Optimal value functions for weakly coupled systems: a posteriori estimates

Oliver Junge[1]* **and Péter Koltai**[1] **

[1] Faculty for Mathematics
Technische Universität München
Boltzmannstr. 3
85748 Garching

Received XXXX, revised XXXX, accepted XXXX
Published online XXXX

Key words List, of, comma, separated, keywords

MSC (2000) 00-xx

We consider weakly coupled LQ optimal control problems and derive estimates on the sensitivity of the optimal value function in dependence of the coupling strength. Our main result is that if a weak coupling suffices to destabilize the closed loop system with the optimal feedback of the uncoupled system then the value function might change drastically with the coupling. As a consequence, it is not reasonable to expect that a weakly coupled system possesses a weakly coupled optimal value function.

1 Introduction

When designing feedback controllers for high dimensional control systems using dynamic programming, the dimension of state space imposes natural limitations for existing methods: In general, the representation of the optimal value function suffers from the curse of dimension. An efficient approximation is only possible if the optimal value function possesses some regularity that we can exploit. Our aim here is to investigate such regularity properties for systems which can be decomposed into a number of small subsystems which – in a proper sense – are only weakly coupled to each other. This scenario is, e.g., motivated by so called *networked control systems*.

More specifically, we consider a system consisting of several weakly coupled subsystems such that an optimal feedback of each subsystem, i.e. without the coupling to the other systems, is easily computable (for example, if it is an LQ system or if its state space dimension is ≤ 3). These feedbacks together form an optimal feedback for the uncoupled system, the *naive feedback*. Not surprisingly, the naive feedback may fail to stabilize the coupled system for particular coupling structures even if the coupling is weak. In this paper, we aim to understand why and how the naive feedback fails, and how the optimal value function changes with increasing coupling strength.

In particular, we consider the time-continuous linear quadratic regulator (LQR) problem, for which the optimal value function is a quadratic function of the form $V(x) = x^T Px$ with a symmetric positive definite matrix P. In Section 3 we recall some existing theory on the sensitivity analysis of the LQR problem. Since the norm in which the sensitivity is measured yields unsatisfactory estimates for the coupling of the optimal value function, in Section 4 the estimates are established in a coupling-adapted norm. Similar estimates are derived in Section 5 for discrete time systems. Section 6 contains a discussion about the obtained estimates. Finally, we turn to the question on what these estimates tell us about the structure of the optimal value function if we can introduce a weak coupling such that the naive feedback fails to stabilize the coupled system. The answer is somewhat surprising: if a coupling (no matter how small) destabilizes the system controlled by the naive feedback, the optimal value function for the coupled system differs massively from the one of the uncoupled system; cf. Section 7.

* Corresponding author E-mail: junge@ma.tum.de, Phone: +49 89 289 17987, Fax: +49 89 289 17985
** E-mail: koltai@ma.tum.de, Phone: +49 89 289 17999, Fax: +49 89 289 17985

2 The optimal value function, uncoupled and weakly coupled systems

We consider a control system

$$\dot{x} = f(x, u), \tag{1}$$

where the right hand side $f : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^n$ is assumed to be C^1 and the state space $\mathcal{X} \subset \mathbb{R}^n$, $0 \in \mathcal{X}$, and the control space $\mathcal{U} \subset \mathbb{R}^m$, $0 \in \mathcal{U}$, are compact. The origin of the uncontrolled system is an unstable equilibrium, i.e. we have that $f(0, 0) = 0$. We are interested in constructing a feedback $u : \mathcal{X} \to \mathcal{U}$ such that the origin of the closed loop system $\dot{x} = f(x, u(x))$ is asymptotically stable. To this end, we consider a continuous cost function $c : \mathcal{X} \times \mathcal{U} \to [0, \infty)$ which fulfills the condition $c(x, u) = 0$ iff $x = 0$. Given an initial value $x_0 \in \mathcal{X}$ and a (measurable) control function $u : [0, \infty) \to \mathcal{X}$, the cost along the associated trajectory $x(t; x_0, u)$ of (1) is

$$J(x_0, u) = \int_0^\infty c(x(t; x_0, u), u(t))\, dt,$$

while the *optimal value function* is

$$V(x) = \inf\{J(x, u) \mid u : [0, \infty) \to \mathcal{X} \text{ measurable}\}.$$

The optimal value function is the solution of the *Hamilton-Jacobi-Bellman equation*

$$\inf_{u \in \mathcal{U}} \{\nabla V(x) \cdot f(x, u) + c(x, u)\} = 0,$$

with the boundary condition $V(0) = 0$. From this equation, an optimal feedback can be constructed by choosing a minimizing u to a given x (assuming that it exists).

We call a control system (1) with cost function c *uncoupled* if one can decompose state and control space into a least two subsystems, $\mathcal{X} = \bigotimes_{i=1}^k \mathcal{X}_i$ and $\mathcal{U} = \bigotimes_{i=1}^k \mathcal{U}_i$, $k \geq 2$, such that f and c can be written as

$$f(x, u) = \begin{bmatrix} f_1(x_1, u_1) \\ \vdots \\ f_k(x_k, u_k) \end{bmatrix}$$

and $c(x, u) = \sum_{i=1}^k c_i(x_i, u_i)$, where $f_i : \mathcal{X}_i \times \mathcal{U}_i \to \mathbb{R}^{n_i}$, $c_i : \mathcal{X}_i \times \mathcal{U}_i \to [0, \infty)$ and $n_1 + \ldots + n_k = n$. Of course, for an uncoupled system, the optimal value function is simply the sum of the optimal value functions of the subsystems:

$$V(x) = \sum_{i=1}^k V_i(x_i).$$

Evidently, if V is C^2, then $\partial_{i,j} V \equiv 0$, but also the converse is true (for the proof, see the appendix):

Proposition 2.1 *For $V \in C^2(\mathcal{X})$ the following statements are equivalent.*

a) There are $V_i \in C^2(\mathcal{X}_i)$, $i = 1, \ldots, k$, such that $V(x) = \sum_{i=1}^k V_i(x_i)$.

b) It holds $\partial_{i,j} V \equiv 0$ for all $i, j \in \{1, \ldots, k\}$, $i \neq j$.

Accordingly, a control system f is called *weakly coupled* if again there is a decomposition of $\mathcal{X}$ and $\mathcal{U}$ as above such that for all $i, j \in \{1, \ldots, k\}$, $i \neq j$,

$$\|\partial_i f_i(x, u)\| \gg \|\partial_j f_i(x, u)\|$$

uniformly in x and u, for a given norm $\|\cdot\|$ (on the $\mathcal{X}_i$). We define the *coupling constant* as

$$\varepsilon_f := \max_{i \neq j} \sup_{x,u} \frac{\|\partial_j f_i(x, u)\|}{\|\partial_i f_i(x, u)\|}.$$

In light of Proposition 2.1 a natural question is whether for weakly coupled systems the mixed derivatives $\partial_{i,j} V$ for the optimal value function are uniformly small.

3 Weakly coupled LQ systems with linear feedback — continuous time

We now consider more specifically linear time invariant systems with quadratic cost function. A weakly coupled linear system is defined by

$$\dot{x} = Ax + Bu, \tag{2}$$

where A is a weakly coupled matrix, i.e. the diagonal blocks dominate the other ones (an implication of the definition above), and B is a block diagonal matrix (i.e. the subsystems are controlled separately). The accumulated cost along a trajectory $x(t) = x(t; x_0, u(\cdot))$ is given by

$$J(x_0, u(\cdot)) = \int_0^\infty x(t)^T Q x(t) + u(t)^T R u(t)\, dt,$$

where Q and R are symmetric positive definite (spd) matrices, both block diagonal to satisfy $c(x,u) = \sum_{i=1}^k c_i(x_i, u_i)$. The optimal value function is then the infimum over all possible control functions. Imposing no restrictions on u, the unique optimal control is realized by the feedback

$$u(x) = -R^{-1} B^T P x,$$

where P is the unique spd solution of the Riccati equation

$$PBR^{-1}B^T P - PA - A^T P - Q = 0, \tag{3}$$

and the optimal value function is given by $V(x) = x^T P x$; see [4], chapter 8.4.

Hence, the question whether the $\partial_{i,j} V$ are small, reduces to the question whether the off-diagonal blocks of P are small compared to the diagonal ones. Our answer will be of asymptotic nature, i.e. by considering small perturbations of an initially block diagonal (uncoupled) system matrix A. Note, that for a block diagonal system matrix and an analogously block diagonal ansatz P equation (3) simplifies to k uncoupled Riccati equations. Then, by uniqueness of the solution is P block diagonal as well.

In order to analyze the effect of perturbations on A, we define the separation of two matrices; see e.g. [6].

Definition 3.1 (Separation of matrices) The separation of two square matrices X and Y, $\mathrm{sep}(X,Y)$, is defined as the smallest singular value of $I \otimes X - Y^T \otimes I$, where I is the identity and $\otimes$ denotes the Kronecker product.[1]

The following estimate plays a central role in our considerations. Although more general results exist [3], a proof of this one is also given, because we use the ideas in it again later.

Theorem 3.2 *Let $P(A)$ denote the solution of* (3) *for fixed B, Q and R, in dependence of A. If $P(A+\delta A) = P(A) + \delta P$, and $A_{cl} = A - R^{-1}B^T P(A)$ denotes the closed-loop system matrix, it holds*

$$\|\delta P\|_{\mathrm{F}} \leq \frac{2\,\|P(A)\|_{\mathrm{F}}}{\mathrm{sep}(A_{cl}, -A_{cl}^T)} \|\delta A\|_{\mathrm{F}} + \mathcal{O}\left(\|\delta A\|_{\mathrm{F}}^2\right). \tag{4}$$

P r o o f. Define the function $r(A,P) = PBR^{-1}B^T P - PA - A^T P - Q$, where P is assumed to be a symmetric matrix. By definition, $r(A, P(A)) = 0$ for all A. We have that $P(\cdot)$ is real-analytic in an appropriate neighborhood of A [1]. It also holds

$$\begin{aligned} \partial_A\, r(A,P)\cdot X &= -PX - X^T P \\ \partial_P\, r(A,P)\cdot X &= XBR^{-1}B^T P + PBR^{-1}B^T X - XA - A^T X. \end{aligned}$$

From the implicit function theorem we have

$$DP(A) = -\partial_P\, r\,(A, P\,(A))^{-1}\, \partial_A\, r(A, P(A)),$$

if $\partial_P\, r\,(A, P\,(A))$ is invertible. By writing it as a linear operator [6], one can see that it is indeed invertible if and only if $\mathrm{sep}(A_{cl}, A_{cl}^T) \neq 0$. Moreover it holds

$$\|DP(A)\|_{\mathrm{F}} \leq \frac{2\,\|P(A)\|_{\mathrm{F}}}{\mathrm{sep}(A_{cl}, -A_{cl}^T)}. \tag{5}$$

A Taylor expansion of P yields the claim. □

[1] It holds $\mathrm{sep}(X,Y) = \mathrm{sep}(Y,X)$, and $\mathrm{sep}(X^T, Y^T) = \mathrm{sep}(X,Y)$.

An immediate consequence of Theorem 3.2 for our situation is that if we are given a weakly coupled linear system (2) and we can decompose $A = A_0 + \delta A$ such that $\|\delta A\|_{\mathrm{F}} \ll \|A_0\|_{\mathrm{F}}$, and $\|DP(A_0)\|_{\mathrm{F}}$ is small, then the optimal value function is “weakly coupled”, i.e. its mixed second derivatives are uniformly small.

Remark 3.3

a) Note that (5) is an a posteriori estimate, i.e. it does not characterize the sensitivity directly in terms of A, B, Q and R, but involves the solution P of the (uncoupled) Riccati equation.

b) In [3] the authors present a non-local perturbation analysis of the Riccati equation (3). This could be used to yield better bounds for the estimates of the mixed second derivatives of the optimal value function. With the present result we have no control about the quality of our estimate, since we have no information about the magnitude of higher order terms. Their result states that $\|\delta P\|_{\mathrm{F}} \le g\left(\|\delta A\|_{\mathrm{F}}\right)$ for $\|\delta A\|_{\mathrm{F}} \in [0, a^*)$, with

$$g(a) = \frac{1}{2d}\left(s - 2a - \left((s-2a)^2 - 8pda\right)^{1/2}\right),$$

where $s = \mathrm{sep}\left(A_{cl}, -A_{cl}^T\right)$, $d = \left\|BR^{-1}B^T\right\|_{\mathrm{F}}$ and $p = \|P(A)\|_{\mathrm{F}}$. The end of the validity region, a^*, is defined as the (smaller) positive root of $2a + 2\sqrt{2dpa} - s = 0$.

4 Coupling-adapted estimates

A somewhat unsatisfactory property of the Frobenius norm $\|\cdot\|_{\mathrm{F}}$ is that $\|\delta A\|_{\mathrm{F}}$ is not closely related to the coupling constant of the system. Much more adequate for measuring the coupling of a system is the following norm, which focuses on the subsystems. Each matrix $X \in \mathbb{R}^{n\times n}$, with $n = \sum_{i=1}^k n_i$, can be partitioned blockwise, according to the subsystems: $X = (X_{ij})_{i,j=1}^k$, where $X_{i,j} \in \mathbb{R}^{n_i \times n_j}$. Define

$$\|X\|_C = \max_{i,j} \|X_{i,j}\|_{\mathrm{F}}.$$

Note that this norm is *not* sub-multiplicative:

$$\|AB\|_C = \max_{i,j} \left\| \sum_{\ell=1}^k A_{i,\ell} B_{\ell,j} \right\|_F \le k\,\|A\|_C\,\|B\|_C.$$

Adapting the proof of Theorem 3.2 to an uncoupled system, we obtain the following sensitivity estimate:

Lemma 4.1 *Consider the situation of Theorem 3.2 and suppose that the system is uncoupled. Then*

$$\|(DP(A)\cdot\delta A)_{i,j}\|_{\mathrm{F}} \le \frac{\|P(A)_{i,i}\|_{\mathrm{F}} + \|P(A)_{j,j}\|_{\mathrm{F}}}{\mathrm{sep}((A_{cl})_{j,j}, -(A_{cl})_{i,i}^T)} \max\left\{ \|\delta A_{i,j}\|_{\mathrm{F}}, \|\delta A_{j,i}\|_{\mathrm{F}} \right\}. \tag{6}$$

Proof. Note, that the system is uncoupled, hence A, $P(A)$ and A_{cl} are all block-diagonal. It holds block-wise

$$(\partial_A\, r(A,P)\cdot\delta A)_{i,j} = -P_{i,i}\delta A_{i,j} - \delta A_{j,i}^T P_{j,j}.$$

Recall that $\partial_P\, r(A,P)\cdot X = -XA_{cl} - A_{cl}^T X =: -Q$ for symmetric matrices X. It holds block-wise

$$X_{i,j}(A_{cl})_{j,j} + (A_{cl})_{i,i}^T X_{i,j} = Q_{i,j} = -P_{i,i}\delta A_{i,j} - \delta A_{j,i}^T P_{j,j}.$$

Solving this for $X_{i,j}$ implies the claim analogously to the proof of Theorem 3.2. □

This also gives us a possibility to measure the effect on the coupling constant of the closed-loop system.

Corollary 4.2 *In the situation of Lemma 4.1 the first order perturbation δA_{cl} of the closed loop system matrix induced by the perturbation δA of the system matrix A satisfies*

$$\|\delta(A_{cl})_{i,j}\|_{\mathrm{F}} \le \left(1 + \left\|B_{i,i}R_{i,i}^{-1}B_{i,i}^T\right\|_{\mathrm{F}} \frac{\|P(A)_{i,i}\|_{\mathrm{F}} + \|P(A)_{j,j}\|_{\mathrm{F}}}{\mathrm{sep}((A_{cl})_{j,j}, -(A_{cl})_{i,i}^T)}\right) \max\left\{ \|\delta A_{i,j}\|_{\mathrm{F}}, \|\delta A_{j,i}\|_{\mathrm{F}} \right\}. \tag{7}$$

Proof. Use $A_{cl} = A - BR^{-1}B^T P(A)$. □

5 Weakly coupled LQ systems with linear feedback — discrete time

The considerations here are analogous to the ones in sections 3 and 4. The discrete time control system is given by

$$x_{m+1} = Ax_m + Bu_m, \quad m = 0, 1, 2, \ldots.$$

We assume A to be weakly coupled and B to be a block diagonal matrix. The cost accumulated along a trajectory $(x_m)_{m\in\mathbb{N}} = (x_m(x_0; (u_k)_{k\in\mathbb{N}}))_{m\in\mathbb{N}}$ is given by

$$J(x_0, (u_k)_{k\in\mathbb{N}}) = \sum_{m=0} x_m^T Q x_m + u_m^T R u_m,$$

where Q and R are block diagonal spd matrices. If there are no restrictions on u the optimal feedback is given by

$$u(x) = -\left(R + B^T P B\right)^{-1} B^T P A x,$$

and the optimal value function is $V(x) = x^T P x$, with P being the unique spd solution of the discrete Riccati equation (see [4], chapter 8.4)

$$P = A^T \left(P - PB\left(R + B^T PB\right)^{-1} B^T P\right) A + Q. \tag{8}$$

In the following we derive a first order perturbation result on P depending on A.

For arbitrary square matrices M_1, M_2 and S the equation

$$X - M_1^T X M_2 = S$$

is solvable if the matrix $I - M_1^T \otimes M_2^T$ is nonsingular, and the absolute condition (w.r.t. the Frobenius norm) of the solution is given by $1/\mathrm{sep}^\#(M)$ with

$$\mathrm{sep}^\#(M_1, M_2) := \sigma_{\min}(I - M_1^T \otimes M_2^T),$$

where $\sigma_{\min}(\cdot)$ denotes the smallest singular value.

Theorem 5.1 *Let the linear discrete time system be uncoupled, and let $P(A)$ denote the solution of* (8)*. Then we have*

$$\|(DP(A)\cdot\delta A)_{i,j}\|_F \le \frac{\|P_{i,i}(A_{cl})_{i,i}\|_F + \left\|(A_{cl})_{j,j}^T P_{j,j}\right\|_F}{\mathrm{sep}^\#\left((A_{cl})_{i,i}, (A_{cl})_{j,j}\right)} \max\{\|\delta A_{i,j}\|_F, \|\delta A_{j,i}\|_F\}. \tag{9}$$

P r o o f. Just as in the proof of Theorem 3.2, we use the implicit function theorem. For this, let

$$g(A, P) = A^T\left(PB\left(R + B^T PB\right)^{-1} B^T P - P\right) A + P - Q.$$

Setting $\bar{A} = A - (R + B^T PB)^{-1} B^T PA$, we obtain

$$\begin{aligned} \partial_A g(A,P)\cdot X &= -X^T P\bar{A} - \bar{A}^T P X \\ \partial_P g(A,P)\cdot X &= X - \bar{A}^T X \bar{A}. \end{aligned}$$

Since all matrices involved are block diagonal, we have in particular the block wise equations

$$\begin{aligned} (\partial_A g(A,P)\cdot X)_{i,j} &= -X_{j,i}^T P_{j,j}\bar{A}_{j,j} - \bar{A}_{i,i}^T P_{i,i} X_{i,j} \\ (\partial_P g(A,P)\cdot X)_{i,j} &= X_{i,j} - \bar{A}_{i,i}^T X_{i,j} \bar{A}_{j,j}. \end{aligned}$$

Since $\bar{A} = A_{cl}$ for $P = P(A)$, considerations as in the proof of Theorem 3.2 and Lemma 4.1 imply the claim. □

Remark 5.2 For the discrete time case as well, in [3] the authors present a non-local perturbation analysis in the Frobenius norm for the solution of (8).

6 Coupling estimates in dependence on the number of subsystems

The estimates (5) and (6) are easier to compare if we see, that the denominators are the same:

Proposition 6.1 *For a block-diagonal matrix we have*

$$\mathrm{sep}(A,-A^T) = \min_{i,j} \mathrm{sep}(A_{i,i}, -A_{j,j}^T).$$

Proof. For block-diagonal A the Lyapunov equation $A^T X + XA = Q$ is decoupled,

$$A_{i,i}^T X_{i,j} + X_{i,j} A_{j,j} = Q_{i,j}.$$

This means that the solution operator of the Lyapunov equation is block-wise decoupled, hence its singular values are the singular values of the blocks. This proves the claim. □

Under the assumption that the denominator in (5) stays constant in the number of subsystems k, we expect the Frobenius norm estimate[2] to increase as $k^{1/2}$. Using (5) to obtain an estimate in $\|\cdot\|_C$, we can use norm equivalence. Having k subsystems, one verifies easily that

$$\|A\|_C \le \|A\|_{\mathrm{F}} \le k\,\|A\|_C\,.$$

Thus, we have

$$\|\delta P\|_C \le \|\delta P\|_{\mathrm{F}} \le \|DP(A)\|_{\mathrm{F}}\,\|\delta A\|_{\mathrm{F}} \le k\,\|DP(A)\|_{\mathrm{F}}\,\|\delta A\|_C\,.$$

In general, this would imply an estimate $\sim k^{3/2}$.

However, the coupling-adapted norm estimate (6) shows that the first order perturbations do not depend on the number of subsystems *at all*. An important conclusion is that the coupling strength of the optimal value function does not increase with the number of subsystems. One easily extends these considerations to the discrete time case.

7 The optimal value function of destabilized systems (continuous time)

In this section we investigate the structure of the optimal value function if a coupling destabilizes the optimally controlled (linear time-continuous) system. We make use of the notion of (real-structured) *pseudospectra* [5]. The set

$$\sigma_\varepsilon^*(A) := \left\{ z\in\mathbb{C} \,\middle|\, \exists \delta A \in \mathbb{C}^{n\times n} : \|\delta A\|_2 \le \varepsilon, z\in\sigma(A+\delta A)\right\}$$

is the (complex) ε-pseudospectrum, while

$$\sigma_\varepsilon(A) := \left\{ z\in\mathbb{C} \,\middle|\, \exists \delta A \in \mathbb{R}^{n\times n} : \|\delta A\|_2 \le \varepsilon, z\in\sigma(A+\delta A)\right\}$$

is called the real-structured ε-pseudospectrum of $A\in\mathbb{R}^{n\times n}$, where $\sigma(\cdot)$ denotes the usual spectrum. The norm of the smallest real perturbation δA that destabilizes A, the so-called *(real) stability radius* of A [2], is given by

$$r_{\mathbb{R}}(A) := \min_{\delta A\in\mathbb{R}^{n\times n}} \left\{ \|\delta A\|_2 \,\middle|\, \exists z\in i\mathbb{R}, z\in\sigma(A+\delta A)\right\}.$$

So far we may conclude that if δA destabilizes A, then $\|\delta A\|_2 \ge r_{\mathbb{R}}(A)$. What can be said about the estimate (4) in these cases?

Proposition 7.1 *It holds*

$$\mathit{sep}(A,-A^T) \le 2 r_{\mathbb{R}}(A). \qquad (10)$$

Proof. From Theorem 3.1 [6] we have

$$\begin{aligned}
\mathrm{sep}(A,-A^T) &\le \mathrm{sep}_\lambda(A,-A^T) := \min_{\varepsilon_1,\varepsilon_2} \left\{ \varepsilon_1+\varepsilon_2 \,\middle|\, \sigma_{\varepsilon_1}^*(A)\cap\sigma_{\varepsilon_2}^*(-A^T)\neq\emptyset\right\}\\
&\le \min_{\varepsilon_1,\varepsilon_2} \left\{ \varepsilon_1+\varepsilon_2 \,\middle|\, \exists z\in i\mathbb{R}, z\in\sigma_{\varepsilon_1}(A), z\in\sigma_{\varepsilon_2}(-A^T)\right\}\\
&= 2r_{\mathbb{R}}(A).
\end{aligned}$$

The inequality in the second line comes from restricting the minimization to real pseudospectra and the imaginary axis; the equation in the third line comes from the fact that complex eigenvalues of a real matrix appear in conjugate pairs, i.e. if $i\omega\in\sigma(A+\delta A)$ for some $\omega\in\mathbb{R}$, then $-i\omega\in\sigma(A+\delta A)$ but also $-i\omega\in\sigma(-A^T-\delta A^T)$. □

[2] For such vague — however, instructive — estimates we assume that the subsystems "are of same order", e.g. all $\|P_{i,i}\|_{\mathrm{F}}$ are approximately of the same magnitude.

This result, together with (4), shows that if a weak coupling (i.e. with a small constant $\varepsilon > 0$) destabilizes a system, we have to take a strong coupling (i.e. with constant of size $\mathcal{O}(1)$) of the optimal value function into account.

Corollary 7.2 *For a uncoupled matrix A we have*

$$sep(A, -A^T) = \min_{i,j} sep(A_{i,i}, -A_{j,j}^T) \leq 2 \min_i r_{\mathbb{R}}(A_{ii}).$$

To put this result into the context of our coupling adapted norm, define

$$r_{\mathbb{R}}^C(A) := \min_{\delta A \in \mathbb{R}^{n \times n}} \left\{ \|\delta A\|_C \;\middle|\; \exists z \in \mathbb{C} : \text{Re } z = 0, z \in \sigma(A + \delta A) \right\},$$

and note $\|A\|_2 \leq k \, \|A\|_C$. It follows $\text{sep}(A, -A^T) \leq 2k r_{\mathbb{R}}^C(A)$.

Remark 7.3 Although the above estimate is sufficient to highlight the effect of a destabilizing weak coupling, we would like to remark here that the worst case is $\text{sep}(A, -A^T) = \mathcal{O}(\varepsilon^2)$ for an ε-coupling.

Examples. We consider two different types of examples. First, where the system matrix is strongly non-normal, so the pseudospectra corresponding to small perturbations reach into the positive complex half plane. Second, where the control part of the cost (i.e. the matrix R) is much bigger than the state part of the cost (i.e. the matrix Q), hence the optimally controlled system is weakly stable in the sense that the spectral abscissa of the closed-loop system matrix is small (compared to the modulus of the corresponding eigenvalues). In all examples we take R to be the identity matrix.

Example 7.4 The optimally controlled system has robust stable eigenvalues, i.e. the spectral abscissa is of order $\mathcal{O}(1)$. A weak coupling destabilizes the closed-loop system.

$$A = \begin{pmatrix} 1 & 1000 & 0 \\ 0 & -0.5 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad Q = I_{3\times 3}.$$

Closed-loop eigenvalues: $-0.5, -\sqrt{2}, -\sqrt{2}$. The coupling matrix

$$\delta A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -0.1 \\ 0.04 & 0 & 0 \end{pmatrix}$$

moves the eigenvalues to $-1.6707 \pm 0.8163i, 0.0130$, and (with two digits precision)

$$P = \begin{pmatrix} 2.4 & 130 & 0 \\ 130 & 93000 & 0 \\ 0 & 0 & 0.41 \end{pmatrix}, \qquad \delta P + P = \begin{pmatrix} 0.62 & 120 & -1.3 \\ 120 & 33000 & -450 \\ -1.3 & -450 & 8.5 \end{pmatrix}$$

Example 7.5 The optimally controlled system has robust stable eigenvalues. A weak coupling destabilizes the closed-loop system.

$$A = \begin{pmatrix} 1 & 50 & 0 & 0 & 0 \\ 0 & -0.5 & 50 & 0 & 0 \\ 0 & 0 & -0.5 & 50 & 0 \\ 0 & 0 & 0 & -0.5 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad Q = I_{5\times 5}.$$

The closed-loop eigenvalues are $-1.11, -1.41, -0.5, -0.5, -0.5$. The coupling matrix

$$\delta A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1 \\ 0.001 & 0 & 0 & 0 & 0 \end{pmatrix}$$

moves the eigenvalues to $0.39 \pm 1.4i, -1.9 \pm 0.38i, -1.1$.

Example 7.6 The optimally controlled system is only weakly stable due to an expensive control. A weak coupling destabilizes the closed-loop system

$$A = \begin{pmatrix} 10^{-2} & 1 & 0 & 0 \\ -1 & 10^{-2} & 0 & 0 \\ 0 & 0 & 10^{-2} & 1 \\ 0 & 0 & -1 & 10^{-2} \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad Q = 10^{-4} I_{4\times 4}.$$

The closed-loop eigenvalues are $-0.012 \pm 1i, -0.012 \pm 1i$. The coupling matrix

$$\delta A = \begin{pmatrix} 0 & 0 & 5\cdot 10^{-2} & 0 \\ 0 & 0 & 0 & 0 \\ 2\cdot 10^{-2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

moves the eigenvalues to $-0.028 \pm 1i, 0.0036 \pm 1i$. The matrices of the associated optimal value functions are

$$P = 10^{-2} \cdot \begin{pmatrix} 4.4 & -0.05 & 0 & 0 \\ -0.05 & 4.4 & 0 & 0 \\ 0 & 0 & 4.4 & -0.05 \\ 0 & 0 & -0.05 & 4.4 \end{pmatrix}, \qquad P + \delta P = 10^{-2} \cdot \begin{pmatrix} 7.6 & -0.08 & 4.6 & -0.05 \\ -0.08 & 7.6 & -0.04 & 4.6 \\ 4.6 & -0.04 & 3.8 & -0.04 \\ -0.05 & 4.6 & -0.04 & 3.8 \end{pmatrix}.$$

Clearly, the off-diagonal blocks of $P + \delta P$ are not small in comparison to the diagonal blocks.

8 The optimal value function of destabilized discrete-time systems

Just as before, we compute a bound for the $\text{sep}^{\#}(\cdot,\cdot)$ term in (9). Recall

$$\text{sep}^{\#}(M_1, M_2) = \min_{\|X\|_F = 1} \|X - M_1^T X M_2\|_F.$$

We would like to bound $\text{sep}^{\#}(A, A)$ in terms of the real stability radius of A, defined by

$$r_{\mathbb{R}}(A) := \min_{\delta A \in \mathbb{R}^{n\times n}} \left\{ \|\delta A\|_2 \,\middle|\, \exists \lambda \in \mathbb{C} : \, |\lambda| = 1, \, \lambda \in \sigma(A + \delta A) \right\}.$$

Now assume we have $\lambda \in \mathbb{C}$, $|\lambda| = 1$, non-negative numbers $\varepsilon_{1,2}$, vectors $\|v_{1,2}\|_2 = 1$ and $\|w_{1,2}\|_2 = \varepsilon_{1,2}$ such that for the real matrices $M_{1,2}$ holds

$$(M_i^T - \lambda I) v_i = w_i, \quad i = 1, 2.$$

It follows (x^* denotes the conjugate transpose of x)

$$\begin{aligned} w_1 w_2^* &= (M_1^T - \lambda I) v_1 v_2^* (M_2 - \bar{\lambda} I) \\ &= M_1^T v_1 v_2^* M_2 - \lambda v_1 v_2^* M_2 - \bar{\lambda} M_1^T v_1 v_2^* + v_1 v_2^* \\ &= M_1^T v_1 v_2^* M_2 - \lambda v_1 (\bar{\lambda} v_2^* + w_2^*) - \bar{\lambda}(\lambda v_1 + w_1) v_2^* + v_1 v_2^* \\ &= M_1^T v_1 v_2^* M_2 - v_1 v_2^* - \lambda v_1 w_2^* - \bar{\lambda} w_1 v_2^*. \end{aligned}$$

With this equation and by setting $X = v_1 v_2^*$ we conclude that $\text{sep}^{\#}(M_1, M_2) \le \varepsilon_1 \varepsilon_2 + \varepsilon_1 + \varepsilon_2$.

Since a $\lambda \in \mathbb{C}$ is in the (complex) ε-pseudospectrum of A if there is a vector v of unit modulus such that $\|(A - \lambda I)v\|_2 = \varepsilon$, and the complex pseudospectreum contains the real one, we can find a $\|v\|_2 = 1$ such that $\|(A - \lambda I)v\|_2 = r_{\mathbb{R}}(A)$.

Setting $v_1 = v_2 = v$ in the above computation we find

Proposition 8.1

$$\text{sep}^{\#}(A, A) \le r_{\mathbb{R}}(A) \left(2 + r_{\mathbb{R}}(A)\right).$$

This result, together with (9), shows that if a weak coupling (i.e. with a small constant $\varepsilon > 0$) destabilizes a system, we have to take a strong coupling (i.e. with constant of size $\mathcal{O}(1)$) of the optimal value function into account.

9 Conclusions

We have shown that even if one can divide an optimal control problem into subsystems which appear to be weakly coupled according to the structure of the Jacobian, the optimal value function still might not inherit this structure. Certainly, more sophisticated ways of defining and detecting "weak coupling" are imaginable which would not shown this undesired effect. Also, it might be possible to define cost functions such that the feedback for the associated uncoupled system still steers the weakly coupled system into some neighborhood of the target set. These are questions for further investigations.

10 Appendix

Proof of Proposition 2.1:

P r o o f. "⇒": Trivial.
"⇐": In order to show the claim, we need following sublemma:
Sublemma: If V is C^2 in the x_1 and x_2 variables (let z denote the other variables), and $\partial_{1,2}V(x_1, x_2, z) \equiv 0$, then $V(x_1, x_2, z) = V_1(x_1, z) + V_2(x_2, z)$, where V_i is C^2 in the x_i variable.
Proof of sublemma: It holds

$$\partial_1 V(x_1, x_2, z) - \partial_1 V(x_1, 0, z) = \int_0^{x_2} \partial_{1,2} V(x_1, s, z)\, ds = 0.$$

Again by the fundamental theorem of calculus we have

$$V(x_1, x_2, z) - V(0, x_2, z) - V(x_1, 0, z) + V(0, 0, z) = 0,$$

i.e. $V(x_1, x_2, z) = V_1(x_1, z) + V_2(x_2, z)$. From $V_1 = V - V_2$, and $\partial_1 V_2 \equiv 0$, the rest of the claim follows immediately.

The proof of the proposition follows by induction. For $k = 2$ the sublemma implies the statement. For the induction step, assume that $V(x) = \sum_{i=1}^{\ell} V_i(x_i, z)$, where $z = (x_{\ell+1}, \ldots, x_k)$ and the V_i are all C^2. It follows for all i

$$0 = \partial_{i,\ell+1} V(x) = \partial_{i,\ell+1} V_i(x_i, x_{\ell+1}, z'),$$

where $z' = (x_{\ell+2}, \ldots, x_k)$. The sublemma implies

$$V_i(x_i, x_{\ell+1}, z') = V_i^{(1)}(x_i, z') + V_i^{(2)}(x_{\ell+1}, z'),$$

$V_i^{(1)}$ and $V_i^{(2)}$ being C^2 in x_i and $x_{\ell+1}$, respectively. Hence we have $V(x) = \sum_{i=1}^{\ell+1} W_i(x_i, z')$, where

$$\begin{aligned} W_i(x_i, z') &= V_i^{(1)}(x_i, z'), \quad i = 1, \ldots, \ell, \\ W_{\ell+1}(x_{\ell+1}, z') &= \sum_{j=1}^{\ell} V_j^{(2)}(x_{\ell+1}, z'). \end{aligned}$$

This completes the proof. □

References

[1] D. F. Delchamps, Analytic stabilization and the algebraic Riccati equation, in: Decision and Control, 1983. The 22nd IEEE Conference on, (dec. 1983), pp. 1396 –1401.
[2] D. Hinrichsen and A. J. Pritchard, Stability radii of linear systems, Syst. Control Lett. **7**(February), 1–10 (1986).
[3] M. M. Konstantinov, P. H. Petkov, and N. D. Christov, Perturbation analysis of the continuous and discrete matrix Riccati equations, in: Proc. 1986 ACC, Seattle, WA, (1986), pp. 636–639.
[4] E. D. Sontag, Mathematical Control Theory: Deterministic Finite Dimensional Systems, 2. edition (Springer, New York, 1998).
[5] L. N. Trefethen and M. Embree, Spectra and pseudospectra: the behavior of nonnormal matrices and operators, 2. edition (Princeton University Press, 2005).
[6] J. M. Varah, On the separation of two matrices, SIAM J. Num. Anal. **16**(2), 216–222 (1979).