RANDOM LAGRANGE MULTIPLIERS AND TRANSVERSALITY

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.

I. A General Case

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 discuss limits on its range of applicability, and then only at the end sketch some arguments as to why the results are true.

We consider a problem of the form

$$\max_{\{C(s)\}_{s=0}^{\infty}} E\left[\sum_{t=0}^{\infty} \beta^t U_t\left(\{C(s)\}_{s=-\infty}^{t}, \{Z(s)\}_{s=-\infty}^{t}\right)\right] \tag{1}$$

subject to

$$g_t\left(\{C(s)\}_{s=-\infty}^{t}, \{Z(s)\}_{s=-\infty}^{t}\right) \le 0,\ t = 0,\ldots,\infty\ . \tag{2}$$

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 $\{Z(s)\}_{s=-\infty}^{t}$, i.e. current and past values of the exogenous variables. Of course implicitly this means that also $\{C(s)\}_{s=-\infty}^{t}$ is known at *t*, since choice of $C(t)$ always must be a function of the information available at *t*. 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 Hamiltonian expression,

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

and then differentiating it to form the FOC's:

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

Notice that:

- In contrast to the deterministic case, the Hamiltonian 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 $C(t)$ with respect to which the FOC is taken.
- Because U_t and g_t 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.
- The β^t term at the left in (4) is superfluous and is usually just omitted.

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 $f(\cdot)$ and a number *a* such that $f(x)>a$ implies that *x* lies outside the constraint set and $f(x)<a$ 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 $g_i(x)\le 0$, 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

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

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[1]. However, if we can find an $\bar{x}$ 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" sufficient: 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 i) V is a continuous, concave function on a finite-dimensional linear space, ii) V is differentiable at $\bar{x}$ with gradient $\frac{\partial V(\bar{x})}{\partial x}$, iii) g_i, $i = 1,\ldots,k$ are convex functions, each differentiable at $\bar{x}$ with gradient $\frac{\partial g_i(\bar{x})}{\partial x}$, iv) there is a set of non-negative numbers λ_i, $i = 1,\ldots,k$ such that $\frac{\partial V(\bar{x})}{\partial x} = \sum_i \lambda_i \frac{\partial g(\bar{x})}{\partial x}$, and v) $g_i(\bar{x}) \le 0$, $i = 1,\ldots,k$, then $\bar{x}$ maximizes V over the set of x's satisfying $g_i(x) \le 0$, $i = 1,\ldots,k$.

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_i 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

---

[1] If you want an example of this, try to use Lagrange multiplier methods to solve the problem of maximizing $-x^2 - (y-1)^2$ subject to $(x-1)^2 + y^2 \le 1$ and $(x+1)^2 + y^2 \le 1$. This problem satisfies all the differentiability, concavity and convexity one might like, yet does not yield to a direct Lagrange multiplier approach because at the optimum first-order expansions of the constraints do not characterize the constraint set. The difficulty goes away if the right-hand-sides of the constraints are changed from 1 to 1.1, say.

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

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

$$E\left(\sum_{t=0}^{\infty}\sum_{s=0}^{t}\beta^{t}\frac{\partial U_t\left(\{\bar{C}_v\}_{v=0}^{t},\{Z_v\}_{v=0}^{t}\right)}{\partial C_s}\cdot C_s\right)=f\left(\{C_s\}_{s=0}^{\infty}\right)$$
$$=E\left(\sum_{t=0}^{\infty}\beta^{t}\lambda_t\sum_{s=0}^{t}\frac{\partial g_t\left(\{\bar{C}_v\}_{v=0}^{t},\{Z_v\}_{v=0}^{t}\right)}{\partial C_s}\cdot C_s\right) \qquad (8)$$

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

$$E\left(\sum_{s=0}^{\infty}\sum_{t=s}^{\infty}\beta^{t}\frac{\partial U_t\left(\{\bar{C}_v\}_{v=0}^{t},\{Z_v\}_{v=0}^{t}\right)}{\partial C_s}\cdot C_s\right)=E\left(\sum_{s=0}^{\infty}\sum_{t=s}^{\infty}\beta^{t}\lambda_t\frac{\partial g_t\left(\{\bar{C}_v\}_{v=0}^{t},\{Z_v\}_{v=0}^{t}\right)}{\partial C_s}\cdot C_s\right), \qquad (9)$$

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. This justification may be hard to understand at this point. It is made explicit in a simple special case at the end of these notes.

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

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

iii) there is a sequence of random variables $\{\bar{C}_t\}$ such that each $\bar{C}_t$ is a function only of information available at *t*, $V\left(\{\bar{C}_t\},\{Z_t\}\right)$ is finite with the partial sums defining it on the right hand side of (i) converging to a limit, and for each $t=1,\ldots,\infty$ $g\left(\bar{C}_t,\bar{C}_{t-1},Z_t\right)\le 0$;

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 $\{\lambda_t\}$, 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) $$\frac{\partial U\left(\bar{C}_t,\bar{C}_{t-1},Z_t\right)}{\partial C_t}+\beta E_t\left[\frac{\partial U\left(\bar{C}_{t+1},\bar{C}_t,Z_{t+1}\right)}{\partial C_t}\right]=\lambda_t\frac{\partial g\left(\bar{C}_t,\bar{C}_{t-1},Z_t\right)}{\partial C_t}+\beta E_t\left[\lambda_{t+1}\frac{\partial g\left(\bar{C}_{t+1},\bar{C}_t,Z_t\right)}{\partial C_t}\right]$$ for all $t\ge 0$ (i.e., the **Euler equations** hold).

vii) **(transversality)** for every feasible C sequence $\{C_t^*\}$

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

Then $\{\bar{C}_t\}$ maximizes V subject to $g\left(C_t,C_{t-1},Z_t\right)\le 0$ for all $t=0,\ldots,\infty$.

Proof: Suppose $\{C_t^*\}$ is a feasible sequence of consumption choice rules that achieves a higher value of V than does $\{\bar{C}_t\}$, despite $\{\bar{C}_t\}$'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 $g\left(\bar{C}_t,\bar{C}_{t-1},Z_t\right)$. By differentiability and by concavity of U and convexity of g, we know that for each t

$$D_1U_t\cdot\left(C_t^*-\bar{C}_t\right)+D_2U_t\cdot\left(C_{t-1}^*-\bar{C}_{t-1}\right)\ge U\left(C_t^*,C_{t-1}^*,Z_t\right)-U_t \tag{10}$$

and similarly

$$D_1g_t\cdot\left(C_t^*-\bar{C}_t\right)+D_2g_t\cdot\left(C_{t-1}^*-\bar{C}_{t-1}\right)\le g\left(C_t^*,C_{t-1}^*,Z_t\right)-g_t \tag{11}$$

Using (10), the definition of V, and our working hypothesis that $\{C_t^*\}$ gives a higher value of V than does $\{\bar{C}_t\}$, we conclude that

$$\lim_{T\to\infty} E\left[\sum_{t=0}^{T} \beta^t \cdot \left(D_1 U_t \cdot \left(C_t^* - \overline{C}_t\right) + D_2 U_t \cdot \left(C_{t-1}^* - \overline{C}_{t-1}\right)\right)\right] > 0\,. \tag{12}$$

But our Euler equations as given in (vi) assure us that (12) 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. In particular, (12) is exactly

$$\lim_{T\to\infty} \left\{ \begin{array}{c} E\left[\sum_{t=0}^{T} \beta^t \lambda_t \cdot \left(D_1 g_t \cdot \left(C_t^* - \overline{C}_t\right) + D_2 g_t \cdot \left(C_{t-1}^* - \overline{C}_{t-1}\right)\right)\right] \\ + E\left[\beta^T \cdot \left(D_1 U_T - \lambda_T D_1 g_T\right) \cdot \left(C_t^* - \overline{C}_t\right)\right] \end{array} \right\}. \tag{13}$$

Since the C^* is by hypothesis feasible, since $\lambda_t \geq 0$, and since $\lambda_t g_t$ is zero with probability one, $\lambda_t \cdot \left(g\left(C_t^*, C_{t-1}^*, Z_t\right) - g_t\right) \leq 0$. The first expectation within curly brackets in (13) 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 (vii). This completes the proof by contradiction: while (12) has to exceed zero if $\left\{C_t^*\right\}$ improves on $\left\{\overline{C}_t\right\}$, the conditions of the theorem guarantee that it is equal to (13), which has to be non-positive.

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

$$\beta^t E\left[\lambda_t \frac{\partial g\left(\overline{C}_t, \overline{C}_{t-1}, Z_t\right)}{\partial C_t} \cdot \overline{C}_t\right] \underset{t\to\infty}{\to} 0\ . \tag{14}$$

Often in economic models the U terms in the true transversality condition as given in (vii) drops out or converges to zero automatically. (14) then guarantees transversality at one particular point, $\left\{C_t^*\right\} = \left\{0\right\}$; though in most economic models the zero sequence is in the feasible set, this need not always be true. 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. There are models in which, if we replaced our condition (vii) by (14), 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 (14), while failing our condition (vii).

REFERENCES

Kushner, Harold J. [1965], "On Stochastic Extremum Problems: Calculus," Journal of Mathematical Analysis and Applications 10, 354-367.

_________________ [1965], "On the Stochastic Maximum Principle: Fixed Time of Control," Journal of Mathematical Analysis and Applications 11, 78-92.