Comments on Chapters 20 and 21 (Calculus of Variations and Optimal Control Theory) of "Introduction to Mathematical Economics" (third edition, 2008) by E.T. Dowling

Erik J. Balder, Mathematical Institute, University of Utrecht, Netherlands

1 Introduction

The textbook "Introduction to Mathematical Economics" (third edition, 2008) by E.T. Dowling, ISBN 0-07-135896 (henceforth IME) is intended for economics and business students who seek to enrich their knowledge of mathematics in sofar as it is immediately applicable to economic and business problems in academics. In this connection IME can be used both for self-study or as a supplement to other texts. It is part of the Schaum Outlines Series, published by McGraw-Hill and, like all other publications in this series, it is characterized by the Outlines' *learning-by-doing philosophy*. Indeed, about each of its subjects IME presents just a little theory, which is then supported by a number of explicitly solved problems (of these the present book contains no less than 1600).[1]

In 2011 I used IME's chapters 20 and 21 (in particular the former) as an initial stepping stone towards optimal control and dynamic programming in my Utrecht course on Optimization. I did this because, for the first time in that course, the students had received very diverse training in mathematics: my audience consisted of a mix of third year Bachelor students of Mathematics and Research Master students of Economics. For this reason my comments below are strictly limited to chapters 20 and 21 of IME. Perhaps, as time progresses, I will have the opportunity to take a similarly sharp look at the other chapters.

Because I found IME to be very useful for my course, I hope that my remarks, which I give in the critical but certainly constructive spirit that is traditional among mathematicians, may help to make its chapters 20-21 even more helpful and clear for its users. While I realize quite well that the average reader of IME is not looking for mathematical precision per se, one main thrust of my remarks below is that certainly when confusion threatens, a modest amount of precision should be exercised. Another such thrust is that at several junctures the material in chapters 20-21 can easily be enriched for the student at a minimum

[1] Unfortunately, per subject these problems would seem to have only very small gradations in difficulty; thus fast learners do not profit enough from it, I observed.

of expansion.[2] Yet another thrust concerns mistakes in the text that should certainly be addressed in the next edition.

2 Major issues, Chapter 20

a. As do many other texts on economics, IME does not specify the domain of integrands such as the function $F(t, x, \dot{x})$ in formula (20.1). Especially for economic applications, where quantities must always be nonnegative, this is not a wise strategy in my opinion.[3]

b1. As do many other texts on the calculus of variations, IME can confuse the novice by its notation. This holds especially for its use of the symbol $\dot{x}$, which can be a real variable at one time and a function at another time (sometimes even in the same expression), and the traditional but lamentable suppression of "(t)"'s, whereas maintaing these could have supported precision and order. For instance, already on p. 462 the meaning of $\dot{x}$ in Euler's equation (20.2a) is very different from one of the $\dot{x}$'s in (20.2b), even though the text declares that (20.2a) can be expressed as (20.2b).

Why not write the function F in real variables t, ξ, η and write Euler's equation (20.2a-b) as $F_\xi(t, x(t), \dot{x}(t)) = \frac{d}{dt} F_\eta(t, x(t), \dot{x}(t))$? I found this approach, carried out systematically, to be very suitable for my class.

b2. As another case of possibly confusing language, consider the sentence written in point 4 at the bottom of p. 463.

c. Chapter 20 focuses entirely on the simplest problem in the CoV and its isoperimetric variant in 20.6 (incidentally, why not add one exercise/example of, say, the case with two isoperimetric constraints to show the proper way to the student?) Why doesn't the chapter include further modifications, such as the free endpoint problem or the problem with endpoint inequalities, in anticipation of those treated in 21.4-21.5? This could already be of benefit for the economic Examples 20.21-20.22, where no statement of initial and end point conditions can be found. And why are there no Bolza-type modifications in neither chapter? (also these can be very relevant for economic growth problems).

In addition, after introducing at least four of such modifications in my own course, I used the opportunity to ask the mathematics students to provide proofs of the corresponding necessary optimality conditions, similar to Example 20.2.

d. Said proof in Example 20.2 ("*To prove* that Euler's equation ...") is not rigorous and a small comment should report this honestly to the reader: the argument for *existence* of $\frac{d}{dt}(\frac{\partial F}{\partial \dot{x}})$, which is a prerequisite for partial integration, is lacking for reasons of convenience.

[2] Of course, a space-saving setup would be to let chapter 21 precede chapter 20, by treating CoV as a special case of Optimal Control, but in my opinion the present order is just right for presentation in class.

[3] A striking example of this bad habit can be found in microeconomics. The fact that the typical Cobb-Douglas function is defined on $\mathbb{R}^n_+$, but is only differentiable on $\mathbb{R}^n_{++}$ has led to erroneous statements in even the most advanced textbooks on microeconomics and/or mathematical economics – see my recent article in Springer LNEMS 655 (2011), pp. 21-38.

e. The initial points 1-2 in 20.5 are a rather correct summary of Arrow's sufficiency theorem (even though I would prefer: "If the function $F(t, \xi, \eta)$ is jointly concave in (ξ, η) ..." (note: I use *function* and not functional!). However, their execution in the subsequent points 1 and 2 falls short. Indeed, significantly more can be said:
1a) .... making the extremal a *unique* global maximum.
1b) .... which is sufficient for a *global* maximum.
2a) .... making the extremal a *unique* global minimum.
2b) .... which is sufficient for a *global* minimum.

The impact of these improvements should also be carried over to various examples. In my own course, I used this opportunity to give students first a full proof of this sufficiency result and then assigned as exercises the proofs of analogous sufficiency results for the modifications mentioned in my remark c above.

f. In 20.1 a mistake has been made, for by itself k_1, which is a square root, can only be nonnegative. The point is that in the relevant expression for k_1 a $\pm$ should have been used. This correction allows for slopes of both signs, as it should be.

g. Figure 20-4 is less than satisfactory, all the more so since the surface formula is not supported by any further explanation.

h. In my opinion 20.32(b) (for instance) could become highly confusing to a student who is trying to solve "minimize $\int_0^1 x^2(t)dt$ subject to $x(0) = 0$ and $x(1) = 1$" ... Of course, this is due to the fact that non-existence of optimal solutions, with its inconvenient consequences for Euler's equation, is not discussed in Chs. 20-21 of IME.

i. What precisely is the significance of "F is neither maximized nor minimized. It is a saddle point." in 20.15? After all, this exercise is about optimization of the integral functional $J(x) := \int_{t_0}^{t_1} F(t, x(t), \dot{x}(t))dt$. At best (and this turns out to be verifiably true in 20.15) the saddle point property of the integrand F can point at non-existence of local minima or maxima of J.

3 Minor matters, Chapter 20

a. The derivation of the length S in 20.2 relies on pure algebraic jugglery with quasi-infinitesimals (dx divided by dt is "calculated" to be $\frac{dx}{dt}$, i.e. the derivative). Instead, it might be preferable to observe in figure 20-1(b) that dx divided by dt, a tangent, is in first approximation equal to the slope $\frac{dx}{dt}$. Personally, I would then also prefer to write Δx instead of dx etc., precisely to avoid such jugglery.

b. In 20.10 what is usually called "local minimum" in Ch. 20 is called "relative minimum". Why? and would it not be advisable to use "relative minimum" throughout? For apparently the index of IME only refers to relative extremum (not to local extremum), a term which is also prominent in previous chapters.

c. Misprints: 1. on p. 463 up to (20.7) three f's should be replaced by F's, 2. in (20.3) there is a needless space in the integrand, 3. ∂m and not ∂_m is intended in line 8 of p. 463, p. 488 "equatiuon" in 20.29.

4 Issues, Chapter 21

Several issues raised in section 2 can be repeated mutatis mutandis and I shall not do this in extenso. I mention notation[4], inclusion of Bolza-type modifications, time-optimal control modifications, infinite horizon modifications, sufficiency results involving concavity/convexity that can be strengthened. Apart from these, there are some new issues:

a. A line following (21.2) states: "Assuming the Hamiltonian is differentiable in y and strictly concave so there is an interior solution and not an endpoint solution". This line is both false and mysterious: what is meant by solution? interior to what? why should strict concavity imply interiority? A follow-up line then proclaims: "If the solution does not (sic) involve an end point, $\partial H/\partial y$ need not (sic) equal zero in the first condition, but H must still be maximized with respect to y ... We shall generally assume interior solutions."

What is true is the following. Next to the deficiency about the *definition domains* of functions, already mentioned in point a of section 2, there is a concomitant deficiency: *ranges* of functions are not mentioned either in IME. But for optimal control theory the specification of such a range (say Ω), i.e., the set of all control points[5], is essential! Only now can the reader grasp what is meant by "there is an interior solution and not an endpoint solution" and by "We shall generally assume interior solutions" on p. 494: at the very least the optimal control function y^* must be supposed to be such that

$$y^*(t) \text{ is an interior point of the set } \Omega. \tag{1}$$

Moreover, the reader now realizes that by "not an endpoint solution" the following is meant: $y^*(t)$, because it is assumed to be an interior point of Ω, cannot be a *boundary point* of Ω. Observe that in this particular context this choice of words can cause confusion: "endpoint" has absolutely nothing to do with the standard endpoint conditions of p. 461 and had better be avoided in my opinion. But the first mysterious sentence as a whole is not clear yet! Could it be that by "interior solution" the author means: the maximizing control point (obtained by maximizing $H(x(t), \theta, \lambda(t), t)$ over θ in Ω) that results from the Pontryagin maximum principle? If so, then it is quite true that *under interiority assumption* (1), combined with the assumption of strict concavity of $H(\xi, \theta, \lambda, t)$ in the control variable θ (plus the assumption that Ω be a convex set), $\partial H/\partial y = 0$ is equivalent to said part of the maximum principle. However, here (1) acts as a condition, and *not* as a consequence (recall IME's "*so* there is an interior

[4] For instance, for pedagocial reasons again I would prefer to write $H(\xi, \theta, \lambda, t) = f(\xi, \theta, t) + \lambda g(\xi, \theta, t)$, etc. Incidentally, why is the end position of the variable t different from its frontal position in F in Ch. 20?

[5] Usually taken to be time-independent for simplicity

solution"). Moreover, we now see that the second mysterious sentence should only have only one "not" and not two: "If the solution involves an end point, $\partial H/\partial y$ need not equal zero in the first condition, but H must still be maximized with respect to y ... We shall generally assume interior solutions."

After these repairs have been made, would it not be fair to the student to indicate in the text, and supported at least by a small example/exercise, that in many optimal control problems the interiority assumption (1) is "too heroic"? Presumably, for many this may be all that they will officially learn about optimal control, so they had better be warned.

b. In 21.5 the symbol x^* is used, apparently for the first time (it might have been better to follow this good notational habit more frequently), to denote the trajectory that is associated to the optimal control *for the free endpoint problem*, which arises from the original optimal control problem by ignoring the end point constraint altogether. It is therefore somewhat mystifying to call $x^*(T)$ simply "the optimal value" instead of "the value at T of the trajectory associated to the optimal control for the free endpoint problem".

c. The three first lines of p. 499 would seem to indicate a contradictio in terminis: "If $x^*(T) < x_{min}$" is incompatible with "when $x^*(T) = x_{min}$"! However, what is meant here is something entirely different: let x^* be as above (i.e., designed for the free endpoint problem) and let x^{**} be the trajectory associated to *another* optimal control, namely the one with endpoint constraint $x(T) = x_{min}$. What is then meant by the first three lines of p. 499 (and this is supported by the sequel on that page) is: "If $x^*(T) < x_{min}$" ... "when $x^{**}(T) = x_{min}$".

d. In Example 6 on pp. 499-500 the above-mentioned three lines on p. 499 have not been properly executed: the check for $\lambda(2) \geq 0$ should be performed explicitly.

e. Misprints: in 3b) on p. 501: $e^{-\rho t}$ should be $e^{-\rho T}$, in D on p. 512 raising to the power $-1.6855t$ should be replaced by multiplying by a factor $e^{-1.6855t}$.