Econ 504.2, Lecture 1: Transversality and Stochastic Lagrange Multipliers

Christopher A. Sims
Princeton University
sims@princeton.edu

March 29, 2001

Example: LQPY

The ordinary LQ permanent income model has agents solving

$$\max_{\{C_s, W_s\}} E\Big[\sum_{t=0}^{\infty} \beta^t (C_t - \tfrac{1}{2}C_t^2)\Big]$$

subject to

$$W_t = R(W_{t-1} - C_{t-1}) + Y_t \qquad (*)$$

$$E[\beta^{.5t} W_t] \underset{t\to\infty}{\to} 0\,. \qquad (**)$$

The solution, for the simple case where $R\beta = 1$ and Y_t is i.i.d. with mean $\bar{Y}$, this model's solution is well known to be

$$C_t = (1-\beta)W_t + \beta\bar{Y}\,.$$

A reasonable modification of LQPY

Suppose we replace $(**)$ by a simple $W_t \geq 0$. This changes makes sense. Where does the limit on the growth rate of W in $(**)$ come from? No economic reasoning seems to underly it. We believe that the agent should see constraints on making W large and negative (i.e., borrowing a lot), but why the constraint on *positive* accumulation at a high rate?

Under either of this modification the problem is no longer LQ. It is clear that the standard solution is not optimal, so long as $\mathrm{Var}(Y_t) > 0$ and $Y_t \geq 0$ with probability one. Note that the standard solution implies

$$W_t = W_{t-1} + Y_t - \bar{Y} \ . \tag{1}$$

that is, W_t is a random walk, and thus a martingale (i.e. satisfies $E_t W_{t+1} = W_t$).

Reasonable modification continued

Theorem: A bounded martingale converges almost surely.

Since the changes in W_t always have nonzero variance, W does not converge. Therefore, by the theorem, it is unbounded — both above and below. In particular, eventually it will get above

$$\bar{W} = \frac{1}{R-1} \ .$$

Once $W_t \geq \bar{W}$, we can set $C_t \equiv 1$, which delivers maximum possible ("satiation level") utility, forever, and we can be sure that no matter how bad our luck in drawing Y_t values, we can avoid violating the $W_t \geq 0$ constraint. This has to be better than continuing with the standard solution, which would at this point push C above 1. This deviation from the standard solution entails W increasing toward infinity at the rate β^{-t}, which is why with $(**)$ imposed we do find the standard solution to be optimal.

Standard TVC and our modified LQPY problem

The Lagrange multiplier on the constraint in this problem is $\lambda_t = 1 - C_t$, and the usual TVC is

$$E_0[\beta^t \lambda_t W_t] = E_0[\beta^t (1 - C_t) W_t] \underset{t\to\infty}{\to} 0 \ .$$

Since W_t is a random walk in this solution and has i.i.d. increments, its second moment is $O(t)$, as is (therefore) $E_0[C_t W_t]$. The conventional TVC is satisfied.

So this is a problem with concave objective function, and convex constraints. The "standard solution" satisfies all the Euler equations and the conventional TVC — but it is not in fact an optimum. In a standard finite-dimensional problem, a concave objective function and convex constraint sets imply that any solution to the FOC's is an optimum. What's wrong here?

Notation: The Most General Setup

- Our practice: things dated t are always "known" — i.e. available for use as arguments of decision functions — at t. This differs from the convention in much of the growth literature, and in the classic Blanchard-Kahn treatment of linear RE models, but is saves much confusion.
- A stochastic optimization problem in general form

$$\max_{\boldsymbol{C}_0^\infty} E\left[\sum_{t=0}^{\infty} \beta^t U_t\left(\boldsymbol{C}_{-\infty}^t, \boldsymbol{Z}_{-\infty}^t\right)\right] \tag{2}$$

subject to

$$g_t\left(\boldsymbol{C}_{-\infty}^t, \boldsymbol{Z}_{-\infty}^t\right) \leq 0,\ t = 0, ..., \infty\ , \tag{3}$$

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

- An implicit constraint: $\{C_t\}$ is **adapted** to $\{Z_t\}$. Each C_t is not a vector of real numbers, but instead a function mapping the information available at t, $\boldsymbol{Z}_{-\infty}^t$, into vectors of real numbers.
- It is possible to eliminate the random variables and expectations from our discussion by considering the simplified special case where at each t there are only finitely many possible values of $\boldsymbol{Z}_{-\infty}^t$. Then the C_t decision function is just a long vector, characterized by the list of values it takes at each possible value for $\boldsymbol{Z}_{-\infty}^t$; expectations are just weighted sums.

Lagrangian and FOC's

- The Lagrangian for this problem

$$E\left[\sum_{t=0}^{\infty}\beta^t U_t\left(\boldsymbol{C}_{-\infty}^t, \boldsymbol{Z}_{-\infty}^t\right)\right.$$
$$\left. - \sum_{t=0}^{\infty}\beta^t \lambda_t g_t\left(\boldsymbol{C}_{-\infty}^t, \boldsymbol{Z}_{-\infty}^t\right)\right], \quad (4)$$

- The FOC's

$$\frac{\partial H}{\partial C(t)} =$$
$$\beta^t E_t\left[\sum_{s=0}^{\infty}\beta^s\frac{\partial U_{t+s}}{\partial C(t)} - \sum_{s=0}^{\infty}\beta^s\frac{\partial g_{t+s}}{\partial C(t)}\lambda_{t+s}\right] = 0,$$
$$t = 0, \ldots, \infty \quad (5)$$

Necessity and Sufficiency?

Separating Hyperplane Theorem If $V(\cdot)$ is a continuous, concave function over a convex constraint set Γ in some linear space, and if there is an x^* with $V(x^*) > V(\bar{x})$, then $\bar{x}$ maximizes V over Γ if and only if 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

$$f(x) = \sum_{i=1}^{n} f_i \cdot x_i \quad (6)$$

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

$$f_i = \frac{\partial V}{\partial x_i}(\bar{x}) \tag{7}$$

and nearly always write

$$f(x) = \sum_j \lambda_j \frac{\partial g_j(\bar{x})}{\partial x} \cdot x \tag{8}$$

with $\lambda_i \ge 0$, all i. The "nearly" is necessary because of what is known as the "constraint qualification".

Kuhn-Tucker Theorem (sufficiency) 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 $\bar{x}$,
- there is a set of non-negative numbers $\lambda_i, i = 1, \ldots, k$ such that

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

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

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

The fly in the ointment: convergence of infinite sums

Interpret V as given by the maximand in (2), $\bar{x}$ as being $\bar{C}$, the optimal C sequence, and x as being a generic C sequence. In our stochastic problem, (6)-(8) become

$$E\left[\sum_{t=0}^{\infty}\sum_{s=0}^{t}\beta^t\frac{\partial U_t\left(\boldsymbol{C}_0^t,\boldsymbol{Z}_0^t\right)}{\partial C_s}\cdot C_s\right]=f\left(\boldsymbol{C}_0^{\infty}\right)$$

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

The version of (9) with orders of summation interchanged (?!)is

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

$$=E\left[\sum_{s=0}^{\infty}\sum_{t=s}^{\infty}\beta^t\lambda_t\frac{\partial g_t\left(\bar{\boldsymbol{C}}_0^t,\boldsymbol{Z}_0^t\right)}{\partial C_s}\cdot C_s\right], \quad (10)$$

from which it is easy to see that (5) follows, if we equate the coefficients on C_s terms on the two sides of the equation.

Some simplifications

- Drop t subscripts on U and g.
- Give U and g each only finitely many arguments.
- I.e., $U_t = U\left(C_t, C_{t-1}, Z_t\right)$ and $g_t = g\left(C_t, C_{t-1}, Z_t\right)$

Infinite-dimensional stochastic Kuhn-Tucker

Infinite-Dimensional Kuhn-Tucker Suppose

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

ii. U is concave and each element of $g(C_t, C_{t-1}, Z_t)$ 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 $\bar{\boldsymbol{C}}_0^{\infty}$ such that each $\bar{C}_t$ is a function only of information available at t, $V(\bar{\boldsymbol{C}}_{-\infty}^{\infty}, \boldsymbol{Z}_{-\infty}^{\infty})$ is finite with the partial sums defining it on the right hand side of (i) converging to a limit, and, for each $t \geq 0$, $g(\bar{C}_t, \bar{C}_{t-1}, Z_t) \leq 0$ with probability one;

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 $\boldsymbol{\lambda}_0^{\infty}$, 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.

$$\frac{\partial U\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right)}{\partial C_t} + \beta E_t\left[\frac{\partial U\left(\bar{C}_{t+1}, \bar{C}_t, Z_{t+1}\right)}{\partial C_t}\right] = \lambda_t \frac{\partial g\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right)}{\partial C_t} + \beta E_t\left[\lambda_{t+1}\frac{\partial g\left(\bar{C}_{t+1}, \bar{C}_t, Z_t\right)}{\partial C_t}\right] \tag{11}$$

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

vii. (**transversality**) for every feasible C sequence $\hat{\boldsymbol{C}}_0^{\infty}$ for which

$$V\left(\bar{\boldsymbol{C}}_{-\infty}^{\infty}, \boldsymbol{Z}_{-\infty}^{\infty}\right) < V\left(\hat{\boldsymbol{C}}_{-\infty}^{\infty}, \boldsymbol{Z}_{-\infty}^{\infty}\right),$$

$$\limsup_{t\to\infty} \beta^t E\left[\left(\frac{\partial U\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right)}{\partial C_t} - \lambda_t \frac{\partial g\left(\bar{C}_t, \bar{C}_{t-1}, Z_t\right)}{\partial C_t}\right) \cdot \left(\hat{C}_t - \bar{C}_t\right)\right] \le 0 \,. \quad (12)$$

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

Simplification to the "standard" TVC

Note that for those elements of the vector of TVC's in (12) that correspond to derivatives with respect elements of the C_t vector that do not appear with a lag in U or g, the E_t terms in the Euler equations (11) drop out, so that the Euler equations guarantee that for these elements of C, the TVC's hold trivially — the expression that is supposed to go to zero in $\limsup$ actually is identically zero. Thus there is only one TVC per "state" variable, if we call any variable that enters with a lag a state.

In many economic models we have the following additional simplifications:

a. The subvector of C that enters with a lag, which we will call "S", for "state vector", can be "solved for", so that the constraints have the form

$$S_t \le h(S_{t-1}, I_t, Z_t) \,,$$

where I_t is notation for the part of the C vector other than S.

b. The state variables are inherently positive, so that we are sure $S_t \ge 0$ for all t for any feasible S sequence, even if there are no explicit constraints to this effect, and $S_t \equiv 0$ is feasible.

c. S_t does not enter the U function at all.

It is easy to check that under these conditions our general TVC (12) greatly simplifies, to become

$$\lim_{t\to\infty} E_0\left[\beta^t \lambda_t \bar{S}_t = 0\right] \,.$$

In other words we can get rid of the $\limsup$ operator, replacing it with an ordinary $\lim$, we get rid of the term depending on U, and we avoid having to consider the alternative sequences $\hat{C}$.

Application to the Linear-Quadratic Permanent Income Example

In the conventional solution, we get from the FOC's

$$C_t = E_t C_{t+1} \; .$$

For the conventional solution to be correct, the constraint must be interpreted as an *equality*, so that to get it into our Kuhn-Tucker framework we must treat as two inequality constraints (both linear, so both convex despite the sign change):

$$W_t \leq RW_{t-1} + Y_t - C_t$$

$$-W_t \leq -(RW_{t-1} + Y_t - C_t) \; .$$

There are then two positive Lagrange multipliers, one that is zero when the first constraint is not binding, the other zero when the second constraint is not binding.

If we ignore the growth constraint ($**$), the solution to the problem is just to set $C_t \equiv 1$, even though apparently Euler equations and TVC are satisfied by the conventional solution. The condition (a) above is not satisfied, however, because instead of having W_t on the left, one of the constraints has $-W_t$ on the left, so the constraints are not "standard". Also, in this problem the state variable W is not inherently positive, violating another of the conditions for application of the standard TVC. In particular, when the constraint that has $-W_t$ on the left is binding, W_t is a "bad", not a "good". It, together with the requirement in the conventional solution that W not grow too fast, is what forces us to consume beyond satiation.

In the LQPY model with $W_t \geq 0$ replacing the growth constraint, W is necessarily non-negative, but the conventional solution violates the requirement, in a model with convex constraints and concave objective, that the Lagrange multipliers all be non-negative. The fact that when $C_t > 1$, $\lambda_t < 0$, is an immediate signal that the conventional solution is not an optimum.