DISCRETIZATION ERROR IN OPTIMAL CONTROL*

William W. Hager
Department of Mathematics
University of Florida
Gainesville, Florida 32611 USA

and

Asen L. Dontchev
Institute of Mathematics
Bulgarian Academy of Sciences
Sofia, Bulgaria

Abstract

Sensitivity analysis is used to estimate the error associated with Euler's discretization to a nonlinear optimal control problem with convex control constraints.

1. Introduction.

In this paper we use a result from sensitivity analysis to estimate the error in Euler's approximation to a nonlinear optimal control problem with convex control constraints. This paper presents some of the key ingredients in the analysis while the complete theory appears in [4]. In earlier papers, Budak *et al.* [1] and Cullum [2] prove convergence of the optimal value associated with discrete approximations to state and control constrained problems. Mordukhovich [8] shows that the discrete optimal cost converges to the true optimal cost if and only if a relaxation of the control problem is stable. Estimates for the error in the optimal control associated with higher order discretizations of unconstrained nonlinear problems are derived by Hager [5]. Dontchev [3] obtains an error estimate for Euler's approximation applied to an optimal control problem with convex cost, linear system dynamics, and linear inequality state and control constraints. In this paper, the assumptions of cost convexity and constraint linearity are dropped -- we consider a problem with nonlinear system dynamics and a general convex control constraint. Our method of analysis makes use of the so-called averaged modulus of smoothness, introduced by Sendov and Popov [9].

*This work was supported by the U.S. Army Research Office Contract DAAL03-89-G-0082 and by the Bulgarian Ministry of Science Contract 127. The research was performed while the first author was a visitor at the University of Florida.

2. Abstract result.

This section states the abstract sensitivity result that is applied to the optimal control problem. We consider a family of equations, each equation depending on a parameter p contained in a metric space P. Associated with each $p \in P$, there is a closed subset Ω_p of a Banach space Z_p, a normed vector space Y_p, and a pair of maps $T_p : Z_p \to Y_p$ and $F_p : \Omega_p \to 2^{Y_p}$. We consider the following problem:

$$\text{Find } z \in \Omega_p \text{ such that } T_p(z) \in F_p(z). \tag{1}$$

For convenience, it is assumed that $0 \in P$. The continuity of the solution map Σ defined by

$$\Sigma(p) = \{ z \in \Omega_p : T_p(z) \in F_p(z) \}$$

is related to stability properties of the following linearized problem:

$$\text{Find } z \in \Omega_p \text{ such that } L_p(z - z_p) + y \in F_p(z), \tag{2}$$

where $L_p : Z_p \to Y_p$ is linear, $z_p \in Z_p$, and $y \in Y$.

Throughout this paper, $\|\cdot\|$ denotes a norm in the appropriate space. Letting $B_r(z)$ denote the closed ball with center z and radius r, we make the following definitions:

$$D_\rho(p) = \sup_{\substack{y,z \in B_\rho(z_p) \cap \Omega_p \\ y \neq z}} \frac{\|T_p(z) - T_p(y) - L_p(z-y)\|}{\|z - y\|}$$

and

$$\delta(p) = \|T_p(z_p) - y_p\|$$

where $y_p \in F_p(z_p)$. (The norms above may depend on p although this dependence is not indicated explicitly.) The following result is a consequence of Corollary 1 in [4]:

THEOREM 1. *Suppose that for some positive σ and γ and for each $y \in B_\sigma(y_p)$,* (2) *has a unique solution, denoted $\Psi_p(y)$, that satisfies the inequality*

$$\|\Psi_p(y_1) - \Psi_p(y_2)\| \leq \gamma \|y_1 - y_2\| \text{ for every } y_1 \text{ and } y_2 \in B_\sigma(y_p). \tag{3}$$

If $D_\rho(p)$ and $\delta(p)$ tend to zero as ρ and p tend to 0, then for each ρ and p in a neighborhood of zero with $\rho \geq \gamma\delta(p)/(1 - \gamma D_\rho(p))$, equation (1) *has a unique solution z such that*

$$\|z_p - z\| \leq \frac{\gamma}{1 - \gamma D_\rho(p)} \|T_p(z_p) - y_p\|.$$

3. Euler's method.

We apply Theorem 1 to Euler's discretization of the following nonlinear control problem with control constraints:

$$
\begin{aligned}
&\text{minimize } \int_I g(x(t),u(t))\,dt \\
&\text{subject to } \dot{x}(t) = f(x(t),u(t)) \text{ and } u(t) \in U \text{ a. e. } t \in I, \\
&x(0) = a, \quad x \in W^{1,\infty}, \quad u \in L^{\infty},
\end{aligned}
\tag{4}
$$

where $f : R^{n+m} \to R^n$, $g : R^{n+m} \to R$, $U \subset R^m$ is nonempty, closed and convex, a is the given starting condition, I is the interval $[0,1]$, L^{∞} is the space of essentially bounded functions, and $W^{1,\infty}$ is the space of Lipschitz continuous functions. We assume that there exists a solution (x^*,u^*) to (4) with u^* Riemann integrable, that there exists a closed set $\Delta \subset R^{n+m}$ where both f and g are twice continuously differentiable, and that there exists $\delta > 0$ such that $(x^*(t),u^*(t)) \in \Delta$ and the distance from $(x^*(t),u^*(t))$ to the boundary of Δ is at least δ for every $t \in I$. When we write $\dot{x}^*$, we mean a function whose values on I coincide with those of $f(x^*,u^*)$.

Let H denote the Hamiltonian defined by

$$
H(x,u,\lambda) = g(x,u) + \lambda^T f(x,u),
$$

and let $\lambda = \lambda^*$ be the solution of the adjoint equation

$$
\dot{\lambda}(t) = -\frac{\partial H(x(t),u(t),\lambda(t))}{\partial x} \quad \text{a. e. } t \in I, \quad \lambda(1) = 0,
$$

associated with $x = x^*$ and $u = u^*$. By the minimum principle [7, p. 134], we have:

$$
\frac{\partial H(x^*(t),u^*(t),\lambda^*(t))}{\partial u}(v - u^*(t)) \geq 0 \quad \text{a. e. } t \in I \text{ and for every } v \in U.
$$

Given a natural number N, let $h = 1/N$ be the mesh spacing, and let x_i and u_i denote approximations to $x(t)$ and $u(t)$ at $t = t_i = ih$. We consider the Euler discretization of (4) given by

$$
\begin{aligned}
&\text{minimize } \sum_{i=0}^{N-1} h\, g(x_i,u_i) \\
&\text{subject to } x_{i+1} = x_i + h\, f(x_i,u_i) \text{ and } u_i \in U, \quad i = 0, 1, \cdots, N-1, \quad x_0 = a.
\end{aligned}
\tag{5}
$$

If (x^h,u^h) denotes a solution to (5), let $\lambda = \lambda^h$ denote the solution of the discrete adjoint equation

$$
\lambda_i = \lambda_{i+1} + \frac{h\,\partial H(x_i,u_i,\lambda_{i+1})}{\partial x}, \quad i = N-1, N-2, \cdots, 0, \quad \lambda_N = 0, \tag{6}
$$

associated with $x = x^h$ and $u = u^h$. By the discrete minimum principle [7, p. 280], we have

$$\frac{\partial H(x_i^h, u_i^h, \lambda_{i+1}^h)}{\partial u_i}(v - u_i^h) \ \geq \ 0 \quad \text{for all } v \in U, \quad i \ = \ 0, 1, \cdots, N-1. \tag{7}$$

In order to estimate the distance between (x^*, u^*) and (x^h, u^h), we need a coercivity type assumption for the discrete problem. Define the following matrices:

$$A(t) = \frac{\partial f^*(t)}{\partial x}, \ B(t) = \frac{\partial f^*(t)}{\partial u}, \ Q(t) = \frac{\partial^2 H^*(t)}{\partial^2 x}, \ R(t) = \frac{\partial^2 H^*(t)}{\partial^2 u}, \ S(t) = \frac{\partial^2 H^*(t)}{\partial x \partial u}.$$

Here $f^*(t)$ and $H^*(t)$ stand for $f(x^*(t), u^*(t))$ and $H(x^*(t), u^*(t), \lambda^*(t))$, respectively. Letting A_i, B_i, Q_i, S_i, and R_i denote the corresponding time varying matrices evaluated at $t = t_i$, we assume that there exists a scalar $\alpha > 0$, α independent of N, such that

$$u^T R_i u \ \geq \ \alpha |u|^2, \ 0 \leq i \leq N-1, \text{ whenever } u = v - w \text{ with } v \text{ and } w \in U, \tag{8}$$

and

$$\sum_{i=0}^{N-1} x_i^T Q_i x_i \ + \ u_i^T R_i u_i \ + \ 2x_i^T S_i u_i \ \geq \ \alpha \sum_{i=0}^{N-1} |u_i|^2 \tag{9}$$

whenever $u_i = v_i - w_i$ for some v_i and $w_i \in U$, and

$$x_{i+1} \ = \ x_i \ + \ hA_i x_i \ + \ hB_i u_i, \ i = 0, 1, \cdots, N-1, \ x_0 \ = \ 0.$$

Obviously, the discrete condition (8) holds if there exists $\alpha > 0$ such that

$$u^T R(t) u \ \geq \ \alpha |u|^2 \text{ for every } t \in I \text{ and for each } u = v - w \text{ with } v \text{ and } w \in U.$$

In Appendix 1 of [4], we show that assumption (9) for the discrete problem can be deduced from an analogous assumption for the continuous problem if u^* is continuous. In analyzing the discrete problem (5), we utilize a discrete L^p norm defined by

$$(\|u\|_{L^p})^p \ = \ \sum_{i=0}^{N-1} h|u_i|^p, \ 1 \leq p < \infty, \quad \text{and} \quad \|u\|_{L^\infty} \ = \ \text{maximum } \{ \ |u_i| : 0 \leq i < N \ \}.$$

If ϕ and v satisfy the finite difference system

$$\phi_{i+1} \ = \ \phi_i \ + \ hA_i \phi_i \ + \ hv_i, \ i = 0, 1, \cdots, N-1, \ \phi_0 \ = \ 0,$$

then there exists a constant c, independent of h, such that

$$|\phi_j| \ \leq \ c\|v\|_{L^1} \ \leq \ c\|v\|_{L^2}. \tag{10}$$

Squaring this inequality, multiplying by h, and summing over j yields

$$\|\phi\|_{L^2} \ \leq \ c\|v\|_{L^2}.$$

Hence, if the coercivity condition (9) holds relative to the control, then the following joint state-control coercivity condition holds: There exists $\alpha > 0$ such that

$$h\sum_{i=0}^{N-1} x_i^T Q_i x_i + u_i^T R_i u_i + 2x_i^T S_i u_i \ge \alpha(\|x\|_{L^2}^2 + \|u\|_{L^2}^2)$$

whenever $u_i = v_i - w_i$ for some v_i and $w_i \in U$, and

$$x_{i+1} = x_i + hA_i x_i + hB_i u_i, \ i = 0, 1, \cdots, N-1, \ x_0 = 0.$$

Our convergence result for the discrete problem is expressed in terms of a modulus of smoothness introduced by Sendov and Popov [9]. The local modulus of continuity $\omega(u;t,h)$ of the function u is defined by

$$\omega(u;t,h) = \sup\{|u(a)-u(b)| : a, b \in [t-h/2, t+h/2] \cap I\},$$

while the average modulus of smoothness τ is given by

$$\tau(u;h) = \int_I \omega(u;t,h)\,dt.$$

In [9, pp. 8–11] it is shown that $\tau(u;h) \to 0$ as $h \to 0$ if and only if the bounded function u is Riemann integrable on I; moreover, $\tau(u;h) = O(h)$ if and only if u has bounded variation on I.

THEOREM 2. *If u^* is Riemann integrable and the coercivity assumptions* (8) *and* (9) *hold, then for all N sufficiently large, there exists a local minimizer (x^h, u^h) of* (5) *such that*

$$\underset{0 \le i \le N-1}{\text{maximum}} |u^*(t_i) - u_i^h| = O(h + \tau(u^*;h)),$$

$$\underset{0 \le i \le N}{\text{maximum}} |x^*(t_i) - x_i^h| = O(h + \tau(u^*;h)),$$

$$\underset{0 \le i \le N}{\text{maximum}} |\lambda^*(t_i) - \lambda_i^h| = O(h + \tau(u^*;h)),$$

$$\underset{0 \le i \le N-1}{\text{maximum}} \left|\dot{x}^*(t_i) - \frac{x_{i+1}^h - x_i^h}{h}\right| = O(h + \tau(u^*;h)).$$

Hence, if u^* has bounded variation, then each of these error estimates is of order h.

Proof. We apply Theorem 1 to the necessary conditions associated with the discrete problem (5). The parameter p of Theorem 1 is identified with the mesh spacing h, the set Ω_p consists of discrete triples (x, u, λ) where $u_i \in U$ for each i. Component i, $0 \le i \le N-1$, of the operators T_p and F_p, denoted T_i^h and F_i^h respectively, is the following:

$$T_i^h(x,u,\lambda) = \begin{bmatrix} \dfrac{\partial H(x_i,u_i,\lambda_{i+1})}{\partial x} + \dfrac{\lambda_{i+1}-\lambda_i}{h} \\ \dfrac{\partial H(x_i,u_i,\lambda_{i+1})}{\partial u} \\ f(x_i,u_i) - \dfrac{x_{i+1}-x_i}{h} \end{bmatrix} \quad \text{and} \quad F_i^h(x,u,\lambda) = \begin{bmatrix} 0 \\ \partial U(u_i) \\ 0 \end{bmatrix},$$

where $\partial U(u_i) = \{ w : w^T(v-u_i) \ge 0 \text{ for every } v \in U \}$ is the normal cone to the set U at u_i. In the discrete space Z_p associated with Ω_p, we use the L^∞ norm for each of the 3 components x, u, and λ. In the discrete space Y_p associated with the range of T_p, we use an L^1 norm for the first and last component and the L^∞ norm for the middle component. That is, if $y = (a,b,c) \in Y_p$, then

$$\|y\|_p = \|a\|_{L^1} + \|b\|_{L^\infty} + \|c\|_{L^1}.$$

The point z_p of Theorem 1 is given by $z_p = (x^I, u^I, \lambda^I)$ where

$$x_i^I = x^*(t_i), \quad u_i^I = u^*(t_i), \quad \lambda_i^I = \lambda^*(t_i).$$

Component i of the point y_p, denoted y_i^h, is the triple

$$y_i^h = \begin{bmatrix} 0 \\ \dfrac{\partial H(x_i^I, u_i^I, \lambda_i^I)}{\partial u} \\ 0 \end{bmatrix}.$$

Observe that with this choice for y_p, we have $y_p \in F(z_p)$. The linear operator L_p is defined in the following way: It acts on a discrete triple (x,u,λ) to produce a vector whose i-th component is

$$L^h(x,u,\lambda)_i = \begin{bmatrix} A_i^T\lambda_{i+1} + Q_i x_i + S_i u_i + \dfrac{\lambda_{i+1}-\lambda_i}{h} \\ R_i u_i + S_i^T x_i + B_i^T \lambda_{i+1} \\ A_i x_i + B_i u_i - \dfrac{x_{i+1}-x_i}{h} \end{bmatrix}.$$

It can be verified that under the smoothness assumptions and for ρ smaller than δ, we have $D_\rho(h) \to 0$ as $h \to 0$. Now consider the term $T_p(z_p) - y_p$:

$$(T_p(z_p) - y_p)_i = \begin{bmatrix} \dfrac{\partial H(x_i^I, u_i^I, \lambda_{i+1}^I)}{\partial x} + \dfrac{\lambda_{i+1}^I - \lambda_i^I}{h} \\ \dfrac{\partial H(x_i^I, u_i^I, \lambda_{i+1}^I)}{\partial u} - \dfrac{\partial H(x_i^I, u_i^I, \lambda_i^I)}{\partial u} \\ f(x_i^I, u_i^I) - \dfrac{x_{i+1}^I - x_i^I}{h} \end{bmatrix}. \tag{11}$$

The middle component of this vector is O(h) since λ^* is Lipschitz continuous. Since the analysis of the first and last component in (11) is similar, we only focus on the last component:

$$\left| f(x_i^I, u_i^I) - \frac{x_{i+1}^I - x_i^I}{h} \right| \le \frac{1}{h} \int_{t_i}^{t_{i+1}} |f(x_i^I, u_i^I) - \dot{x}^*(t)|\, dt \le \tag{12}$$

$$\frac{1}{h} \int_{t_i}^{t_{i+1}} |f(x_i^I, u_i^I) - f(x^*(t), u^*(t))|\, dt \le \frac{c}{h} \int_{t_i}^{t_{i+1}} h + \omega(u^*; t, 2h)\, dt$$

where c denotes a generic constant, independent of h. Multiplying (12) by h, summing over i, and exploiting the inequality $\tau(u; kh) \le k\tau(u; h)$ for each natural number k (see [9, p. 11]), it follows that

$$\|T_p(z_p) - y_p\|_p = O(h + \tau(u^*; h)).$$

Next, we need to analyze the linearization and establish the existence of a constant γ satisfying (3). The analysis essentially parallels that of [6] except that continuous norms are replaced by their discrete analogues. We need to examine how the solution to the following system depends on the perturbations q_i, r_i, and s_i:

$$\begin{gathered} A_i^T\lambda_{i+1} + Q_i x_i + S_i u_i + \frac{\lambda_{i+1} - \lambda_i}{h} + q_i = 0, \quad \lambda_N = 0, \\ (R_i u_i + S_i^T x_i + B_i^T \lambda_{i+1} + r_i)(v - u_i) \ge 0 \ \text{ for every } v \in U, \\ A_i x_i + B_i u_i - \frac{x_{i+1} - x_i}{h} + s_i = 0, \quad x_0 = a, \end{gathered} \tag{13}$$

$i = 0, 1, \cdots, N-1$. Note that the system (13) constitutes the first-order necessary conditions (see [7, p. 280]) associated with the following quadratic program:

$$\text{minimize } h \sum_{i=0}^{N-1} \frac{1}{2} x_i^T Q_i x_i + \frac{1}{2} u_i^T R_i u_i + x_i^T S_i u_i + q_i^T x_i + r_i^T u_i \tag{14}$$

subject to $x_{i+1} = x_i + hA_i x_i + hB_i u_i + hs_i$ and $u_i \in U$, $0 \le i \le N-1$, $x_0 = a$.

By Lemma 2 of [6], there is a one-to-one correspondence between a solution to (14) and a solution to (13) when (9) holds.

Now consider the perturbations (q^i, r^i, s^i) for $i = 1$ and 2. Let (x^i, u^i, λ^i) denote the associated solutions to (13). Referring to Section 2 of [6] and replacing continuous norms by the corresponding discrete norms, we have

$$\|u^1 - u^2\|_{L^2} \le c(\|q^1 - q^2\|_{L^1} + \|r^1 - r^2\|_{L^2} + \|s^1 - s^2\|_{L^1}).$$

Utilizing (10), we also conclude that

$$\|x^1 - x^2\|_{L^\infty} + \|\lambda^1 - \lambda^2\|_{L^\infty} \le c(\|q^1 - q^2\|_{L^1} + \|r^1 - r^2\|_{L^2} + \|s^1 - s^2\|_{L^1}). \tag{15}$$

Finally, by (8) and Lemma 1 of [6], we have

$$\|u^1 - u^2\|_{L^\infty} \le c(\|r^1 - r^2\|_{L^\infty} + \|x^1 - x^2\|_{L^\infty} + \|\lambda^1 - \lambda^2\|_{L^\infty}).$$

Combining this with (15) yields

$$\|x^1 - x^2\|_{L^\infty} + \|u^1 - u^2\|_{L^\infty} + \|\lambda^1 - \lambda^2\|_{L^\infty} \le$$
$$c(\|q^1 - q^2\|_{L^1} + \|r^1 - r^2\|_{L^\infty} + \|s^1 - s^2\|_{L^1}).$$

Hence, there exists a constant γ such that (3) holds with $\sigma = \infty$.

By Theorem 1, there exists a solution to the discrete necessary conditions (6) and (7) associated with (5) which satisfies the first 3 estimates of Theorem 2. The discrete and continuous state equations along with the previously established error estimates imply that

$$\left|\dot{x}^*(t_i) - \frac{x_{i+1}^h - x_i^h}{h}\right| = |f(x^*(t_i), u^*(t_i)) - f(x_i^h, u_i^h)| = O(h + \tau(u^*; h)),$$

which gives the last estimate of Theorem 2. The fact that x^h and u^h are local minimizers of (5) is established in [4].

REFERENCES

[1] B. M. Budak, E. M. Berkovich, and E. N. Solov'eva, *The convergence of difference approximations in optimal control problems*, Zh. Vychisl. Math. i Mat. Fiz., 9 (1969), pp. 522–547.

[2] J. Cullum, *Finite-dimensional approximations of state-constrained continuous optimal control problems*, SIAM J. Control, 10 (1972), pp. 649–670.

[3] A. L. Dontchev, *Error estimates for a discrete approximations to constrained control problems*, SIAM J. Numer. Anal., 13 (1981), pp. 500–514.

[4] A. L. Dontchev and W. W. Hager, *Sensitivity in nonlinear optimal control*, July 4, 1990, submitted.

[5] W. W. Hager, *Rate of convergence for discrete approximations to unconstrained control problems*, SIAM J. Numer. Anal., 13 (1076), pp. 449–471.

[6] W. W. Hager, *Multiplier methods for nonlinear optimal control*, SIAM J. Numer. Anal., 27 (1990), pp. 1061–1080.

[7] A. D. Ioffe and V. M. Tihomirov, *Theory of extremal problems*, North Holland, Amsterdam, 1979.

[8] B. Sh. Mordukhovich, *On difference approximations of optimal control systems*, Appl. Math. Mech., 42 (1978), pp. 452–461.

[9] Bl. Sendov and V.A. Popov, *The averaged moduli of smoothness*, J. Wiley & Sons, New York, 1988.