NUMERICAL STABILIZATION OF BILINEAR CONTROL SYSTEMS

LARS GRÜNE†

Abstract. For bilinear control systems with constrained control values extremal Lyapunov exponents are computed numerically by solving discounted optimal control problems. Based on this computation a numerical algorithm to calculate stabilizing control functions is developed.

Key words. stabilization, bilinear control systems, Lyapunov exponents, discounted optimal control problems, Hamilton Jacobi Bellman equation

AMS subject classifications. 93D22, 49L25

1. Introduction. In this paper we present numerical algorithms for the calculation of extremal Lyapunov exponents and stabilization of bilinear control systems in $\mathbb{R}^d$, i.e. systems of the form

$$\dot{x}(t) = \Big(A_0 + \sum_{i=1}^{m} u_i(t) A_i\Big) x(t), \quad x(0) = x_0 \in \mathbb{R}^d \setminus \{0\} \tag{1.1}$$

with $A_j \in \mathbb{R}^{d\times d}$, $j = 0, \ldots, m$, $u(\cdot) \in \mathcal{U} := \{u : \mathbb{R} \to U,\ u \text{ measurable}\}$ with a compact and convex set of control values $U \subset \mathbb{R}^m$ with nonvoid interior. The Lyapunov exponent of (1.1) with respect to an initial value $x_0 \in \mathbb{R}^d$ and a control function $u(\cdot) \in \mathcal{U}$ is given by

$$\lambda(x_0, u(\cdot)) := \limsup_{t\to\infty} \frac{1}{t} \ln \|x(t, x_0, u(\cdot))\|.$$

where $x(t, x_0, u(\cdot))$ denotes the trajectory of (1.1).

Bilinear control systems arise e.g. by linearization of a nonlinear control system with a common fixed point x^* for all control values $u \in U$ with respect to x. They were first systematically studied by R. Mohler [18] in 1973. Lyapunov exponents were introduced by A.V. Lyapunov in 1892 (under the name of order numbers) as a tool to study nonlinear differential equations via their linearizations along trajectories. Recent results about the Lyapunov spectrum of families of time varying matrices (cfr. F. Colonius, W. Kliemann [11]) made it possible to characterize the domain of null controllability of bilinear systems using Lyapunov exponents (cfr. F. Colonius, W. Kliemann [10]). A basic property of the Lyapunov exponents is that $\lambda(x, u(\cdot)) < 0$ iff $x(t, x_0, u(t))$ converges to zero faster than any exponential e^{at} with $\lambda(x_0, u(\cdot)) < a < 0$. As an easy consequence $\inf_{u(\cdot)\in\mathcal{U}} \lambda(x_0, u(\cdot)) < 0$ implies that there exists a control function such that the corresponding trajectory converges to zero. The domain of null controllability — the set of all points x_0 with negative minimal Lyapunov exponent — may be only a part of $\mathbb{R}^d$ and as a consequence stabilization may only be possible for subsets of $\mathbb{R}^d$. Null controllability in this context always means asymptotical null controllability since the origin is not reachable in finite time from any other point of the state space. This implies that an approach via the mimimum time function (cfr. e.g. M. Bardi, M. Falcone [1]) does not apply here.

In contrast to the direct approach to this stabilization problem via Lyapunov functions (cfr. e.g. R. Chabour, G. Sallet, J. Vivalda [5]) the method developed here

†Institut für Mathematik, Universität Augsburg, Universitätsstr. 8, 86135 Augsburg, Germany, E-Mail: Lars.Gruene@Math.Uni-Augsburg.de

is in some sense an indirect approach:
First a numerical approximation of the extremal Lyapunov exponents of (1.1) is calculated. This enables us to characterize the stability properties of (1.1). Once this approximation is known we stabilize the system (i.e. we find control functions such that the corresponding trajectories converge to zero) by searching for control functions such that the corresponding Lyapunov exponent is close to the minimal exponent or at least negative. In §2 these problems are discussed in terms of optimal control theory. We show that the problem of calculating extremal Lyapunov exponents — which can be expressed as an average yield optimal control problem — can be approximated by discounted optimal control problems.

If we look at the uncontrolled system with $U = \{0\}$ it turns out that the Lyapunov exponents are just the real parts of the eigenvalues of A_0. Together with the corresponding eigenspaces they determine the stability properties of the system. For the controlled system we need suitable generalizations of eigenspaces associated with the Lyapunov exponents. The basic ideas of this concept are presented in §3 followed by an interpretation of the results of §2 in terms of calculating extremal Lyapunov exponents.
Section 4 presents algorithms to solve discounted optimal control problems numerically based on a discretization scheme by I. Capuzzo Dolcetta [2], [4] and M. Falcone [12], [13] connected to the framework of dynamic programming (cfr. [3]). Section 5 contains several numerical examples calculated with these algorithms.

2. Discounted and average cost optimal control problem. In this section we will show that average yield optimal control problems can be approximated by discounted optimal control problems.
Consider a control system on a *connected n-dimensional C^∞-manifold M* given by

$$\dot{x}(t) = X(x(t), u(t)) \quad \text{for all } t \in \mathbb{R} \tag{2.1}$$

$$x(0) = x_0 \in M \tag{2.2}$$

$$u(\cdot) \in \mathcal{U} := \{u : \mathbb{R} \to U \mid u \text{ measurable}\} \tag{2.3}$$

with

$$U \subseteq \mathbb{R}^m \text{ compact} \tag{2.4}$$

$$X(\cdot, u) \text{ is a } C^\infty\text{−vector field on } M, \text{ continuous on } M \times U \tag{2.5}$$

$$\text{for all } x \in M,\ u(\cdot) \in \mathcal{U} \text{ the trajectory } \varphi(t, x, u(\cdot)) \text{ exists for all } t \in \mathbb{R} \tag{2.6}$$

We now consider the following two optimal control problems given by the control system (2.1)–(2.6) and a cost function g satisfying

$$g : M \times U \to \mathbb{R} \text{ continuous on } M \times U \tag{2.7}$$

$$|g(x, u)| \le M_g \quad \forall (x, u) \in M \times U \tag{2.8}$$

The δ-discounted cost for $\delta > 0$ and the average cost are defined by

$$J_\delta(x, u(\cdot)) := \int_0^\infty e^{-\delta t} g(\varphi(t, x, u(\cdot)), u(t)) dt \tag{2.9}$$

$$J_0(x, u(\cdot)) := \limsup_{T \to \infty} \frac{1}{T} \int_0^T g(\varphi(t, x, u(\cdot)), u(t)) dt; \tag{2.10}$$

The associated optimal value functions are

$$v_\delta(x) := \inf_{u(\cdot)\in\mathcal{U}} J_\delta(x,u(\cdot)) \tag{2.11}$$

$$v_0(x) := \inf_{u(\cdot)\in\mathcal{U}} J_0(x,u(\cdot)). \tag{2.12}$$

A basic property of the discounted optimal value function is Bellman's optimality principle: for any $t>0$ we have

$$v_\delta(x) = \inf_{u(\cdot)\in\mathcal{U}} \left\{ \int_0^t e^{-\delta s} g(\varphi(s,x,u(\cdot)),u(s))ds \;+\; e^{-\delta t} v_\delta(\varphi(t,x,u(\cdot))) \right\} \tag{2.13}$$

For the average cost a similar estimate is valid: for any $t>0$ we have

$$v_0(x) = \inf_{u(\cdot)\in\mathcal{U}} \{v_0(\varphi(t,x,u(\cdot))\} \tag{2.14}$$

Results about the relation between discounted and average cost optimal control problems as the discount rate tends to 0 have been developed by F. Colonius [6] and F. Wirth [20]. Here we will first show the relation between the value of δJ_δ and J_0 along certain trajectories. Then we will use similar techniques as in [6] and [20] to obtain convergence results for the optimal value functions. The first theorem shows that J_0 is bounded if δJ_δ is bounded. Since J_0 has an infinite time horizon it is not sufficient that δJ_δ is bounded for the initial value. It has to be bounded for all $\varphi(t,x,u(\cdot))$, $t>0$ and the corresponding shifted control function.

THEOREM 2.1. (Approximation Theorem I) *Consider optimal control systems on M given by* (2.1)–(2.6) *and* (2.7)–(2.10)*, a discount rate $\delta>0$, $x\in M$, $u(\cdot)\in\mathcal{U}$, $C\in\mathbb{R}$, and $\alpha>0$, such that*
$\delta J_\delta(\varphi(t,x,u(\cdot)),u(t+\cdot)) \le C-\alpha$ *for all* $t\ge 0$*. Then*

$$J_0(x,u(\cdot)) < C.$$

Proof. We may assume $C=0$ by using $g-C$ instead of g. In the first step we show that for every $t>0$ there exists a $\tilde\tau(t)$, such that

$$\int_t^{\tilde\tau(t)} g(\varphi(s,x,u(\cdot)),u(s))ds \le -\frac{\alpha}{2\delta}. \tag{2.15}$$

Abbreviate $f(s) := e^{-\delta(s-t)} g(\varphi(s,x,u(\cdot)),u(s))$. Obviously there exists a $\tilde\tau(t)$ such that (2.15) is true for the shifted discounted functional $\int_t^{\tilde\tau(t)} f(s)ds \le -\frac{\alpha}{2\delta}$. Choose $\tilde\tau(t)$ minimal with this property. Since g is bounded there exist constants $a,b>0$ such that $\tilde\tau(t)-t\in[a,b]$ $\forall t>0$, $a=\frac{\alpha}{2\delta M_g}$. In case of $\int_t^{\tilde\tau(t)} f^+(s)ds = 0$ (2.15) is immediately implied.
In the case that $\int_t^{\tilde\tau(t)} f^+(s)ds > 0$ it follows that $\int_t^{\tilde\tau(t)} f^-(s)ds < -\frac{\alpha}{2\delta}$ and we can choose

$\gamma > 0$ maximal such that $\int_t^{t+\gamma} f^-(s)ds = -\frac{\alpha}{2\delta}$. Hence we have

$$\int_t^{\tau} f^+(s)ds - \int_{t+\gamma}^{\tau} f^-(s)ds > 0 \text{ for all } \tau \in [t+\gamma, \tau\tilde{(t)})$$

and

$$\int_t^{\tilde{\tau}(t)} f^+(s)ds - \int_{t+\gamma}^{\tilde{\tau}(t)} f^-(s)ds = 0.$$

Fixing $\varepsilon > 0$ we can define a monotone increasing sequence (τ_i), $i \in \mathbb{N}$ by $\tau_1 := t$, $\tau_2 := t+\gamma$,

$$\tau_{i+1} := \max\{\tau \in [\tau_i, \tilde{\tau}(t) \,|\, \int_{\tau_{i-1}}^{\tau_i} f^+(s)ds = \int_{\tau_i}^{\tau_{i+1}} f^-(s)ds\}.$$

From the construction of this sequence it follows that τ_i converges to $\tilde{\tau}(t)$ and we may truncate the sequence by choosing $k \in \mathbb{N}$ such that $|\tau_{k-1} - \tilde{\tau}(t)| < \varepsilon$ and set $\tau_k := \tilde{\tau}(t)$. Now we can estimate

$$\begin{aligned}
\int_t^{\tilde{\tau}(t)} g(\varphi(s,x,u(\cdot)),u(s))ds &= \int_t^{\tilde{\tau}(t)} e^{\delta(s-t)} f(s)ds \\
&\le \sum_{i=2}^{n-1} \left(\int_{\tau_{i-1}}^{\tau_i} e^{\delta(s-t)} f^+(s)ds - \int_{\tau_i}^{\tau_{i+1}} e^{\delta(s-t)} f^-(s)ds \right) + M_g\varepsilon - \frac{\alpha}{2\delta} \\
&\le \sum_{i=2}^{n-1} \left(\underbrace{\int_{\tau_{i-1}}^{\tau_i} e^{\delta(\tau_i-t)} f^+(s)ds - \int_{\tau_i}^{\tau_{i+1}} e^{\delta(\tau_i-t)} f^-(s)ds}_{=0} \right) + M_g\varepsilon - \frac{\alpha}{2\delta} \\
&= M_g\varepsilon - \frac{\alpha}{2\delta}
\end{aligned}$$

which proves (2.15) since $\varepsilon > 0$ was arbitrary.

To prove the theorem we first fix $T > 0$ and define a sequence $(\tilde{\tau}_i)$, $1 \le i \le k$ by $\tilde{\tau}_0 := 0$, $\tilde{\tau}_{i+1} := \tilde{\tau}(\tilde{\tau}_i)$, as long as $\tilde{\tau}(\tilde{\tau}_i) \le T$, $\tilde{\tau}_k := T$.
Then we have $a \le \tilde{\tau}_{i+1} - \tilde{\tau}_i \le b \;\; \forall i = 0, \dots, k-1$ and hence $\frac{T}{b} \le k \le \frac{T}{a}$. By definition of $\tilde{\tau}(t)$ it follows that $\int_{\tilde{\tau}_i}^{\tilde{\tau}_{i+1}} g(\varphi(t,x,u(\cdot)),u(t))dt < -\frac{\alpha}{2\delta} \;\; \forall i = 0, \dots, k-2$. This yields

$$\int_0^T g(\varphi(t,x,u(\cdot)),u(t))dt$$

$$= \sum_{i=0}^{k-2} \int_{\tilde{\tau}_i}^{\tilde{\tau}_{i+1}} g(\varphi(t,x,u(\cdot)),u(t))dt + \int_{\tilde{\tau}_{k-1}}^{\tilde{\tau}_k} g(\varphi(t,x,u(\cdot)),u(t))dt$$

$$\leq -\frac{k\alpha}{2\delta} + (\tilde{\tau}_k - \tilde{\tau}_{k-1})M_g \quad \leq \quad -\frac{T\alpha}{2b\delta} + bM_g \tag{2.16}$$

and as a conclusion

$$\limsup_{T\to\infty} \frac{1}{T} \int_0^T g(\varphi(t,x,u(\cdot)),u(t))dt \leq \limsup_{T\to\infty} \; -\frac{\alpha}{2b\delta} + \frac{bM_g}{T} = -\frac{\alpha}{2b\delta} < 0$$

which finishes the proof. □

Note that it is possible just to replace $\leq$ by $\geq$ and $-\alpha$ by $+\alpha$ to obtain the analogous result for a lower bound of J_0.

THEOREM 2.2. (Approximation Theorem II) *Consider optimal control systems on M given by* (2.1)–(2.6) *and* (2.7)–(2.10).
Assume there exists a control function $u(\cdot) \in \mathcal{U}$ such that $J_0(x,u(\cdot)) \leq C - \alpha$ for constants $C \in \mathbb{R}$, $\alpha > 0$. Then there exists a constant $R = R(x,u(\cdot),\alpha) > 0$ such that

$$\delta J_\delta(x,u(\cdot)) < C \quad \text{for all } \delta < R.$$

Proof. We may again assume $C = 0$. Hence it follows that there exists $T_0 \geq 0$ such that

$$\int_0^T g(\varphi(t,x,u(\cdot)),u(t))dt \leq T(-\frac{\alpha}{2}) \quad \forall T \geq T_0. \tag{2.17}$$

Now assume that $\delta J_\delta(x,u(\cdot)) \geq 0$ for arbitrarily small $\delta > 0$. The first step of the proof of Theorem 2.1 for the opposite inequality with $t = 0$ applied to $g + \frac{\alpha}{2}$ then yields that there exist arbitrarily large times $\tilde{T} > 0$ such that

$$\int_0^{\tilde{T}} g(\varphi(t,x,u(\cdot)),u(t))dt + \tilde{T}\frac{\alpha}{2} > 0$$

which contradicts (2.17). Hence the assertion follows. □

In contrast to the first Approximation Theorem here it is not possible simply to replace $\leq$ by $\geq$ and $-\alpha$ by $+\alpha$ to obtain a analogous result for the lower bound. Estimate (2.17) does only hold for the reverse inequality if in (2.10) the lim sup is replaced by the lim inf.

We will now combine these two theorems with controllability properties to obtain results about the relation between δv_0 and v_δ as δ tends to 0. To do this we first introduce some definitions.

DEFINITION 2.3. *The* positive orbit of $x \in M$ up to the time T *is defined by*

$$O_T^+(x) := \{y \in M \,|\, \text{there is } 0 \leq t \leq T \text{ and } u(\cdot) \in \mathcal{U}, \text{ such that } \varphi(t,x,u(\cdot)) = y\}.$$

The positive orbit *of $x \in M$ is defined by*

$$O^+(x) := \bigcup_{T\geq 0} O_T^+(x).$$

The negative orbits $O_T^-(x)$ and $O^-(x)$ are defined similarly by using the time reversed system.

For a subset $D \subset M$ we define $O_T^+(D) := \bigcup_{x\in D} O_T^+(x)$ and $O^+(D)$, $O_T^-(D)$, $O^-(D)$ analogously.

DEFINITION 2.4. *A subset $D \subseteq M$ is called a* control set*, if:*

(i) *$D \subseteq \overline{O^+(x)}$ for all $x \in D$*

(ii) *for every $x \in D$ there is $u(\cdot) \in \mathcal{U}$ such that the corresponding trajectory $\varphi(t, x, u(\cdot))$ stays in D for all $t \geq 0$*

(iii) *D is maximal with the properties (i) and (ii)*

A control set C is called invariant*, if*

$$\overline{C} = \overline{O^+(x)} \ \forall x \in C.$$

A non invariant control set is called variant*.*

In order to avoid degenerate situations we need the following setup:
Let $L = \mathcal{LA}\{X(\cdot, u),\, u \in U\}$ denote the Lie-algebra generated by the vector fields $X(\cdot, u)$. Let Δ_L denote the distribution generated by L in TM, the tangent space of M. Assume that

$$\dim \Delta_L(x) = \dim M \ \text{ for all } \ x \in M. \tag{2.18}$$

This assumption guarantees that the positive and negative orbit of any point $x \in M$ up to any time $T \neq 0$ have nonvoid interior. Note that the definition of control sets demands only approximate reachability (i.e. existence of controls steering into any neighbourhood of a given point); as a consequence of assumption (2.18) we have exact controllability in the interior of control sets, more precisely $\operatorname{int} D \subset O^+(x)$ for all $x \in D$.

The following proposition shows — as an extension of [7, Proposition 2.3] — that we have exact controllability in *finite time* on certain compact subsets:

PROPOSITION 2.5. *Consider a control system on M given by* (2.1)–(2.6) *and satisfying* (2.18)*.*
Let $D \subset M$ be a control set and consider compact sets $K_1 \subset O^-(D)$, $K_2 \subset \operatorname{int} D$. Then there exists a constant $r > 0$ such that for every $x \in K_1$, $y \in K_2$ there exists a control function $u(\cdot) \in \mathcal{U}$ with $\varphi(t_0, x, u(\cdot)) = y$ for some $t_0 \leq r$.

Proof. (i) We first show that for every $x \in K_1$, $z \in K_2$ there is an open neighborhood $U(x)$ such that all $y \in U(x)$ can be steered to z in bounded time t_0.
By (2.18) there is $T < \infty$ and $z_1 \in \operatorname{int} D \cap O^-_{\leq T}(z)$ and an open neighborhood $U(z_1) \subset \operatorname{int} D \cap O^-_{\leq T}(z)$. For $x \in K$ there exists a control $u(\cdot) \in \mathcal{U}$ and a time $t_1 < \infty$ such that $\varphi(t_1, x, u(\cdot)) = z_1$ (as a consequence of exact controllability in the interior of control sets). Since the solutions of the system depend continuously on the initial value, there is an open neighborhood $U(x)$ whith $\varphi(t_1, x_1, u(\cdot)) \in U(z_1)$ for all $x_1 \in U(x)$. Putting this together yields $U(x) \subset O^-_{\leq t_1+T}(y)$ which proofs the assertion with $t_0 \leq t_1 + T$.

(ii) For $x \in K_1$, $y \in K_2$ we now show that there exists a time $t_y < \infty$ such that all z in some open neighborhood of y can be reached from x in time t_y.
Let $x_1 \in \operatorname{int} D$ and $u_1(\cdot) \in \mathcal{U}$, $t_1 < \infty$ such that $\varphi(t_1, x, u(\cdot)) = x_1$ (the existence of x_1, $u_1(\cdot), t_1$ follows from (2.18)). Again by (2.18) there exists $T < \infty$ and $y_1 \in \operatorname{int} D \cap \mathcal{O}^-_{\leq T}(x_1)$, let $U(y_1)$ be an open neighborhood of y_1 contained in $\operatorname{int} D \cap \mathcal{O}^-_{\leq T}(x_1)$. Now because of the exact controllability there exists $u_2(\cdot) \in \mathcal{U}$, $t_2 < \infty$ with $\varphi(t_2, y_1, u_2) = y$. Since the solution of the control system using the control $u_2(\cdot)$ defines a semigroup

of homeomorphisms on M, the open neighborhood $U(y_1)$ is mapped onto some open neighborhood $U(y)$ and $U(y) \subset \mathcal{O}^+_{\leq t_1+T+t_2}(x)$. This means that all $z \in U(y)$ can be reached from x in time $t_y = t_1 + T + t_2$.

(iii) Because of the compactness of K_1 and K_2 now the proof of the Proposition follows. □

The following proposition summarizes the consequences of these controllability properties for the optimal value functions.

PROPOSITION 2.6. *Consider optimal control systems on M given by* (2.1)–(2.6), (2.7)–(2.10) *and satisfying* (2.18).
Let $D \subset M$ be a control set and consider compact sets $K_1 \subset O^-(D)$, $K_2 \subset \mathrm{int}D$.
Then the following estimates hold:

(i) $v_0(x) = v_0(y)$ for all $x, y \in \mathrm{int}D$
(ii) $v_0(x) \leq v_0(y)$ for all $x \in O^-(D)$, $y \in \mathrm{int}D$
(iii) $|\delta v_\delta(x) - \delta v_\delta(y)| \leq \varepsilon(\delta)$ for all $x, y \in K_2$
(iv) $\delta v_\delta(x) \leq \delta v_\delta(y) + \varepsilon(\delta)$ for all $x \in K_1$, $y \in K_2$

and $\varepsilon(\delta) \to 0$ as δ tends to 0.

Proof. Just combine (2.13) and (2.14) with the controllability properties stated above. □

Now we can formulate the results about the relation between the optimal value functions.

PROPOSITION 2.7. *Consider optimal control systems on M given by* (2.1)–(2.6), (2.7)–(2.10) *and satisfying* (2.18). *Then*

$$\limsup_{\delta \to 0} \delta v_\delta(x) \leq v_0(x) \text{ for all } x \in M.$$

Proof. Fix $\varepsilon > 0$. Choose a control function $u(\cdot)$ such that $|v_0(x) - J_0(x, u(\cdot))| \leq \frac{\varepsilon}{2}$. Using Theorem 2.2 with $\alpha = \frac{\varepsilon}{2}$ yields a $R_1 > 0$ such that for all $\delta \in (0, R_1]$: $\delta v_\delta(x) \leq \delta J_\delta(x, u(\cdot)) \leq J_0(x, u(\cdot)) + \frac{\varepsilon}{2} \leq v_0(x) + \varepsilon$. It follows that $\limsup_{\delta \to 0} \delta v_\delta(x) \leq v_0(x)$ since $\varepsilon > 0$ was arbitrary. □

PROPOSITION 2.8. *Consider optimal control systems on M given by* (2.1)–(2.6), (2.7)–(2.10) *and satisfying* (2.18).
Let $D \subseteq M$ be a control set. Then for every compact $Q \subset \mathrm{int}D$ and every $\varepsilon > 0$ there exists a $R_0 > 0$ such that

$$\delta v_\delta(x) \leq v_0(x) + \varepsilon \text{ for all } \delta \in (0, R_0], \ \ x \in Q.$$

Proof. Fix $x_0 \in Q$. Using Proposition 2.7 we know that there exists a constant $R_1 > 0$ such that $\delta v_\delta(x_0) \leq v_0(x_0) + \frac{\varepsilon}{2}$ for all $\delta \in (0, R_1]$. Now choose $R_2 > 0$ such that Proposition 2.6, (iii) holds with $\varepsilon(\delta) < \frac{\varepsilon}{2}$ for $\delta < R_2$. Since v_0 is constant on Q now the assertion holds for all $x \in Q$ with $R_0 := \min\{R_1, R_2\}$. □

LEMMA 2.9. **(Pointwise convergence)** *Consider optimal control systems on M given by* (2.1)–(2.6) *and* (2.7)–(2.10).
Assume there exists $x \in M$, $R \in \mathbb{R}$ and a set $B \subset M$ such that $\delta v_\delta(y) \leq \delta v_\delta(x) + \alpha(\delta)$ for all $y \in B$, $\delta \in (0, R]$ and constants $\alpha(\delta) \geq 0$. Assume there exist optimal controls $u_\delta(\cdot) \in \mathcal{U}$ for all $\delta \in (0, R]$ such that $\varphi(t, x, u_\delta(\cdot)) \in B$ for all $t \geq 0$.
Then for every $\varepsilon > 0$ there exists $R_0 > 0$ such that

$$|\delta v_\delta(x) - v_0(x)| \leq \max\{\varepsilon, \alpha(\delta)\}, \quad \text{for all } \delta \in (0, R_0].$$

In particular if $\alpha(\delta) \to 0$ *as* $\delta \to 0$ *the convergence* $\delta v_\delta(x) \to v_0(x)$ *is implied.*

Proof. From Theorem 2.1 it is clear that $v_0(x) \le \delta v_\delta(x) + \alpha(\delta)$ for all $\delta < R$. Now choose a control function $u(\cdot)$ such that $|v_0(x) - J_0(x, u(\cdot))| \le \frac{\varepsilon}{2}$. Using Theorem 2.2 with $\alpha = \frac{\varepsilon}{2}$ yields a $R_0 > 0$ such that for all $\delta < R_0$: $\delta v_\delta(x) \le \delta J_\delta(x, u(\cdot)) \le J_0(x, u(\cdot)) + \frac{\varepsilon}{2} \le v_0(x) + \varepsilon$. Combining these inequalities finishes the proof. □

Using the estimate of proposition 2.6 two results on uniform convergence can be obtained.

THEOREM 2.10. (Uniform convergence) *Consider optimal control systems on* M *given by* (2.1)–(2.6), (2.7)–(2.10) *and satisfying* (2.18).
Let $D \subseteq M$ *be a control set and assume there exist* $x_0 \in \mathrm{int}D$, *a compact subset* $K \subseteq D$ *and optimal controls* $u_\delta(\cdot)$, *such that*

$$\varphi(t, x_0, u_\delta(\cdot)) \in K, \quad \text{for all } t \ge 0, \text{ for all } \delta \in (0, R]$$

for some constant $R > 0$. *Then*

$$\delta v_\delta \to v_0 \text{ uniformly on compact subsets of } \mathrm{int}D.$$

Proof. By Proposition 2.6, (iii) on any compact subset Q of int$D$ we have $|\delta v_\delta(x) - \delta v_\delta(y)| \le \varepsilon(\delta) \to 0$ uniformly for all $x, y \in Q$ as δ tends to 0. By Proposition 2.6, (iv) x_0 and K fulfill the conditions of Lemma 2.9 with $\alpha(\delta) = \varepsilon(\delta)$ since $K \subseteq D \subseteq O^-(D)$. Hence pointwise convergence follows. Since v_0 is constant on int$D$ uniform convergence on Q follows. □

THEOREM 2.11. (Uniform convergence in compact invariant control sets) *Consider optimal control systems on* M *given by* (2.1)–(2.6), (2.7)–(2.10) *and satisfying* (2.18).
Let $C \subseteq M$ *be a compact invariant control set. Then for* $\delta \to 0$

(i) $\delta v_\delta(x) \to v_0(x)$ for all $x \in \mathrm{int}C$
(ii) $\delta v_\delta \to v_0$ *uniformly on compact subsets of* int$C$
(iii) *if* M *is compact and* C *is the unique invariant control set we have* $\sup_{x \in M} \delta v_\delta(x) \to \sup_{x \in M} v_0(x)$

Proof. Since C is a compact subset of C and no trajectory can leave C the conditions of Theorem 2.10 (with $K = C$) are fulfilled. Hence the assertions (i) and (ii) follow.

If M is compact and C is the unique invariant control set it follows that $O^-(C) = M$ [16, Proof of Lemma 2.2 (i)].

From Proposition 2.6, (ii) and (iv) and the compactness of $M = O^-(C)$ it follows for any compact subset $Q \subset \mathrm{int}C$ that $v_0(x) \le v_0(y)$ and $\delta v_\delta(x) \le \delta v_\delta(y) + \varepsilon(\delta)$ for all $x \in M$, $y \in Q$. Since we have uniform convergence on Q the assertion (iii) is proved. □

Remark 2.12. Note that these results are not valid in general for the corresponding maximization problems, since the second Approximation Theorem is not valid for the reverse inequality. However some of the results remain valid and others are valid under additional conditions:

(i) The application of the results to the maximization problems is possible if the lim sup in (2.10) can be replaced by a lim inf without changing the value of v_0. This is possible if there exist approximately optimal trajectories and controls — with respect to the maximization problem — such that the lim sup is a limit. From [20, proof of Proposition 1.4, (a)] it is clear that this is the fact if there exist approximately optimal trajectories and controls which are periodic. A sufficient condition for this is that there

exists an optimal trajectory that stays inside some compact subset $K \subset \mathrm{int}D$ (cfr. [20, Proposition 2.7]).

(ii) Adding this condition to the assumptions of Theorem 2.10 we obtain Theorem 2.10 from F. Wirth [20] under the weaker condition that the optimal trajectories with respect to the discounted problems stay inside a compact subset of a control set instead of a compact subset of the *interior* of a control set.

(iii) For invariant control sets C we can use [7, Corollary 4.3] to conclude that for any initial value $x_0 \in \mathrm{int}C$ there exist approximately optimal periodic control functions and trajectories. Hence Theorem 2.11 remains valid for the maximization problem without any additional assumptions.

3. Lyapunov exponents of bilinear control systems. We will now return to the *bilinear control systems* in $\mathbb{R}^d$, i.e. systems of the form

$$\dot{x}(t) = \Big(A_0 + \sum_{i=1}^{m} u_i(t)A_i\Big)x(t), \quad x(0) = x_0 \in \mathbb{R}^d \setminus \{0\} \tag{3.1}$$

with $A_j \in \mathbb{R}^{d\times d}$, $j = 0,\dots,m$, $u(\cdot) \in \mathcal{U} := \{u : \mathbb{R} \to U,\ u \text{ measurable}\}$ with a compact and convex set of control values $U \subset \mathbb{R}^m$ with nonvoid interior.

We denote the unique trajectory for any initial value $x_0 \in \mathbb{R}^d$ and any control function $u(\cdot) \in \mathcal{U}$ by $x(t, x_0, u(\cdot))$.

In order to characterize the exponential growth rate of the solutions of (3.1) we define the Lyapunov exponent of a solution by

$$\lambda(x_0, u(\cdot)) := \limsup_{t\to\infty} \frac{1}{t} \ln \|x(t, x_0, u(\cdot))\|. \tag{3.2}$$

The minimal Lyapunov exponent with respect to $x_0 \in \mathbb{R}^n \setminus \{0\}$ ist defined by

$$\lambda^*(x_0) := \inf_{u(\cdot)\in\mathcal{U}} \lambda(x_0, u(\cdot)) \tag{3.3}$$

and the extremal Lyapunov exponents of the control system by

$$\kappa^* := \inf_{x_0\neq 0} \inf_{u(\cdot)\in\mathcal{U}} \lambda(x_0, u(\cdot)) \tag{3.4}$$

$$\kappa := \sup_{x_0\neq 0} \sup_{u(\cdot)\in\mathcal{U}} \lambda(x_0, u(\cdot)) \tag{3.5}$$

$$\tilde{\kappa} := \sup_{x_0\neq 0} \inf_{u(\cdot)\in\mathcal{U}} \lambda(x_0, u(\cdot)) \tag{3.6}$$

The Lyapunov exponent can be interpreted as a measure for the exponential growth of trajectories. Our aim is to calculate numerical approximations of the minimal and maximal Lyapunov exponents with respect to the initial values. If $\lambda^*(x_0) < 0$ the system can be steered asymptotically to the origin from x_0. Using the approximation of the Lyapunov exponents we then are able to calculate controls that stabilize the system.

For a bilinear control system (3.1) the following identity is obvious:

$$\lambda(x_0, u(\cdot)) = \lambda(\alpha x_0, u(\cdot)) \text{ for all } x_0 \in \mathbb{R}^d \setminus \{0\},\ \alpha \in \mathbb{R} \setminus \{0\},\ u \in \mathcal{U}.$$

Due to this observation we can identify all $x \neq 0$ lying on a straight line through the origin. Hence it is sufficient to consider initial values s_0 in $\mathbb{P}^{d-1}$, the real projective

space. To calculate the Lyapunov exponents we can project the system onto the unit sphere $\mathbb{S}^{d-1}$ via $s_0 := x_0/\|x_0\|$. This yields the projection onto $\mathbb{P}^{d-1}$ by identifying opposite points. A simple application of the chain rule shows that the projected system can be written as

$$\dot{s}(t) = h_0(s(t)) + \sum_{i=1}^{m} u_i(t) h_i(s(t)) \tag{3.7}$$

where

$$h_i(s) = [A_i - s^t A_i s \cdot \mathrm{Id}]s \quad \forall i = 0, \ldots, m.$$

The Lyapunov exponent (3.2) with respect to $s_0 = x_0/\|x_0\|$ can be written as

$$\lambda(x_0, u(\cdot)) = \lambda(s_0, u(\cdot)) = \limsup_{t\to\infty} \frac{1}{t} \int_0^t q(s(\tau, s_0, u(\cdot)), u(\tau)) d\tau \tag{3.8}$$

where

$$q(s, u) = s^t \Big(A_0 + \sum_{i=0}^{m} u_i A_i \Big) s. \tag{3.9}$$

We recall some facts about projected bilinear control systems and their Lyapunov exponents.

For the projected bilinear system assumption (2.18) reads

$$\dim \Delta_L(p) = d - 1 \text{ for all } p \in \mathbb{P}^{d-1}, \quad L = \mathcal{LA}\{h(\cdot, u), u \in U\}$$

where $h(\cdot, u) := h_0(\cdot) + \sum_{i=1}^{m} u_i h_i(\cdot)$. Under this assumption the following facts hold (cfr. [9, Corollary 4.4], [8, Theorem 3.10]):

If κ_1 denotes the maximal Lyapunov exponent of the original system and κ_2^* the minimal exponent of the time reversed system the identity $\kappa_1 = -\kappa_2^*$ holds.

For the projected system there exist k control sets with nonvoid interior where $1 \le k \le d$. These are called the main control sets. They are linearly ordered by
$D_i < D_j \quad \Leftrightarrow \quad$ there exists $p_i \in D_i$, $p_j \in D_j$, $t > 0$ and $u(\cdot) \in \mathcal{U}$ such that $\varphi(t, p_i, u) = p_j$.

The control set D_1 is open, the control set $C := D_k$ is closed and invariant. All other control sets are neither open nor closed. Furthermore we have $O^-(p) = \mathbb{P}^{d-1}$ for all $p \in \mathrm{int}C$.

The linear order of the control sets implies a linear order on the minimal Lyapunov exponents (which can easily be proved using Proposition 2.6):
$\lambda^*(p_i) \le \lambda^*(p_j)$ for $p_i \in D_i$, $p_j \in D_j$ and $i < j$.
Furthermore $\lambda^*(p)$ is constant on the interior of control sets.

Under the following condition there is a stronger relation between the control sets of the projected and the Lyapunov exponents of the bilinear system: Considering the set of control values $\rho U := \{\rho u \,|\, u \in U\}$ for $\rho \ge 0$ and the corresponding set of control functions $\mathcal{U}^\rho$ we assume the following ρ–ρ' inner pair condition:

> For all $0 \le \rho \le \rho'$ and all $(u(\cdot), p) \in \mathcal{U}^\rho \times \mathbb{P}^{d-1}$ there exist $T > 0$ and $S > 0$ such that $\varphi(T, p, u(\cdot)) \in \mathrm{int} O^{\rho'+}_{S+T}(p)$ (the positive orbit corresponding to $\mathcal{U}^{\rho'}$)

Let D^ρ be a main control set corresponding to $\mathcal{U}^\rho$. We define the *Lyapunov spectrum of (3.1) over* $\overline{D^\rho}$ by

$$\Sigma_{Ly}^\rho(\overline{D^\rho}) := \{\lambda(p,u(\cdot)) \,|\, \varphi(t,p,u) \in \overline{D^\rho} \text{ for all } t \geq T \text{ for some } T \geq 0\}$$

and the *Lyapunov spectrum of (3.1) by*

$$\Sigma_{Ly}^\rho := \{\lambda(p,u(\cdot)) \,|\, u(\cdot) \in \mathcal{U},\, p \in \mathbb{P}^{d-1}\}.$$

Under the ρ–ρ' inner pair condition we know that

$$\Sigma_{Ly} = \bigcup_{i=1}^{k(\rho)} \Sigma_{Ly}(\overline{D_i^\rho}) \tag{3.10}$$

for all exept at most countably many $\rho < \rho'$, where $k(\rho)$ is the number of main control sets D_i^ρ corresponding to $\mathcal{U}^\rho$ ([11, Corollary 5.6])

Furthermore $\Sigma_{Ly}^\rho(\overline{D_i^\rho})$ are closed intervals and thus it is sufficient to calculate the minima and the maxima of $\Sigma_{Ly}(\overline{D_i^\rho})$ to obtain the whole Lyapunov spectrum of the system. These maxima and minima can be approximated by periodic trajectories with initial values in $intD_i^\rho$.

In the case $d = 2$ these results hold for all $\rho > 0$ without assuming the ρ–ρ' inner pair condition ([11, Corollary 4.9]).

We will now give an interpretation of the results of §2 in terms of calculating Lyapunov exponents and stabilization. Since we are going to solve the discounted optimal control problem numerically we cannot expect to calculate optimal control functions but only ε-optimal control functions. We call a control function $u_x(\cdot) \in \mathcal{U}$ *uniformly ε-optimal* with respect to $x \in M$ iff $|\delta J_\delta(\varphi(t,x,u_x(\cdot)),u_x(t+\cdot)) - \delta v_\delta(\varphi(t,x,u_x(\cdot)))| < \varepsilon$ for all $t \geq 0$.

THEOREM 3.1. *Consider a bilinear control system* (3.1) *and the related optimal control system on* $\mathbb{P}^{d-1}$ *given by* (3.7) *and* (3.8) *with cost function* q *from* (3.9). *Assume* (2.18) *is satisfied.*
Let

$$v_\delta(x) := \inf_{u(\cdot)\in\mathcal{U}} J_\delta(x,u(\cdot)) \quad \text{and} \quad \bar{v}_\delta(x) := \sup_{u(\cdot)\in\mathcal{U}} J_\delta(x,u(\cdot)).$$

Then the following estimates hold with $\varepsilon \to 0$ as δ tends to 0.

(i) $\delta v_\delta(x) \leq \lambda^*(x) + \varepsilon$ for all $x \in M$

(ii) $\delta v_\delta(x) \leq \lambda^*(x) + \varepsilon$ *uniformly on compact subsets Q of the interior of control sets*

(iii) $|\delta v_\delta(x) - \lambda^*(x)| \leq \varepsilon$ *uniformly on compact subsets Q of the interior of control sets under the conditions of Theorem* 2.10

(iv) $|\delta v_\delta(x) - \lambda^*(x)| \leq \varepsilon$ *uniformly on compact subsets Q of the interior of the invariant control set*

(v) $\sup_{x\in M} \delta v_\delta(x) \to \tilde{\kappa}$ *as δ tends to 0*

(vi) $\inf_{x\in M} \delta \bar{v}_\delta(x) \to \kappa$ *as δ tends to 0.*

(vii) *If $\tilde{\kappa} < 0$ and $u_s(\cdot)$ is uniformly ε-optimal with respect to s then $\varphi(t,x,u_s(\cdot))$ is asymptotically stable for all $x \in \mathbb{R}^d$ with $s = \frac{x}{\|x\|}$ provided δ and ε are sufficiently small.*

(viii) *If $\lambda^* < 0$ in the interior of some control set D and $u_s(\cdot)$ is uniformly ε-optimal with respect to s and $\varphi(t,s,u_s(\cdot))$ stays inside a compact subset of $O^-(D)$ for all times then $\varphi(t,x,u_s(\cdot))$ is asymptotically stable for all $x \in \mathbb{R}^d$ with $s = \frac{x}{\|x\|}$ provided δ and ε are sufficiently small.*

Proof. All assertions follow directly from the results in §2. Assertion (iv) is true since the invariant control set of the projected system is compact. Assertions (v) and (vi) are proved using the fact that the projective space is compact and that there exists a unique invariant control set for the projected system. □

Remark 3.2. Knowing the facts cited in this section we can see that even more can be calculated:

(i) Property (vi) can be used to calculate κ^* by calculating κ of the time reversed system. Hence it is possible to approximate κ, κ^* and $\tilde{\kappa}$ for any bilinear control system satisfying (2.18) by solving discounted optimal control problems.

(ii) For all main control sets D_i we can approximate the minimal Lyapunov exponent over $intD_i$ as follows: Proposition 2.8 yields that $\delta v_\delta < \lambda^* + \varepsilon$ uniformly on compact subsets of $intD_i$. If we find control functions as described in Theorem 3.1 (viii) for $\varepsilon > 0$ we know that there exists a Lyapunov exponent $\lambda^* < \delta v_\delta + \varepsilon$, hence $\lambda^* \in [\delta v_\delta - \varepsilon, \delta v_\delta + \varepsilon]$. However, the existence of such control functions is not guaranteed; nevertheless for all examples discussed in §5 it was possible to find them.

(iii) For systems with $d = 2$ or systems with $d > 2$ satisfying the ρ–ρ' inner pair condition we are also able to compute $\Sigma^\rho_{Ly}(\overline{D})$ for $D = C$ and $D = D_1$ at least for all but countably many $\rho > 0$, since in this cases the upper and lower bounds of this intervalls coincide with κ and $\tilde{\kappa}$ of the original or of the time reversed system, respectively. For all other main control sets we can apply the technique from (ii) to both the original and the time reversed system to calculate $\Sigma^\rho_{Ly}(\overline{D_i})$.

(iv) In the case that $d > 2$ and $\rho > 0$ is one of the (at most countably many) exeptional points of the spectrum (3.10) we can use the monotonicity of v_δ and Σ^ρ_{Ly} in ρ. This implies that there exist values $\rho_1 < \rho < \rho_2$ arbitrarily close to ρ such that the approximated spectrum contains $\Sigma^{\rho_1}_{Ly}$ and is contained in $\Sigma^{\rho_2}_{Ly}$.

4. Numerical solution of the discounted optimal control problem. A discretization scheme to solve discounted optimal control problems in $\mathbb{R}^n$ has been developed by I. Capuzzo Dolcetta and M. Falcone [2], [3], [4], [12], [13]. The algorithm used here to solve these problems is based on this discretization. We will first describe this discretization scheme and then present the modifications for our case, where the system is given on $\mathbb{P}^{d-1}$ instead of $\mathbb{R}^n$.

Hence we first assume that we have a discounted optimal control problem defined by (2.1)–(2.6) and (2.8) with $M = \mathbb{R}^n$. In addition we need the following conditions on X und g:

$$\|X(x,u) - X(y,u)\| \le L_X \|x - y\| \quad \forall x, y \in W \ \forall u \in U \text{ for a } L_X \in \mathbb{R} \tag{4.1}$$

$$\|X(x,u)\| \le M_X \quad \forall (x,u) \in W \times U \text{ for a } M_X \in \mathbb{R} \tag{4.2}$$

$$|g(x,u) - g(y,u)| \le L_g \|x - y\| \quad \forall x, y \in W \ \forall u \in U \text{ for a } L_g \in \mathbb{R} \tag{4.3}$$

The δ discounted cost functional J_δ and the optimal value function v_δ are defined as in (2.9) and (2.11).

Under the assumptions made above the value function v_δ satisfies

$$|v_\delta(x)| \le \frac{M_g}{\delta} \quad \text{and} \quad |v_\delta(x) - v_\delta(y)| \le C|x - y|^\gamma \tag{4.4}$$

for all x, $y \in \mathbb{R}^n$ (cfr. [4], the second estimate can be proved by using [4, Lemma 4.1]). For small $\delta > 0$ we have $C = \frac{M}{\delta}$ for a constant M independent on δ and γ is a constant satisfying $\gamma = 1$ for $\delta > L_X$, $\gamma = \frac{\delta}{L_X}$ for $\delta < L_X$ and $\gamma \in (0,1)$ arbitrary for $\delta = L_X$.

Furthermore (cfr. [17]) v_δ is the unique bounded and uniformly continuous viscosity solution of the Hamilton-Jacobi-Bellman equation

$$\sup_{u\in U}\{\delta v_\delta(x_0) - g(x_0,u) - Dv_\delta(x_0)X(x_0,u)\} = 0 \tag{4.5}$$

The first discretization step is a discretization in time. By replacing Dv_δ by the difference quotient with time step h one obtains

$$\sup_{u\in U}\{v_h(x) - \beta v_h(x + hX(x,u)) - hg(x,u)\} = 0 \tag{4.6}$$

with $\beta := 1 - \delta h$.

It turns out that the unique bounded solution of this equation is the optimal value function v_h of the *discretized optimal control system* with respect to the space $\mathcal{U}_h$ of all controls constant on each intervall $[jh, (j+1)h)$, $j \in \mathbb{N}$:

$$x_0 := x, \quad x_{j+1} := x_j + hX(x_j, u_j), \quad j = 0,\, 1,\, 2, \dots , \tag{4.7}$$

with running cost

$$J_h(x, u(\cdot)) := h\sum_{j=0}^{\infty} \beta^j g(x_j, u_j).$$

Furthermore for all $p \in \mathbb{N}$ v_h satisfies

$$v_h(x) = \inf_{u(\cdot)\in\mathcal{U}_h} \left\{ h\sum_{j=0}^{p-1} \beta^j g(x_j, u_j) + \beta^p v_h(x_p) \right\} \tag{4.8}$$

and the estimates (4.4) also apply to v_h.

The discretization error can be estimated as follows ([4, Theorem 3.1]):

$$\sup_{x\in\mathbb{R}^n} |(v_\delta - v_h)(x)| \le Ch^{\frac{\gamma}{2}} \tag{4.9}$$

for all $h \in (0, \frac{1}{\delta})$. Here we have $C = \frac{M}{\delta^2}$ for small $\delta > 0$ and γ is the constant from (4.4).

The discretization error of the functionals for any $u(\cdot) \in \mathcal{U}_h$ can be estimated as

$$\sup_{x\in\mathbb{R}^n,\, u(\cdot)\in\mathcal{U}_h} |J_h(x, u(\cdot)) - J_\delta(x, u(\cdot))| \le Ch^\gamma \tag{4.10}$$

where $C = \frac{M}{\delta}$ for small $\delta > 0$ and γ as above ([4, Lemma 4.1]).

In order to reduce (4.6) to a finite dimensional problem we apply a finite difference technique. To do this we assume the existence of an open, bounded and convex subset Ω of the state space $\mathbb{R}^n$ which is invariant for (2.1). Thus a triangulation of Ω into a finite number P of simplices S_j with N nodes x_i can be constructed (cfr. [13, Proposition 2.5]) such that $\Omega^k := \cup_{j=1,\dots,P} S_j$ is invariant with respect to the discretized trajectories (4.7). Here $k := \sup\{\|x - y\| \,|\, x \text{ and } y \text{ are nodes of } S_j,\, j = 1 \dots, P\}$. We are now looking for a solution of (4.6) in the space of piecewise affine functions $\mathcal{W} := \{w \in C(\Omega^k) \mid Dw(x) = c_j \text{ in } S_j\}$.

Every point $x_i + hf(x_i, u)$ can be written as a convex combination of the nodes of the simplex containing it with coefficients $\lambda_{ij}(u)$. Let $\Lambda(u) := [\lambda_{ij}(u)]_{i,j=1,\dots,N}$ be

the matrix containing this coefficients and $G(u) := [g(x_i, u)]_{i=1,\dots,N}$ an N-dimensional vector containing the values of g with control value u at the nodes of the triangulation. Now we can rewrite (4.6) as a fixed point equation

$$V = T_h^k(V), \quad T_h^k(V) := \inf_{u\in U} \Big(\beta\Lambda(u)V + hG(u)\Big) \tag{4.11}$$

It follows that T_h^k is a contraction in $\mathbb{R}^N$ with contraction factor $\beta := 1-\delta h$ and therefore has a unique fixed point V^*. If v_h^k denotes the function obtained by $v_h^k(x_i) := [V^*]_i$ and linear interpolation between the nodes the discretization error can be estimated by

$$\sup_{x\in\Omega^k} |(v_h^k - v_h)(x)| \le C\frac{k^\gamma}{h} \tag{4.12}$$

with γ as in (4.4) and $C = \frac{M}{\delta^2}$ for small $\delta > 0$ (cfr. [13, corrigenda]).

For the whole discretization error we obtain the following estimate:

$$\sup_{x\in\Omega^k} |(v_h^k - v_\delta)(x)| \le C(h^{\frac{\gamma}{2}} + \frac{k^\gamma}{h}) \tag{4.13}$$

with the constants from (4.9) and (4.12).

Remark 4.1. These results have been improved by R.L.V. Gonzales and M.M. Tidball. From [14, Lemma 3.4] in connection with [4, Lemma 4.1] it follows that

$$\sup_{x\in\Omega^k} |(v_h^k - v_h)(x)| \le C\left(\frac{k}{\sqrt{h}}\right)^\gamma, \tag{4.14}$$

[14, Theorem 3.1] yields

$$\sup_{x\in\Omega^k} |(v_h^k - v_\delta)(x)| \le C\left(\sqrt{h} + \frac{k}{\sqrt{h}}\right)^\gamma \tag{4.15}$$

with similar constants C and γ.

Remark 4.2. Note that the convergence becomes slow if the discount rate δ becomes small. For the approximation of the average cost functional as described in §2 it is nevertheless necessary to calculate v_δ for small $\delta > 0$. This means that for this purpose we need a fine discretization in time and space to get reliable results.

If one uses estimate (4.13) we obtain as an additional condition that k should be smaller that h, using (4.15) convergence for the case $k = h$ is guaranteed.

To handle the optimal control problem on $\mathbb{P}^{d-1}$ we use the following modifications on this scheme:

We first consider the optimal control problem on $\mathbb{S}^{d-1}$ defined by the projected system (3.7). The optimal value function v_δ then again satisfies (4.4) and is the unique bounded and uniformly continuous viscosity solution of (4.5). This can be proved exactly the same way as in the $\mathbb{R}^n$ case by using the metric on $\mathbb{S}^{d-1}$ induced by the norm on $\mathbb{R}^d$.

We have seen that the discretization in time of (4.5) corresponds to the Euler discretization of the control system. Hence here we use the following Euler method on $\mathbb{S}^{d-1}$; for $h > 0$ and any $s \in \mathbb{S}^{d-1}$ we define

$$\Phi_h(s,u) := \frac{s + hX(s,u)}{\|s + hX(s,u)\|} \tag{4.16}$$

i.e. we perform an Euler step in $\mathbb{R}^d$ and project the solution back to $\mathbb{S}^{d-1}$. With this (4.6) reads

$$\sup_{u\in U}\{v_h(s) - \beta v_h(\Phi_h(s,u)) - hg(s,u)\} = 0. \tag{4.17}$$

and (4.7) translates to

$$s_0 := s, \quad s_{j+1} := \Phi_h(s_j, u), \quad j = 0,\, 1,\, 2, \ldots \tag{4.18}$$

The estimates (4.8)–(4.10) remain valid; again all proofs from the $\mathbb{R}^n$ case apply by using the metric on $\mathbb{S}^{d-1}$ induced by the norm on $\mathbb{R}^d$.

We will now use the fact that this discrete time control system on $\mathbb{S}^{d-1}$ defines a (well defined) control system on $\mathbb{P}^{d-1}$ by identifying s and $-s$ on $\mathbb{S}^{d-1}$. Let $W \subset \mathbb{S}^{d-1}$ be an open set in $\mathbb{S}^{d-1}$ such that it contains the upper half of the sphere. Any discrete time trajectory $(s_i)_{i\in\mathbb{N}_0} \subset \mathbb{S}^{d-1}$ as defined in (4.18) can be mapped on a trajectory $(\tilde{s}_i)_{i\in\mathbb{N}_0} \subset W$ by $\tilde{s}_i := s_i$ if $s_i \in W$, $\tilde{s}_i := -s_i$ if $s_i \notin W$. Since $X(s,u) = -X(-s,u)$ this mapping is well defined and $g(s,u) = g(-s,u)$ implies that v_h does not change if we only consider trajectories in W. Hence we can define a discrete time optimal control problem on W via

$$\tilde{\Phi}_h(s) = \begin{cases} \Phi_h(s), & \Phi_h(s) \in W \\ -\Phi_h(s), & \Phi_h(s) \notin W \end{cases}$$

without changing v_h.

To obtain a region $\Omega \subset \mathbb{R}^{d-1}$ suitable for the space discretization we use a parametrization Ψ of $\mathbb{S}^{d-1}$ which is invertible on W such that Ψ^{-1} maps W to an open and bounded set $\Omega \subset \mathbb{R}^{d-1}$. (The parametrizations used in our examples are given in §5.) Now we can project the system on W to a system on Ω and compute v_h on Ω. The system on Ω is then given by

$$\Phi_{h,\Omega}(x,u) := \Psi^{-1}(\tilde{\Phi}_h(\Psi(x),u)), \quad g_\Omega(x,u) := g(\Psi(x),u)$$

and by definition of $\tilde{\Phi}_h$ the set Ω is invariant for this discrete time system. We can rewrite (4.17) by using $\Phi_{h,\Omega}$ and g_Ω and denoting the solution by $v_{h,\Omega}$. This solution satisfies $v_h(\Psi(x)) = v_{h,\Omega}(x)$ and since Ψ is Lipschitz continuous estimate (4.4) remains valid for $v_{h,\Omega}$.

Thus we can proceed as in the $\mathbb{R}^n$ case described above. Keeping in mind that there exists a one-to-one relation between the system on W and the system on Ω we can simplify the notation by writing Φ_h, g and v_h instead of $\Phi_{h,\Omega}$, g_Ω and $v_{h,\Omega}$.

We will now turn to the problem how the fixed point equation (4.11) can be solved numerically. In order to do this it is possible to use the contraction T_h^k to construct an iteration scheme but since the contraction factor $\beta = 1 - \delta h$ is close to 1 this iteration converges rather slow. An acceleration method for this iteration scheme has been proposed by M. Falcone [12]. Falcone uses the set $\mathcal{V}$ of monotone convergence of T_h^k given by $\mathcal{V} := \{V \in \mathbb{R}^N \mid T_h^k(V) \geq V\}$ where "$\geq$" denotes the componentwise order. A simple computation shows that $\mathcal{V}$ is a convex closed subset of $\mathbb{R}^N$. Given a $V_0 \in \mathcal{V}$ the operator T_h^k is used to determine an initial direction. The algorithm follows this direction until it crosses the boundary of $\mathcal{V}$, then determines a new direction using T_h^k and continues the same way.

A different algorithm to calculate V^* can be developed by observing that V^* is the componentwise maximum of $\mathcal{V}$ and that $\mathcal{V}$ can be written as

$$\mathcal{V} = \left\{ V \in \mathbb{R}^N \;\middle|\; [V]_i \le \min_{u\in U} \left\{ \frac{\beta \sum_{\substack{j=1,\dots,N \\ j\neq i}} \lambda_{ij}(u)[V]_j + hG_i(u)}{1-\beta\lambda_{ii}(u)} \right\} \quad \forall i \in \{1,\dots,N\} \right\} \tag{4.19}$$

Note that the fraction on the right side does not depend on $[V]_i$. Thus we can construct the *increasing coordinate algorithm*:

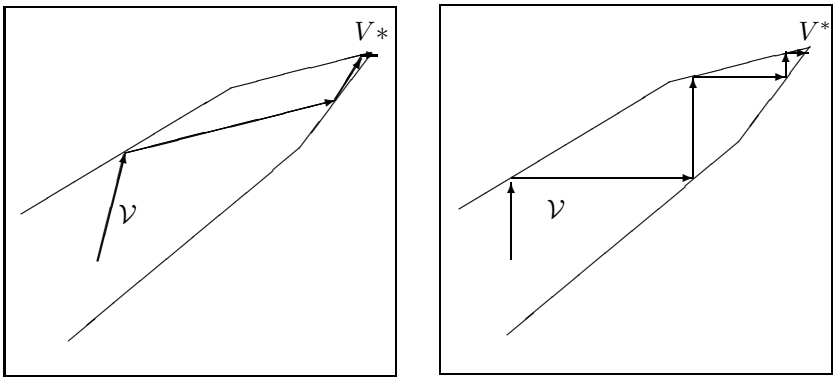
Step 1: take $V \in \mathcal{V}$ (e.g. $V = \left(-\frac{M_g}{\delta},\dots,-\frac{M_g}{\delta}\right)^T$)

Step 2: compute sequentially

$$[V]_i = \min_{u\in U} \left\{ \frac{\beta \sum_{\substack{j=1,\dots,N \\ j\neq i}} \lambda_{ij}(u)[V]_j + hG_i(u)}{1-\beta\lambda_{ii}(u)} \right\} \quad \forall i \in \{1,\dots,N\}.$$

Step 3: continue with Step 2 and the new vector V.

Figure 4.1 shows an illustration of the algorithms for $N = 2$.



Falcone's accelerated method — Increasing coordinate algorithm

FIG. 4.1. *Algorithms*

Note that for every arrow in the left picture the intersection between the initial direction and the boundary of $\mathcal{V}$ has to be determined. To do this — e.g. by bisection as in the implementation used here — the operator T_h^k has to be evaluated several times to decide if a point is inside or outside $\mathcal{V}$. In the increasing coordinate algorithm N arrows (i.e. two arrows in this figure) are calculated by N evaluations of the fraction in step 2. These N evaluations are about as expensive as one evaluation of T_h^k. This means that one iteration in the increasing coordinate algorithm corresponds to one evaluation of T_h^k in the acceleration method.

The convergence of this algorithm is guaranteed by the following lemma:

LEMMA 4.3. *Let V_1 be the vector obtained by applying step 2 for $i = 1,\dots,N$ to a vector $V_0 \in \mathcal{V}$. Then*

$$[V_1]_i - [V_0]_i \;\ge\; [T_h^k(V_0)]_i - [V_0]_i.$$

Proof. Because of $V_0 \in \mathcal{V}$ and (4.19) it follows $[V_1]_i \geq [V_0]_i \ \forall i = 1, \dots, N$. Hence

$$\begin{aligned}
[V_1]_i - [V_0]_i &= \min_{u \in U} \left\{ \frac{\beta \sum_{\substack{j=1,\dots,N \\ j \neq i}} \lambda_{ij}(u)[V_1]_j + hG_i(u) - (1 - \beta\lambda_{ii}(u))[V_0]_i}{1 - \beta\lambda_{ii}(u)} \right\} \\
&\geq \min_{u \in U} \left\{ \beta \sum_{\substack{j=1,\dots,N \\ j \neq i}} \lambda_{ij}(u)[V_0]_j + hG_i(u) - (1 - \beta\lambda_{ii}(u))[V_0]_i \right\} \\
&= \min_{u \in U} \left\{ \beta \sum_{j=1}^{N} \lambda_{ij}(u)[V_0]_j + hG_i(u) - [V_0]_i \right\} \ = \ [T_h^k(V_0)]_i - [V_0]_i
\end{aligned}$$

□

The convergence of the increasing coordinate algorithm therefore is a consequence of the monotone convergence of the iteration scheme using the contraction T_h^k.

All iteration methods described here have in common that during the iteration a minimum over all $u \in U$ has to be calculated. The following lemma shows that this can be done by minimizing over a finite set $U_\varepsilon \subset U$.

LEMMA 4.4. *Assume that X and g are uniformly Lipschitz continuous in the control $u \in U$ with Lipschitz constant L_u. Let $U_\varepsilon \subset U$ such that for all $u \in U$ there exists $\bar{u} \in U_\varepsilon$ with $\|u - \bar{u}\| < \varepsilon$. Let $\mathcal{U}_\varepsilon$ denote the corresponding set of control functions. Then for all $s \in \mathbb{S}^{d-1}$ it holds that*

$$\| \inf_{u(\cdot) \in \mathcal{U}} J_\delta(s, u(\cdot)) - \inf_{\bar{u}(\cdot) \in \mathcal{U}_\varepsilon} J_\delta(s, \bar{u}(\cdot)) \| < C\varepsilon^\eta$$

where for $\delta < L_X + 1$ we have $\eta = \frac{L_X + 1}{\delta}$.

Proof. For all $u(\cdot) \in \mathcal{U}$ there exists $\bar{u}(\cdot) \in \mathcal{U}_\varepsilon$ such that $\|u(t) - \bar{u}(t)\| < \varepsilon$ for almost all $t \in \mathbb{R}$. Hence we have

$$\|\varphi(t, s, u(\cdot)) - \varphi(t, s, \bar{u}(\cdot))\| < L_u \varepsilon t + \int_0^t L_X \|\varphi(\tau, s, u(\cdot)) - \varphi(\tau, s, \bar{u}(\cdot))\| d\tau$$

where $\|\cdot\|$ denotes the norm on $\mathbb{R}^d$. Now the Gronwall Lemma and [4, Lemma 4.1] can be used to estimate this integral equation and the assertion follows. □

For the projected bilinear control system with cost function $g = q$ the assumptions of Lemma 4.4 are fulfilled and hence we may use a finite set of control values to calculate v_h^k.

Once v_h^k is calculated it can be used to construct ε-optimal control functions:

Step 1: Let $x_0 = x$, $n = 0$.

Step 2: Choose a control value $\tilde{u}_{x_n,h}^k \in U$, such that
$\beta v_h^k(\Phi_h(x_n, \tilde{u}_{x_n,h}^k)) + hg(x_n, \tilde{u}_{x_n,h}^k)$ becomes minimal.

Step 3: Let $u_{x,h}^k(t) = \tilde{u}_{x_n,h}^k$ for all $t \in [nh, (n+1)h]$.

Step 4: Let $x_{n+1} = \Phi_h(x_n, \tilde{u}_{x_n,h}^k)$, $n = n + 1$ and continue with Step 2.

In step 2 a unique $\tilde{u}_{x_n,h}^k \in U$ may be found e.g. by using a lexicographic order on U.

THEOREM 4.5. *Let $u^k_{x,h}$ denote the control function defined above. Then for every $\varepsilon > 0$ there exist $H > 0$, $K(h) > 0$, such that for all $h < H$, $k \le K(h)$:*

$$|J_\delta(x, u^k_{x,h}(\cdot)) - v_\delta(x)| \le \varepsilon \quad \forall x \in \Omega.$$

Proof. Using (4.12) or (4.14) and the definition of $u^{k,i}_{x,h} := u^k_{x,h}|_{[ih,(i+1)h)}$ we have for sufficiently small k and x_i from (4.7):

$$\begin{aligned} hg(x_i, u^{k,i}_{x,h}) + \beta v^k_h(\Phi_h(x_i, u^{k,i}_{x,h})) &\ge hg(x_i, u^{k,i}_{x,h}) + \beta v_h(\Phi_h(x_i, u^{k,i}_{x,h})) - \frac{\varepsilon}{2} \\ &\ge v_h(x_i) - \frac{\varepsilon}{2} \;\; \ge \;\; v^k_h(x_i) - \varepsilon \end{aligned}$$

and with $u^{0,i}_{x,h} \in U$ denoting the value, where $hg(x_i, u) + \beta v_h(\Phi_h(x_i, u))$, $u \in U$ attains its minimum:

$$\begin{aligned} hg(x_i, u^{k,i}_{x,h}) + \beta v^k_h(\Phi_h(x_i, u^{k,i}_{x,h})) &\le hg(x_i, u^{0,i}_{x,h}) + \beta v^k_h(\Phi_h(x_i, u^{0,i}_{x,h})) \\ &\le hg(x_i, u^{0,i}_{x,h}) + \beta v_h(\Phi_h(x_i, u^{0,i}_{x,h})) + \frac{\varepsilon}{2} \\ &= v_h(x_i) + \frac{\varepsilon}{2} \;\; \le \;\; v^k_h(x_i) + \varepsilon. \end{aligned}$$

Putting this together yields

$$|hg(x_i, u^{k,i}_{x,h}) + \beta v^k_h(\Phi_h(x_i, u^{k,i}_{x,h})) \;\; - \;\; v^k_h(x_i)| \;\; \le \;\; \varepsilon \quad \forall x \in \overline{\Omega^k} \tag{4.20}$$

By induction we can conclude that for every $\varepsilon > 0$, $p \in \mathbb{N}$, $h > 0$ there exists $k > 0$ such that:

$$|h \sum_{j=0}^{p} \beta^j g(x_j, u^{k,j}_{x,h}) + \beta^{p+1} v^h_k(x_{p+1}) \quad - \quad v^k_h(x)| \;\; \le \;\; \frac{\varepsilon}{2} \quad \forall x \in \overline{\Omega^k} \tag{4.21}$$

Since $\beta < 1$ for all $h > 0$ and g and v^k_h are bounded on $\overline{\Omega^k}$, for every $\varepsilon > 0$ we may find a $p_h \in \mathbb{N}$ such that

$$|h \sum_{j=0}^{\infty} \beta^j g(x, u^{k,j}_{x,h}) - h \sum_{j=0}^{p_h} \beta^j g(x, u^{k,j}_{x,h}) - \beta^{p_h+1} v^k_h(x)| < \frac{\varepsilon}{2} \quad \forall x \in \overline{\Omega^k}, u \in \mathcal{U}_h. \tag{4.22}$$

Combining (4.12) or (4.14), (4.21) and (4.22) yields

$$|J_h(x, u^k_{x,h}(\cdot)) - v_h(x)| \le \varepsilon \quad \forall x \in \Omega.$$

Using estimates (4.10) and (4.9) the assertion follows. □

Remark 4.6. The proof also shows how k and h have to be chosen: First choose h such that (4.10) and (4.9) hold for the desired accuracy, then choose k dependend on p_h from (4.22) such that (4.21) is fulfilled.

To construct a control function that is uniformly ε-optimal we can put together the ε-optimal control functions according to the following definition and lemma.

DEFINITION 4.7. *Let $u_x(\cdot) \in \mathcal{U}$ be control functions for every $x \in \overline{\Omega^k}$. Let $(\tau_i)_{i \in \mathbb{N}}$ be a real sequence of* switching times *satisfying $\tau_1 = 0$, $\tau_{i+1} > \tau_i$ and $a \le \tau_{i+1} - \tau_i \le b$*

$\forall i \in \mathbb{N}$ for positive constants a, $b \in \mathbb{R}$, $a \leq b$.
Then we define control functions $\bar{u}_x(\cdot) \in \mathcal{U}$ by:

$$\bar{u}_x|_{[\tau_i,\tau_{i+1})} \equiv u_{\varphi(x,\tau_i,\bar{u}_x(\cdot))}|_{[0,\tau_{i+1}-\tau_i)} \quad \forall i \in \mathbb{N}$$

LEMMA 4.8. *Assume for every $x \in \overline{\Omega^k}$ there exists a control function $u_x(\cdot) \in \mathcal{U}$ such that*
$|J_\delta(x, u_x(\cdot)) - v_\delta(x)| < \varepsilon$. Then for $\bar{u}_x(\cdot) \in \mathcal{U}$ from Definition 4.7 the following estimate holds:

$$J_\delta(\varphi(\sigma, x, \bar{u}_x(\cdot)), \bar{u}_x(\sigma + \cdot)) \leq v_\delta(\varphi(\sigma, x, \bar{u}_x(\cdot))) + \frac{e^{\delta b}}{\delta a}\varepsilon \quad \forall \sigma \geq 0.$$

Proof. For all $t > 0$ it holds that

$$\begin{aligned} v_\delta(x) &\geq J_\delta(x, u_x(\cdot)) - \varepsilon \\ &\geq \int_0^t e^{-\delta\tau} g(\varphi(x, \tau, u_x(\cdot)), u_x(\tau))d\tau + e^{-\delta t} v_\delta(\varphi(x, t, u_x(\cdot)) - \varepsilon \end{aligned} \tag{4.23}$$

By induction with $t = \tau_i$ it follows that

$$J_\delta(x, \bar{u}_x(\cdot)) \leq v_\delta(x) + \sum_{i=0}^{\infty} e^{-\delta\tau_i}\varepsilon$$

and for $0 < 1 - \delta a < 1$ this sum can be estimated by

$$\sum_{i=0}^{\infty} e^{-\delta\tau_i} \leq \sum_{i=0}^{\infty} e^{-\delta a i} \leq \sum_{i=0}^{\infty} (1 - \delta a)^i \leq \frac{1}{\delta a}.$$

Together with the definition of the $\bar{u}_x(\cdot)$ this implies

$$J_\delta(\varphi(\tau_i, x, \bar{u}_x(\cdot)), \bar{u}_x(\tau_i + \cdot)) \leq v_\delta(\varphi(\tau_i, x, \bar{u}_x(\cdot))) + \frac{\varepsilon}{\delta a}$$

for all $i \in \mathbb{N}$.

For the times in between let $\sigma > 0$, $\tilde{\varepsilon} > 0$ and consider $u_{x_0}(\cdot) \in \mathcal{U}$ such that $|J_\delta(x_0, u_{x_0}(\cdot)) - v_\delta(x_0)| \leq \tilde{\varepsilon}$:

$$\begin{aligned} v_\delta(x_0) + \tilde{\varepsilon} &\geq \int_0^\infty e^{-\delta t} g(\varphi(t, x_0, u_{x_0}(\cdot)), u_{x_0}(t))dt \\ &= \int_0^\sigma e^{-\delta t} g(\varphi(t, x_0, u_{x_0}(\cdot)), u_{x_0}(t))dt \\ &\quad + e^{-\delta\sigma} \int_0^\infty e^{-\delta t} g(\varphi(t, \varphi(\sigma, x_0, u_{x_0}(\cdot)), u_{x_0}(\sigma + \cdot)), u_{x_0}(\sigma + t))dt \\ &= \int_0^\sigma e^{-\delta t} g(\varphi(t, x_0, u_{x_0}(\cdot)), u_{x_0}(t))dt \end{aligned}$$

$$
\begin{aligned}
&+\ e^{-\delta\sigma} J_\delta(\varphi(\sigma, x_0, u_{x_0}(\cdot)), u_{x_0}(\sigma+\cdot)) \\
&\geq \int_0^\sigma e^{-\delta t} g(\varphi(t, x_0, u_{x_0}), u_{x_0}(t))dt \ + \ e^{-\delta\sigma} v_\delta(\varphi(\sigma, x_0, u_{x_0})) \\
&\geq v_\delta(x_0).
\end{aligned}
$$

From this inequality it follows that

$$
|v_\delta(\varphi(\sigma, x_0, u_{x_0}(\cdot))) - J_\delta(\varphi(\sigma, x_0, u_{x_0}(\cdot)), u_{x_0}(\sigma+\cdot))| \ \leq \ e^{\delta\sigma}\tilde{\varepsilon}.
$$

Choosing $i \in \mathbb{N}$ maximal with $\tau_i \leq \sigma$ and $x_0 := \varphi(\tau_i, x, \bar{u}_x(\cdot))$ it follows that:

$$
|v_\delta(\varphi(\sigma, x, \bar{u}_x(\cdot))) - J_\delta(\varphi(\sigma, x, \bar{u}_x(\cdot)), \bar{u}_x(\sigma+\cdot))| \leq e^{\delta(\sigma-\tau_i)}\frac{1}{\delta a}\varepsilon \leq e^{\delta b}\frac{1}{\delta a}\varepsilon = \frac{e^{\delta b}}{\delta a}\varepsilon
$$

which finishes the proof. □

Remark 4.9. This lemma does not answer the question which switching times τ_i are optimal. In estimate (4.23) we have to assume the worst case, i.e. that the error up to the time t

$$
\varepsilon(t) := |v_\delta(x) - \int_0^t e^{-\delta\tau} g(\varphi(x, \tau, u_x(\cdot)), u_x(\tau))d\tau - e^{-\delta t} v_\delta(\varphi(x, t, u_x(\cdot))|
$$

may be equal to ε for all $t > 0$ and hence the error becomes large if $a = \min(\tau_{i+1} - \tau_i)$ becomes small. The numerical examples discussed in the next section show that good results can be obtained for small a.

Using the results from Theorem 3.1 we can use the control functions constructed here to develop an *algorithm to stabilize bilinear control systems*:

Step 1: Calculate v_h^k, the approximation of the optimal value function for small discount rate $\delta > 0$ to approximate the minimal Lyapunov exponents of the systems (under the assumptions of Theorem 3.1).

Step 2: Given an initial value $x \in \mathbb{R}^d$ with $\lambda^*(x) < 0$ compute the control function that is ε-optimal along its trajectory according to Definition 4.7 (using the projected system). The trajectory of the bilinear system using this control is asymptotically stable under the assumptions of Theorem 3.1 provided h and k are small enough.

Note that the main numerical expense lies in the calculation of the approximated optimal value function v_h^k. Once this function is known the algorithm to calculate the control functions is numerically simple and quite fast.

For this algorithm only the information $x(t, x_0, u(\cdot))/\|x(t, x_0, u(\cdot))\|$ of the bilinear system is needed. In particular the calculated control functions are exactly the same for all $x_1, x_2 \in \mathbb{R}^d$ with $x_1/\|x_1\| = x_2/\|x_2\|$ and hence the algorithm works for arbitrarily large or small $\|x\|$. It is not necessary to discretize the trajectory of the bilinear system or to lift the discretized solution from $\mathbb{S}^{d-1}$ to $\mathbb{R}^d$ which then would imply that small discretization errors on $\mathbb{S}^{d-1}$ could become large in $\mathbb{R}^d$.

The value function and the corresponding optimal control values for each point can also be used to "verify" the assumptions of Theorem 3.1, (viii) numerically: if there exists a set such that $v_h^k < 0$ and this set is invariant with respect to the numerically computed optimal controls the corresponding trajectory will tend to 0 for any initial value from this set, provided the discretization is fine enough (see also Remark 3.2).

Remark 4.10. The way the stabilizing control functions are constructed leads to the question if v_h^k can be used to construct a stabilizing feedback for the bilinear control system. This question is closely related to the optimal switching times τ_i. If it is possible to choose $(\tau_{i+1} - \tau_i)$ arbitrarily small it could also be possible to obtain an ε-optimal feedback e.g. by linear interpolation or averaging of the feedback for the discrete time system.

The main problem in proving this property of the switching times lies in the fact that the Euler method yields only linear convergence in h, hence quadratic convergence for one time step. Thus the difference between $v_h(\varphi(h, x, u))$ (the value that can be reached after the first time step) and $v_h(\Phi_h(x, u))$ (the value that is supposed to be reached) is of the order $h^{2\gamma}$. For $\gamma < 1/2$ this error will accumulate and convergence is no longer guaranteed. However, there is hope to overcome this difficulty by using a higher order method to calculate $\Phi_h(x, u)$ which then will require a different proof of the convergence of v_h.

5. Numerical examples. In this section we will present some numerical examples calculated with the algorithm developed in the previous sections. All examples were computed on an IBM6000 Workstation.

The first example is a bilinear control system in $\mathbb{R}^2$, the two-dimensional linear oscillator given by

$$\ddot{x} + 2b\dot{x} + (1 + u)x = 0$$

or written as a two-dimensional system by $x_1 = x,\ \ x_2 = \dot{x}$

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ -1-u & -2b \end{pmatrix}}_{=:A(u)} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{5.1}$$

The projection of the system to $\mathbb{S}^1$ by $s = \frac{x}{\|x\|}$ reads

$$\dot{s} = \begin{pmatrix} s_2(1 + us_1^2 + 2bs_1s_2) \\ -(1+u)s_1 - 2bs_2 + s_2^2(us_1 + 2bs_2) \end{pmatrix} \tag{5.2}$$

For the one-dimensional sphere we may use the parametrization via polar coordinates $\Psi(\varphi) = (\cos\varphi, \sin\varphi)$ where $\Psi^{-1}(s) = \arcsin(s_2)$, $s_2 \in [-\pi/2, \pi/2]$, $\Psi^{-1}(s) = \arcsin(\pi - s_2)$, $s_2 \in [\pi/2, 3\pi/2]$. In polar coordinates the cost function reads $g(\varphi, u) = -\sin\varphi(u\cos\varphi + 2b\sin\varphi)$ and we can choose $\Omega = (0 - \varepsilon, \pi + \varepsilon)$ to cover the whole projective space (identified with one half of the sphere).

h	ops_1	ops_2
1.0	13	11477
0.1	42	11477
0.01	51	11477

TABLE 5.1
Dependence on the time step h ($k = 0.032$, $\delta = 1.0$)

The Tables 5.1–5.3 show the number of iterations in the increasing coordinate algorithm (ops_1) and the number of evaluations of the operator T_h^k in the accelerated algorithm (ops_2) depending on certain parameters with damping parameter $b = 1.5$. Remember that one iteration in the increasing coordinate algorithm corresponds to

k	ops_1	ops_2
0.063	28	5918
0.032	42	11477
0.0063	233	49625

TABLE 5.2
Dependence on the space discretization k ($h = 0.1$, $\rho = 1.0$)

δ	V_0	ops_1	ops_2
5.0	-0.66	16	2001
2.0	-1.66	35	5543
1.0	-3.32	42	11477
0.1	-33.23	194	121187
0.01	-332.27	1707	-
0.001	-3322.72	16836	-

TABLE 5.3
Dependence on the discount rate δ ($h = 0.1$, $k = 0.032$)

one evaluation of T_h^k. The used set of control values was ρU with $U = \{-1, 1\}$ and $\rho = 0.5$.

Using the techniques described in Remark 3.2 the whole Lyapunov spectrum for this system was computed for $\rho = \{0.1, 0.2, \ldots, 1.3\}$ with parameters $h = 0.01$, $k = 0.006$ and $\delta = 0.01$ and locally refined grid with $k = 0.0016$ around the variant control set for $\rho \leq 0.5$. (For $\rho = 0.0$ the exponents are just the eigenvalues of A). For $\rho \leq 1.2$ there exist two control sets D_1 and D_2 and therefore two intervals of Lyapunov exponents. For $\rho = 1.3$ there is only one control set and thus only one interval. For this system a finer discretization of U does not yield different values for v_h^k; it is sufficient to minimize over the extremal control values.

ρ	$\min(D_1)$	$\max(D_1)$	$\min(D_2)$	$\max(D_2)$
0.0	-2.61	-2.61	-0.38	-0.38
0.1	-2.65	-2.58	-0.42	-0.35
0.2	-2.69	-2.52	-0.47	-0.31
0.3	-2.73	-2.47	-0.52	-0.25
0.4	-2.77	-2.42	-0.57	-0.22
0.5	-2.81	-2.37	-0.63	-0.19
0.6	-2.85	-2.31	-0.69	-0.14
0.7	-2.89	-2.24	-0.75	-0.11
0.8	-2.91	-2.18	-0.82	-0.07
0.9	-2.96	-2.09	-0.90	-0.06
1.0	-2.99	-2.00	-0.99	0.00
1.1	-3.00	-1.90	-1.10	0.03
1.2	-3.03	-1.74	-1.27	0.06
1.3	-3.03	–	–	0.10

TABLE 5.4
Lyapunov spectrum for system (5.1) with b=1.5

For $\rho = 0.5$ the system is asymptotically stable for all control functions since the maximal Lyapunov exponent is negative. But as the Lyapunov exponents corresponding to D_1 are much smaller than those of the control set D_2 it can be expected that the optimal trajectories with initial value inside D_1 tend to zero much faster.

Figure 5.2 shows that this is exactly what happens. In this figure the dotted lines correspond to the boundaries of D_1, the dashed lines to the boundary of D_2.

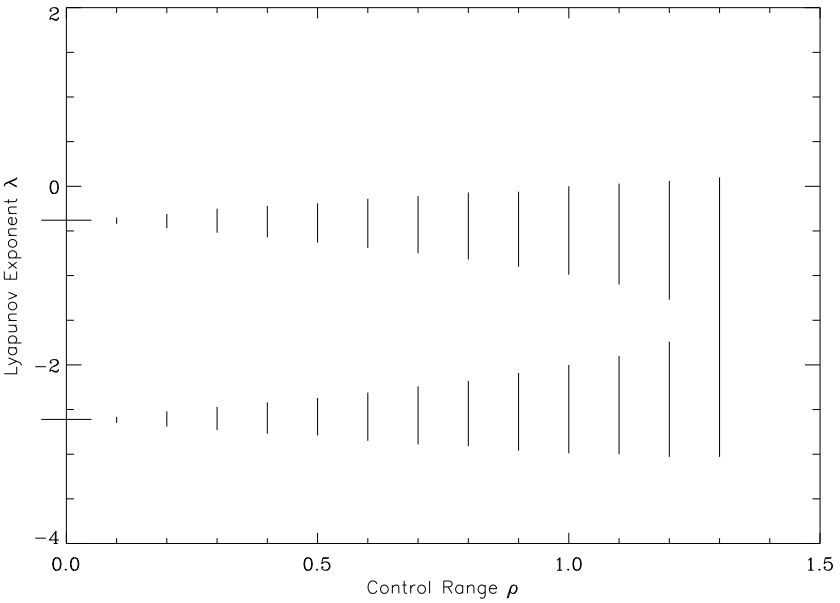


FIG. 5.1. *Lyapunov spectrum of system* (5.1) *with b=1.5*

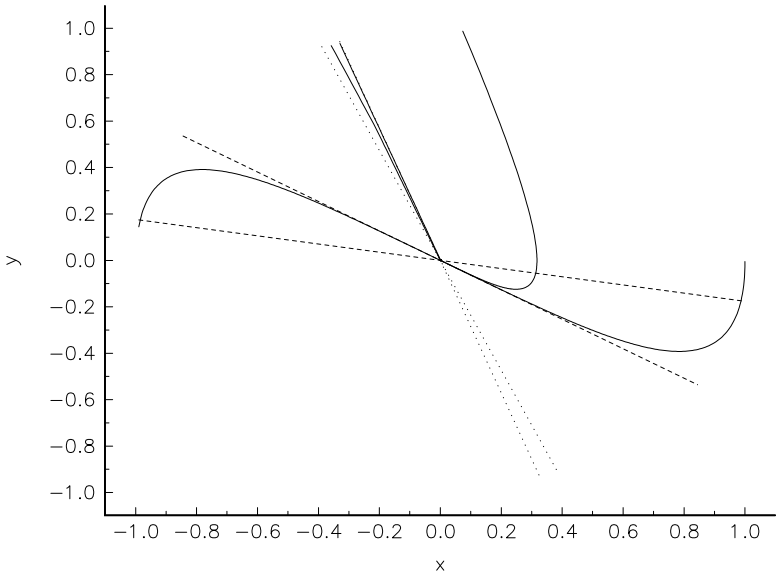


FIG. 5.2. *Trajectories for* $b = 1.5$, $\rho = 0.5$

All trajectories in this section were computed using the extrapolation method for ordinary differential equations by Stoer and Bulirsch [19, §7.2.14]. The parameter a from Definition 4.7 was chosen as $a = h$ (see Remark 4.9).

The second example is the three-dimensional linear oscillator given by

$$\dddot{y} + a\ddot{y} + b\dot{y} + (c+u)y = 0 \tag{5.3}$$

or written as a three-dimensional system by

$$\begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -(c+u) & -b & -a \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \tag{5.4}$$

with $a, b, c \in \mathbb{R}$ and $u \in U$.
The projected system on $\mathbb{S}^2$ reads

$$\dot{s} = \begin{pmatrix} s_2 - s_1(-\tilde{u}s_1s_3 + s_1s_2 + (1-b)s_2s_3 - as_3^2) \\ s_3 - s_2(-\tilde{u}s_1s_3 + s_1s_2 + (1-b)s_2s_3 - as_3^2) \\ -\tilde{u}s_1 - bs_2 - as_3 - s_3(-\tilde{u}s_1s_3 + s_1s_2 + (1-b)s_2s_3 - as_3^2) \end{pmatrix} \tag{5.5}$$

with $\tilde{u} := c + u$.

For $\mathbb{S}^2$ the parametrization by spherical coordinates is not suitable since this parametrization maps two opposite points to a line and hence it is not invertible on one half of the sphere. Thus the stereographic projection is used instead; it is given by

$$\Psi(x) = \left(\frac{2x_1}{1+\|x\|^2}, \frac{2x_2}{1+\|x\|^2}, \frac{2}{1+\|x\|^2} - 1 \right)$$

and

$$\Psi^{-1}(s) = \left(\frac{1}{1+s_3} s_1, \frac{1}{1+s_3} s_2 \right).$$

The cost function reads

$$g(x,u) = -(c+u)\Psi_1(x)\Psi_3(x) + \Psi_1(x)\Psi_2(x) + (1-b)\Psi_2(x)\Psi_3(x) - a\Psi_3(x)^2$$

with $\Psi = (\Psi_1, \Psi_2, \Psi_3)$.
The set Ω was chosen as $\Omega = (-1-\varepsilon, 1+\varepsilon) \times (-1-\varepsilon, 1+\varepsilon)$ to cover the whole $\mathbb{P}^2$ (identified with the upper half of $\mathbb{S}^2$).

All values given have been checked according to Remark 3.2 (ii); in all cases it was possible to find trajectories that realized the values as Lyapunov exponents. Hence the calculated values at least give an approximation of the minimal Lyapunov exponents over the interior of the control sets. To apply the results of Remark 3.2 (iii), i.e. to make sure that this is indeed the Lyapunov spectrum we have to check the $\rho - \rho'$-inner pair condition described in Section 3. Unfortunately up to now it is not known how to check this condition analytically. However, the program CS2DIM from Gerhard Häckl [15] has been used to calculate reachable sets for the system for different ρ-parameters numerically. Since they turned out to be strictly increasing in this example there is strong evidence that the condition is fulfilled.

For the following figures spherical coordinates ($s_1 = \sin\theta\cos\varphi$, $s_2 = \sin\theta\sin\varphi$, $s_3 = \cos\theta$), $x = \theta$, $y = \varphi$ were used and the system was transformed by $z(t) := e^{\frac{1}{3}at}y(t)$.

The first parameters considered for this system were $a = 1$, $b = 0$, $c = 0.5$ and $U = \{-0.3, -0.25, \ldots, 0.25, 0.3\}$. Figure 5.3 shows the two control sets of this system. The control sets were computed again using the program CS2DIM [15].

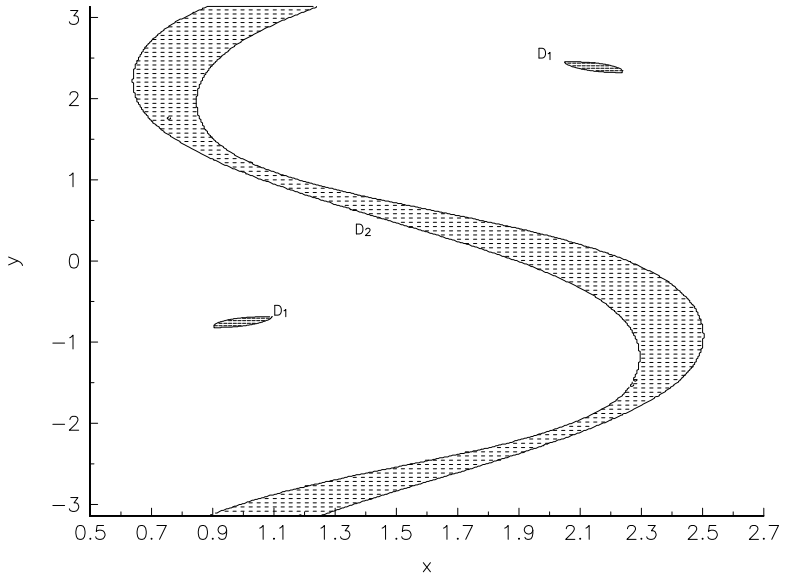


FIG. 5.3. *Control sets of system* (5.5) *with* $a = 1$, $b = 0$, $c = 0.5$

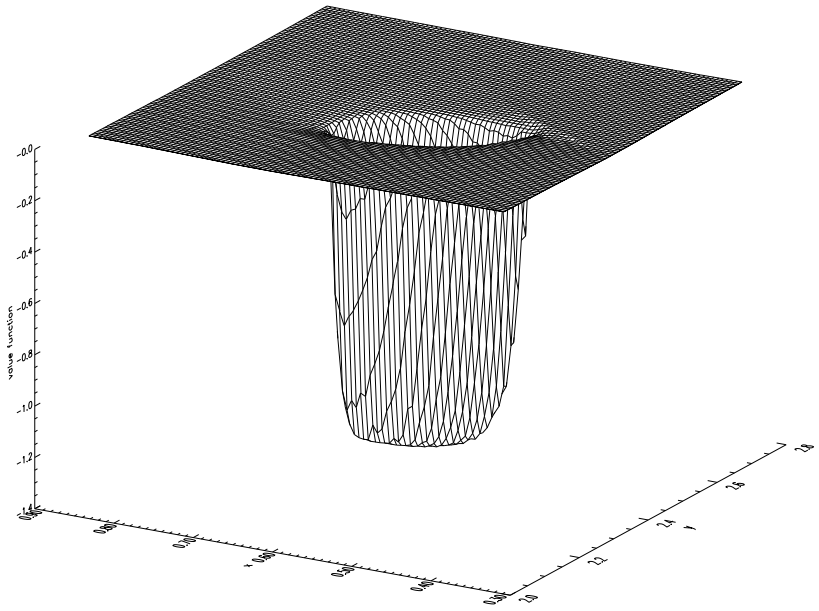


FIG. 5.4. *Value function around* D_1

The numerical parameters used for this example are $k = 0.003$ around D_1, $k = 0.09$ elsewhere, $h = 0.05$ and $\delta = 0.01$. The discounted value function of this system around D_1 is shown in Figure 5.4. The calculated minimal Lyapunov exponent over

D_1 is -1.25, the maximal exponent is -1.15. The calculated minimal and maximal exponents over D_2 are 0.019 and 0.24 and the value function is constant outside D_1. Figure 5.5 shows two trajectories of the projected system with initial values inside D_1. Table 5.5 shows the values of one corresponding trajectory in $\mathbb{R}^3$.

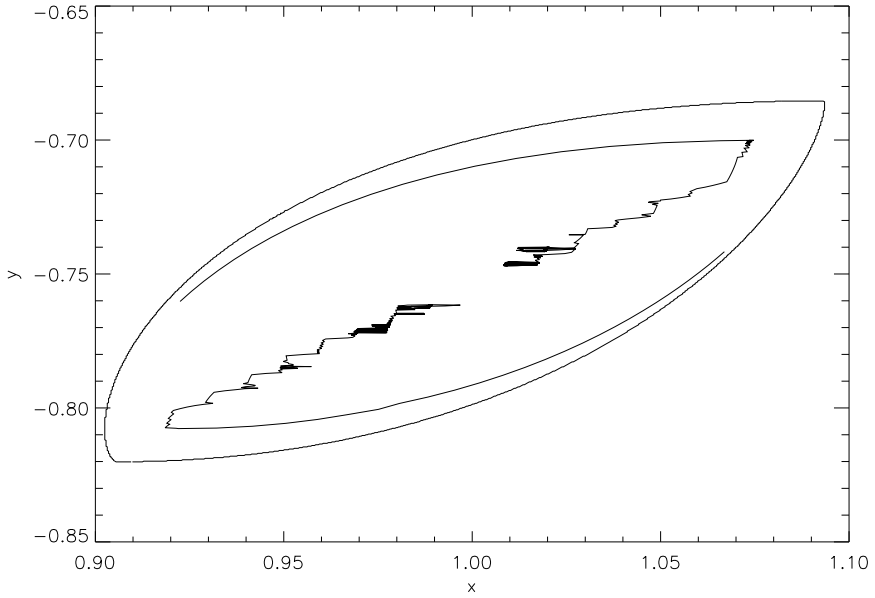


FIG. 5.5. *Optimal trajectories in* D_1

t	x_1	x_2	x_3
1	0.124609	-0.169914	0.219449
2	0.031318	-0.043096	0.060307
3	0.008062	-0.010800	0.014665
4	0.002129	-0.002818	0.003757
5	0.000569	-0.000750	0.000986
6	0.000153	-0.000201	0.000264
7	0.000041	-0.000054	0.000071
8	0.000011	-0.000014	0.000019
9	0.000003	-0.000004	0.000005
10	0.000000	-0.000001	0.000001
11	0.000000	0.000000	0.000000

TABLE 5.5
Stabilized trajectory for system (5.4) *with* $a = 1$, $b = 0$, $c = 0.5$

The second set of parameters considered for this system is $a = -1$, $b = -3$, $c = 0.5$ and $U = \{-1.0, -0.9, \ldots, 0.9, 1.0\}$. Figure 5.6 shows the three control sets of the projected system, the domain of attraction of D_2 (denoted by $A^-(D_2)$) and the domain of attraction of D_2 of the time reversed system (denoted by $A^+(D_2)$).

Here the numerical parameters were $k = 0.002$ around D_1, $k = 0.045$ elsewhere, $h = 0.05$ and $\delta = 0.01$. Figure 5.7 shows the discounted optimal value function around D_1.

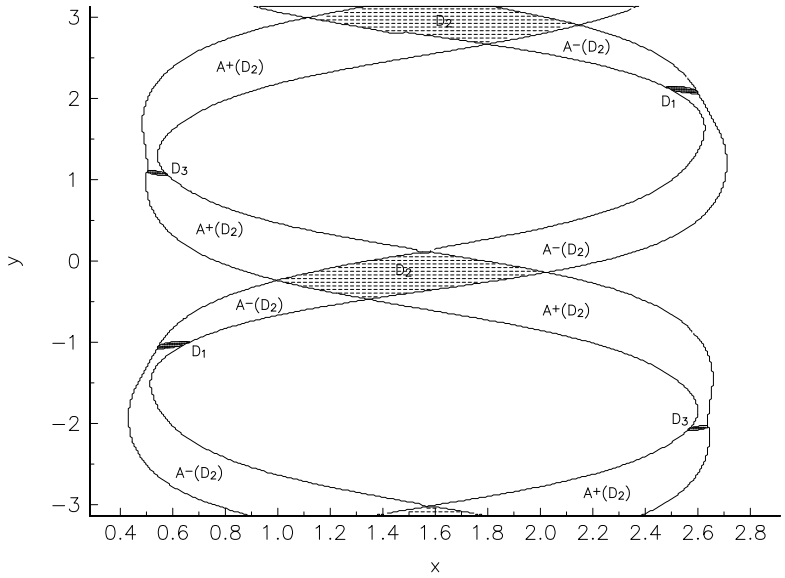


FIG. 5.6. *Control sets of system* (5.5) *with* $a = -1.0$, $b = -3.0$, $c = 0.5$

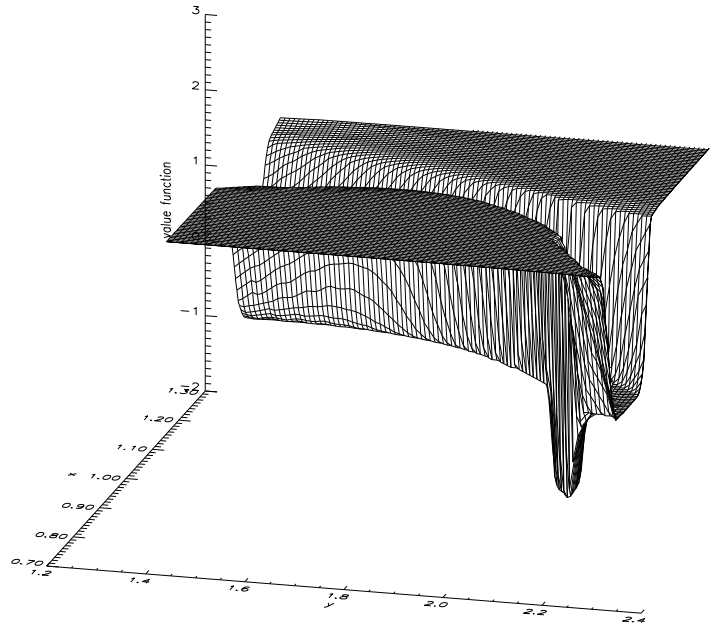


FIG. 5.7. *Value function of the system*

The calculated spectrum for this example is $\lambda(D_1) = [-1.47, -1.17]$, $\lambda(D_2) = [-0.10, 0.43]$ and $\lambda(D_3) = [2.07, 2.36]$.

Figure 5.8 shows an optimal trajectory in $\mathbb{P}^2$, starting in the domain of attraction

of D_2. Table 5.6 shows the corresponding trajectory (x_1, x_2, x_3) in $\mathbb{R}^3$ and another trajectory (y_1, y_2, y_3) in $\mathbb{R}^3$ with projected initial value in D_1. This trajectory tends to 0 much faster which is exactly what one would expect since the minimal Lyapunov exponent inside D_1 is much smaller.

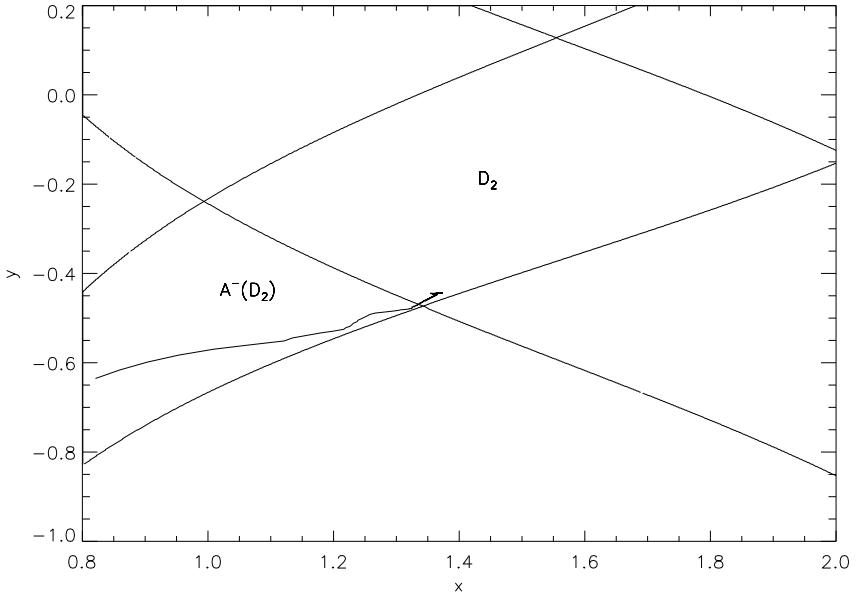


FIG. 5.8. *Optimal trajectory starting in* $A^-(D_2)$

t	x_1	x_2	x_3	y_1	y_2	y_3
1	0.576395	-0.119011	0.071718	0.096972	-0.141621	0.200267
5	0.293857	-0.044627	0.007731	0.000260	-0.000384	0.000562
10	0.142984	-0.020156	0.003083	0.000000	-0.000000	0.000000
15	0.070691	-0.009962	0.001172	0.000000	-0.000000	0.000000
20	0.034949	-0.004924	0.000754	0.000000	-0.000000	0.000000
25	0.017279	-0.002434	0.000373	0.000000	-0.000000	0.000000
30	0.008543	-0.001204	0.000184	0.000000	-0.000000	0.000000
35	0.004224	-0.000595	0.000091	0.000000	-0.000000	0.000000
40	0.002088	-0.000294	0.000045	0.000000	-0.000000	0.000000
45	0.001032	-0.000146	0.000022	0.000000	-0.000000	0.000000
50	0.000510	-0.000072	0.000008	0.000000	-0.000000	0.000000

TABLE 5.6
Stabilized trajectories for system (5.4) *with* $a = -1$, $b = -3$, $c = 0.5$

Acknowledgement: I would like to thank Fritz Colonius for his constant help and many useful discussions as well as a number of anonymous referees for their detailled advice.

REFERENCES

[1] M. BARDI AND M. FALCONE, *An approximation scheme for the minimum time function*, SIAM J. Control Optim., 28 (1990), pp. 950–965.

[2] I. Capuzzo Dolcetta, *On a discrete approximation of the Hamilton-Jacobi equation of dynamic programming*, Appl. Math. Optim., 10 (1983), pp. 367–377.
[3] I. Capuzzo Dolcetta and M. Falcone, *Discrete dynamic programming and viscosity solutions of the Bellman equation*, Ann. Inst. Henri Poincaré, Anal. Non Linéaire, 6 (supplement) (1989), pp. 161–184.
[4] I. Capuzzo Dolcetta and H. Ishii, *Approximate solutions of the Bellman equation of deterministic control theory*, Appl. Math. Optim., 11 (1984), pp. 161–181.
[5] R. Chabour, G. Sallet, and J. Vivalda, *Stabilization of nonlinear systems: A bilinear approach*, Math. Control Signals Syst., 6 (1993), pp. 224–246.
[6] F. Colonius, *Asymptotic behaviour of optimal control systems with low discount rates*, Math. Oper. Res., 14 (1989), pp. 309–316.
[7] F. Colonius and W. Kliemann, *Infinite time optimal control and periodicity*, Appl. Math. Optim., 20 (1989), pp. 113–130.
[8] ———, *Linear control semigroups acting on projective space*, J. Dyn. Differ. Equations, 5 (1993), pp. 495–528.
[9] ———, *Maximal and minimal Lyapunov exponents of bilinear control systems*, J. Differ. Equations, 101 (1993), pp. 232–275.
[10] ———, *Asymptotic null controllability of bilinear systems*, in "Geometry in Nonlinear Control and Differential Inclusions", Banach Center Publications Vol. 32, Warsaw, 1995, pp. 139–148.
[11] ———, *The Lyapunov spectrum of families of time varying matrices*, Trans. Am. Math. Soc., (1996). to appear.
[12] M. Falcone, *Numerical solution of deterministic control problems*, in Proceedings of the International Symposium on Numerical Analysis, Madrid, 1985.
[13] ———, *A numerical approach to the infinite horizon problem of deterministic control theory*, Appl. Math. Optim., 15 (1987), pp. 1–13. *Corrigenda*, ibid. 23 (1991), 213–214.
[14] R. L. V. Gonzáles and M. M. Tidball, *On the rates of convergence of fully discrete solutions of Hamilton-Jacobi equations*. INRIA Rapports de Recherche Nr. 1379, 1991.
[15] G. Häckl, *Numerical approximation of reachable sets and control sets*, Random Comput. Dyn., 1 (1992–1993), pp. 371–394.
[16] W. Kliemann, *Recurrence and invariant measures for degenerate diffusions*, Ann. Probab., 15 (1987), pp. 690–707.
[17] P. L. Lions, *Generalized solutions of Hamilton-Jacobi equations*, Pitman, London, 1982.
[18] R. Mohler, *Bilinear Control Processes*, Academic Press, New York, 1973.
[19] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer Verlag, New York, 1980.
[20] F. Wirth, *Convergence of the value functions of discounted infinite horizon optimal control problems with low discount rates*, Math. Oper. Res., 18 (1993), pp. 1006–1019.