International Mathematical Forum, 2, 2007, no. 30, 1477 - 1498

Classical and Relaxed Optimization Methods for Optimal Control Problems

I. Chryssoverghi, J. Coletsos and B. Kokkinis

Department of Mathematics, School of Appl. Mathematics and Physics
National Technical University of Athens
Zografou Campus, 15780 Athens, Greece
ichris@central.ntua.gr

Abstract. We consider an optimal control problem for systems governed by nonlinear ordinary differential equations, with control and state constraints, including pointwise state constraints. The problem is formulated in the classical and in the relaxed form. Various necessary/sufficient conditions for optimality are first given for both problems. For the numerical solution of these problems, we then propose a penalized gradient projection method generating classical controls, and a penalized conditional descent method generating relaxed controls. Using also relaxation theory, we study the behavior in the limit of sequences constructed by these methods. Finally, numerical examples are given.

Keywords: Optimal control, nonlinear systems, state constraints, gradient projection method, conditional descent method, penalty method, relaxed controls.

Mathematics Subject Classification: 49M

1 Introduction

We consider an optimal control problem for systems governed by nonlinear ordinary differential equations, with control and state constraints, including pointwise state constraints. The problem is formulated in the classical form, and also in the relaxed form using Young measures. Various necessary/sufficient conditions for optimality are first given for both problems. For the numerical solution of these problems, we then propose a penalized gradient projection method generating classical controls, and a penalized conditional descent method generating relaxed controls. Under appropriate assumptions, we prove that relaxed (resp. strong classical) limits of subsequences (resp. sequences) constructed by the classical method are admissible and weakly extremal relaxed (resp. classical) for the relaxed (resp. classical) problem, and that relaxed limits of subsequences of controls constructed by the relaxed method are admissible and strongly extremal for the relaxed problem. Finally, several numerical examples are given. For classical and relaxed optimization and approximation methods applied to optimal control problems, see e.g. [2-9], [11-12], [14], and the references therein.

2 Classical and relaxed optimal control problems

Consider the following optimal control problem. The state equation is given by

$$y'(t) = f(t, y(t), w(t)),\ t \in I := [0,T],\quad y(0) = y^0,$$

where $y(t) \in IR^d$, the constraints on the control w are $w(t) \in U$, for $t \in I$, where U is a compact subset of $IR^{d'}$, the constraints on the state $y := y_w$ are

$$G_1(w) = \bar{g}_1(y(T)) + \int_0^T g_1(t, y(t), w(t))dt = 0,$$

$$G_2(w) = \bar{g}_2(y(T)) + \int_0^T g_2(t, y(t), w(t))dt \le 0,$$

$$G_3(w)(s) = g_3(s, y(s)) \le 0, \quad \text{for } s \in I,$$

where the vector functions $\bar{g}_l, g_l$ take values in IR^{m_l}, $l = 1,2$, and g_3 in IR^{m_3}, and the cost functional is

$$G_0(w) = \bar{g}_0(y(T)) + \int_0^T g_0(t, y(t), w(t))dt.$$

Defining the set of *classical controls*

$$W = \{w : I \to U \mid w \text{ measurable}\} \subset L^2(I, IR^{d'}),$$

the classical optimal control problem is to minimize $G_0(w)$ subject to $w \in W$ and to the above state constraints.

It is well known that, even if the set U is convex, the classical problem may have no solutions. The existence of such a solution is usually proved under strong, often unrealistic for nonlinear systems, convexity assumptions (such as the Cesari property). Reformulated in the so-called relaxed form, the problem is convexified in some sense and has a solution in a larger space under weaker assumptions.

Next, we define the set of *relaxed controls* (Young measures; for the relevant theory, see [13], [10]) by

$$R = \{r : I \to M_1(U) \mid r \text{ weakly measurable}\} \subset L^\infty_w(I, M(U)) \equiv L^1(I, C(U))^*,$$

where $M(U)$ (resp. $M_1(U)$) is the set of Radon (resp. probability) measures on U. The set W (resp. R) is endowed with the relative strong (resp. weak star) topology, and R is convex, metrizable and compact. If each classical control $w(\cdot)$ is identified with its associated Dirac relaxed control $r(\cdot) := \delta_{w(\cdot)}$, then W may be *also* considered as a subset of R, and W is thus dense in R. For a given $\phi \in L^1(I; C(U; IR^n))$ (or equivalently $\phi \in B(I, U; IR^n)$, where B is the set of Caratheodory functions in the sense of Warga [13]) and $r \in L^\infty_w(I, M(U))$ (in particular, for $r \in R$), we shall use for simplicity the notation

$$\phi(x, r(t)) := \int_U \phi(t,u) r(t)(du),$$

and $\phi(t, r(t))$ is thus linear (under convex combinations, for $r \in R$) in r. A sequence (r_k) converges to $r \in R$ in R if and only if

$$\lim_{k\to\infty}\int_I \phi(t, r_k(t))dt = \int_I \phi(t, r(t))dt,$$

for every $\phi \in L^1(I; C(U; IR^n))$, or $\phi \in B(I, U; IR^n)$, or $\phi \in C(I \times U; IR^n)$.

The relaxed optimal control problem is then defined by replacing w by r (with the above notation) and W by R in the classical problem.

We define the norms $\|x\| = (\sum_{i=1}^{p} x_i^2)^{1/2}$, $\|x\|_1 = \sum_{i=1}^{p} |x_i|$, $\|x\|_\infty = \max_{i=1,\ldots,p} |x_i|$, in IR^p, and denote by $\|\cdot\|_{L^2}$, $\|\cdot\|_{L^1}$, $\|\cdot\|_{L^\infty}$, $\|\cdot\|_{L^\infty}$ the corresponding usual norms in $L^2(I)^p$, $L^1(I)^p$, $L^\infty(I)^p$, $C(I)^p$, respectively. We denote by $M(I) \equiv C(I)^*$ the set of finite regular measures on I, and by $\|\cdot\|_*$ the norm in $M(I)$ defined by $\|\mu\|_* = \int_I |\mu|(dt)$ (with $|\mu| = \mu$ if μ is positive). The order relations between vectors, functions or vector functions, are defined componentwise and/or pointwise.

We suppose that the function f is defined on $I \times IR^d \times U$, measurable for y, u fixed, continuous for t fixed, and satisfies

$$\|f(t, y, u)\| \le \psi(t)(1 + \|y\|), \quad \text{for every } (t, y, u) \in I \times IR^d \times U, \quad \text{with } \psi \in L^1(I),$$

$$\|f(t, y_1, u) - f(t, y_2, u)\| \le L\|y_1 - y_2\|, \quad \text{for every } (t, y_1, y_2, u) \in I \times IR^{2d} \times U.$$

The following result is standard (see [13]).

Theorem 2.1 For every relaxed (or classical, as $W \subset R$) control $r \in R$, the state equation has a unique absolutely continuous solution $y := y_r$. Moreover, there exists a constant b such that $\|y_r\|_\infty \le b$ for every $r \in R$.

Let B denote the closed ball in IR^d with center 0 and radius b (see Theorem 2.1). We suppose now in addition that the functions g_l, $l = 0,1,2$, are defined on $I \times B \times U$, measurable for fixed y, u, continuous for fixed t, and satisfy

$$\|g_l(t, y, u)\| \le \zeta_l(t), \quad \text{for every } (t, y, u) \in I \times B \times U,$$

with $\zeta_l \in L^1(I)$, the function g_3 is continuous on $I \times B$, and that the functions $\bar{g}_l$ are continuous on B. The results of the following theorem are proved in [13].

Theorem 2.2 The mappings $y \to y_w$, from L^2 to $C(I)^d$, $y \to y_r$, from R to $C(I)^d$, and $G_l : W \text{ or } R \to \mathbf{R}^{m_l}$, $l = 0,1,2$, $G_3 : W \text{ or } R \to C(I)^{m_3}$, are continuous. If the relaxed problem is feasible, then it has a solution.

Note that in the classical problem we have $y'(t) \in f(t, y(t), U)$ (velocity set), while in the relaxed problem $y'(t) \in \mathrm{co}(f(t, y(t), U))$. Since $W \subset R$, we have in general

$$c_R := \min_{\text{constraints on } r} G_0(r) \le \inf_{\text{constraints on } w} G_0(w) := c_W,$$

where the equality holds, in particular, if there are no state constraints, as W is dense in R. Since usually approximation methods slightly violate the state constraints, approximating an optimal relaxed control by a relaxed or a classical control, hence the relaxed optimal cost c_R, is not a drawback in practice (see [13], p. 248). Note also that approximating sequences of classical controls may converge to relaxed ones.

In order to state the necessary conditions for optimality, we suppose in addition that the functions f, g_l, f_y, f_u, g_{ly}, g_{lu}, $l = 0,1,2$, are defined on $I \times B' \times U'$, where B' (resp. U') is a open set containing B (resp. U), measurable on I for fixed $(y,u) \in B \times U$, continuous on $B \times U$ for fixed $t \in I$, and satisfy

$$\|f_y(t,y,u)\| \le \xi(t), \quad \|f_u(t,y,u)\| \le \eta(t),$$

$$\|g_{ly}(t,y,u)\| \le \zeta_{l1}(t), \quad \|g_{lu}(t,y,u)\| \le \zeta_{l2}(t),$$

for every $(t,y,u) \in I \times B \times U$, with $\xi, \eta, \zeta_{l1}, \zeta_{l2} \in L^1(I)$, and that the functions $\overline{g}_{ly}$ are defined on B' and continuous on B.

The two following theorems can be proved by using the techniques of [13].

Theorem 2.3 (i) If U is convex, the directional derivative of the mapping $G_l : W \to \mathbf{R}^{m_l}$, for $l = 0,1,2$, is given by

$$DG_l(w, w'-w) = \lim_{\alpha \to 0^+} \frac{G_l(w + \alpha(w'-w)) - G_l(w)}{\alpha}$$

$$= \int_0^T [z_l(t) f_u(t, y(t), w(t)) + g_{lu}(t, y(t), w(t))][w'(t) - w(t)]dt, \quad \text{for } w, w' \in W,$$

where $y := y_w$, and the adjoint state $z_l := z_{lw}$, a row vector function ($l = 0$), or a matrix function ($l = 1,2$), is the solution of the classical linear adjoint equation

$$z_l'(t) = -z_l(t) f_y(t, y(t), w(t)) - g_{ly}(t, y(t), w(t)), \quad t \in I,$$

$$z_l(T) = \overline{g}_{ly}(y(T)), \quad \text{with } y := y_w,$$

the controls being regarded here as classical. The directional derivative of $G_3 : W \to C(I)^{m_3}$ is given by the matrix function

$$DG_3(w, w'-w)(s)$$

$$= g_{3y}(s, y(s)) Z(s)^{-1} \int_0^s Z(t) f_u(t, y(t), w(t))[w'(t) - w(t)]dt, \quad s \in I,$$

where the matrix function $Z := Z_w$ satisfies the fundamental matrix equation

$$Z'(t) = -Z(t) f_y(t, y(t), w(t)), \quad t \in I,$$

$$Z(T) = E \quad (E \text{ identity matrix}).$$

(ii) The directional derivative of the mapping $G_l : R \to IR^{m_l}$, for $l = 0,1,2$, is given by

$$DG_l(r, r'-r) := \lim_{\alpha \to 0^+} \frac{G_l(r + \alpha(r'-r)) - G_l(r)}{\alpha}$$

$$= \int_I [z_l(t) f(t, y(t), r'(t) - r(t)) + g_l(t, y(t), r'(t) - r(t))]dt, \quad \text{for } r, r' \in R,$$

where $y = y_r$, and the relaxed adjoint $z_l := z_{lr}$ is the solution of the relaxed linear adjoint equation

$$z_l'(t) = -z_l(t) f_y(t, y(t), r(t)) - g_{ly}(t, y(t), r(t)), \quad t \in I,$$

$$z_l(T) = \overline{g}_{ly}(y(T)), \quad \text{with } y := y_r.$$

The directional derivative of $G_3 : R \to C(I)^{m_3}$ is given by

$$DG_3(r,r'-r)(s) = g_{3y}(s,y(s))Z(s)^{-1}\int_0^s Z(t)\,f_u(t,y(t),r'(t)-r(t))dt\ , \quad s\in I\ ,$$

where $Z := Z_r$ is defined as in (i), but with w replaced by r.

(iii) The following mappings are continuous

$z \mapsto z_w$, from W to $C(I)^d$, $\quad z \mapsto z_r$, from R to $C(I)^d$,

$(w,w') \mapsto DG_l(w,w'-w)$, $l = 0,1,2,3$, from $W\times W$ to $IR^{m_l}, l = 0,1,2$, $C(I)^{m_3}$,

$(r,r') \mapsto DG_l(r,r'-r)$), $l = 0,1,2,3$, from $R\times R$ to IR^{m_l}, $l = 0,1,2$, $C(I)^{m_3}$.

In the notations of DG, it is understood, depending on the arguments, w or r, that the directional derivative is taken in the corresponding space, W or R, on which G is defined. The following theorem gives various necessary conditions for optimality (the weak relaxed minimum principle is proved similarly to [7]).

Theorem 2.4 (i) We suppose that U is convex. If $w \in W$ is optimal for the classical problem, then w is weakly extremal classical, i.e. there exist multipliers

$\lambda_0 \in IR$, $\lambda_1 \in IR^{m_1}$, $\lambda_2 \in IR^{m_2}$, $\lambda_3 \in [C(I)^{m_3}]^* \equiv M(I)^{m_3}$,

with $\lambda_0 \ge 0$, $\lambda_2 \ge 0$, $\lambda_3 \ge 0$, $\sum_{l=0}^{2}\|\lambda_l\| + \|\lambda_3\|_* = 1$, where $\|\lambda_3\|_* = \sum_{j=1}^{m_3}\|\lambda_{j3}\|_{M_1}$,

such that

$$\sum_{l=0}^{2}\lambda_l DG_l(w,w'-w) + \int_0^T \lambda_3(ds)DG_3(w,w'-w)(s)$$
$$= \sum_{l=0}^{2}\lambda_l \int_0^T [z_l(t)\,f_u(t,y(t),w(t)) + g_{lu}(t,y(t),w(t))][w'(t)-w(t)]dt$$
$$+\int_0^T \lambda_3(ds)\ g_{3y}(s,y(s))Z(s)^{-1}\int_0^s Z(t)\,f_u(t,y(t),w(t))[w'(t)-w(t)]dt$$
$$= \int_0^T \{\sum_{l=0}^{2}\lambda_l[z_l(t)\,f_u(t,y(t),w(t)) + g_{lu}(t,y(t),w(t))]$$
$$+\left(\int_t^T \lambda_3(ds)g_{3y}(s,y(s))Z(s)^{-1}\right)Z(t)\,f_u(t,y(t),w(t))\}[w'(t)-w(t)]dt \ge 0,$$

for every $w' \in W$,

and

$$\lambda_2 G_2(w) = 0\ ,\ \int_0^T \lambda_3(ds)G_3(w)(s) = 0 \quad \text{(classical transversality conditions).}$$

The above inequality condition is equivalent to the pointwise weak classical minimum principle

$$\{\sum_{l=0}^{2}(\lambda_l[z_l(t)\,f_u(t,y(t),w(t)) + g_{lu}(t,y(t),w(t))]$$
$$+[\int_t^T \lambda_3(ds)g_{3y}(s,y(s))Z(s)^{-1}]Z(t)\,f_u(t,y(t),w(t))\}\ w(t)$$
$$= \min_{u\in U}\{\{\sum_{l=0}^{2}\lambda_l[z_l(t)\,f_u(t,y(t),w(t)) + g_{lu}(t,y(t),w(t))]$$
$$+[\int_t^T \lambda_3(ds)g_{3y}(s,y(s))Z(s)^{-1}]Z(t)\,f_u(t,y(t),w(t))\}\ u\}, \quad \text{for a.a. } t\in I\ .$$

(ii) If $r \in R$ is optimal for either the relaxed or the classical problem, then r is strongly extremal relaxed, i.e. there exist multipliers as in (i), such that

$$\sum_{l=0}^{2} \lambda_l DG_l(r, r'-r) + \int_0^T \lambda_3(ds) DG_3(r, r'-r)(s)$$
$$= \sum_{l=0}^{2} \lambda_l \int_0^T [z_l(t) f(t, y(t), r'(t)-r(t)) + g_l(t, y(t), r'(t)-r(t))] dt$$
$$+ \int_0^T \lambda_3(ds) g_{3y}(s, y(s)) Z(s)^{-1} \int_0^s Z(t) f(t, y(t), r'(t)-r(t)) dt$$
$$= \int_0^T \{ \sum_{l=0}^{2} \lambda_l [z_l(t) f(t, y(t), r'(t)-r(t)) + g_l(t, y(t), r'(t)-r(t))]$$
$$+ [\int_t^T \lambda_3(ds) g_{3y}(s, y(s)) Z(s)^{-1}] Z(t) f(t, y(t), r'(t)-r(t)) \} dt \geq 0,$$

for every $r' \in R$,

and

$$\lambda_2 G_2(r) = 0, \quad \int_0^T \lambda_3(ds) G_3(r)(s) = 0 \quad \text{(relaxed transversality conditions).}$$

The above inequality condition is equivalent to the pointwise strong relaxed minimum principle

$$\sum_{l=0}^{2} \lambda_l [z_l(t) f(t, y(t), r(t)) + g_l(t, y(t), r(t))]$$
$$+ [\int_t^T \lambda_3(ds) g_{3y}(s, y(s)) Z(s)^{-1}] Z(t) f(t, y(t), r(t))$$
$$= \min_{u \in U} \{ \sum_{l=0}^{2} \lambda_l [z_l(t) f(t, y(t), u) + g_l(t, y(t), u)]$$
$$+ [\int_t^T \lambda_3(ds) g_{3y}(s, y(s)) Z(s)^{-1}] Z(t) f(t, y(t), u) \}, \quad \text{for a.a. } t \in I.$$

If in addition U is convex, then this minimum principle implies the pointwise weak relaxed minimum principle

$$\{ \sum_{l=0}^{2} \lambda_l [z_l(t) f_u(t, y(t), r(t)) + g_{lu}(t, y(t), r(t))]$$
$$+ [\int_t^T \lambda_3(ds) g_{3y}(s, y(s)) Z(s)^{-1}] Z(t) f_u(t, y(t), r(t)) \} \, r(t)$$
$$= \min_{\phi} \{ \{ \sum_{l=0}^{2} \lambda_l [z_l(t) f_u(t, y(t), r(t)) + g_{lu}(t, y(t), r(t))]$$
$$+ [\int_t^T \lambda_3(ds) g_{3y}(s, y(s)) Z(s)^{-1}] Z(t) f_u(t, y(t), r(t)) \} \, \phi(t, r(t)) \}, \quad \text{for a.a. } t \in I,$$

where the minimum is taken over the set $B(I, U; U)$ of Caratheodory functions (see [13]) $\phi : I \times U \to U$, which in turn implies the global weak relaxed condition

$$\int_0^T \{ \sum_{l=0}^{2} \lambda_l [z_l(t) f_u(t, y(t), r(t)) + g_{lu}(t, y(t), r(t))]$$
$$+ [\int_t^T \lambda_3(ds) g_{3y}(s, y(s)) Z(s)^{-1}] Z(t) f_u(t, y(t), r(t)) \} [\phi(t, r(t)) - r(t)] dt \geq 0,$$

for every $\phi \in B(I, U; U)$.

A control r satisfying this condition and the above relaxed transversality conditions is called weakly extremal relaxed.

The following theorem gives sufficient conditions for optimality.

Theorem 2.5 With the derivatives in u omitted (resp. included) in our last assumptions, we suppose in addition that the data are such that G_0, G_2, G_3 are convex and that G_1 is affine. If $r \in R$ (resp. $w \in W$, with U convex) is admissible and strongly extremal relaxed (resp. weakly extremal classical) for the relaxed (resp. classical) problem, with $\lambda_0 > 0$, then r is optimal for this problem.

Proof. (Relaxed case, the classical case is similar) The assumptions imply that the functional

$$G(r) := \sum_{l=0}^{2} \lambda_l G_l(r) + \int_I \lambda_3(ds) G_3(r)(s)$$

is convex. The necessary inequality condition of Theorem 2.4 is then satisfied if and only if r minimizes G on R. Suppose now that r does not minimize G_0, in which case there exists $r' \in R$ satisfying the state constraints and such that $G_0(r') < G_0(r)$. Using the state constraints and the relaxed transversality conditions, we obtain

$$G(r') := \sum_{l=0}^{2} \lambda_l G_l(r') + \int_I \lambda_3(ds) G_3(r')(s)$$

$$\leq \lambda_0 G_0(r') < \lambda_0 G_0(r) = \sum_{l=0}^{2} \lambda_l G_l(r) + \int_I \lambda_3(ds) G_3(r)(s) = G(r),$$

i.e. r does not minimize G, a contradiction.

3 Classical and relaxed optimization methods

Let (M_l^m), $l = 1,2,3$, be nonnegative increasing sequences such that $M_l^m \to \infty$ as $m \to \infty$, $\gamma > 0$, $b, c \in (0,1)$, and (β_k), (ζ_k) positive sequences, with (β^m) decreasing and converging to zero, and $\zeta_k \leq 1$.

Define first the penalized functionals on W

$$G^m(w) := G_0(w) + \frac{1}{2}\{M_1^m \sum_{j=1}^{m_1} \left|G_{1j}(w)\right|^2 + M_2^m \sum_{j=1}^{m_2} [\max(0, G_{2j}(w))]^2$$

$$+ M_3^m \sum_{j=1}^{m_3} \int_I [\max(0, G_{3j}(w)(s))]^2 ds\}.$$

It can be easily shown that the directional derivative of G^m is given by

$$DG^m(w, w'-w) = DG_0(w, w'-w)$$

$$+ M_1^m \sum_{j=1}^{m_1} G_{1j}(w) DG_{1j}(w, w'-w) + M_2^m \sum_{j=1}^{m_2} \max(0, G_{2j}(w)) DG_{2j}(w, w'-w)$$

$$+ M_3^m \sum_{j=1}^{m_3} \int_I \max(0, G_{3j}(w)(s)) DG_{3j}^n(w, w'-w)(s) ds.$$

The classical penalized gradient projection method is described by the following Algorithm, where U is assumed to be convex.

Algorithm 1

Step 1. Set $k := 0$, $m := 1$, and choose an initial control $w_0^1 \in W$.

Step 2. Find $v_k^m \in W$ such that

$$e_k := DG^m(w_k^m, v_k^m - w_k^m) + (\gamma/2)\left\|v_k^m - w_k^m\right\|_{L^2}^2$$
$$= \min_{v' \in W}[DG^m(w_k^m, v' - w_k^m) + (\gamma/2)\left\|v' - w_k^m\right\|_{L^2}^2],$$

and set $d_k := DG^m(w_k^m, v_k^m - w_k^m)$.

Step 3. If $|d_k| \le \beta^m$, set $w^m := w_k^m$, $v^m := v_k^m$, $e^m := e_k$, $d^m := d_k$, $m := m+1$, and then go to Step 2.

Step 4. (Modified Armijo step search) Find the lowest integer value s (positive or not), say $\bar{s}$, such that $\alpha = c^s \zeta_k \in (0,1]$ and α satisfies the inequality

$$G^m(w_k^m + \alpha_k(v_k^m - w_k^m)) - G^m(w_k^m) \le \alpha_k b d_k,$$

and then set $\alpha := c^{\bar{s}} \zeta_k$.

Step 5. Set $w_{k+1}^m := w_k^m + \alpha_k(v_k^m - w_k^m)$, $k := k+1$, and go to Step 2.

One can see by "completing the square" that Step 2 amounts to finding the projection v_k^l of the function

$$u_k^m(t) := w_k^m(t) - (1/\gamma)\{\sum_{l=0}^{m2} \lambda_l^m [z_{lk}^m(t) f_u(t, y_k^m(t), w_k^m(t)) + g_{lu}(t, y_k^m(t), w_k^m(t))]$$
$$+\left(\int_t^T \lambda_3^m(ds) g_{3y}(s, y_k^m(s)) Z(s)^{-1}\right) Z(t) f_u(t, y_k^m(t), w_k^m(t))\},$$

onto W, which in turn reduces to finding the corresponding pointwise projection onto U for a.a. $t \in I$. On the other hand, by the definition of the directional derivative and since $b, c \in (0,1)$, clearly the Armijo step α_k in Step 4 can be found for every k. The parameter γ is chosen experimentally to yield a good rate of convergence.

A (strongly or weakly, classical or relaxed) extremal control is called *abnormal* if there exist multipliers as in the corresponding optimality conditions, with $\lambda_0 = 0$. A control is admissible and abnormal extremal in rather exceptional situations (see [13]).

With w^m as defined in Step 3, define the sequences of multipliers

$$\lambda_1^m := M_1^m G_1(w^m), \quad \lambda_2^m := M_2^m \mathbf{max}(0, G_2(w^m)),$$
$$\lambda_3^m(s) := M_3^m \mathbf{max}(0, G_3(w^m)(s)) = M_3^m \mathbf{max}(0, g_3(s, y^m(s))),$$

where **max** denotes a vector of maximum values.

Theorem 3.1 We suppose here that U is convex.

(i) In the presence of state constraints, if the whole sequence $(w_k^{m(k)})$ generated by Algorithm 1 converges to some $w \in W$ in L^2 strongly and the sequences (λ_l^m), $l = 1,2,3$, (λ_3^m) in $L^1(I)^{m_3}$, are bounded, then w is admissible and weakly extremal classical for the classical problem. In the absence of state constraints, if a subsequence

$(w_k)_{k\in K}$ (no index m) converges to some $w\in W$ in L^2 strongly, then w is weakly extremal classical for the classical problem.

(ii) In the presence of state constraints, if a subsequence $(w^m)_{m\in M}$ (regarded as a sequence in R) of the sequence generated by Algorithm 1 in Step 3 converges to some r in R and the sequences (λ_l^{nm}), $l=1,2,3$, (λ_3^{nm}) in $L^1(I)^{m_3}$, are bounded, then r is admissible and weakly extremal relaxed for the relaxed problem. In the absence of state constraints, if a subsequence $(w_k)_{k\in K}$ (no index m) converges to some r in R, then r is weakly extremal relaxed for the relaxed problem.

(iii) In any of the convergence cases (i) or (ii) with state constraints, suppose that the classical, or relaxed, problem has no admissible, abnormal extremal, controls. If the limit control is admissible, then the sequences of multipliers are bounded, and this control is extremal as above.

Proof. (State constraints present) We shall first show that $m\to\infty$ in the Algorithm. Suppose on the contrary that m remains constant after a finite number of iterations in k, and so we *drop* here the index m. Consider case (i). Let us show that then $d_k\to 0$. Since $w_k\to w$ in L^2 strongly, using also Proposition 2.1 in [1], we can deduce that $u_k\to u$ in L^2 strongly, where

$$u(t):=w(t)-(1/\gamma)\{\sum_{l=0}^{2}\lambda_l[z_l(t)f_u(t,y(t),w(t))+g_{lu}(t,y(t),w(t))]$$
$$+\left(\int_t^T\lambda_3(ds)g_{3y}(s,y(s))Z(s)^{-1}\right)Z(t)f_u(t,y(t),w(t))\}.$$

Since v_k is the projection of u_k onto W, we have also $v_k\to v$ in L^2 strongly, where v is the projection of u onto W. Clearly, by Step 2, $d_k\le e_k\le 0$ for every k, hence, by Theorem 2.3

$$e:=\lim_{k\to\infty}e_k=DG(w,v-w)+(\gamma/2)\|v-w\|^2\le 0,$$
$$d:=\lim_{k\to\infty}d_k=DG(w,v-w)\le\lim_{k\to\infty}e_k=e\le 0.$$

Suppose that $d<0$. The function $\Phi(\alpha):=G(w+\alpha(v-w))$ is continuous on $[0,1]$, and since the directional derivative $DG(w,v-w)$ is linear in $v-w$, Φ is differentiable on $(0,1)$ and has derivative $\Phi'(\alpha)=DG(w+\alpha(v-w),v-w)$. By the mean value theorem, for each $\alpha\in(0,1]$ there exists $\alpha'(\alpha)\in(0,\alpha)$ such that

$$G(w_k+\alpha(v_k-w_k))-G(w_k)=\alpha DG(w_k+\alpha'(\alpha)(v_k-w_k),v_k-w_k).$$

Therefore, by Theorem 2.3

$$G(w_k+\alpha(v_k-w_k))-G(w_k)=\alpha(d+\varepsilon_{k\alpha}),\quad \text{for } \alpha\in[0,1],$$

where $\varepsilon_{k\alpha}\to 0$ as $k\to\infty$, $k\in K$, and $\alpha\to 0^+$. Now, we have $d_k=d+\eta_k$, where $\eta_k\to 0$ as $k\to\infty$, and since $b\in(0,1)$

$$d+\varepsilon_{k\alpha}\le b(d+\eta_k)=bd_k,\quad \text{for } \alpha\in[0,\bar\alpha],\ k\ge\bar k,$$

for some $\bar\alpha>0$ and $\bar k$. Hence

$$G(w_k+\alpha(v_k-w_k))-G(w_k)\le\alpha bd_k,\quad \text{for } \alpha\in[0,\bar\alpha],\ k\ge\bar k.$$

It follows from the choice of the Armijo step α_k in Step 4 that $\alpha_k\ge c\bar\alpha$, for $k\ge\bar k$. Hence

$$G(w_{k+1})-G(w_k)=G(w_k+\alpha_k(v_k-w_k))-G(w_k)\le \alpha_k bd_k \le c\bar{\alpha}bd_k \le c\alpha bd/2,$$

for $k\ge \bar{k}$. This contradicts the fact that $G(w_k)\to G(w)$, by Theorem 2.2. Therefore, we must have $d=e=0$, $d_k\to 0$, and $e_k\to 0$. But Step 3 then implies that $m\to\infty$, which is a contradiction. Therefore, $m\to\infty$ in Algorithm 1.

Let us show now that $d_k\to 0$ in case (ii). Since R is compact, let $(w_k)_{k\in K}$, $(v_k)_{k\in K}$ be subsequences, regarded as sequences in R, of the sequences generated by Algorithm 1 that converge to some $\tilde{r}\in R$, $r'\in R$, respectively. Then, by the continuity of the relaxed state and adjoint operators (Theorems 2.2-3), using Proposition 2.1 in [1], and since $d_k\le e_k\le 0$, we have

$$e:=\lim_{k\to\infty,\,k\in K} e_k=\lim_{k\to\infty,\,k\in K}[DG(w_k,v_k-w_k)+(\gamma/2)\|v_k-w_k\|^2]$$

$$=d+(\gamma/2)\int_\Omega [r'(x)-\tilde{r}(x)]^2\,dt\le 0,$$

$$d=\lim_{k\to\infty,\,k\in K} d_k=\lim_{k\to\infty,\,k\in K} DG(w_k,v_k-w_k)\le \lim_{k\to\infty,\,k\in K} e_k=e\le 0,$$

with

$$d:=\int_0^T\{\sum_{l=0}^{2}\lambda_l[z_l(t)f_u(t,y(t),\tilde{r}(t))+g_{lu}(t,y(t),\tilde{r}(t))]$$

$$+\left(\int_t^T\lambda_3(ds)g_{3y}(s,y(s))Z(s)^{-1}\right)Z(t)f_u(t,y(t),\tilde{r}(t))\}[r'(t)-\tilde{r}(t)]dt,$$

where $\lambda_0:=1$, $y:=y_{\tilde{r}}$, $z:=z_{\tilde{r}}$. Note that d_k is, but d *is not*, a directional derivative in this case. Suppose that $d<0$. Since we have also, for $\alpha'(\alpha)\in(0,\alpha)$

$$\lim_{k\to\infty,\,k\in K,\,\alpha\to 0^+} DG(w_k+\alpha'(\alpha)(v_k-w_k),v_k-w_k)=d,$$

we get as above

$$G(w_{k+1})-G(w_k)\le c\alpha bd/2,\quad \text{for } k\ge\bar{k},\ k\in K,$$

for some $\bar{k}$. Since the whole sequence $(G(w_k))$ is non-increasing by Steps 4 and 5, it follows that $G(w_k)\to-\infty$ as $k\to\infty$, $k\in K$, which leads to a contradiction, as above. Therefore, $d=e=0$, $d_k\to 0$, $e_k\to 0$, $k\in K$, and this holds also for the whole sequences by the uniqueness of the limit 0. We conclude as above that $l\to\infty$ in Algorithm 1.

(i) Suppose that the sequences (λ_l^m) are bounded and, up to subsequences, that $\lambda_l^m\to\lambda_l$, $l=1,2$, and $\lambda_3^m\to\lambda_3$ in $M(I)^{m_3}$ weak star. By Theorem 2.2, since $w^m\to w$ in L^2 strongly, we have

$$0=\lim_{m\to\infty}\frac{\lambda_1^m}{M_1^m}=\lim_{m\to\infty}G_1(w^m)=G_1(w),$$

$$0=\lim_{m\to\infty}\frac{\lambda_2^m}{M_2^m}=\lim_{m\to\infty}[\mathbf{max}(0,G_1(w^m))]=\mathbf{max}(0,G_2(w)),$$

$$0=\lim_{m\to\infty}\frac{\|\lambda_3^m\|_{L^1}}{M_3^m}=\lim_{m\to\infty}\frac{1}{M_3^m}\int_I\lambda_3^m(s)ds=\lim_{m\to\infty}\int_I\mathbf{max}(0,G_3(w^m)(s))ds$$

$$=\int_I\mathbf{max}(0,G_3(w)(s))ds,$$

which show that w is admissible. Now, let any $w' \in W$ and let $(w' \in W)$ be a subsequence converging to w' in L^2 strongly. By Steps 2 and 3, we have

$$DG^m(w^m, w'-w^m) + (\gamma/2)\|w'-w^m\|_{L^2}^2$$
$$= DG_0(w^m, w'-w^m) + \lambda_1^m DG_1(w^m, w'-w^m) + \lambda_2^m DG_2(w^m, w'-w^m)$$
$$+\int_I \lambda_3^m(s) DG_3(w^m, w'-w^m)(s)ds + (\gamma/2)\|w'-w^m\|_{L^2}^2 \geq d^m .$$

By Theorem 2.3, the derivatives DG_l, $l = 0,1,2,3$ (DG_3 with values in $C(I)^{m_3}$) are continuous and $\lambda_3^m \to \lambda_3$ in $M(I)^{m_3}$ weak star. Since also $w^m \to w$ in L^2 strongly, we can pass to the limit in the above inequality and obtain

$$DG_0(w, w'-w) + \lambda_1 DG_1(w, w'-w) + \lambda_2 DG_2(w, w'-w)$$
$$+\int_0^T \lambda_3(ds) DG_3(w, w'-w)(s) + (\gamma/2)\|w'-w\|_{L^2}^2 \geq 0,$$

which holds for every $w' \in W$. Replacing w' by $w + \theta(w'-w)$, $\theta \in (0,1]$, dividing by θ, and taking the limit as $\theta \to 0$, we obtain the same inequality, but without the term containing the norm. On the other hand, by the definition of λ_2^m and Theorem 2.2, if $G_{2j}(w) < 0$, for some index $j \in \{1,...,m_2\}$, then $\lambda_{2j}^m = 0$ for m sufficiently large, hence $\lambda_{2j} = 0$, which shows that $\lambda_2 G_2(w) = 0$. Now, since w is admissible, for each $j = 1,...,m_3$ fixed, we have $G_{3j}(w) \leq 0$, i.e.

$$g_{3j}(s, y(s)) \leq 0, \quad s \in I, \quad \text{with } y = y_w .$$

Let $\varepsilon > 0$ be given, and define the sets

$$S_{\varepsilon j} = \{s \in I \,|\, -\varepsilon \leq g_{3j}(s, y(s)) \leq 0,\},$$
$$S'_{\varepsilon j} = \{s \in I \,|\, g_{3j}(s, y(s)) \leq -\varepsilon\}.$$

Let m' be such that

$$|g_{3j}(s, y^m(s)) - g_{3j}(s, y(s))| \leq \varepsilon, \quad s \in I, \quad \text{for } m \geq m',$$

By the definition of λ_3^m, we have

$$\lambda_{3j}^m(s) = 0, \quad s \in S'_{\varepsilon j}, \quad \text{for } m \geq m'.$$

It follows that

$$\eta_j^m := \left|\int_0^T \lambda_{3j}^m(s) g_{3j}(s), y^m(s))ds\right| = \left|\int_{S_{\varepsilon j}} \lambda_{3j}^m(s) g_{3j}(s), y^m(s))ds\right|$$
$$\leq 2\varepsilon \|\lambda_{3j}^m\|_{L^1(S_{\varepsilon j})} \leq 2\varepsilon \|\lambda_{3j}^m\|_{L^1(I)} \leq 2c\varepsilon, \quad \text{for } m \geq m'.$$

Therefore, by the involved weak star and uniform convergences

$$0 = \lim_{m\to\infty} \eta_j^m = \left|\int_0^T \lambda_{3j}(ds) g_{3j}(s, y(s))\right|, \quad j = 1,...,m_3,$$

or equivalently (since $\lambda_3 \geq 0$ and $g_3(s, y(s)) \leq 0$, in the limit)

$$\int_0^T \lambda_3(ds) g_3(s, y(s)) = 0 .$$

On the other hand, since $\lambda_3^m \geq 0$ and $\lambda_3^m \to \lambda_3$ weak star, we have

$$1 + \sum_{l=0}^{2} \|\lambda_l^m\| + \int_I \lambda_3^m(s)ds = 1 + \sum_{l=0}^{2} \|\lambda_l^m\| + \int_0^T \{\sum_{j=1}^{m_3} [1 \cdot \lambda_{3j}^m(s)]\}ds$$

$$\to 1+\sum_{l=0}^{2}\|\lambda_l\|+\int_0^T\{\sum_{j=1}^{m_3}[1\cdot\lambda_{j3}(ds)]\}=1+\sum_{l=0}^{2}\|\lambda_l\|+\|\lambda_3\|_*=a\neq 0 .$$

We clearly have also $\lambda_0=1$, $\lambda_2\geq 0$, and since $\lambda_3^m\geq 0$ and $\lambda_3^m\to\lambda_3$ in $M(I)^{m_3}$ weak star, we have also $\lambda_3\geq 0$. Dividing all λ_l by a, w is thus weakly extremal classical.

(ii) Let (w^m) be a subsequence (same notation) of the sequence generated in Step 3 that converges in R to some r in R. The admissibility of r is proved as in (i). Suppose as in (i) that $\lambda_l^m\to\lambda_l$, $l=1,2$, and $\lambda_3^m\to\lambda_3$ in $M(I)^{m_3}$ weak star. By Step 2, we have

$$DG^m(w^m,w'-w^m)+(\gamma/2)\left\|w'-w^m\right\|_{L^2}^2$$
$$=DG_0(w^m,w'-w^{nm})+\lambda_1^m DG_1(w^m,w'-w^m)+\lambda_2^m DG_2(w^m,w'^n-w^m)$$
$$+\int_I\lambda_3^m(s)DG_3(w^m,w'-w^m)(s)ds+(\gamma/2)\left\|w'-w^m\right\|_{L^2}^2\geq d^m ,$$

for every $w'\in W$, which can be written

$$\sum_{l=0}^{2}\lambda_l^m\int_0^T[z_l^m(t)f_u(t,y^m(t),w^m(t))+g_{lu}(t,y^m(t),w^m(t))][w'(t)-w^m(t)]dt$$
$$+\int_0^T\lambda_3^m(ds)g_{3y}(s,y^m(s))Z(s)^{-1}\left(\int_0^s Z(t)f_u(t,y^m(t),w^m(t))[w'(t)-w^m(t)]dt\right)$$
$$+(\gamma/2)\int_0^T\left\|w'(t)-w^m(t)\right\|^2dt\geq d^m ,$$

for every $w'\in W$. Choosing now any Caratheodory function $\phi\in B(I\times U;U)$ and setting $w'(t):=\phi(t,w(t))$ in this inequality, by the continuity of the relaxed state and adjoint operators (Theorems 2.2-3) and using Proposition 2.1 in [1] (also for each fixed s - the integrals $\int_0^s$ are equicontinuous and converge for each s, hence uniformly), and the convergences $\lambda_3^m\to\lambda_3$ in $M(I)^{m_3}$ weak star and $w^m\to r$ in R, we can pass to the limit in this inequality and obtain

$$\sum_{l=0}^{2}\lambda_l\int_0^T[z_l(t)f_u(t,y(t),r(t))+g_{lu}(t,y(t),r(t))][\phi(t,r(t))-r(t)]dt$$
$$+\int_0^T\lambda_3(ds)g_{3y}(s,y(s))Z(s)^{-1}\left(\int_0^s Z(t)f_u(t,y(t),r(t))[\phi(t,r(t))-r(t)]dt\right)$$
$$+(\gamma/2)\int_0^T\left\|\phi(t,r(t))-r(t)\right\|^2dt\geq 0,\quad\text{for every such }\phi,$$

which implies, by a θ-argument similar to (i), the same inequality, but without the last integral term, and with multipliers as in the optimality conditions. The transversality conditions are proved similarly to (i).

(iii) In any of the above cases (i) or (ii), suppose that the limit control is admissible and that the classical, or relaxed, problem has no admissible, abnormal extremal, controls. Suppose that the multipliers are not all bounded. Then, dividing the corresponding inequality resulting from Step 2 by the greatest multiplier norm and passing to the limit for a subsequence, we readily see that we obtain an extremality inequality where the first multiplier is zero, and that the limit control is abnormal extremal, a contradiction. Therefore, the sequences of multipliers are bounded, and by (i) or (ii), this limit control is extremal as above.

The proofs in the absence of state constraints are similar.

If the additional assumptions of Theorem 2.4 are also satisfied (classical case), then Algorithm 1 computes optimal classical controls in case (i).

Next, we define the penalized functionals on R

$$G^m(r) = G_0(r) + \frac{1}{2}\{M_1^m \sum_{j=1}^{m_1} \left|G_{1j}(r)\right|^2 + M_2^m \sum_{j=1}^{m_2} [\max(0, G_{2j}(r))]^2$$

$$+ M_3^m \sum_{j=1}^{m_3} \int_I [\max(0, G_{3j}(r)(s))]^2 ds\} .$$

The directional derivative of G^m is given by

$$DG^m(r, r'-r) = DG_0(r, r'-r)$$

$$+ M_1^m \sum_{j=1}^{m_1} G_{1j}(r) DG_{1j}(r, r'-r) + M_2^m \sum_{j=1}^{m_2} \max(0, G_{2j}(r)) DG_{2j}(r, r'-r)$$

$$+ M_3^m \sum_{j=1}^{m_3} \int_I \max(0, G_{3j}(r)(s)) DG_{3j}^n(r, r'-r)(s) ds .$$

The relaxed penalized conditional descent method is described by the following Algorithm, where U is not necessarily convex.

Algorithm 2

Step 1. Set $k := 0$, $m := 1$, and choose an initial control $r_0^1 \in R$.

Step 2. Find $\overline{r}_k^m \in R$ such that

$$d_k := DG^m(r_k^m, \overline{r}_k^m - r_k^m) = \min_{r' \in R} DG^m(r_k^m, r' - r_k^m).$$

Step 3. If $\left|d_k\right| \le \beta^m$, set $r^m := r_k^m$, $\overline{r}^m := \overline{r}_k^m$, $d^m := d_k$, $m := m+1$, and go to Step 2.

Step 4. Find the lowest integer value s (positive or not), say $\overline{s}$, such that $\alpha(s) = c^s \zeta_k \in (0,1]$ and $\alpha(s)$ satisfies the inequality

$$G^m(r_k^m + \alpha(s)(\overline{r}_k^m - r_k^m)) - G^m(r_k^m) \le \alpha(s) b d_k ,$$

and then set $\alpha_k := \alpha(\overline{s})$.

Step 5. Choose any $r_{k+1}^m \in R$ such that

$$G^l(r_{k+1}^l) \le G^l(r_k^l + \alpha_k(\overline{r}_k^l - r_k^l)),$$

set $k := k+1$, and go to Step 2.

With r^m as defined in Step 3, define the sequences of multipliers

$$\lambda_1^m := M_1^m G_1(r^m), \quad \lambda_2^m := M_2^m \mathbf{max}(0, G_2(r^m)),$$

$$\lambda_3^m(s) := M_3^m \mathbf{max}(0, G_3(r^m)(s)) = M_3^m \mathbf{max}(0, g_3(s, y^m(s))) .$$

Theorem 3.2 We suppose that the derivatives in u are excluded in the last assumptions of Section 2.

(i) In the presence of state constraints, if a subsequence $(r^l)_{l \in L}$ of the sequence generated by Algorithm 2 in Step 3 converges to some r in R and the sequences (λ_l^m), $l = 1,2,3$, (λ_3^m) in $L^1(I)^{m_3}$, are bounded, then r is admissible and strongly

extremal relaxed for the relaxed problem. In the absence of state constraints, if a subsequence $(r_k)_{k\in K}$ (no index m) converges to some r in R, then r is strongly extremal relaxed for the relaxed problem.
(ii) In case (i) with state constraints, suppose that the relaxed problem has no admissible, abnormal extremal, controls. If r is admissible, then the sequences (λ_m^m) are bounded and r is also strongly extremal relaxed for the relaxed problem.
Proof. (State constraints present) Suppose by contradiction, as in the proof of Theorem 3.1, that m remains constant after a finite number of iterations in k, and so we drop here the index m. Since R is compact, let $(r_k)_{k\in K}$, $(\overline{r}_k)_{k\in K}$ be subsequences of the sequences generated by Algorithm 2 that converge to some $\tilde{r}\in R$, $\tilde{\overline{r}}\in R$, respectively. By Theorem 2.3, we have

$$d := \lim_{k\to\infty,\, k\in K} d_k = \lim_{k\to\infty,\, k\in K} DG(r_k,\overline{r}_k - r_k) = DG(\tilde{r},\tilde{\overline{r}}-\tilde{r}) \le 0.$$

Suppose that $d<0$. The function $\Phi(\alpha) := G(r+\alpha(r'-r))$ is continuous on $[0,1]$, and since the directional derivative $DG(r,r'-r)$ is linear in $r'-r$, Φ is differentiable on $(0,1)$ and has derivative $\Phi'(\alpha) = DG(r+\alpha(r'-r),r'-r)$. By the mean value theorem, for each $\alpha\in(0,1]$ there exists $\alpha'(\alpha)\in(0,\alpha)$ such that

$$G(r_k+\alpha(\overline{r}_k - r_k)) - G(r_k) = \alpha DG(r_k+\alpha'(\alpha)(\overline{r}_k - r_k), \overline{r}_k - r_k)\,.$$

Therefore, by Theorem 2.3

$$G(r_k+\alpha(\overline{r}_k - r_k)) - G(r_k) = \alpha(d+\varepsilon_{k\alpha}), \quad \text{for } \alpha\in[0,1],$$

where $\varepsilon_{k\alpha}\to 0$ as $k\to\infty$, $k\in K$, and $\alpha\to 0^+$. We then show, similarly to Theorem 3.1, that $d=0$, hence $d_k\to 0$, which leads also to a contradiction. Therefore, $m\to\infty$ in Algorithm 2.
(i) Let (r^m) be a subsequence (same notation) of the sequence generated in Step 3 of Algorithm 2, that converges to some $r\in R$ as $m\to\infty$. The admissibility of r is shown similarly to Theorem 3.1. Suppose that the sequences (λ_m^m) are bounded and, up to subsequences, that $\lambda_m^m \to \lambda_m$. Now, by Steps 2 and 3 we have, for any $r'\in R$

$$DG^m(r^m,r'-r^m) = \sum_{l=0}^{2}\lambda_l^m DG_l(r^m,r'-r^m) + \int_I \lambda_3^m(s) DG_3(r^m,r'-r^m)(s)ds \ge d^m\,.$$

Since $\left|d^m\right| \le \beta^m \to 0$ by Step 3, using the above convergences and Theorem 2.3, we can pass to the limit and obtain

$$\sum_{l=0}^{2}\lambda_l DG_l(r,r'-r) + \int_I \lambda_3(s) DG_3(r,r'-r)(s)ds \ge 0, \quad \text{for every } r'\in R,$$

with multipliers λ_l as in the optimality conditions. Finally, we find in the limit the transversality and other conditions similarly to Theorem 3.1. Therefore, r is also weakly extremal relaxed for the relaxed problem.

The proof in the absence of state constraints is similar.
(ii) The proof is similar to that of Theorem 3.1 (iii).

If the additional assumptions of Theorem 2.4 are also satisfied (relaxed case), then Algorithm 2 computes optimal relaxed controls.

One can see (using Filippov's selection theorem, see [13]) that a *classical* (measurable) control $\bar{r}_k^m$ in Step 2 can be found for every k by appropriately minimizing $H(t,\bar{y}^m(t),\bar{z}^m(t),u)$ on U, for each $t\in I$ (in practice, this is trivially done for discrete, e.g. piecewise constant, controls).

Algorithm 2 can be implemented as follows. Suppose for simplicity that there are no state constraints. We first choose the initial discrete control in Step 1 to be of Gamkrelidze type, i.e. equal for each t to a convex combination of $d+1$ Dirac measures at $d+1$ points of U, where d is the dimension of the system. Suppose, by induction, that the control r_k^m computed in Algorithm 2 is of Gamkrelidze type. Since the control $\bar{r}_k^m$ in Step 2 is chosen to be classical, i.e. piecewise Dirac, the control $\tilde{r}_k^m := r_k^m + \alpha(\bar{r}_k^m - r_k^m)$ in Step 5 is piecewise equal to a convex combination of $d+2$ Dirac measures. Using now a known property of convex hulls of finite vector sets, we can construct a Gamkrelidze control r_{k+1}^m equivalent to $\tilde{r}_k^m$, i.e. such that

$$f(t,\bar{\tilde{y}}_k^m(t),r_{k+1}^m(t)) = f(t,\bar{\tilde{y}}_k^m(t),\tilde{r}_k^m(t)) \in IR^d, \quad t\in I,$$

where $\tilde{y}_k^m$ corresponds to $\tilde{r}_k^m$, by selecting only $d+1$ appropriate points in U among the $d+2$ ones defining $\tilde{r}_k^m$, for each t. The control r_{k+1}^m yields then the same state, values of functionals and G^m as $\tilde{r}_k^m$. Therefore, the constructed control r_k^m is of Gamkrelidze type for every k.

In practice, by choosing in Algorithms 1 and 2 moderately growing sequences (M_l^m) and a sequence (β^m) relatively fast converging to zero, the resulting sequences of multipliers (λ_l^m) are often kept bounded. One can choose a fixed $\zeta_k := \zeta \in (0,1]$ in Step 4; a usually faster and adaptive procedure is to set $\zeta_0 := 1$, and then $\zeta_k := \alpha_{k-1}$, for $k \ge 1$.

Gamkrelidze Formulation Approach

When directly applied to nonconvex optimal control problems whose solutions are non-classical relaxed controls, the classical method yields often very poor convergence (due to highly oscillating extremal controls). For this reason, we describe here an alternative approach, assuming that U is convex, following [7], that uses the Gamkrelidze formulation. For simplicity, we consider only the case without state constraints. Consider the following relaxed problem, with state equation

$$y'(t) = f(t,y(t),r(t)),\ t\in I, \quad y(0)=y^0,$$

control constraint $r\in R$, and cost functional

$$G(r) := \bar{g}(y(T)).$$

For each t fixed, the vector $f(t,y(t),r(t))$ in IR^d belongs to the convex hull of the set

$f(t,y(t),U)\subset IR^d$. Hence, we can write

$$f(t,y(t),r(t)) = \sum_{j=1}^{d+1} v_j(t) f(t,y(t),w_j(t)), \quad \text{with } 0\le v_j(t)\le 1,\ \sum_{j=1}^{d+1} v_j(t) = 1,$$

and by Filippov's selection theorem, we can suppose that these functions v_j, w_j are measurable. Therefore, the control r yields the same state y as the Gamkrelidze control $r_G := \sum_{j=1}^{d+1} v_j(t)\delta_{w_j(t)}$. Conversely, every such a control r_G is clearly a relaxed control r that yields the same state. Therefore, the above relaxed control problem is equivalent to the following extended classical problem, with state equation

$$y'(t) = \sum_{j=1}^{d+1} v_j(t) f(t, y(t), w_j(t)) \quad \text{in } I, \quad y(0) = y^0,$$

classical controls $\mathbf{v} = (v_j)$, $\mathbf{w} = (w_j)$, convex control constraints

$$\sum_{j=1}^{d+1} v_j(t) = 1, \; 0 \le v_j(t) \le 1, \; w_j(t) \in U, \; j = 1,\dots,d+1,$$

and cost functional $\mathbf{G}(\mathbf{v},\mathbf{w}) := \overline{g}(y(T))$. We can therefore apply the classical method (Algorithm 1) to this problem. The main disadvantage of this approach is that the dimension of the control space is rapidly increased; it can be successfully applied for relatively small dimensions d, d'. In the general case, i.e. if U is not convex, one can use Algorithm 2 to solve such highly nonconvex problems.

Finally, Gamkrelidze relaxed controls (in practice discrete ones) computed as above, or by Algorithm 2, can then be approximated, and simulated, by classical controls using a standard procedure, see [2], [4].

4 Numerical examples

Let $I := [0,1]$.

Example 1. Define the reference state and control

$$\overline{y}_1(t) = e^{-t}, \; \overline{y}_2(t) = e^{-2t}, \; \overline{y}_3(t) = e^{-3t},$$
$$\overline{w}(t) = \min(1, -1 + 2.5t),$$

and consider the following optimal control problem, with state equations

$$y_1' = -y_1 + y_3 - e^{-3t} + \sin y_1 - \sin \overline{y}_1 + w_1 - \overline{w},$$
$$y_2' = y_1 - 2y_2 - e^{-t} + w_2 - \overline{w},$$
$$y_3' = y_2 - 3y_3 - e^{-2t} + w_3 - \overline{w},$$
$$t \in I,$$
$$y_1(0) = y_2(0) = y_3(0) = 1,$$

control constraint set $U = [-1,1]$, and cost functional

$$G_0(w) := 0.5 \int_0^1 \{\sum_{i=1}^{3} [(y_i - \overline{y}_i)^2 + (w_i - \overline{w})^2]\} dt.$$

The optimal control and state are clearly $w^* := (\overline{w}, \overline{w}, \overline{w})$ and $y^* := (\overline{y}_1, \overline{y}_2, \overline{y}_3)$. Algorithm 1, without penalties, was applied to this example, using the midpoint scheme (step size 0.002) for solving the differential equations, with piecewise constant classical controls, and $\gamma = 0.5$. After 15 iterations in k, we obtained

$$G_0(w_K) = 1.366 \cdot 10^{-12}, \; d_k = -4.979 \cdot 10^{-15}, \; \varepsilon_k = 4.316 \cdot 10^{-7}, \; \zeta_k = 4.789 \cdot 10^{-7},$$

where d_k was defined in Step 2 of Algorithm 1, ε_k is the max-error for the state at the nodes, and ζ_k the max-error for the control at the midpoints of the intervals. Figure 1 shows the first component of the last computed control.

Example 2. With the same state equations, cost, and parameters as in Example 1, but with constraint set $U = [-0.75,1]$, additional pointwise state constraints

$$G_3(w)(t) = 0.8 - 0.4t - y_1(t) \le 0, \quad t \in I,$$

and applying here the penalized Algorithm 1, we obtained after 85 iterations in k

$$G_0(w_k) = 3.765340220 \cdot 10^{-3}, \quad d_k = -4.502 \cdot 10^{-6},$$

the maximum state constraint violation

$$\theta_k := \max_i [\max(0,\ 0.8 - 0.4t_i - y_{1i,k})] = 1.444 \cdot 10^{-5},$$

and the first control and state components shown in Figures 3 and 4.

Example 3. Consider the nonconvex problem, with state equations

$$y_1' = -y_1 + w, \quad y_2' = 0.5(y_1 - \bar{y})^2 - w^2, \quad t \in I,$$

$$y_1(0) = 1, \quad y_2(0) = 0,$$

where $\bar{y}(t) = e^{-t}$, control constraint set $U = [-1,1]$, and cost $G(w) = y_2(1)$. The unique optimal relaxed control is clearly $r^*(t) = (\delta_{-1} + \delta_1)/2$ (δ_{-1}, δ_1 Dirac measures), with optimal state $y^* = \bar{y}$ and optimal cost $G(r^*) = -1$. Note that the optimal relaxed cost can be approximated as closely as desired with a classical control, but cannot be attained for such a control. Since here the set $f(t, y, U)$ is a continuous arc, hence a connected set, the Gamkrelidze formulation involves only three controls v, u, w

$$y_1' = -y_1 + vu + (1-v)w, \quad y_2' = \frac{1}{2}(y_1 - \bar{y})^2 - vu^2 - (1-v)w^2, \quad t \in I,$$

$$y_1(0) = 1, \quad y_2(0) = 0,$$

with $v \in [0,1]$ and $u, w \in [-1,1]$. Applying to this problem Algorithm 1 without penalties, we obtained after 15 iterations the control $v_k \approx 0.5$ with max-error $\approx 3 \cdot 10^{-6}$ at the midpoints, the controls $u_k = -1$, $w_k = 1$ exactly, the optimal state with max-error $\approx 3.8 \cdot 10^{-6}$ at the nodes, the approximate cost $G(v_k, u_k, w_k) = -0.999999999997$, and $d_k = -1.134 \cdot 10^{-6}$.

Example 4. Consider the following problem, with state equations

$$y_1' = -y_2 + w_1, \quad y_2' = -y_1 + w_2, \quad t \in [0, 0.5),$$

$$y_1' = -y_2 + w_1 - t + 0.5, \quad y_2' = -y_1 + w_2 - t + 0.5, \quad t \in [0.5, 1],$$

$$y_1(0) = y_2(0) = 1,$$

nonconvex control constraint set

$$U = \{(u_1, 0) \in IR^2 \mid 0 \le u_1 \le 1\} \cup \{(0, u_2) \in IR^2 \mid 0 \le u_2 \le 1\},$$

and cost functional

$$G_0(w) = 0.5\int_0^1 [(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2] dt,$$

where $\bar{y} = e^{-t}$. Clearly, the unique optimal relaxed control is

$$r^*(t) = \delta_{(0,0)}, \quad t \in [0, 0.5) \quad \text{(classical)}$$

$$r^*(t) = 2(1-t)\delta_{(0,0)} + (t-0.5)\delta_{(1,0)} + (t-0.5)\delta_{(0,1)}, \quad t \in [0.5,1] \quad \text{(non-classical)},$$

where δ denotes Dirac measures, which yields the optimal state $(y_1^*, y_2^*) = (\bar{y}, \bar{y})$, and cost $G_0(r^*) = 0$. Algorithm 2, without penalties, was applied here, using also the midpoint scheme (step size 0.002), here with piecewise constant relaxed controls. After 120 iterations in k, we obtained

$$G_0(r_k) = 2.552 \cdot 10^{-6}, \quad d_k = -3.393 \cdot 10^{-5}.$$

Figures 4 and 5 show the last relaxed control probability functions

$$p_0(t) = r_k(t)(\{(0,0)\}) \approx \begin{cases} 1, \ t \in [0,0.5) \\ 2(1-t), \ t \in [0.5,1) \end{cases}$$

$$p_1(t) = r_k(t)(\{(1,0)\}) \approx \begin{cases} 0, \ t \in [0,0.5) \\ t-0.5, \ t \in [0.5,1) \end{cases}$$

We also obtained $p_2(t) = r_k(t)(\{(0,1)\}) = 1 - p_0(t) - p_1(t)$.

Example 5. Consider the following problem, with state equation

$$y' = -y + w, \quad t \in I, \quad y(0) = 1,$$

control constraint set $U = [-1,1]$, state constraints

$$G_1(w) = y(1) - 0.5 = 0, \quad G_3(w)(s) = 0.3 - y(s) \le 0, \ s \in I,$$

and nonconvex cost functional

$$G_0(w) = \int_0^1 (0.5y^2 - w^2)dt.$$

Writing the solution of the state equation in closed form via the fundamental solution, first forward with initial condition $y(0) = 1$ and then backward with terminal condition $y(1) = 0.5$, we can see that the unique optimal relaxed control and state are

$$r^*(t) = \begin{cases} \delta_{-1}, \ t \in [0,\rho) \quad \text{(classical)} \\ 0.35\delta_{-1} + 0.65\delta_1, \ t \in [\rho,\sigma) \quad \text{(non-classical)} \\ \delta_1, \ t \in [\sigma,1] \quad \text{(classical)} \end{cases}$$

$$y^*(t) = \begin{cases} -1 + 2e^{-t}, \ t \in [0,\rho) \\ 0.3, \ t \in [\rho,\sigma) \\ 1 - 0.5e^{1-t}, \ t \in [\sigma,1] \end{cases}$$

where $\rho = -\ln 0.65 \approx 0.43$ is such that $-1 + 2e^{-\rho} = 0.3$, and $\sigma = 1 - \ln 1.4 \approx 0.66$ such that $1 - 0.5e^{1-\sigma} = 0.3$. The optimal cost is $G_0(r^*) \approx -0.868398913624$. Applying here the penalized Algorithm 2, we obtained after 200 iterations in k the results

$$G_0(r_k^m) = -0.868200567558, \quad G_1(r_k^m) = -9.732 \cdot 10^{-6},$$

$$\max_i[\max(0, 0.3 - y_{1i,k}^m)] = 2.962 \cdot 10^{-4}, \quad d_k = -1.236 \cdot 10^{-3}.$$

Figure 6 shows the last relaxed control probability function

$$p_1(t) = r_k^m(t)(\{(1)\}) \approx \begin{cases} 0, \ t \in [0,\rho) \\ 0.65, \ t \in [\rho,\sigma) \\ 1, \ t \in [\sigma,1] \end{cases}$$

and we also obtained $p_0(t) = r_k^m(t)(\{-1\}) = 1 - p_1(t)$. Figure 7 shows the last state.

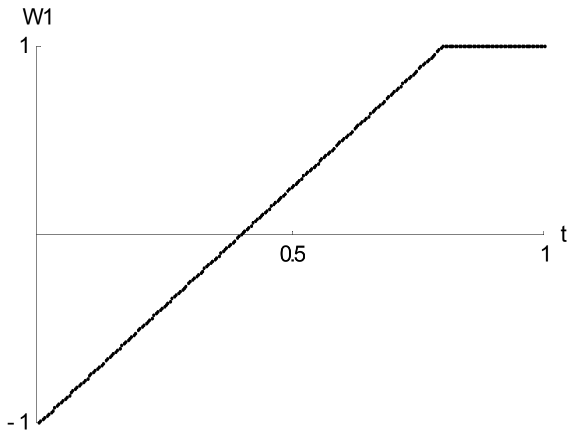

Figure 1. Example 1: Last control (1^{st} component)

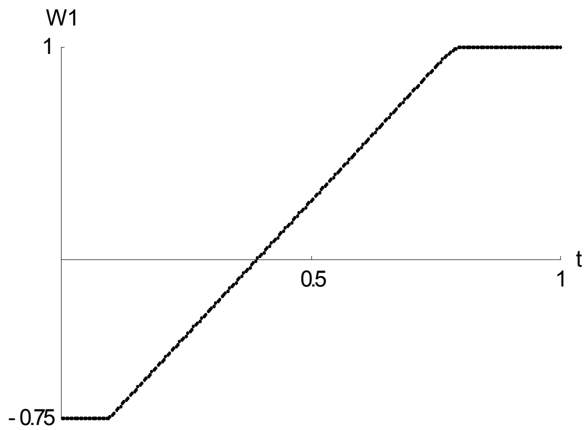

Figure 2. Example 2: Last control (1^{st} component)

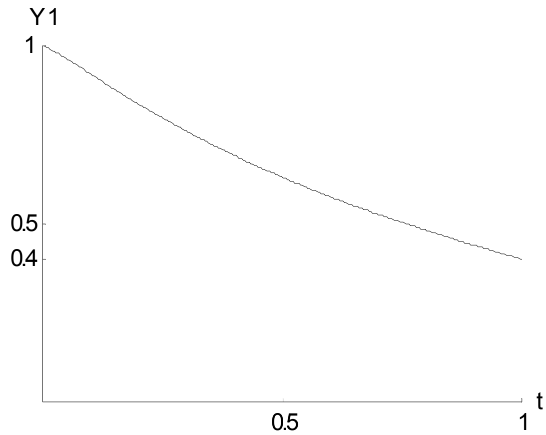

Figure 3. Example 2: Last state (1st component)

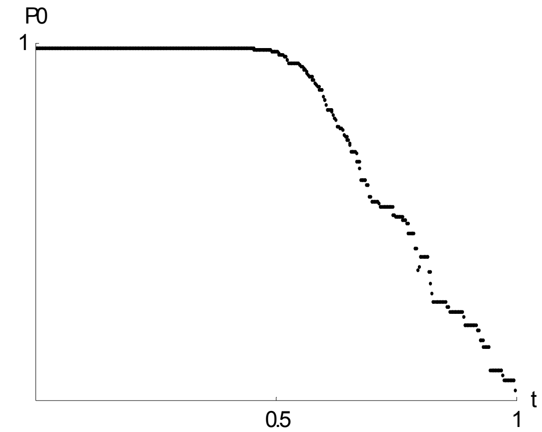

Figure 4. Example 4: Last relaxed control probability p_0

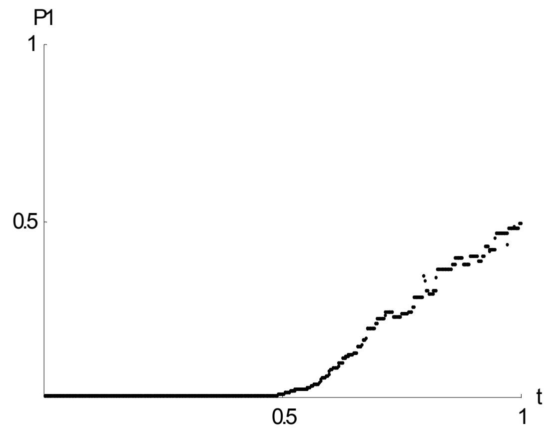

Figure 5. Example 4: Last relaxed control probability p_1

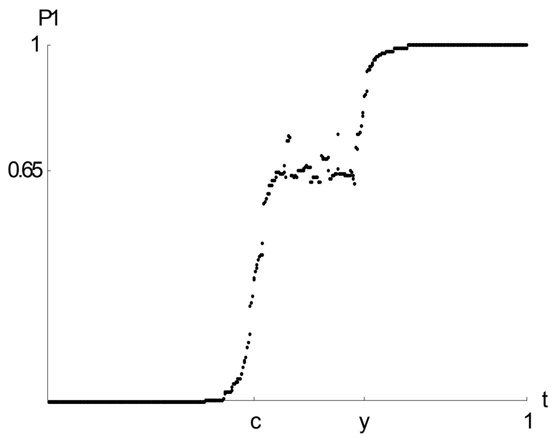

Figure 6. Example 5: Last relaxed control probability p_1

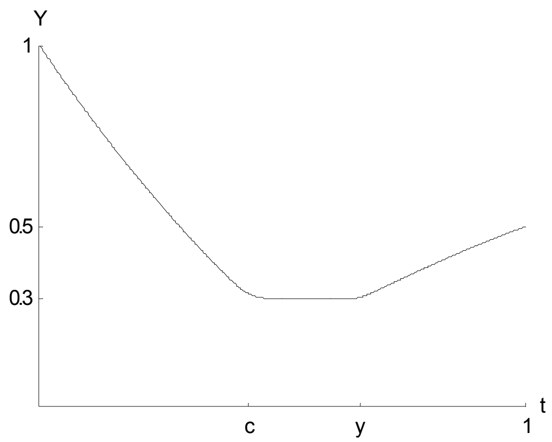

Figure 7. Example 5: Last state

References

[1] I. Chryssoverghi, Nonconvex optimal control of nonlinear monotone parabolic systems, Systems and Control Letters, 8 (1986), 55-62.

[2] I. Chryssoverghi, A. Bacopoulos, Discrete approximation of relaxed optimal control problems, Journal of Optimization Theory and Applications, 65, 3 (1990), 395-407.

[3] I. Chryssoverghi, A. Bacopoulos, J. Coletsos and B. Kokkinis, Discrete approximation of nonconvex hyperbolic optimal control problems with state constraints, Control and Cybernetics, 27, 1 (1998), 29-50.

[4] I. Chryssoverghi, J. Coletsos and B. Kokkinis, Discrete relaxed method for semilinear parabolic optimal control problems, Control and Cybernetics, 28, 2 (1999), 157-176.

[5] I. Chryssoverghi, J., Coletsos, B., Kokkinis, Approximate relaxed descent method for optimal control problems, Control and Cybernetics, 30, 4 (2001), 385-404.

[6] I. Chryssoverghi, Discrete gradient projection method with general Runge-Kutta schemes and control parameterizations for optimal control problems, Control and Cybernetics, 34, (2005), 425-451.
[7] I. Chryssoverghi, Discretization methods for semilinear parabolic optimal control problems, International Journal of Numerical Analysis and Modeling, 3 (2006), 437-458.
[8] A.L. Dontchev, W.W. Hager and V.M. Veliov, Second-order Runge-Kutta approximations in control constrained optimal control, SIAM Journal on Numerical Analysis, 38, 1 (2000), 202-226.
[9] T. Roubíček, A Convergent computational method for constrained optimal relaxed control problems, Journal of Optimization Theory and Applications, 69 (1991), 589-603.
[10] T. Roubíček, Relaxation in Optimization Theory and Variational Calculus, Walter de Gruyter, Berlin (1997).
[11] V.M. Veliov, On the time-discretization of control systems, SIAM Journal on Control and Optimization, 35, 5 (1997), 1470-1486.
[12] A. Schwartz and E. Polak, Consistent approximations for optimal control problems based on Runge-Kutta integration, SIAM Journal on Control and Optimization, 34, 4 (1996), 1235-1269.
[13] J. Warga, Optimal Control of Differential and Functional Equations, Academic Press, New York (1972).
[14] J. Warga, Steepest Descent with Relaxed Controls, SIAM Journal on Control and Optimization, 15 (1977) 674-682.

Received: May 18, 2006