t g_t = 0$ with probability one,

$$\lambda_t \cdot \left(g\left(\hat{C}_t, \hat{C}_{t-1}, Z_t\right) - g_t\right) \leq 0\,.$$

The first expectation within curly brackets in (16) is therefore less than or equal to zero for every T, by convexity of g. Thus the first term has a lim sup less than or equal to zero. The non-positivity of the lim sup of the second term in the curly brackets is exactly what we assumed in our transversality condition (11). This completes the proof by contradiction: while (14) has to exceed zero if $\hat{C}$ improves on

[3]Note that it is in this last step that we use the fact that the FOC with respect to C_t has "E_t" in front of it. If we had only an "E" in front of it, we would not be able to apply the law of iterated expectations here. The argument we are making would go through if the Euler equations were written without expectations, since these much stronger conditions would imply the Euler equations with E_t in front. But of course weaker sufficient conditions are more useful than stronger ones.

$\bar{C}$, the conditions of the theorem guarantee that it is equal to (16), which has to be non-positive. □

Necessity of Transversality: While there are necessity results for transversality conditions in some contexts, it appears difficult to obtain one for the setup here. When we specialize to the case of dynamic programming, we will get necessary conditions.

Note that the transversality condition (11) is not in quite the usual form. The usual form would simply assert

$$\beta^t E\left[\lambda_t \frac{\partial g\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right)}{\partial C_t} \cdot \bar{C}_t\right] \underset{t\to\infty}{\longrightarrow} 0 \tag{17}$$

Often in economic models the U terms in the true transversality condition as given in (11) drop out or converge to zero automatically. (17) then guarantees transversality at one particular point, $\hat{C}_0^\infty = 0$ which, though it is feasible in most economic models, need not always be feasible. The conventional transversality condition is also too strong in that it requires actual convergence, rather than only that the lim inf be non-negative. It is too weak in that it checks only one point in the feasible set.

This means that there are models in which, if we replaced our condition (11) by (17), there would be C sequences that satisfy all the conditions of the modified theorem that are not in fact optima. A leading example of such a model is the linear-quadratic permanent income model with a borrowing constraint replacing the usual bound on the rate of growth of wealth. The standard linear decision rule is not optimal in such a case, but it satisfies the standard transversality condition (17), while failing our condition (11).

There are also models in which there is an optimum, satisfying the Euler equations, but the standard transversality condition does not hold at the optimum. An example is the original Ramsey growth model, without discounting. This example is explained in (Barro and Sala-I-Martin, 1995, p.507-8).[4]

REFERENCES

BARRO, R. J., AND X. SALA-I-MARTIN (1995): *Economic Growth*. McGraw-Hill, New York.

KUSHNER, H. J. (1965a): “On Stochastic Extremum Problems: Calculus,” *Journal of Mathematical Analysis and Applications*, 10, 354–367.

——— (1965b): “On the Stochastic Maximum Principle: Fixed Time of Control,” *Journal of Mathematical Analysis and Applications*, 11, 78–92.

[4]Note, though, that Barro and Sala-I-Martin’s claim that in two references they cite the standard transversality condition has been shown to be necessary “when there is time discounting and the objective function converges" is misleading. The references they cite assume regularity conditions that rule out utility functions and technologies that appear in standard macroeconomic growth models.