Newton's Method for Problems of Optimal Control of Heterogeneous Systems*

Vladimir M. VELIOV†

1 Introduction

This paper deals with numerical approximation of optimal control problems for heterogeneous systems by generalized Newton's method (SQP method, see e.g. [1, 3]). The main concern is to establish error estimates, and moreover, such estimates that reveal the advantage of using appropriate second order discretization schemes. In this respect the main result in the paper is new even in the special case of an ordinary control system, but the presence of integral relations in the heterogeneous systems creates essential additional difficulties.

Problems of control of heterogeneous system arise in many areas, especially in population biology, economics and social sciences. A heterogeneous control system can be viewed as a parameterized (infinite) family of ordinary control systems, coupled together by some aggregated quantities (called *externalities* in economic context) on which the dynamics of the individual control systems depends. Formally, these are infinite-dimensional control systems with a nonlocal dependence on the spacial variable (the latter being the parameter of heterogeneity). The meaning of the parameter of heterogeneity could be: age or duration (in population or vintage capital models), propensity to risky behavior (in social and epidemiplogical models), etc. We refer to [9] for an extended bibliography.

A general heterogeneous control problem $\mathcal{P}$ is formulated in Section 2, where also a necessary and a sufficient optimality conditions are presented. As usually these conditions have the form of variational inequalities (generalized equations) $\mathcal{E}$ to which

*This research was partly supported by the Austrian Science Foundation under contract N0. 15618-OEK

†Institute for Econometrics, Operations Research and Systems Theory, Vienna University of Technology, Vienna, Austria, vveliov@eos.tuwien.ac.at

we apply the Newton method. The sufficient condition implies also Lipschitz stability of the problem $\mathcal{P}$ which is a key supposition for the error analysis.

In Section 3 we present a version of the Newton method for abstract generalized equations and a corresponding error estimate adapted from [5]. The principle conditions for *consistency* and *stability* are involved.

Section 4 describes the implementation of the generalized Newton method to our control problem. The differential equation is discretized by a second order Runge-Kutta scheme, which leads to a discrete-time optimal control problem $\mathcal{P}_N$. In order to obtain a discrete version, $\mathcal{E}_N$, of the generalized equations $\mathcal{E}$ we use appropriate approximations to the adjoint differential equation and to the variational part of the optimality conditions $\mathcal{E}$. These approximations are designed in such a way that $\mathcal{E}_N$ represents exactly the system of necessary conditions (discrete maximum principle) for $\mathcal{P}_N$. This allows us to deduce in Section 5 the Lipschitz stability of $\mathcal{E}_N$ (required by the abstract convergence result) as a consequence of the Lipschitz stability of $\mathcal{P}$. Following ideas from [6] we estimate the order of consistency of the approximation $\mathcal{E}_N$ to $\mathcal{E}$, which turns out to be higher than first (typically second), under mild regularity conditions. In particular, the optimal control is assumed only Lipschitz continuous and may have points of nondifferentiablity, which usually exist in presence of control constraints. In Section 5 we formulate also our main result – the error estimate of the method.

2 The General Model and the Optimality Conditions

The general model of a heterogeneous optimal control systems that we consider in this paper has the form

$$\min \int_P l(x(T,p))\,\mathrm{d}p \;+\; \int_0^T \int_P L(x(t,p),y(t,p),u(t,p))\,\mathrm{d}p\,\mathrm{d}t, \tag{1}$$

$$\begin{aligned} \dot{x}(t,p) &= f(x(t,p),y(t,p),u(t,p)), \quad x(0,p)=x_0(p), && (2)\\ y(t,p) &= \int_P g(p,x(t,q),y(t,q),u(t,q))\,\mathrm{d}q, && (3)\\ & u(t,p)\in U, \end{aligned}$$

where $t\in[0,T]$ is interpreted as time, and "dot" means differentiation with respect to t, the parameter $p\in P$ is finite-dimensional, $x\in\mathbf{R}^n$ is the state variable, $y\in\mathbf{R}^m$ is an aggregated state variable, $u\in\mathbf{R}^r$ is a control, x_0 is a given initial condition. The functions f, g, l, and L are defined in the respective spaces.

We mention that an explicit dependence on t and p of the functions involved can be easily included by introducing two additional state variables.

We shall use the following function spaces:
$\mathcal{X}$ – the space of functions $x : [0,T] \times P \mapsto \mathbf{R}^n$ such that for a.e. p the function $x(\cdot,p)$ is Lipschitz continuous uniformly in p, and for every t the function $x(t,\cdot)$ belongs to $L_\infty(P;\mathbf{R}^n)$;
$\mathcal{X}_0$ – the subset of $\mathcal{X}$ consisting of those x which satisfy $x(0,\cdot) = x_0(\cdot)$;
$\mathcal{Y}$ – the space $L_\infty([0,T] \times P;\mathbf{R}^m)$;
$\mathcal{U} = \{u \in L_\infty([0,T] \times P;\mathbf{R}^r) \,:\, u(t,p) \in U \text{ for a.e. } (t,p)\}$;
$\mathcal{S} = \mathcal{X}_0 \times \mathcal{Y} \times \mathcal{U}$.

Suppositions:
(A1) U is convex and compact, P is compact and Lebesgue measurable, $\mathrm{meas}(P) > 0$, $x_0 \in L_\infty(P)$;
(A2) all derivatives up to order two of the functions f, g, l and L, with respect to x, y and u, exist and are locally Lipschitz continuous (uniformly in p, if g is concerned); g is differentiable with respect to p and the derivative is locally Lipschitz, locally uniformly in (x,u);
(A3) For every $x \in \mathcal{X}$, $u \in \mathcal{U}$, equation (3) has a unique solution[1] $y(t,p)$
$= \Upsilon(t,p,x(t,\cdot),u(t,\cdot))$ in $\mathcal{Y}$, where Υ depends Lipschitz continuously on x and u.

Given $u \in \mathcal{U}$, it can be proved similarly as in [14] or [2, Lemma 5.3] that a solution to (2),(3) exists locally in t, and can be extended as long as $\|x(t,\cdot)\|_\infty$ does not escape to infinity. If for $u \in \mathcal{U}$ the solution (x,y) exists on $[0,T]$ then $s = (x,y,u) \in \mathcal{S}$ is called a *control-trajectory triplet*. The corresponding value of the objective function will be denoted by $J(u)$.

For a given reference control-trajectory triplet $s = (x,y,u) \in \mathcal{S}$ we define the *adjoint equation*

$$-\dot\xi(t,p) = \xi(t,p) f_x(s(t,p)) + \int_P \eta(t,q) g_x(q,s(t,p))\,\mathrm{d}q + L_x(s(t,p)), \qquad (4)$$
$$\xi(T,p) = l_x(x(T,p)),$$
$$\eta(t,p) = \xi(t,p) f_y(s(t,p)) + \int_P \eta(t,q) g_y(q,s(t,p))\,\mathrm{d}q + L_y(s(t,p)). \qquad (5)$$

Everywhere in the paper subscripts that have no meaning of natural numbers denote differentiation with respect to the specified variable. Obviously the above system can be viewed as a distributed parameter system, where the state space is $L_\infty(P;\mathbf{R}^n) \times$

[1] In most of the applications the author knows, the function g is either independent of y, or has a cascade form: the first component of g does not depend on y, the second may depend on y_1 only, and so on. In this case (A3) is certainly fulfilled.

$L_\infty(P;\mathbf{R}^m)$. We take this into account in the next definition. For $z=(s,\xi,\eta)\in L_\infty(P;\mathbf{R}^n\times\mathbf{R}^m\times\mathbf{R}^r)\times L_\infty(P;\mathbf{R}^n)\times L_\infty(P;\mathbf{R}^m)\mapsto\mathbf{R}$ we define the *Hamiltonian*

$$H(z)=\int_P\left[\xi(p)f(s(p))+\int_P\eta(q)g(q,s(p))\,\mathrm{d}q+L(s(p))\right]\,\mathrm{d}p.$$

Here and below $\xi(p)$ and $\eta(p)$ are considered as vector-rows while the primal variables $x(p)$, $y(p)$ and $u(p)$ are columns, so that the multiplication in the above expression is, in fact, the scalar product.

For z and Δz as in the definition of H we introduce the following notations:

$$H_s(z)(p)=\xi(p)f_s(s(p))+\int_P\eta(q)g_s(q,s(p))\,\mathrm{d}q+L_s(s(p)),$$

$$H_\xi(z)(p)=f(s(p));\quad H_\eta(z)(p)=\int_P g(p,s(q))\,\mathrm{d}q,$$

$$(H_z(z)\Delta z)(p)=H_s(z)(p)\Delta s(p)+\Delta\xi(p)H_\xi(z)(p)+\Delta\eta(p)H_\eta(z)(p),$$

and also

$$(H_{ss}(z)\Delta s)(p):=H_{ss}(z)(p)\Delta s(p)$$

$$=\left[\xi(p)f_{ss}(s(p))+\int_P\eta(q)g_{ss}(q,s(p))\,\mathrm{d}q+L_{ss}(s(p))\right]\Delta s(p),$$

$$(H_{\xi s}(z)\Delta s)(p)=f_s(s(p))\Delta s(p),\quad (H_{s\xi}(z)\Delta\xi)(p)=\Delta\xi(p)f_s(s(p)),$$

$$(H_{\eta s}(z)\Delta s)(p)=\int_P g_s(p,s(q))\Delta s(q)\,\mathrm{d}q,\quad (H_{s\eta}(z)\Delta s)(p)=\int_P\Delta\eta(q)g_s(q,s(p))\,\mathrm{d}q,$$

$$H_{zs}\Delta z=H_{ss}\Delta s+H_{\xi s}\Delta\xi+H_{\eta s}\Delta\eta,\quad H_{z\xi}\Delta z=H_{s\xi}\Delta s,\quad H_{z\eta}\Delta z=H_{s\eta}\Delta s.$$

Clearly, the above notations can be interpreted also as derivative of H in the respective function spaces. In particular, the following (formal) expansion holds

$$H(z+\Delta z)=H(z)+\int_P(H_z(z)\Delta z)(p)\,\mathrm{d}p$$

$$+\frac{1}{2}\int_P\left[\langle(H_{zs}(z)\Delta z)(p),\Delta s(p)\rangle+\langle(H_{z\xi}(z)\Delta z)(p),\Delta\xi(p)\rangle+\langle(H_{z\eta}(z)\Delta z)(p),\Delta\eta(p)\rangle\right]\,\mathrm{d}p.$$

Proposition 1 *Let $s=(x,y,u)\in\mathcal{S}$ be a given reference control-trajectory triplet. Then there exists a number $\nu>0$ such that for every admissible control $\tilde u=u+\Delta u\in\mathcal{U}$ with $\|\Delta u\|_\infty\le\nu$*
(i) Equations (2),(3) have a (unique) solution $\tilde x$, $\tilde y$ on $[0,T]\times P$;
(ii) One can represent $\tilde x=x+\Delta x+o(\|\Delta u\|_\infty^2)$, $\tilde y=y+\Delta y+o(\|\Delta u\|_\infty^2)$, where $\Delta s=(\Delta x,\Delta y,\Delta u)$ satisfies the linearized equations

$$\Delta x_t(t,p) = f_s(s(t,p))\Delta s(t,p), \tag{6}$$

$$\Delta y(t,p) = \int_P g_s(p,s(t,q))\Delta s(t,q)\,\mathrm{d}q; \tag{7}$$

(iii) The adjoint system (4),(5) corresponding to s has a unique solution (ξ,η), and

$$\begin{aligned} J(u+\Delta u)-J(u) &= \int_0^T\int_P H_u(z(t,\cdot))(p)\Delta u(t,p)\,\mathrm{d}p\,\mathrm{d}t \\ &+ \frac{1}{2}\int_0^T\int_P \langle H_{ss}(z(t,\cdot))(p)\Delta s(t,p),\Delta s(t,p)\rangle\,\mathrm{d}p\,\mathrm{d}t \\ &+ \frac{1}{2}\int_P \langle l_{xx}(x(T,p))\Delta x(T,p),\Delta x(T,p)\rangle\,\mathrm{d}p \\ &+ o(\|\Delta u\|_{L_2}^2). \end{aligned} \tag{8}$$

The first claim of the proposition, as well as the estimation

$$\|\Delta x\|_\infty + \|\Delta y\|_\infty \le C\|\Delta u\|_\infty$$

(from which it follows, having in mind the local existence) can be proven similarly as [9, Proposition 1]. The proof of the existence and uniqueness of the solution of the adjoint system employs a standard fixed-point argument and is similar to that of [2, Lemma 5.2] or [9, Proposition 2]. The proof of (8) follows the line of proof of the similar claim concerning usual optimal control systems, therefore is omitted.

As a direct consequence we obtain a necessary optimality condition in the form of local maximum principle[2].

Theorem 1 *If $s\in\mathcal{S}$ is optimal, then the adjoint system corresponding to s has a unique solution $(\xi,\eta)\in\mathcal{X}\times\mathcal{Y}$, and it satisfies for a.e. (t,p)*

$$H_u(z(t,\cdot))(p)\in -N_U(u(t,p)), \tag{9}$$

where

$$N_U(v)=\begin{cases}\emptyset & \text{if } v\notin U,\\ \{\nu\,:\,\langle\nu,w-v\rangle\le 0 \text{ for all } w\in U\} & \text{if } v\in U.\end{cases}$$

is the normal cone to U at v.

Coercivity: We say that the system is coercive at s if there exists $\rho>0$ such that for every $\Delta u\in\mathcal{U}-\mathcal{U}$ and corresponding solution $\Delta s=(\Delta x,\Delta y,\Delta u)$ of (6),(7) with $\Delta x(0,\cdot)=0$ it holds that

$$\int_P \langle l_{xx}(x(T,p))\Delta x(T,p),\Delta x(T,p)\rangle\,\mathrm{d}p + \int_0^T\int_P \langle H_{ss}(z(t,\cdot))(p)\Delta s(t,p),\Delta s(t,p)\rangle\,\mathrm{d}p\,\mathrm{d}t \ge \rho\|\Delta u\|_{L_2}^2.$$

The following is another consequence of Proposition 1.

[2]In fact a global maximum principle also holds and can be proven by the method of needle variations for infinite dimensional control systems, see e.g. [7, 8].

Theorem 2 *Let (A1)–(A3) and Coercivity hold at $z \in \mathcal{S}$. Let, moreover, z satisfy the minimum principle (9). Then s is locally (in L_∞) optimal.*

The primal/dual system and the maximum principle can be rewritten as the following generalized equations[3] with respect to $s \in \mathcal{S}$, $(\xi, \eta) \in \mathcal{X} \times \mathcal{Y}$:

$$\begin{aligned} 0 &= -\dot{x}(t,p) + H_\xi(z(t,\cdot))(p), && (10)\\ 0 &= -y(t,p) + H_\eta(z(t,\cdot))(p), && (11)\\ 0 &= \dot{\xi}(t,p) + H_x(z(t,\cdot))(p), && (12)\\ 0 &= -\xi(T,p) + l_x(x(T,p)), && (13)\\ 0 &= -\eta(t,p) + H_y(z(t,\cdot))(p), && (14)\\ 0 &\in H_u(z(t,\cdot))(p) + N_U(u(t,p)). && (15) \end{aligned}$$

These generalize equations will be approached by a version of the Newton method that will be presented in an abstract setting in the next section.

3 Newton's Method for Generalized Equations

In this section we adapt some of the results from [5] concerning the convergence of the Newton method applied to generalized equations.

Let Z and D be two subsets of Banach spaces, and let Z be closed and convex. We consider the generalized equation

$$0 \in F(z) + W(z), \tag{16}$$

where $F : Z \mapsto D$ is a single-valued and $W : Z \Rightarrow D$ is a set-valued mapping, respectively. The Newton method will not be applied directly to this inclusion, rather, to an "approximate" inclusion, presumably in finite-dimensional spaces. For this reason we consider sequences of subsets Z_N and D_N of linear spaces, with closed and convex Z_N, sequences of mappings F_N and W_N, and a "projection" $\pi_N : Z \mapsto Z_N$, as the following diagram shows:

$$\begin{array}{ccc} Z & \xrightarrow{F,W} & D \\ \pi_N \downarrow & & \\ Z_N & \xrightarrow{F_N,W_N} & D_N. \end{array}$$

The Newton method applied to the generalized equation

$$0 \in F_N(z) + W_N(z) \tag{17}$$

[3]The term "generalized" refers to the fact that the last relation is an inclusion rather than an equation.

consists of the following: if z^k is the current iterate, the next iterate z^{k+1} is found from the "linearized" equation

$$0 \in F_N(z^k) + F'_N(z^k)(z^{k+1} - z^k) + W_N(z^{k+1}), \quad k = 0, 1, \ldots. \tag{18}$$

Such a sequence will be called *Newton sequence*.

We shall use the notation $B_\beta(z) := \{z' \in Z : \|z' - z\| \leq \beta\}$, for a ball in Z, and similarly in the other spaces involved. Moreover, denote

$$G_N(\bar{z}; z) := F_N(\bar{z}) + F'_N(\bar{z})(z - \bar{z}).$$

We shall suppose the following.

(B1) (*Existence*) Inclusion (16) has a solution $\hat{z}$;
(B2) (*Smoothness*) For every N the function F_N is Frechét differentiable in $B_\beta(\pi_N(\hat{z}))$ (for some $\beta > 0$ and the derivative F'_N is locally Lipschitz with a constant M independent of N;
(B3) (*Consistency*) There is a sequence $d_N \in F_N(\pi_N(\hat{z})) + W_N(\pi_N(\hat{z}))$ such that $\|d_N\| \longrightarrow 0$ with $N \to +\infty$.

Define the set-valued map $\Gamma_N : D_N \Rightarrow Z_N$ as the inverse of the map $z \longrightarrow G_N(\pi_N(\hat{z}); z) + W_N(z)$. That is,

$$z \in \Gamma_N(d) \iff d \in G_N(\pi_N(\hat{z}), z) + W_N(z).$$

Clearly $\pi_N(\hat{z}) \in \Gamma_N(d_N)$. The next supposition requires existence of *Lipschitz localization* [12] of Γ_N around $(d_N, \pi_N(\hat{z}))$.

(B4) (*Stability*) There exist positive numbers β, δ, and L such that the map $d \longrightarrow \Gamma_N(d) \cap B_\beta(\pi_N(\hat{z}))$ is single-valued and Lipschitz continuous with constant L in $B_\delta(d_N)$.

The next theorem follows from the results in [5, Section 3].

Theorem 3 *For every $L' > L$ and $q \in (0, 1)$ there exist $\delta > 0$ and N_0 such that if $N \geq N_0$, then (17) has a unique solution $\hat{z}_N$ in $B_\delta(\pi_N(\hat{z}))$, and*

$$\|\hat{z}_N - \pi_N(\hat{z})\| \leq L'\|d_N\|.$$

Moreover, for every $z_N^0 \in B_\delta(\pi_N(\hat{z}))$ there is a unique Newton sequence $z_N^k \in B_\delta(\pi_N(\hat{z}))$, namely satisfying

$$0 \in G_N(z_N^k; z_N^{k+1}) + W_N(z_N^{k+1}),$$

and it holds that

$$\|z_N^k - \hat{z}_N\| \leq q^{2^k - 1}\|z_N^0 - \hat{z}_N\|.$$

4 Implementation for Control of Heterogeneous Systems

As a consequence of Theorem 3 we obtain the inequality

$$\|z_N^k - \Pi_N(\hat{z})\| \le q^{2^k-1}\|z_N^0 - z_N\| + L'\|d_N\|. \tag{19}$$

The first term in the right-hand side converges quadratically to zero, while the second one depends on the order of consistency of the approximation and on the choice of N. In the implementation below N will be the cardinality of the discretization mesh along the time and along each of the dimensions of the parameter space P. Then the size of the problem (17) is proportional to $N^{\dim(p)+1}$. Therefore one should look for an approximation with high order of consistency in order to achieve a reasonable accuracy $\|d_N\|$ for numerically still tractable values of N. As we shall argue in the end of the next section, for general control constrained problems one cannot achieve better order than second (that is, $\|d_N\| \le \text{const}/N^2$) by means of Runge-Kutta-type discretization schemes. A second order consistency, however, could be achieved by a variety of Runge-Kutta schemes. To keep the exposition shorter we shall demonstrate this using a particular second order Runge-Kutta scheme known as Heun scheme[4]. In addition, we shall suppose that g is independent of y. In the case where g has a cascade form (see footnote 1) the construction of the approximating equations below is exactly the same, but in the general case equations (3) and (5) require respective solvers. Also, we take $L = 0$, which is not a restriction, since a nonzero L can be included in this framework by introducing in a standard way an additional variable. We formulate these additional suppositions as

(A3′) g is independent of y and $L = 0$.

Having in mind the generalized equations in the end of Section 2 we define the spaces

$$Z = \{z = (x, y, u, \xi, \eta) \in \mathcal{X}_0 \times \mathcal{Y} \times \mathcal{U} \times \mathcal{X} \times \mathcal{Y}\}, \quad D = L_\infty^n \times L_\infty^m \times L_\infty^n \times \mathbf{R}^n \times L_\infty^m \times L_\infty^r,$$

where the spaces L_∞ are all taken on $[0, T] \times P$. As above we abbreviate $s = (x, y, u)$.

Now we shall define the discrete space Z_N and the projection π_N. We take a natural number N and denote $h = T/N$ and $t_i = ih$. Moreover we fix a net $P_N \subset P$ (presumably an h-net) that will be used for numerical integration over P. For an element $z = (x, y, u, \xi, \eta) \in Z$ we denote for $i = 0, \dots N-1$ and $p \in P_N$

$$\begin{aligned} z_i(p) &= (x_i(p), y_i(p), u_i'(p), u_i''(p), \xi_i(p), \eta_i(p)) \\ &\stackrel{def}{=} (x(t_i, p), y(t_i, p), u(t_i, p), u(t_{i+1}, p), \xi(t_{i+1}, p), \eta(t_i, p)). \end{aligned}$$

[4] As we show in [6], not every Runge-Kutta scheme that is consistent of second order with smooth differential equations provides second order consistency with "smooth" optimal control problems. Additional requirements for the coefficients of the scheme arise in the control case, which however, are satisfied for the Heun scheme.

We denote also $z_N(p) = x_N(p) \stackrel{def}{=} x(t_N, p)$. Thus each $z_i(\cdot)$, $i = 0, \ldots, N$ is a vector function defined on P_N. Then we define for $z \in Z$

$$\pi_N(z) = (z_0(\cdot), \ldots, z_N(\cdot)). \tag{20}$$

The set Z_N will consist of all vector functions on P_N with the structure of $\pi_N(z)$. That is, Z_N contains the image of π_N, but also the elements like in (20) in which $u_i'(p)$ is not necessarily equal to $u_{i-1}'(p)$. Clearly, this is a finite dimensional space since P_N is a finite set[5]. We endow Z_N with the discrete L_∞-norm.

We shall postpone the definition of the space D_N and the respective norm to the next section. First we give the reader a better intuition, by informally discretizing the problem (1),(2),(3) using a "heuristic" argument.

We introduce a linear "integrator" based on P_N, namely I^N is a linear mapping from the functions $g : P \mapsto \mathbf{R}^m$ to $\mathbf{R}^m$ depending only on the values of g at the points in P_N, and for every integrable function $g : P \mapsto \mathbf{R}^m$

$$I^N(g) = \sum_{p \in P_N} \alpha_p g(p) \approx \int_P g(q)\, \mathrm{d}q,$$

where $\alpha_p > 0$. The meaning of "$\approx$" will be specified in the next section, but I^N is presumably a cubature formula based on P_N.

The Heun scheme applied to a differential equation $\dot{\lambda} = b(t, \lambda)$ is given by

$$\lambda_{i+1} = \lambda_i + 0.5h[b(t_i, \lambda_i) + b(t_{i+1}, \lambda_i + hb(t_i, \lambda_i)].$$

In the case of equation (2) the calculation of the second term would involve the value $y(t_{i+1}, p) \approx y_{i+1}(p)$, which requires the values of $x_{i+1}(\cdot)$. This implicit feature can be avoided by using a predictor Euler step for x which allows calculation of $x_{i+1}(\cdot)$ with (formal) error $O(h^2)$. The resulting formulas become: for $i = 0, \ldots, N-1$

$$\begin{aligned} x_{i+1}(p) &= x_i(p) + 0.5h[f(s_i(p)) + f(\tilde{s}_i(p))], & (21)\\ y_i(p) &= I^N(g(p, s_i(\cdot)), \quad x_0(\cdot) \text{ – given}, & (22) \end{aligned}$$

where we use the abbreviations

$$s_i(p) = \begin{pmatrix} x_i(p) \\ I^N(g(p, x_i(\cdot), u_i'(\cdot)) \\ u_i'(p) \end{pmatrix}, \quad \tilde{s}_i(p) = \begin{pmatrix} x_i(p) + hf(s_i(p)) \\ I^N(g(p, x_i(\cdot) + hf(s_i(\cdot)), u_i''(\cdot)) \\ u_i''(p) \end{pmatrix}.$$

[5]For a more general Runge-Kutta scheme the number of independent discrete controls $u_i'(p), u_i''(p), \ldots$ arising in one time step $[t_i, t_{i+1}]$ and for a fixed $p \in P_N$ equals the number of different intermediate times that the scheme involves (they all should belong to $[t_i, t_{i+1}]$). For the Heun scheme these are t_i and t_{i+1}, therefore two discrete control values are associate to each $[t_i, t_{i+1}]$ and each $p \in P_N$.

Equations (21), (22) will be the approximate versions of equations (10) and (11). To obtain appropriate approximations to the rest of the (generalized) equations in (10)–(15) we proceed as follows. We consider the discrete-time discrete-parameter problem

$$\text{minimize } I^N(l(x^N(\cdot))) \tag{23}$$

subject to (21), (22) and $u_i'(p),\ u_i''(p) \in U$ for $p \in P_N$. Applying the discrete minimum principle (and having in mind the linearity of the operator I^N) one can obtain the following adjoint equations for $i = 0, \ldots, N-1$

$$\begin{aligned}
\xi_{i-1}(p) &= \xi_i(p) + \frac{h}{2}[\xi_i(p)\left(f_x(s_i(p)) + f_x(\tilde{s}_i(p))(1 + hf_x(s_i(p)))\right) \qquad (24)\\
&\quad + I^N\left(\xi_i(\cdot)f_y(\tilde{s}_i(\cdot))g_x(\cdot, \tilde{s}_i(p))\right)\left(1 + hf_x(s_i(p))\right) + 2I^N\left(\eta_i(\cdot)g_x(\cdot, s_i(p))\right)],\\
\xi_{N-1}(p) &= I^N(l_x(x_N(\cdot))), \qquad (25)\\
\eta_i(p) &= 0.5\left[\xi_i(p)\left(1 + hf_x(\tilde{s}_i(p))\right) + \ hI^N\left(\xi_i(\cdot)f_y(\tilde{s}_i(\cdot))g_x(\cdot, \tilde{s}_i(p))\right)\right] f_y(s_i(p)), \qquad (26)
\end{aligned}$$

and conditions for minimum for $i = 0, \ldots, N-1$

$$\begin{aligned}
&0.5\left(\xi_i(p)(1 + hf_x(\tilde{s}_i(p)) + hI^N(\xi_i(\cdot)f_y(\tilde{s}_i(\cdot))g_x(\cdot, \tilde{s}_i(p)))\right) f_u(s_i(p)) \qquad (27)\\
&\qquad + I^N\left(\eta_i(\cdot)g_u(\cdot, s_i(p))\right) \in N_U(u_i'(p)),\\
&\xi_i(p)f_u(\tilde{s}_i(p)) + I^N\left(\xi_i(\cdot)f_y(\tilde{s}_i(\cdot))g_u(\cdot, \tilde{s}_i(p))\right) \in N_U(u_i''(p)), \qquad (28)
\end{aligned}$$

where "1" means the identity matrix with an appropriate dimension.

Remark 1 Notice, that given the discrete control $(u_i'(\cdot), u_i''(\cdot))$, $i = 0, \ldots, N-1$ and $x_0(\cdot)$, one can successively determine (left to right) the discrete trajectories $x_i(\cdot)$ and $y_i(\cdot)$ for all i. Then one can successively determine (right to left) the corresponding dual variables. Then from the conditions for minimum one can determine "next" controls (or the gradient of the discrete objective function with respect to the control). This makes it possible to apply an iterative (or gradient) method for solving the discrete problem, but in this paper we are interested in the Newton method and this observation is not of principle importance, as it is if a gradient method would be applied.

The above system of generalized equations can be formulated in a compact form using the following discrete-time Hamiltonian H^N defined for the components $z = (x_i, y_i, u_i', u_i'', \xi_i, \eta_i)$, $i = 0, \ldots, N-1$ of any element $z_i \in Z_N$:

$$H^N(z_i(\cdot)) = 0.5I^N\left(\xi_i(\cdot)(f(s_i(\cdot)) + f(\tilde{s}_i(\cdot))) + \eta_i(\cdot)I^N(g(\cdot, s_i(\cdot\cdot))\right),$$

where "$\cdot$" is the dummy argument of the outer summation I^N, and "$\cdot\cdot$" is the dummy argument of the inner summation I^N. The equations (21),(22),(24)–(28) obtained

above for $z \in Z_N$ can be rewritten in the following way: for $i = 0, \dots, N-1$ and $p \in P_N$

$$
\begin{aligned}
0 &= -\frac{x_{i+1}(p) - x_i(p)}{h} + \frac{1}{\alpha_p}\frac{\partial}{\partial \xi(p)} H^N(z_i(\cdot)), && (29)\\
0 &= -y_i(p) + \frac{1}{\alpha_p}\frac{\partial}{\partial \eta(p)} H^N(z_i(\cdot)), \quad x^0(\cdot) \text{ – given}, && (30)\\
0 &= \frac{\xi_i(p) - \xi_{i-1}(p)}{h} + \frac{1}{\alpha_p}\frac{\partial}{\partial x(p)} H^N(z_i(\cdot)), && (31)\\
0 &= -\xi_{N-1}(p) + l_x(x_N(p)) && (32)\\
0 &= -\frac{1}{\alpha_p}\eta_i(p) + \frac{1}{\alpha_p}\frac{\partial}{\partial y(p)} H^N(z_i(\cdot)), && (33)\\
0 &\in -\frac{\partial}{\partial u'(p)} H^N(z_i(\cdot)) + N_U(u_i'(p)), && (34)\\
0 &\in -\frac{\partial}{\partial u''(p)} H^N(z_i(\cdot)) + N_U(u_i''(p)). && (35)
\end{aligned}
$$

It will be convenient to use the notations:

$$
H^N_{x(p)}(z_i(\cdot)) = \frac{1}{\alpha_p}\frac{\partial}{\partial x(p)} H^N(z_i(\cdot)), \quad \dots, \quad H^N_{\eta(p)}(z_i(\cdot)) = \frac{1}{\alpha_p}\frac{\partial}{\partial \eta(p)} H^N(z_i(\cdot)),
$$

while the second subscript of H^N that appears below means true differentiation.

The above system represents the generalized equation (17) in the general formulation of the Newton method. According to Section 3, the generalized Newton method consists of the following: given the current iteration $z^k(\cdot) \in Z_N$ the next iteration z^{k+1} is found as a solution of the following "linearized" inclusion (18):

$$
\begin{aligned}
0 &= -(x_{i+1}(p) - x_i(p))/h + H^N_{\xi(p)}(z_i^k(\cdot)) + H^N_{\xi(p),z(p)}(z_i^k(\cdot))(z_i(p) - z_i^k(p)),\\
0 &= -y_i(p) + H^N_{\eta(p)}(z_i^k(\cdot)) + H^N_{\eta(p),z(p)}(z_i^k(\cdot))(z_i(p) - z_i^k(p)),\\
0 &= (\xi_i(p) - \xi_{i-1}(p))/h + H^N_{x(p)}(z_i^k(\cdot)) + H^N_{x(p),z(p)}(z_i^k(\cdot))(z_i(p) - z_i^k(p)),\\
0 &= -\xi_{N-1}(p) + l_x(x_N^k(p)) + l_{xx}(x_N^k(p))(x_N(p) - x_N^k(p)),\\
0 &= -\eta_i(p) + H^N_{y(p)}(z_i^k(\cdot)) + H^N_{y(p),z(p)}(z_i^k(\cdot))(z_i(p) - z_i^k(p)),\\
0 &\in H^N_{u'(p)}(z_i^k(\cdot)) + H^N_{u'(p),z(p)}(z_i^k(\cdot))(z_i(p) - z_i^k(p)) + N_U(u_i'(p)),\\
0 &\in H^N_{u''(p)}(z_i^k(\cdot)) + H^N_{u''(p),z(p)}(z_i^k(\cdot))(z_i(p) - z_i^k(p)) + N_U(u_i''(p)).
\end{aligned}
$$

As above, the subscripts of H^N denote partial differentiation. As we see in the next section, the last system coincides with the necessary (and sufficient, under the suppositions in the next section) condition for optimality for a linear-quadratic discrete-time optimal control problem, it could be solved by linear-quadratic programming. But the main concern of this paper is the error analysis, therefore we omit the details of the algorithmic realization.

5 Error Analysis

In this section we shall ensure that conditions (B1)–(B4) of the abstract Theorem 3 are satisfied, and shall give (rather non-restrictive) conditions under which the discretization error (the second term in the right-hand side of (19) is of second order with respect to the step size.

First of all we specify the space D_N. In accordance with (29)–(35) we denote

$$d_i = (d_i^x, d_i^y, d_i^\xi, d_i^\eta, d_i^{u'}, d_i^{u''})$$

and define

$$D_N = \{d = (d_0, \dots, d_{N-1}, d_N^\xi) \ :$$
$$d_i \in \mathcal{L}^n_{1,\infty} \times \mathcal{L}^n_{\infty,\infty} \times \mathcal{L}^n_{1,\infty} \times \mathcal{L}^n_{\infty,\infty} \times \mathcal{L}^n_{\infty,\infty} \times \mathcal{L}^n_{\infty,\infty}, i = 0, \dots, N-1 \text{ and } d_N^\xi \in \mathcal{L}^n_\infty\},$$

where the discrete $\mathcal{L}_{1,\infty}$-spaces (we use calligraph $\mathcal{L}$ to distinguish from the the continuous L-spaces) is endowed with the norm

$$\|(d_0^x(\cdot), \dots, d_N^x(\cdot))\|_{1,\infty} = h \sum_{i=1}^{N} \max_{p \in P_N} |d_i^x(p)|,$$

Verification of (B1). To ensure (B1) we suppose

Existence. The problem (1)–(3) has a solution $(\hat{x}, \hat{y}, \hat{u}) = \hat{z} \in Z$.

Verification of (B2). These conditions follow from (A1) since the mapping F is defined by (the single-valued part of) the right-hand side of (10)–(15).

Verification of (B3). We have to estimate what is the residual d in (29)–(35) if we substitute $(z_0(\cdot), \dots, z_N(\cdot)) = \pi_N(\hat{z})$. To do this we need the next notion and supposition.

For a function $v : [0, T] \mapsto \mathbf{R}^r$ we denote by $\omega(v, [0, T]; t, h)$ the modulus of continuity

$$\omega(v, [0, T]; t, h) = \sup\{|v(t') - v(t'')| \ : \ t', \, t'' \in [t - h/2, t + h/2] \cap [0, T]\}.$$

The average modulus of smoothness of v on $[0, T]$ is defined as

$$\tau(v; h) = \int_0^T \omega(v, [0, T]; t, h) \, \mathrm{d}t.$$

If v is defined almost everywhere, then by definition $\tau(v; h)$ is the infimum of the averaged moduli of all extensions of v to $[0, T]$.

In the next supposition we restrict ourselves to the case of an one-dimensional parameter p. The higher dimensional case requires more complicated considerations from the

theory of cubature formulae in which also the geometry of P should be involved in the regularity conditions below.

Regularity. For every $p \in P$ the function $\hat{u}(\cdot, p)$ is Lipschitz continuous uniformly in p. Moreover, $P = [0, \kappa] \subset \mathbf{R}$ and for every $t \in [0, T]$ the function $\hat{u}(t, \cdot)$ is Lipschitz continuous uniformly in t. Accordingly, we suppose that P_N is the mesh in P that splits it into N equal intervals, and that I^N is any composite linear quadrature formula which is exact for linear functions.

We denote

$$\tau^t(\hat{u}_t; h) = \sup_{p \in P} \tau(\hat{u}_t(\cdot, p); h), \quad \tau^p(\hat{u}_p; h) = \sup_{t \in [0,T]} \tau(\hat{u}_p(t, \cdot); h),$$

where u_t and u_p are the derivatives with respect to t and p.

The next lemmas, which follow from results in [13], are used repeatedly in the analysis of the residual.

Lemma 1 *There exists c such that for every $v \in W^{1,\infty}([0, T]; \mathbf{R}^r)$ and for every $h \in (0, T]$*

$$\left| \int_0^h v(s)\, \mathrm{d}s - \frac{h}{2}(v(0) + v(h)) \right| \leq ch \int_0^h \omega(\dot{v}, [0, h]; t, h)\, \mathrm{d}t.$$

Lemma 2 *There exists c such that for every $v \in W^{1,\infty}([0, T]; \mathbf{R}^r)$ and for every $h \in (0, T]$*

$$\left| I^N(v) - \int_P v(q)\, \mathrm{d}q \right| \leq c\tau(\dot{v}; h).$$

The next lemma verifies (B3) and estimates the residual d_N in (19).

Lemma 3 *Let $z = \pi_N(\hat{z})$ be substituted in the right-hand side of (29)–(35). Then there is a constant c (independent of N) such that the image d satisfies*

$$\|d\| \leq ch(h + \tau^t(\hat{u}_t; h) + \tau^p(\hat{u}_p; h)).$$

Proof. Clearly (32) follows from (13), therefore $d_N^\xi = 0$. Moreover, the estimation for d^y follows from (11) and (2). Each of the five remaining residuals has to be estimated

separately. The next two estimates follow from the first order Taylor expansion and Lemma 2 and are repeatedly used in the proof:

$$|\tilde{s}_i(p) - s_{i+1}(p)| \le ch(h + \tau^p(\hat{u}_p; h)), \tag{36}$$

$$|\xi(t_i,p) - \xi(t_{i+1},p)(1 + hf_x(s(t_{i+1},p)) - h\int_P \xi(t_{i+1},q)f_y(s(t_{i+1},q))g_x(q,s(t_{i+1},p))\,\mathrm{d}q$$
$$\le ch(h + \tau^p(\hat{u}_p; h)). \tag{37}$$

We shall present the estimation for d^ξ and skip the rest of the calculations. In the next formulas we shall skip the "hat" in the notation of the optimal solution. Using the explicit version (24) of (32), the estimation (36) and Lemma 2 we have

$$\begin{aligned}
d_i^\xi &= (-\xi_{i-1}(p) + \xi_i(p))/h + \frac{1}{2}\xi_i(p)\left(f_x(s_i(p)) + f_x(\tilde{s}_i(p))(1 + hf_x(s_i(p)))\right.\\
&\quad \left. + \frac{1}{2}I^N\left(\xi_i(\cdot)f_y(\tilde{s}_i(\cdot))g_x(\cdot,\tilde{s}_i(p))\right)(1 + hf_x(s_i(p)))\right) + \frac{1}{2}I^N\left(\eta_i(\cdot)g_x(\cdot,s_i(p))\right)\\
&= (\xi(t_{i+1},p) - \xi(t_i,p))/h\\
&+ \frac{1}{2}\left[\xi(t_{i+1},p)f_x(s(t_i,p)) + \xi(t_{i+1},p)f_x(s(t_{i+1},p)) + h\xi(t_{i+1},p)f_x(s(t_{i+1},p))f_x(s(t_i,p))\right.\\
&\quad + \int_P \xi(t_{i+1},q)f_y(s(t_{i+1},q))g_x(q,s(t_{i+1},p))\,\mathrm{d}q\\
&\quad + h\int_P \xi(t_{i+1},q)f_y(s(t_{i+1},q))g_x(q,s(t_{i+1},p))\,\mathrm{d}q f_x(s(t_i,p))\\
&\quad \left. + \int_P \eta(t_i,q)g_x(q,s(t_i,p))\,\mathrm{d}q + O_N\right]\\
&= \frac{1}{h}\int_{t_i}^{t_{i+1}} \dot{\xi}(t,p)\,\mathrm{d}t - \frac{1}{2}\dot{\xi}(t_{i+1},p) + \frac{1}{2}\left[\xi(t_{i+1},p)(1 + hf_x(s(t_{i+1},p))\right.\\
&\quad \left. + h\int_P \xi(t_{i+1},q)f_y(s(t_{i+1},q))g_x(q,s(t_{i+1},p))\,\mathrm{d}q\right] f_x(s(t_i,p))\\
&\quad + \frac{1}{2}\int_P \eta(t_i,q)g_x(q,s(t_i,p))\,\mathrm{d}q + O_N,
\end{aligned}$$

where O_N is estimated by $ch(h+\tau^p(\hat{u}_p;h))$. Then using (37) and Lemma 2 we conclude that

$$\begin{aligned}
d_i^\xi &= \frac{1}{h}\int_{t_i}^{t_{i+1}} \dot{\xi}(t,p)\,\mathrm{d}t - \frac{1}{2}\dot{\xi}(t_{i+1},p) + \frac{1}{2}\left[\xi(t_i,p)f_x(s(t_i,p)) + \int_P \eta(t_i,q)g_x(q,s(t_i,p))\,\mathrm{d}q\right] + O_N\\
&= \frac{1}{h}\int_{t_i}^{t_{i+1}} \dot{\xi}(t,p)\,\mathrm{d}t - \frac{1}{2}[\dot{\xi}(t_{i+1},p) + \dot{\xi}(t_i,p)] + O_N = O'_N,
\end{aligned}$$

where

$$|O'_N| \le ch(h + \tau^p(\hat{u}_p; h)) + c\int_{t_i}^{t_{i+1}} \omega(\hat{u}_t, [t_i, t_{i+1}]; t, h)\,\mathrm{d}t,$$

according to Lemma 1. Summing in i we obtain the claim of the lemma. Q.E.D.

Verification of (B4). We only sketch this part, since the details become too long. Below we denote $\bar{z}_i(p) = \pi_N(\hat{z})_i(p)$ and abbreviate

$$\bar{\sigma}_i(p) = (\bar{x}_i(p), \bar{y}_i(p), \bar{u}'_i(p), \bar{u}''_i(p)), \quad \sigma_i(p) = (x_i(p), y_i(p), u'_i(p), u''_i(p)).$$

Following the argument in [4, Lemma 11] and [6, Lemma 6.1] one can prove (and this is the main difficulty) that *Coercivity* implies similar property for the discretized system. Namely, for all sufficiently small $h > 0$ the inequality

$$\sum_{p\in P_N}\left[\langle l_{xx}(\bar{x}_N(p))\Delta x_N(p), \Delta x_N(p)\rangle + h\sum_{i=0}^{N-1}\langle H^N_{s(p),s(p)}(\bar{z}_i(\cdot))\Delta s_i(p), \Delta s_i(p)\rangle\right] \tag{38}$$

$$\geq \frac{1}{2}\rho\sum_{p\in P_N}\sum_{i=0}^{N-1}((\Delta u'_i(p))^2 + (\Delta u'_i(p))^2).$$

holds for every discrete controls $\Delta u'_i(p), \Delta u''_i(p) \in U - U$ and the corresponding solution $\Delta x_i(p), \Delta y_i(p)$ of the linear discrete-time system

$$\begin{aligned}(\Delta x_{i+1}(p) - \Delta x_i(p))/h &= H^N_{\xi(p),\sigma(p)}(\bar{z}_i(\cdot))\Delta\sigma_i(p),\\ \Delta y_i(p) &= H^N_{\eta(p),\sigma(p)}(\bar{z}_i(\cdot))\Delta\sigma_i(p)\end{aligned}$$

where $\Delta\sigma_i = (\Delta x_i, \Delta y_i, \Delta u'_i, \Delta u''_i)$.

Then we utilize the approach from [11] and [4]. In order to prove (B4) we have to consider the system in the end of Section 4 with $z^k_i(\cdot)$ replaced by $\bar{z}_i(\cdot)$ and with a disturbance $d \in D_N$ instead of 0. (For further reference we call this system (DLS).) One can verify that the obtained in this way system represents the necessary optimality condition for the following linear-quadratic problem:

$$\min\left\{\sum_{p\in P}\left\langle l_{xx}(\bar{x}_N(p))(x_N(p) - \bar{x}_N(p)) + l_x(\bar{x}_N(p)) - d^\xi_N,\ x_N(p) - \bar{x}_N(p)\right\rangle\right.$$

$$\left. + \sum_{i=0}^{N-1}\sum_{p\in P}\left\langle H^N_{\sigma(p),\sigma(p)}(\bar{z}_i(\cdot))(\sigma_i(p) - \bar{\sigma}_i(p)) + H^N_{\sigma(p)}(\bar{z}_i(\cdot)) - d^\sigma_i(p),\ \sigma_i(p) - \bar{\sigma}_i(p)\right\rangle\right\},$$

subject to

$$\begin{aligned}(x_{i+1}(p) - x_i(p))/h &= H^N_{\xi(p),\sigma(p)}(\bar{z}_i(\cdot))(\sigma_i(p) - \bar{\sigma}_i(p)) + H^N_{\xi(p)}(\bar{z}_i(\cdot)) - d^x_i(p),\\ y_i(p) &= H^N_{\eta(p),\sigma(p)}(\bar{z}_i(\cdot))(\sigma_i(p) - \bar{\sigma}_i(p)) + H^N_{\eta(p)}(\bar{z}_i(\cdot)) - d^y_i(p),\\ &\quad u'_i(p),\ u''_i(p) \in U.\end{aligned}$$

According to [11, Lemma 4] the coercivity condition (38) implies that the primal/dual solution $z \in Z_N$ of the above linear-quadratic problem depends in a Lipschitz way on the disturbance $d \in D_N$. Moreover, under the coercivity this primal/dual solution is exactly the solution of the generalized equation (DLS), which proves (B4).

Getting all together we obtain the following theorem.

Theorem 4 *Suppose that (A1), (A2), (A3'), Existence, Regularity and Coercivity (see Section 2 for the last) are fulfilled. Let $q \in (0,1)$. There exist $\delta > 0$ and N_0 such that if $N > N_0$ and the initial iteration z^0 satisfies $\|z^0 - \pi_N(\hat{z})\| \le \delta$, then the generalized Newton scheme introduced in Section 4 generates a unique sequence $z^k = (z_0^k(\cdot), \dots, z_N^k(\cdot))$ and*

$$\max_{p\in P_N} |x_N^k(p) - \hat{x}(T,p)| + \max_{i=0,\dots,N-1} \max_{p\in P_N} \Big[|x_i^k(p) - \hat{x}(t_i,p)| + |y_i^k(p) - \hat{y}(t_i,p)| + |u_i'^k - \hat{u}(t_i)| + |u_i''^k - \hat{u}(t_{i+1})| \Big]$$
$$\le q^{2^k-1} \|z^0 - \pi_N(\hat{z})\| + Ch(h + \tau^t(\hat{u}_t; h) + \tau^p(\hat{u}_p; h)).$$

We mention that from [13, Sect. 1.3] it follows that $\tau^t(\hat{u}_t; h)$ and $\tau^p(\hat{u}_p; h)$ converge to zero with h if the derivatives $\hat{u}_t$ and $\hat{u}_p$ are Riemann integrable, and $\tau^t(\hat{u}_t; h) + \tau^p(\hat{u}_p; h) = O(h)$ if the two derivatives are of bounded variation. In the latter case the estimate in Theorem 4 is $O(h^2)$. This is the usual case that arises in the applications (excluding pathological examples). On the other hand, more regularity of the optimal control than existence (almost everywhere) of a derivative with bounded variation is a too strong supposition that typically fails in the case of constrained control. Therefore higher than second order accuracy cannot be achieved in general by Runge-Kutta discretization schemes.

References

[1] W. Alt. Discretization and mesh-independence of Newton's method for generalized equations. In *Mathematical programming with data perturbations*, 1–30, Lecture Notes in Pure and Appl. Math., 195, Dekker, New York, 1998.

[2] M. Brokate. Pontryagin's principle for control problems in age-dependent population dynamics. *J. Math. Biology.*, 23:75–101, 1985.

[3] A.L. Dontchev. Lipschitzian stability of Newton's method for variational inclusions. In *System modelling and optimization*, (Cambridge, 1999), 119–147, Kluwer Acad. Publ., Boston, 2000,

[4] A.L. Dontchev, W.W. Hager. Lipschitz stability in nonlinear control and optimization. *SIAM J. Control and Optim.*, **31**:569–603, 1993.

[5] A.L. Dontchev, W.W. Hager, and V.M. Veliov. Uniform convergence and mesh independence of the Newton method in optimal control. *SIAM J. Control and Optim.*, **39**(3):961–980, 2000.

[6] A.L. Dontchev, W.W. Hager, and V.M. Veliov. Second-order Runge-Kutta approximations in control constrained optimal control, *SIAM J. Numerical Anal.*, **38**(1):202–226, 2000.

[7] H.O. Fattorini. A unified theory of necessary conditions for nonlinear and nonconvex control systems. *Applied Mathematics and Optimizations*, 15, 141–185, 1987.

[8] H.O. Fattorini and H. Frankowska. Necessary conditions for infinite-dimensional control problems. *Math. Control Signals Systems*, 4(1), 41–67, 1991.

[9] G. Feichtinger, G. Tragler, and V.M. Veliov. Optimality conditions for age-structured control systems. *To appear in J. Math. Anal. Appl.*

[10] G. Feichtinger, Ts. Tsachev, and V.M. Veliov. Age and duration structured control model for optimal prevention and treatment of HIV. (Submitted.)

[11] W.W. Hager. Multiplier methods for nonlinear optimal control. *SIAM J. Numer. Anal.*, **27**:1061–1080, 1990.

[12] R. T. Rockafellar, R. J.-B. Wets, *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften 317, Springer-Verlag, Berlin, 1998.

[13] B. Sendov and V. Popov. *The Averaged Moduli of Smoothness.* John Wiley, N.Y., 1988.

[14] G.F. Webb. *Theory of Nonlinear Age-Dependent Population Dynamics.* Marcel Dekker, New York, 1985.