

# ECON 504.2, LECTURE 1: TRANSVERSALITY AND STOCHASTIC LAGRANGE MULTIPLIERS

CHRISTOPHER A. SIMS  
PRINCETON UNIVERSITY  
SIMS@PRINCETON.EDU

## 1. EXAMPLE: LQPY

The ordinary LQ permanent income model has agents solving

$$\max_{\{C_t, W_t\}} E \left[ \sum_{t=0}^{\infty} \beta^t (C_t - \frac{1}{2} C_t^2) \right]$$

subject to

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

$$E[\beta^{-5t} W_t] \xrightarrow{t \rightarrow \infty} 0. \quad (**)$$

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

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

## 2. A REASONABLE MODIFICATION OF LQPY

- Where does the limit on the growth rate of  $W$  in  $(**)$  come from? 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?
- So replace  $(**)$  by  $\liminf E[R^{-t} W_t] \geq 0$ , a standard form for a “no-Ponzi” condition. Then the problem is no longer LQ, and the standard solution is not optimal, so long as  $\text{Var}(Y_t) > 0$  and  $Y_t \geq 0$  with probability one.

## 3. WHY IS THE STANDARD SOLUTION NOT OPTIMAL?

It implies

$$W_t = W_{t-1} + Y_t - \bar{Y}. \quad (1)$$

So  $E_t W_{t+1} = W_t$ , i.e.  $W_t$  is a martingale.

**Theorem:** A bounded martingale converges almost surely.

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

$$W^* = \frac{1}{R-1}.$$

Once  $W_t \geq W^*$ , 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  $W_t \geq 0$ .

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  $\beta^{-t}$ , which is why with (\*\*\*) imposed we do find the standard solution to be optimal.

#### 4. 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

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

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?

#### 5. NOTATION: THE MOST GENERAL SETUP

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

$$\max_{C_0^\infty} E \left[ \sum_{t=0}^{\infty} \beta^t U_t (C_{-\infty}^t, Z_{-\infty}^t) \right] \tag{2}$$

subject to

$$g_t (C_{-\infty}^t, Z_{-\infty}^t) \leq 0, t = 0, \dots, \infty, \tag{3}$$

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

#### 6.

- 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  $t$ ,  $Z_{-\infty}^t$ , 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  $Z_{-\infty}^t$ . Then the  $C_t$  decision function is just a long vector, characterized by the list of values it takes at each possible value for  $Z_{-\infty}^t$ ; expectations are just weighted sums.

## 7. LAGRANGIAN AND FOC'S

- The Lagrangian for this problem

$$E \left[ \sum_{t=0}^{\infty} \beta^t U_t (C_{-\infty}^t, Z_{-\infty}^t) - \sum_{t=0}^{\infty} \beta^t \lambda_t g_t (C_{-\infty}^t, Z_{-\infty}^t) \right], \quad (4)$$

- The FOC's

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

## 8. NECESSITY AND SUFFICIENCY?

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

## 9.

In a finite-dimensional problem with  $x \in \mathbb{R}^n$ , 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  $g_i(x) \leq 0$ , then it will also be true that we can always pick

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

and nearly always write

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

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

10.

**Kuhn-Tucker Theorem (sufficiency):** If

- $V$  is a continuous, concave function on a finite-dimensional linear space,
- $V$  is differentiable at  $\bar{x}$ ,
- $g_i, i = 1, \dots, k$  are convex functions, each differentiable at  $\bar{x}$ ,
- there is a set of non-negative numbers  $\lambda_i, i = 1, \dots, k$  such that

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

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

## 11. 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

$$\begin{aligned} E \left[ \sum_{t=0}^{\infty} \sum_{s=0}^t \beta^t \frac{\partial U_t(C_0^t, Z_0^t)}{\partial C_s} \cdot C_s \right] &= f(C_0^\infty) \\ &= E \left[ \sum_{t=0}^{\infty} \beta^t \lambda_t \sum_{s=0}^t \frac{\partial g_t(\bar{C}_0^t, Z_0^t)}{\partial C_s} \cdot C_s \right] \end{aligned} \quad (9)$$

12.

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

$$\begin{aligned} E \left[ \sum_{s=0}^{\infty} \sum_{t=s}^{\infty} \beta^t \frac{\partial U_t(\bar{C}_0^t, Z_0^t)}{\partial C_s} \cdot C_s \right] \\ = E \left[ \sum_{s=0}^{\infty} \sum_{t=s}^{\infty} \beta^t \lambda_t \frac{\partial g_t(\bar{C}_0^t, Z_0^t)}{\partial C_s} \cdot C_s \right], \end{aligned} \quad (10)$$

Using the law of iterated expectations, together with the fact that  $C_s$  is a function of information known at  $s$ , we can expand this expression to

$$\begin{aligned} E \left[ \sum_{s=0}^{\infty} E_s \left[ \sum_{t=s}^{\infty} \beta^t \frac{\partial U_t(\bar{C}_0^t, Z_0^t)}{\partial C_s} \right] \cdot C_s \right] \\ = E \left[ \sum_{s=0}^{\infty} E_s \left[ \sum_{t=s}^{\infty} \beta^t \lambda_t \frac{\partial g_t(\bar{C}_0^t, Z_0^t)}{\partial C_s} \right] \cdot C_s \right]. \end{aligned} \quad (11)$$

Since  $C_s$  can be any function of  $Z_0^s$  for which the objective function is defined, it is clear that we cannot guarantee this equality for all candidate  $C_s$  sequences unless the coefficients on  $C_s$  on both sides of the equation are equal with probability one. Imposing this condition gives us the Euler equations.

13. SOME SIMPLIFICATIONS

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

14. INFINITE-DIMENSIONAL STOCHASTIC KUHN-TUCKER

**Infinite-Dimensional Kuhn-Tucker:** Suppose

- (i)  $V(C_{-\infty}^{\infty}, Z_{-\infty}^{\infty}) = \liminf_{T \rightarrow \infty} E_0 \left[ \sum_{t=0}^T \beta^t U(C_t, C_{t-1}, Z_t) \right]$ ;
- (ii)  $U$  is concave and each element of  $g(C_t, C_{t-1}, Z_t)$  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  $\bar{C}_0^{\infty}$  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 (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;

15.

- (v) There is a sequence of non-negative random vectors  $\lambda_0^{\infty}$ , with each  $\lambda_t$  in the corresponding information set at  $t$ , and satisfying  $\lambda_t g(\bar{C}_t, \bar{C}_{t-1}, Z_t) = 0$  with probability one for all  $t$ ;
- (vi)

$$\begin{aligned} \frac{\partial U(\bar{C}_t, \bar{C}_{t-1}, Z_t)}{\partial C_t} + \beta E_t \left[ \frac{\partial U(\bar{C}_{t+1}, \bar{C}_t, Z_{t+1})}{\partial C_t} \right] \\ = \lambda_t \frac{\partial g(\bar{C}_t, \bar{C}_{t-1}, Z_t)}{\partial C_t} + \beta E_t \left[ \lambda_{t+1} \frac{\partial g(\bar{C}_{t+1}, \bar{C}_t, Z_t)}{\partial C_t} \right] \end{aligned} \quad (12)$$

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

16.

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

$$V(\bar{C}_{-\infty}^{\infty}, Z_{-\infty}^{\infty}) < V(\hat{C}_{-\infty}^{\infty}, Z_{-\infty}^{\infty}),$$

or

$\limsup_{t \rightarrow \infty} \beta^t$

$$E \left[ \left( \frac{\partial U(\bar{C}_t, \bar{C}_{t-1}, Z_t)}{\partial C_t} - \lambda_t \frac{\partial g(\bar{C}_t, \bar{C}_{t-1}, Z_t)}{\partial C_t} \right) \cdot (\hat{C}_t - \bar{C}_t) \right] \leq 0. \quad (13)$$

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  $C_{-1}$ .

**Necessity:** The Euler equations are always necessary conditions. There are regularity conditions that make transversality part of the necessary conditions, but specifying these regularity conditions gets us into deeper mathematical waters, so we will not take this up.

### 17. SIMPLIFICATION TO THE "STANDARD" TVC

Note that for those elements of the vector of TVC's in (13) 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 (12) 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. For elements of the  $C_t$  vector that enter *only* with a lag, the corresponding TVC components are identically zero. Thus there is only one non-trivial TVC per "state" variable, if we label as a state any variable that enters both unlagged and with a lag.

### 18. RESTRICTIONS

In many economic models we have the following additional simplifications:

- (a) The subvector of  $C$  that enters both currently and with a lag, which we will call "S", for "state vector", can be "solved for", so that the constraints have the form

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

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

- (b) Paths with  $E_0[\liminf \beta^t \lambda_t S_t < 0]$  are not feasible (usually because of a no-Ponzi condition or an inherently positive  $S_t$ ), while paths in which  $\lim \beta^t \lambda_t S_t = 0$  are feasible (often obvious because  $S_t \equiv 0$  is feasible); here the  $\lambda_t$  sequences are those emerging from the candidate optimum path  $\bar{S}_t$ ;
- (c)  $S_t$  does not enter the  $U$  function at all.

### 19. THE SIMPLIFIED CONDITION

It is easy to check that under these conditions our general TVC (13) greatly simplifies, to become

$$\lim_{t \rightarrow \infty} E_0 [\beta^t \lambda_t \bar{S}_t = 0].$$

In other words we can get rid of the lim sup 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  $\hat{C}$ .

Condition (b) above usually is guaranteed by the fact that, as part of our hypothesis about the candidate optimum, all the  $\lambda_t$ 's are non-negative, while we have a lower bound on  $S_t$ . If these common conditions are met, then the "dot-product" form of the TVC equivalent to the requirement that each  $E_0[\beta^t \lambda_{it} \bar{S}_{it}]$  separately converges to zero, so we can check the transversality condition one variable at a time.

## 20. 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):

$$\begin{aligned} \mu: & \quad W_t \leq R(W_{t-1} - C_{t-1} + Y_t) \\ \nu: & \quad -W_t \leq -(R(W_{t-1} - C_{t-1}) + Y_t). \end{aligned}$$

There are then two positive Lagrange multipliers,  $\mu$  and  $\nu$ .

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". 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 version of the model with the no-Ponzi condition replacing the growth constraint, the problem is again non-standard, because still one of the constraints has a  $-W_t$  on the right-hand side.

To see that the full TVC is violated in the conventional solution if there is no  $W$ -growth constraint, observe that the TVC is

$$\limsup_{t \rightarrow \infty} \beta^t E [(1 - \bar{C}_t)(\hat{C}_t - \bar{C}_t) - (\mu_t - \nu_t)(\hat{W}_t - \bar{W}_t)] \leq 0.$$

The Euler equations for  $C$  and  $W$  allow us to conclude that  $\mu_t - \nu_t = 1 - C_t$ . In the standard solution,  $1 - \bar{C}_t$  is a random walk, so it becomes positive infinitely often and negative infinitely often. It is feasible, as we have seen, to choose consumption equal to 1 in every period that the standard solution would make it exceed one, and to leave consumption equal to its standard-solution value at all other times. This yields higher utility than  $\bar{C}_t$  and it implies that eventually  $\hat{W}_t = O(R^t)$ . With our assumption that  $R\beta = 1$ , we see then that there are feasible  $W$ 's for which the lim sup in the  $W$  component of the TVC is in fact positive. Also, since this  $\hat{C}_t$  makes  $\hat{C}_t - \bar{C}_t$  negative at exactly those dates when  $1 - \bar{C}_t$  is negative, the  $C$  component of the TVC must also have a non-negative lim sup.

When the constraint is the no-Ponzi condition  $\liminf_{T \rightarrow \infty} R^{-T} E W_T \geq 0$ , we have to convert the limit constraint into one that we can enter in the Lagrangian. The usual way to do this is to use the no-Ponzi condition to derive the “intertemporal budget constraint”:

$$\begin{aligned}
 E_0 \left[ \sum_{t=0}^{\infty} R^{-t} C_t - \sum_{t=1}^{\infty} R^{-t} Y_t \right] \\
 \leq \lim_{T \rightarrow \infty} E_0 \left[ \sum_{t=0}^{T-1} R^{-t} C_t - \sum_{t=1}^T R^{-t} Y_t + R^{-T} W_T \right] \\
 = W_0.
 \end{aligned}$$

We then enter this constraint into the Lagrangian with its own Lagrange multiplier (which does not vary over time). Note that our theorem about sufficiency of Euler equations plus transversality does not cover this case, as we did not allow for constraints that involve infinite sums.

Finally, another version of this problem uses  $C_t \leq W_t$ , all  $t$ , as the constraints that prevent excessive borrowing. These constraints of course get their own Lagrange multipliers, which do have time subscripts. Since the constraints will bind only at occasional values of  $t$ , the Lagrange multipliers on them are zero at most  $t$ .