Proceedings of the 5th WSEAS Int. Conf. on SIMULATION, MODELING AND OPTIMIZATION, Corfu, Greece, August 17-19, 2005 (pp399-406)

Discretization-Optimization Methods for Optimal Control Problems

ION CHRYSSOVERGHI
Department of Mathematics, School of Applied Mathematics and Physics
National Technical University of Athens
Zografou Campus, 15780 Athens
GREECE

Abstract: - We consider an optimal control problem described by nonlinear ordinary differential equations, with control and state constraints. Since this problem may have no classical solutions, it is also formulated in relaxed form. The classical control problem is then discretized by using the implicit midpoint scheme for state approximation, while the controls are approximated by piecewise constant classical ones. We first study the behavior in the limit of properties of discrete optimality, and of discrete admissibility and extremality. We then apply a penalized gradient projection method to each discrete classical problem, and also a corresponding progressively refining discretization-optimization method to the continuous classical problem, thus reducing computing time and memory. We show that accumulation points of sequences generated by these methods are admissible and extremal for the corresponding discrete or continuous, classical or relaxed, problem. For nonconvex problems whose solutions are non-classical, we show that we can apply the above methods to the problem formulated in the Gamkrelidze form. Finally, numerical examples are given.

Key-Words: - Optimal control, discretization, midpoint scheme, piecewise constant controls, penalized gradient projection method, relaxed controls.

1 Introduction

In this paper we propose discretization-optimization methods generating classical controls, instead of relaxed controls (see [1], [2], [5]), for solving optimal control problems, and study their behavior in the limit in the frameworks of classical and relaxation theories.

We consider an optimal control problem described by nonlinear ordinary differential equations, with control and end-point state constraints, and end-point cost. Since this problem may have no classical solutions, it is also formulated in relaxed form. The classical control problem is then discretized by using implicit midpoint schemes for state and adjoint approximation, the midpoint integration rule for approximation of the integrals involved in the derivatives of functionals, while the controls are approximated by piecewise constant classical ones. We first give various necessary conditions for optimality for the continuous classical and relaxed problems, and for the discrete problem. Next, we show that strong accumulation points in L^2 of sequences of optimal (resp. admissible and extremal) discrete controls are optimal (resp. admissible and weakly extremal) for the continuous classical problem, and that relaxed accumulation points of sequences of optimal (resp. admissible and extremal) discrete controls are optimal (resp. admissible and weakly extremal) for the continuous relaxed problem. We then apply a penalized gradient projection method to each discrete classical problem, and also a corresponding mixed discretization-optimization method to the continuous classical problem that progressively refines the discretization during the iterations, thus reducing computing time and memory, especially for large systems. We prove that accumulation points of sequences generated by the fixed discretization method are admissible and extremal for each discrete problem, and that strong classical (resp. relaxed) accumulation points of sequences of discrete controls generated by the progressively refining method are admissible and weakly extremal for the continuous classical (resp. relaxed) problem. For nonconvex problems whose solutions are non-classical, we show that we can apply the above methods to the problem formulated in the Gamkrelidze form. Using a standard procedure, the computed Gamkrelidze controls can then be approximated by classical ones. Finally, numerical examples are given.

Problems involving pointwise state constraints have been studied in [4], [5]. For various discretization and optimization methods in optimal control, see [1-7], [9], [10], and references there.

Proceedings of the 5th WSEAS Int. Conf. on SIMULATION, MODELING AND OPTIMIZATION, Corfu, Greece, August 17-19, 2005 (pp399-406)

2 The continuous problems

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

$$y'(t) = f(t, y(t), w(t)), \quad \text{for a.a. } t \in I := [0,T],$$

$$y(0) = y^0, \quad y(t) \in \mathbb{R}^d,$$

the constraints on the control w are

$$w(t) \in U, \quad \text{for a.a. } t \in I,$$

where U is a compact subset of $\mathbb{R}^{d'}$, the constraints on the state $y := y_w$ are

$$G_1(w) := g_1(y(T)) = 0, \quad G_2(w) := g_2(y(T)) \le 0,$$

where the vector functions g_1, g_2 take values in $\mathbb{R}^{m_1}, \mathbb{R}^{m_2}$, respectively, and the cost functional to be minimized is

$$G_0(w) := g_0(y(T)).$$

We define the set of *classical controls* by

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

and the set of *relaxed controls* (for the relevant theory, see [11] and [8]) 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 considered as a subset of R, and W is thus dense in R. For a given function $\phi \in B(I, U; \mathbb{R}^n)$, where B is the set of Caratheodory functions in the sense of Warga [11], and $r \in R$, we use the simplified notation

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

We can now define the relaxed problem. The state equation is

$$y'(t) = f(t, y(t), r(t)), \quad \text{for a.a. } t \in I,$$

$$y(0) = y^0, \quad \text{with } y := y_r,$$

the control constraint is $r \in R$, and the state constraints and cost are defined as in the classical problem, but with w replaced by r, according to the above notation.

We denote by $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^n$, $n \ge 1$. We suppose in the sequel that the function f is defined on $I \times \mathbb{R}^d \times U$, measurable for y, u fixed, continuous for t fixed, and satisfies

$$\|f(t, y, u)\| \le \psi(t) + \beta \|y\|,$$

for every $(t, y, u) \in I \times \mathbb{R}^d \times U$,

with $\psi \in L^1(I)$, $\beta \ge 0$,

$$\|f(t, y_1, u) - f(t, y_2, u)\| \le L \|y_1 - y_2\|,$$

for every $(t, y_1, y_2, u) \in I \times \mathbb{R}^{2d} \times U$.

Theorem 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 control $r \in R$.

Let B denote the closed ball in $\mathbb{R}^d$ with center 0 and radius b, defined in Theorem 1. We suppose now in addition that the functions g_l, $l = 0,1,2$, are continuous on B.

Theorem 2 The mappings $G_l : W \text{ (resp. } R) \to \mathbb{R}^{m_l}$, $l = 0,1,2$, are continuous on W (resp. R). 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 one $y'(t) \in \text{co}[f(t, y(t), U)]$. The classical problem may have no classical solution, and because $W \subset R$, 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, since W is dense in R. Usually, numerical methods slightly violate the state constraints; so, approximating an optimal relaxed control by a relaxed or a classical one, hence the relaxed optimal cost c_R, is not a drawback in practice (see [11], p. 248). Note also that approximating sequences of classical controls may converge to relaxed ones.

In order to state the various necessary conditions for optimality, we suppose in addition that the functions f, f_y, f_u are defined on $I \times B' \times U'$, where B' (resp. U') is an 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 such that

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

for every $(t, y, u) \in I \times B \times U$,

with $\xi, \eta \in L^1(I)$, and that the functions g_l, g_{ly}, $l = 0,1,2$, are defined on B' and continuous on B.

Theorem 3 (i) If U is convex, then for $w, w' \in W$ the directional derivative of the mapping G_l, $l = 0,1,2$, defined on W, is given by

$$DG_l(w, w'-w) := \lim_{\alpha \to 0^+} \frac{G_l(w + \alpha(w'-w)) - G_l(w)}{\alpha}$$
$$= \int_0^T z_l(t) f_u(t, y(t), w(t))[w'(t) - w(t)]dt,$$

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 defined by the linear adjoint equation

$$z_l'(t) = -z_l(t) f_y(t, y(t), w(t)), \quad \text{for a.a. } t \in I,$$
$$z_l(T) = g_{ly}(y(T)), \quad \text{with } y := y_w,$$

where the controls are considered as classical ones.
(ii) For $r, r' \in R$, the directional derivative of the mapping G_l, $l = 0,1,2$, defined on R, is given by

$$DG_l(r, r'-r) := \lim_{\alpha \to 0^+} \frac{G_l(r + \alpha(r'-r)) - G_l(r)}{\alpha}$$
$$= \int_0^T z_l(t) f(t, y(t), r'(t) - r(t))dt,$$

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

$$z_l'(t) = -z_l(t) f_y(t, y(t), r(t)), \quad \text{for a.a. } t \in I,$$
$$z_l(T) = g_{ly}(y(T)), \quad \text{with } y := y_r.$$

(iii) The mappings

$$(w, w') \mapsto DG_l(w, w'-w)$$
(resp. $(r, r') \mapsto DG_l(r, r'-r)$), $l = 0,1,2,$

are continuous on $W \times W$ (resp. $R \times R$).

Theorem 4 (i) If U is convex and the control $w \in W$ is optimal for the classical problem, then w is *weakly extremal classical*, i.e. there exist multipliers $\lambda_0 \in \mathbb{R}$, $\lambda_1 \in \mathbb{R}^{m_1}$, $\lambda_2 \in \mathbb{R}^{m_2}$, with $\lambda_0 \geq 0$, $\lambda_2 \geq 0$, $\sum_{l=0}^{2} \|\lambda_l\| = 1$, such that

(1) $\displaystyle\sum_{l=0}^{2} \lambda_l DG_l(w, w'-w)$
$$= \sum_{l=0}^{2} \lambda_l \int_0^T z_l(t) f_u(t, y(t), w(t))[w'(t) - w(t)]dt$$
≥ 0, for every $w' \in W$,

(2) $\lambda_2 G_2(w) = 0$ (*transversality condition*).

The condition (1) is equivalent to the pointwise weak classical minimum principle

$$\sum_{l=0}^{2} \lambda_l z_l(t) f_u(t, y(t), w(t)) w(t)$$
$$= \min_{u \in U} \sum_{l=0}^{2} \lambda_l z_l(t) f_u(t, y(t), w(t)) u, \quad \text{for a.a. } t \in I.$$

(ii) If the control $r \in R$ 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

(3) $\displaystyle\sum_{l=0}^{2} \lambda_l DG_l(r, r'-r)$
$$= \sum_{l=0}^{2} \lambda_l \int_0^T z_l(t) f(t, y(t), r'(t) - r(t))dt \geq 0,$$
for every $r' \in R$,

(4) $\lambda_2 G_2(r) = 0$.

The condition (3) is equivalent to the pointwise strong relaxed minimum principle

$$\sum_{l=0}^{2} \lambda_l z_l(t) f(t, y(t), r(t))$$
$$= \min_{u \in U} \sum_{l=0}^{2} \lambda_l z_l(t) f(t, y(t), u), \quad \text{for a.a. } t \in I.$$

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

$$\sum_{l=0}^{2} \lambda_l z_l(t) f_u(t, y(t), r(t))\, r(t)$$
$$= \min_{\phi} \sum_{l=0}^{2} \lambda_l z_l(t) f_u(t, y(t), r(t))\, \phi(t, r(t)),$$
for a.a. $t \in I$,

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

$$\sum_{l=0}^{2} \lambda_l \int_0^T z_l(t) f_u(t, y(t), r(t))[\phi(t, r(t)) - r(t)]dt$$
≥ 0, for every $\phi \in B(I, U; U)$.

A control r satisfying this condition and (4) is called *weakly extremal relaxed*.

3 The discrete problems

In the sequel, we suppose that the functions f, f_y, f_u and g_l, g_{ly}, g_{lu}, $l = 0,1,2$, are continuous in all their arguments. Let $(N^n)_{n \geq 0}$ be an increasing sequence of positive integers such that $N^n \to \infty$. We set

$$N := N^n, \quad N' := N'^n, \quad h^n := T / N,$$
$$t_i^n := i h^n, \quad i = 0, \dots, N,$$
$$I_i^n := [t_{i-1}^n, t_i^n), \quad i = 1, \dots, N-1, \quad I_N^n := [t_{N-1}^n, t_N^n].$$

We define the set of *discrete classical controls*

$$W^n := \{ w^n \in W \mid w^n(t) := w_i^n \in U \text{ in } \overset{o}{I_i^n}, \; i = 1, \dots, N \}.$$

For a given discrete control $w^n \in W^n$, the discrete state $y^n := y^n_{w^n} = (y^n_0, ..., y^n_N)$ is the solution of the implicit midpoint scheme

$$y^n_i = y^n_{i-1} + h^n f(\bar{t}^n_i, \bar{y}^n_i, w^n_i), \quad i = 1, ..., N,$$

$$y^n_0 := y^0,$$

with $\bar{y}^n_i := (y^n_{i-1} + y^n_i)/2$, $\bar{t}^n_i := (t^n_{i-1} + t^n_i)/2$.

Theorem 5 If $h^0 < 2/L$, then the discrete state y^n is uniquely defined, and there exists a constant b' such that $\|y^n_i\| \le b'$, $i = 0, ..., N$, for every n and $w^n \in W^n$.

From now on, we suppose that $h^0 < 2/L$. The discrete state equation can then be solved numerically, for each $i = 1, ..., N$, by the standard predictor-corrector method.

The discrete control constraint is $w^n \in W^n$. Define the discrete mappings

$$G^n_l(w^n) := g_l(y^n_N), \quad l = 0, 1, 2.$$

The discrete state constraints are *either* of the two following ones

Case (a) $\|G^n_1(w^n)\| \le \varepsilon^n_1$,

Case (b) $G^n_1(w^n) = \varepsilon^n_1$,

and

$$G^n_2(w^n) \le \varepsilon^n_2,$$

where the *admissibility perturbations* ε^n_l are appropriate positive numbers or vectors converging to zero, to be defined later. The discrete cost to be minimized is $G^n_0(w^n)$.

Theorem 6 The mappings $w^n \to y^n$, $w^n \to G^n_l(w^n)$ are continuous on W^n. If any of the two above discrete problems is feasible, then it has a solution.

Theorem 7 If U is convex, then for $w^n, w'^n \in W^n$ the directional derivative of the mapping G^n_l, $l = 0, 1, 2$, defined on W^n, is given by

$$DG^n_l(w^n, w'^n - w^n)$$

$$= h^n \sum_{i=1}^{N} z^n_{li} f_u(\bar{t}^n_i, \bar{y}^n_i, w^n_i)(w'^n_i - w^n_i),$$

where the adjoint state z^n_l is given by the linear implicit scheme

$$z^n_{l,i-1} = z^n_{li} + h^n \bar{z}^n_{li} f_y(\bar{t}^n_i, \bar{y}^n_i, w^n_i), \quad i = N, ..., 1,$$

$$z^n_{lN} = g_{ly}(y^n_N), \quad \text{with } y^n := y^n_{w^n}.$$

The mappings $(w^n, w'^n) \mapsto DG_l(w^n, w'^n - w^n)$, for $l = 0, 1, 2$, are continuous on $W^n \times W^n$.

We now state the discrete necessary conditions for optimality.

Theorem 8 If U is convex and w^n is optimal for the discrete problem with state constraints, Case (b), then w^n is *discrete extremal classical*, i.e. there exist multipliers $\lambda^n_0 \in \mathbb{R}$, $\lambda^n_1 \in \mathbb{R}^{m_1}$, $\lambda^n_2 \in \mathbb{R}^{m_2}$, with $\lambda^n_0 \ge 0$, $\lambda^n_1 \ge 0$, $\sum_{l=0}^{2} \|\lambda_l\| = 1$, such that

(5) $\sum_{l=0}^{2} \lambda^n_l DG^n_l(w^n, w'^n - w^n) \ge 0$,

for every $w'^n \in W^n$,

(6) $\lambda^n_2 [G^n_2(w^n) - \varepsilon^n_2] = 0$.

The condition (5) is equivalent to the discrete pointwise weak classical minimum principle

$$\sum_{l=0}^{2} \lambda^n_l z^n_{li} f_u(\bar{t}^n_i, \bar{y}^n_i, w^n_i) w^n_i$$

$$= \min_{u \in U} \sum_{l=0}^{2} \lambda^n_l z^n_{li} f_u(\bar{t}^n_i, \bar{y}^n_i, w^n_i) u, \quad i = 1, ..., N.$$

4 Behavior in the limit

In this section we study the behavior in the limit of properties of discrete optimality, and of discrete admissibility and extremality. Define the piecewise constant functions

$$\bar{y}^n(t) := (y^n_{i-1} + y^n_i)/2, \quad t \in I^n_i, \ i = 1, ..., N,$$

and the piecewise linear functions

$$\hat{y}^n(t) := y^n_{i-1} + (t - t^n_{i-1}) f(\bar{t}^n_i, \bar{y}^n_i, w^n_i), \quad t \in I^n_i,$$

$$i = 1, ..., N.$$

Theorem 9 (Consistency of states and functionals)
(i) Let $(w^n \in W^n)$ be a sequence such that $w^n \to w$ in L^2 strongly. Then $w \in W$, $\hat{y}^n \to y$, $\bar{y}^n \to y$ uniformly, where $y := y_w$, and

$$G^n_l(w^n) \to G_l(w), \quad l = 0, 1, 2.$$

(ii) Let $(w^n \in W^n \subset R)$ be a sequence such that $w^n \to r$ in R. Then $\hat{y}^n \to y$, $\bar{y}^n \to y$ uniformly, where $y := y_r$, and

$$G^n_l(w^n) \to G_l(r), \quad l = 0, 1, 2.$$

Theorem 10 (Consistency of adjoints and derivatives of functionals)
(i) If $(w^n \in W^n)$ is a sequence such that $w^n \to w$ in L^2 strongly (resp. $w^n \to r$ in R), then $\hat{z}^n_l \to z_l$, $\bar{z}^n_l \to z_l$, uniformly, where $z_l := z_{lw}$ (resp. $z_l := z_{lr}$).

(ii) If $(w^n \in W^n)$, $(w'^n \in W'^n)$ are sequences such that $w^n \to w$, $w'^n \to w'$ in L^2 strongly, then

$$DG_l^n(w^n, w'^n - w^n) \to DG_l(w, w' - w).$$

Theorem 11 (Control approximation)
(i) For every $w \in W$, there exists a sequence $(w^n \in W^n)$ that converges to w in L^2 strongly.
(ii) For every $r \in R$, there exists a sequence $(w^n \in W^n \subset R)$ that converges to r in R.

We suppose in the sequel that each considered continuous classical or relaxed problem is feasible. The following theorem addresses the behavior in the limit of optimal discrete controls.

Theorem 12 If there are state constraints, we suppose that the sequences (ε_l^n) in the discrete state constraints, Case (a), converge to zero and satisfy

$$\left\|G_1^n(\tilde{w}^n)\right\| \le \varepsilon_1^n, \quad G_2^n(\tilde{w}^n) \le \varepsilon_2^n, \quad \varepsilon_2^n \ge 0,$$

for every n, where $(\tilde{w}^n \in W^n)$ is some sequence converging in L^2 (resp. in R) to an optimal control, if it exists (resp. which exists) $\tilde{w} \in W$ (resp. $\tilde{r} \in R$) of the classical (resp. relaxed) problem. For each n, let w^n be optimal for the discrete problem, Case (a). Then every accumulation point of (w^n) in L^2 (resp. R) is optimal for the continuous classical (resp. relaxed) problem.

Next, we consider the discrete problems with state constraints, Case (b). We first construct sequences of perturbations (ε_l^n), converging to zero and such that the discrete problem is feasible for every n, as follows. For each n, let $w'^n \in W^n$ be a solution of the following auxiliary minimization problem without state constraints

$$c^n = \min_{w^n \in W^n} \{\sum_{j=1}^{m_1} [G_{j1}^n(w^n)]^2 + \sum_{j=1}^{m_2} [\max(0, G_{j2}^n(w^n))]^2\}.$$

Then set

$$\varepsilon_{j1}^n := G_{j1}^n(w'^n), \quad j = 1,...,m_1,$$

$$\varepsilon_{j2}^n := \max(0, G_{j2}^n(w'^n)), \quad j = 1,...,m_2.$$

Using our assumptions, it can be shown that $c^n \to 0$, hence $\varepsilon_1^n \to 0$ and $\varepsilon_2^n \to 0$.

The following theorem addresses the behavior in the limit of admissible and extremal discrete controls.

Theorem 13 For each n, let w^n be admissible and extremal for the discrete problem, Case (b), with the perturbations (ε_l^n) constructed as above. Then every accumulation point of (w^n) in L^2 (if it exists) is admissible and weakly extremal classical for the continuous classical problem, and every accumulation point in R (which always exists) is admissible and weakly extremal relaxed for the continuous relaxed problem.

5 Discretization-optimization methods

We suppose here that U is convex. Let (M_1^m), (M_2^m) be nonnegative increasing sequences such that $M_l^m \to \infty$ as $m \to \infty$, and define the penalized discrete functionals

$$G^{nm}(w^n) := G_0^n(w^n) + \frac{1}{2}\{M_1^m \sum_{j=1}^{m_1} \left|G_{1j}^n(w^n)\right|^2$$

$$+ M_2^m \sum_{j=1}^{m_2} [\max(0, G_{2j}^n(w^n))]^2.$$

Let $\gamma \ge 0$, $b, c \in (0,1)$, and let (β^m), (ζ^m) with $\zeta^m \le 1$, be positive decreasing sequences that converge to zero. The algorithm described below contains various options. In the case of the progressively refining version, we suppose that either $N(n+1) = N(n)$ or $N(n+1) = \mu N(n)$, for some integer $\mu \ge 2$. In this case, we have $W^n \subset W^{n+1}$, and thus a control $w^n \in W^n$ may be considered also as belonging to W^{n+1}, and therefore the computation of states, adjoints and derivatives of functionals for this control, but with the possibly finer discretization $n+1$, makes sense.

Algorithm
Step 1. Set $k := 0$, $m := 1$, choose a value of n and an initial control $w_0^{n1} \in W^n$.
Step 2. Find $v_k^{nm} \in W^n$ such that

$$e_k := DG^{nm}(w_k^{nm}, v_k^{nm} - w_k^{nm}) + (\gamma/2)\left\|v_k^{nm} - w_k^{nm}\right\|_{L^2}^2$$

$$= \min_{v'^n \in W^n} [DG^{nm}(w_k^{nm}, v'^n - w_k^{nm})$$

$$+ (\gamma/2)\left\|v'^n_k - w_k^{nm}\right\|_{L^2}^2],$$

and set $d_k := DG^{nm}(w_k^{nm}, v_k^{nm} - w_k^{nm})$.
Step 3. If $\left|e_k\right| \le \beta^m$, set $w^{nm} := w_k^{nm}$, $v^{nm} := v_k^{nm}$, $e^m := e_k$, $d^m := d_k$, $m := m+1$, $[n := n+1]$, and go to Step 2.
Step 4. (*Armijo step search*) Find the lowest integer value $s \in \mathbb{Z}$, say $\bar{s}$, such that $\alpha = c^s \zeta^m \in (0,1]$ and α satisfies the inequality

$$G^{nm}(w_k^{nm} + \alpha_k(v_k^{nm} - w_k^{nm})) - G^{nm}(w_k^{nm}) \le \alpha_k b e_k,$$

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

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

In this Algorithm, we consider two versions:

Version A. "$n := n+1$" is skipped in Step 3: n is a constant integer chosen in Step 1, i.e. we choose a *fixed discretization* and replace the discrete functionals G_l^n by the perturbed ones.

Version B. "$n := n+1$" is not skipped in Step 3: in this case, it can be shown that $n \to \infty$, i.e. we have a *progressively refining* discrete method and we can take $n = 1$ in Step 1, hence $n = m$ in the Algorithm.

The progressively refining version has the advantage of reducing computing time and memory, and also of avoiding the computation of minimum feasibility perturbations (see Section 4). It is justified by the fact that finer discretizations become progressively more efficient as the iterate gets closer to an extremal control, while relatively coarser ones in the early iterations have not much influence on the final results.

If $\gamma > 0$ (*penalized gradient projection method*), one can see by "completing the square" that Step 2 reduces to finding, for each i, the projection of a vector onto U. If $\gamma = 0$ (*penalized conditional gradient method*), Step 2 reduces to the minimization of a linear function on U, for each i.

A (continuous strongly or weakly, classical or relaxed, or a discrete) extremal control is called *abnormal* if there exist multipliers as in the corresponding optimality conditions, with $\lambda_0 = 0$ (or $\lambda_0^n = 0$). A control is admissible and abnormal extremal in exceptional, degenerate, situations. Define the sequences of multipliers

$$\lambda_1^{nm} := M_1^m G_1^n(w^{nm}),$$

$$\lambda_2^{nm} := M_2^m \mathbf{max}(0, G_2^n(w^{nm})),$$

where **max** denotes a vector of max values, and w^{nm} is defined in Step 3 of the Algorithm.

Theorem 14 (i) In Version B, let (w^{nm}) be a subsequence (if it exists) of the sequence generated by the Algorithm in Step 3 that converges to some $w \in W$ in L^2 strongly as $m \to \infty$ (hence $n \to \infty$). If the sequences (λ_l^{nm}) are bounded, then w is admissible and weakly extremal classical for the continuous classical problem.

(ii) In Version B, let (w^{nm}) be a subsequence of the sequence generated by the Algorithm in Step 3 that converges to some r in R as $m \to \infty$ (hence $n \to \infty$). If the sequences (λ_l^{nm}) are bounded, then r is admissible and weakly extremal relaxed for the continuous relaxed problem.

(iii) In Version A, let $(w^{nm} \in W^n)$, n fixed, be a subsequence generated by the Algorithm in Step 3 that converges to some $w^n \in W^n$ as $m \to \infty$. If the sequences (λ_l^{nm}) are bounded, then w^n is admissible and extremal for the fixed discrete problem.

(iv) In any of the above convergence cases (i), (ii), (iii), suppose that the (discrete or continuous) limit 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.

In practice, by choosing moderately growing sequences (M_m^l) and a sequence (β^l) relatively fast converging to zero, the resulting sequences of multipliers (λ_m^{nl}) are often kept bounded.

When directly applied to nonconvex optimal control problems whose solutions are non-classical relaxed controls, the classical methods often yield very poor convergence. For this reason, we describe now another approach that uses the Gamkrelidze formulation of the problem. For simplicity, we consider the case without state constraints. We suppose that U is convex. Consider the relaxed problem, with state equation

$$y'(t) = f(t, y(t), r(t)), \text{ for a.a. } t \in I, \quad y(0) = y^0,$$

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

$$G(r) := g(y(T)).$$

For each $t \in I$ fixed, the vector $f(t, y(t), r(t))$ belongs to $\mathrm{co}[f(t, y(t), U)] \subset \mathbb{R}^d$, hence

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

$$\text{with } v_j(t) \in [0,1], \quad \sum_{j=1}^{d+1} v_j(t) = 1,$$

and by Filippov's selection theorem (see [11]), we can suppose that 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 one, with state equation

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

$$y(0) = y^0,$$

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

$$\sum_{j=1}^{d+1} v_j(t) = 1, \quad v_j(t) \in [0,1], \quad w_j(t) \in U,$$

$$j = 1,...,d+1,$$

and cost $\mathbf{G}(\mathbf{v},\mathbf{w}) = g(y(T))$. Consequently, we can apply the methods described above to this problem. The main disadvantage of this approach is that the dimension of the control space is rapidly increased. It can therefore be successfully applied for relatively small dimensions d, d'. The Gamkrelidze relaxed controls computed thus can then be approximated by classical controls using a standard procedure (see [1]). If U is not convex, one can use methods generating relaxed controls to solve such strongly nonconvex problems (see [1], [2], [5]).

6 Numerical examples

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

a) Define the reference state $\bar{y}(t) := e^{-t}$ and control

$$\bar{w}(t) := \begin{cases} -1, \; t \in [0, 0.25) \\ -0.8 + 1.8s^2(2-s), \; t \in [0.25, 1] \end{cases}$$

with $s = (t - 0.25)/0.75$. Consider the following problem, with state equations

$$y_1' = -y_1 + \sin y_1 - \sin \bar{y} + w - \bar{w},$$

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

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

control constraint set $U = [-1,1]$, and cost $G_0(w) = y_2(1)$. Clearly, the optimal control and state are $\bar{w}$ and $\bar{y}$. The discrete gradient projection method, without penalties, was applied to this example, with $\gamma = 0.5$, $N = 128$, and zero initial control. After 9 iterations in k, we obtained the control shown in Fig.1 and the following results

$$G_0^n(w_k^n) = 8.198 \cdot 10^{-11}, \quad e_k = -7.644 \cdot 10^{-11},$$

$$\eta_k = 3.314 \cdot 10^{-6}, \quad \zeta_k = 8.614 \cdot 10^{-5},$$

where η_k is the discrete max state error at the end-points of the intervals I_i^n and ζ_k the discrete max control error at the midpoints of these intervals.

b) With the constraint set $U = [-0.7, 0.3]$, the control constraints being now strictly active for the method and for the problem, we obtained after 9 iterations in k the control shown in Fig.2 and the results

$$G_0^n(w_k^n) = 3.446229100842155 \cdot 10^{-2},$$

$$e_k = 1.033 \cdot 10^{-13}.$$

c) With the first state equation replaced by $y_1' = -y_1 + w$, the constraint set $U = [-0.5, 0.8]$, the additional state constraint $G_1(w) = y_1(T) - 0.5 = 0$, and applying here the discrete penalized gradient projection method, we obtained after 96 iterations in k the control shown in Fig.3, the state shown in Fig.4, and the results

$$G_0^n(w_k^{nm}) = 8.032566544564755 \cdot 10^{-2},$$

$$G_1^n(w_k^{nm}) = -2.570 \cdot 10^{-4}, \quad e_k = -6.367 \cdot 10^{-5}.$$

d) Consider the following nonconvex problem, with state equations

$$y_1' = -y_1 + w, \quad y_2' = 0.5(y_1 - \bar{y})^2 - w^2,$$

$$y_1(0) = 1, \quad y_2(0) = 0, \quad \text{with } \bar{y}(t) = e^{-t},$$

control constraint set $U = [-1,1]$, and cost $G(w) = y_2(1)$. The unique optimal relaxed control is clearly $r^*(t) = (\delta_{-1} + \delta_1)/2$, 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 velocity set $f(t, y, U)$ is a continuous arc in $\mathbb{R}^2$, hence a connected set in $\mathbb{R}^2$, the Gamkrelidze formulation involves only three controls v, u, w

$$y_1' = -y_1 + vu + (1-v)w,$$

$$y_2' = 0.5(y_1 - \bar{y})^2 - vu^2 - (1-v)w^2,$$

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

with $v \in [0,1]$ and $u, w \in [-1,1]$. Applying, without penalties, the discrete conditional gradient method (i.e. with $\gamma = 0$), which yielded a better convergence for this special problem, with $N = 128$ and initial controls

$$v_0 := 0.5 + 0.3t,$$

$$u_0 := -0.7 - 0.3t, \quad w_0 := 0.7 + 0.3t,$$

we obtained after 10 iterations in k the control $v_k^n \approx 0.5$ with max error $\leq 6 \cdot 10^{-5}$, the controls $u_k^n = -1$, $w_k^n = 1$ exactly, the optimal state with max error $\leq 8 \cdot 10^{-5}$, $e_k = -2.355 \cdot 10^{-5}$, and the cost

$$\mathbf{G}^n(v_k^n, u_k^n, w_k^n) = -0.999999998746912.$$

Finally, the progressively refining version of the methods where also applied to the above problems, with successive step sizes 1/32, 1/64, 1/128, in three nearly equal periods, and yielded results of practically similar accuracy, but required here about half the computing time.

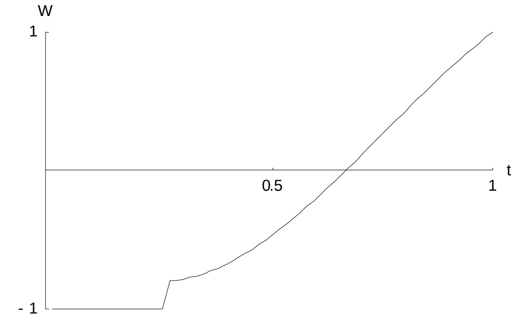

Fig.1

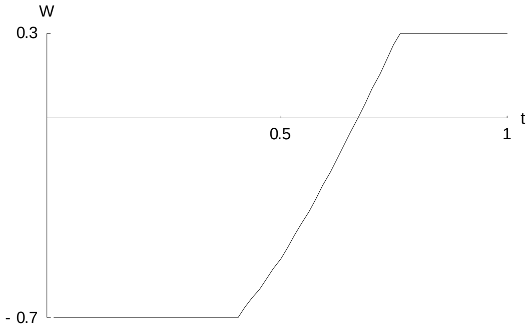

Fig.2

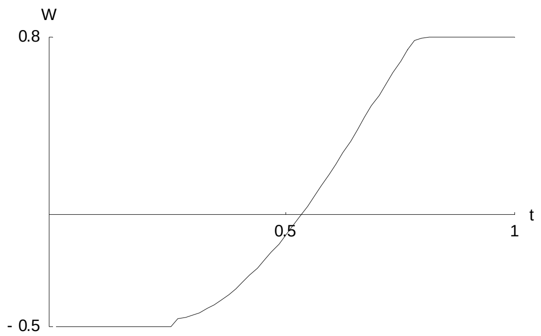

Fig.3

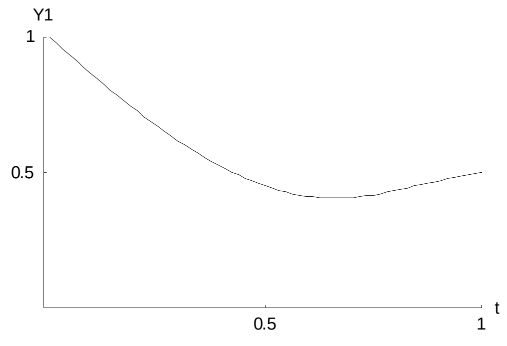

Fig.4

References:

[1] I. Chryssoverghi, A. Bacopoulos, B. Kokkinis and J. Coletsos, Mixed Frank-Wolfe Penalty Method with Applications to Nonconvex Optimal Control Problems, *Journal of Optimization Theory and Applications*, Vol.94, No.2, 1997, pp. 311-334.

[2] I. Chryssoverghi, J. Coletsos and B. Kokkinis, Discrete Relaxed Method for Semilinear Parabolic Optimal Control Problems, *Control and Cybernetics*, Vol.28, No.2, 1999, pp. 157-176.

[3] I. Chryssoverghi, Approximate Gradient Projection Method with General Runge-Kutta Schemes and Piecewise Polynomial Controls for Optimal Control Problems, to appear in *Control and Cybernetics*, 2005.

[4] I. Chryssoverghi, J. Coletsos and B. Kokkinis, Discretization Methods for Optimal Control Problems with State Constraints, to appear in *Journal of Computational and Applied Mathematics*, 2005.

[5] I. Chryssoverghi, J. Coletsos and B. Kokkinis, Discretization Methods for Nonconvex Optimal Control Problems with State Constraints, to appear in *Numerical Functional Analysis and Optimization*, 2005.

[6] I. Chryssoverghi, Discretization Methods for Semilinear Parabolic Optimal Control Problems, to appear in *International Journal of Numerical Analysis and Modeling*, 2005.

[7] E. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer, 1997.

[8] T. Roubíček, *Relaxation in Optimization Theory and Variational Calculus*, Walter de Gruyter, 1997.

[9] V. M. Veliov, On the Time-discretization of Control Systems, *SIAM Journal on Control and Optimization*, Vol.35, No.5, 1997, pp. 1470-1486.

[10] A. Schwartz and E. Polak, Consistent Approximations for Optimal Control Problems Based on Runge-Kutta Integration, *SIAM Journal on Control and Optimization*, Vol. 34, No. 4, 1996, pp. 1235-1269.

[11] J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, 1972.