

RANDOM LAGRANGE MULTIPLIERS

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. Discrete Time

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

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

subject to

$$g_t \left(\{C(s)\}_{s=-\infty}^t, \{Z(s)\}_{s=-\infty}^t \right) \leq 0, \quad t = 0, \dots, \infty . \quad (2)$$

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,

$$E \left[\sum_{t=0}^{\infty} b^t U_t(\{C(s)\}_{s=-\infty}^t, \{Z(s)\}_{s=-\infty}^t) - \sum_{t=0}^{\infty} b^t \lambda_{t+s}(\{C(s)\}_{s=-\infty}^t, \{Z(s)\}_{s=-\infty}^t) \right], \quad (3)$$

and then differentiating it to form the FOC's:

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

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 b^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. That is, 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 we can always write any such f as

$$f(x) = \sum f_i x_i \quad (5)$$

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

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

and always write

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

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. It is to rule out these problems that we impose on infinite horizon problems what are called *transversality conditions*.

To apply the discussion of the separating hyperplane principle 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

$$\begin{aligned} E \left(\sum_{t=0}^{\infty} \sum_{s=0}^t b^t \frac{\partial U_t(\{\bar{C}_v\}_{v=0}^t, \{Z_v\}_{v=0}^t)}{\partial C_s} \cdot C_s \right) &= f(\{C_s\}_{s=0}^{\infty}) \\ &= E \left(\sum_{t=0}^{\infty} b^t l_t \sum_{s=0}^t \frac{\partial g_t(\{\bar{C}_v\}_{v=0}^t, \{Z_v\}_{v=0}^t)}{\partial C_s} \cdot C_s \right) \end{aligned} \quad (8)$$

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

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

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 sure that all the infinite sums involved converge. Checking 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, and in all dynamic programming problems, U has no lags in its arguments. To get versions of transversality

that are closer to what is commonly discussed in economic models, we now specialize to the case where $U_t = U(C_t, C_{t-1}, Z_t)$ and $g_t = g(C_t, C_{t-1}, Z_t)$. In this case (8) becomes

$$E \left[\sum_{t=0}^{\infty} b^t (D_1 U(\bar{C}_t, \bar{C}_{t-1}, Z_t) C_t + D_2 U(\bar{C}_t, \bar{C}_{t-1}, Z_t) C_{t-1}) \right] \\ = E \left[\sum_{t=0}^{\infty} b^t (D_1 g(\bar{C}_t, \bar{C}_{t-1}, Z_t) C_t + D_2 g(\bar{C}_t, \bar{C}_{t-1}, Z_t) C_{t-1}) \right] \quad (10)$$

and (9) becomes

$$E \left[\sum_{t=0}^{\infty} b^t (D_1 U(\bar{C}_t, \bar{C}_{t-1}, Z_t) + b D_2 U(\bar{C}_{t+1}, \bar{C}_t, Z_{t+1})) C_t \right] \\ = E \left[\sum_{t=0}^{\infty} b^t (D_1 g(\bar{C}_t, \bar{C}_{t-1}, Z_t) + b D_2 g(\bar{C}_{t+1}, \bar{C}_t, Z_{t+1})) C_t \right] \quad (11)$$

In general, all four of the infinite sums in (10) and (11) are candidates for failure of convergence, and thereby for undermining the validity of the Euler equations (given by (4)), obtained by simple coefficient-matching. Once we have imposed the Euler equations, the terms in the sums on the left and right hand sides of (11) are identical, so if one converges, so does the other.

Of course the convergence of these sums depends on the argument sequence $\{C\}$ as well as on the coefficients. The usual practice, which certainly checks a necessary condition for application of the separating hyperplanes theorem, is to verify convergence at $\{C\} = \{\bar{C}\}$. However, the convergence has to hold at least over the entire feasible set of $\{C\}$'s. In some problems, therefore, it turns out that following usual practice and checking convergence only at $\{C\} = \{\bar{C}\}$ is not enough.

II. Proof for a simple case

A. Probability Space and Information Structure

We simplify by supposing a probability space Ω containing finitely many points, for which we will ordinarily use the symbols w or ω . Each point w in Ω is a sequence $w = \{Z(0), \dots, Z(T)\}$ of finitely many vectors $Z(t)$, with T , the length of the sequence, the same for each w in Ω . We postulate a probability function $p(\cdot)$ on Ω , with $p(w)$ interpreted as the probability that the true state of the world is w .¹

¹ If you've had some measure theory, it may help to recognize that we are sidestepping the need to consider a sigma-field of events on which probabilities are defined in addition to the probability space itself. When Ω contains finitely many points, the only interesting sigma-field is the class of all subsets of Ω and probabilities can always be generated from the probabilities of individual points.

As usual, a k -dimensional **random vector** is defined as a function from the probability space Ω to k -dimensional Euclidean space \mathfrak{R}^k . Since each ω in Ω is a sequence of Z 's, we can define a set of $T+1$ random vectors $[Z(0;\cdot), \dots, Z(T;\cdot)]$ by

$$\omega = \{Z(0;\omega), \dots, Z(T;\omega)\}. \quad (12)$$

As before we model this process by saying that the information available at time t is the values of $Z(s;\omega)$ for $s < t$. For any ω in Ω , we will denote by ω_t the vector of random variables

$$\omega_t = \{Z(0;\omega), \dots, Z(t;\omega)\}. \quad (13)$$

A random variable x is then said to be known at t or determined by information available at t if and only if for any points ω, η in Ω , $\omega_t = \eta_t$ implies $x(\omega) = x(\eta)$. Note that for the information structure to be nontrivial, there must be points ω, η in Ω with $\omega \neq \eta$ but $\omega_t = \eta_t$ for some $t < T$.

B. Solving the Optimization Problem

We consider the same optimization problem as at the start, defined in (1) and (2) above. To convert this to a non-stochastic problem to which we can apply the usual first order conditions, we need to get rid of the expectation operator and the infinite horizon in (1). This produces

$$\sum_{\omega} p(\omega) \sum_{t=0}^T b^t U_t \left(\{C(s;\omega_s)\}_{s=0}^t, \{Z(s;\omega_s)\}_{s=0}^t \right) \quad (14)$$

Assuming the usual regularity conditions (including differentiability of U and the g 's in their C arguments for each ω with non-zero probability), the first order necessary conditions for an optimum in the problem defined by (14) and (2) are

$$\sum_{\omega \in \Omega} p(\omega) \sum_{s=0}^{T-t} b^s \left(\frac{\partial U_{t+s} \left(\{C(u;\omega_u), Z(u;\omega_u)\}_{u=0}^{t+s} \right)}{\partial C(t;\eta_t)} - \lambda_{t+s}(\omega) \frac{\partial g_{t+s} \left(\{C(u;\omega_u), Z(u;\omega_u)\}_{u=0}^{t+s} \right)}{\partial C(t;\eta_t)} \right) = 0 \quad (15)$$

Now notice that the partial derivatives in (15) are zero whenever ω is not in η_t . This means that we can replace the sum over all of Ω at the far left by a sum over ω in η_t . Furthermore, we can divide by $p(\eta_t)$. This makes the weighted sums expectations conditional on η_t , according to the usual elementary definition of conditional probability and conditional expectation. And finally notice that the Lagrange multiplier $\lambda_{t+s}(\omega)$ multiplies a partial derivative of g_{t+s} , which depends only on ω_{t+s} . Therefore, without affecting the value of the expression we can replace every

occurrence of $\lambda_{t+s}(\omega)$ with $\lambda_{t+s}^*(\omega_{t+s}) = \sum_{\omega \in \omega_{t+s}} \frac{p(\omega)}{p(\omega_{t+s})} \lambda_{t+s}(\omega)$. This converts (15) to (4),

which has been our objective.

Obviously this argument does not justify the uses we will make of these first order conditions, in which the probability space will generally have a continuum of elements. The references below discuss these issues more carefully.

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.