Stabilization of Large Linear Systems

Chunyang He [1] Volker Mehrmann [1]

Abstract

We discuss numerical methods for the stabilization of large linear multi-input control systems of the form $\dot{x} = Ax + Bu$ via a feedback of the form $u = Fx$. The method discussed in this paper is a stabilization algorithm that is based on a subspace splitting. This splitting is done via the matrix sign-function method. Then a projection into the unstable subspace is performed followed by a stabilization technique via the solution of an appropriate algebraic Riccati equation. There are several possibilities to deal with the freedom in the choice of the feedback as well as in the cost functional used in the Riccati equation. We discuss several optimality criteria and show that in special cases the feedback matrix F of minimal spectral norm is obtained via the Riccati equation with the zero constant term. A theoretical analysis about the distance to instability of the closed loop system is given and furthermore numerical examples are presented that support the practical experience with this method.

1. Introduction

Consider a linear control system

$$\dot{x} = Ax + Bu, \quad x(0) = x_0, \tag{1}$$

where A is a real $n \times n$ matrix and B a real $n \times m$ matrix. We discuss the problem of chosing a real $m \times n$ feedback matrix F such that the feedback $u = Fx$ stabilizes the system, i.e. $A + BF$ has all eigenvalues in the open left half plane. Stabilization is an important task in many applications. Apart from the obvious applications in control a similar problem also arises in the construction of methods for the solution of parabolic partial differential equations [1].

Stabilizing feedback matrices can be chosen in several different ways. A method that is often used is pole placement, see [2] and the references therein. This method observes a lot of difficulties in the numerical solution.

To illustrate the difficulties consider the following example which, like all other examples in this paper, were performed in MATLAB version 4.1 on an HP 715-33 workstation, with machine epsilon $2.22 * 10^{-16}$.

Example 1 Let $A = \mathrm{diag}(.1, .2, .3, .4, .5, .6)$, $B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}^T$ Suppose that we wish to assign the eigenvalues $-6, -5, -4, -3, -1.1, -1$. Consider the well-known pole-placement methods suggested in [3] and [4]. When computing the eigenvalues of $A + BF$ one finds that in both cases the assigned eigenvalues have only one correct digit. One reason for this bad result is that the spectral norm of the feedback matrix is very large, $\|F\|_2 = 3.7879 * 10^6$.

In a multi-input system, the situation becomes somewhat better.

Example 2 Let $A = \mathrm{diag}(.1, .2, .3, .4, .5, .6, .7, .8)$ and $B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 & 7 \\ 3 & 4 & 5 & 6 & 7 & 8 & 7 & 6 \end{bmatrix}^T$. Suppose we wish to obtain the eigenvalues $-8, -7, -6, -5, -4, -3, -1.1, -1$. The pole-placement method of [3] produces eigenvalues with one correct digit and $\|F\|_2 = 8.0802 * 10^6$ and the method in [4] yields three correct digits and $\|F\|_2 = 3.3004 * 10^5$.

The above examples are not exceptional. When the dimension of the system is larger than 10, then most closed loop systems obtained via pole assignment have eigenvalues which are very sensitive to perturbations.

In an extensive test we took A as a 10×10 random matrix, B as a 10×1 random matrix in MATLAB and $-1, -2, \ldots, -10$ as eigenvalues to be assigned. In 100 testruns we found that in 86 cases the norm of feedback matrix was larger than 10^7 and in the other 14 examples the norm of feedback matrix was larger than 10^6.

This suggests that the pole placement problem is probably intrinsically ill-conditioned, see also [2] and for large scale control problems, the situation becomes even more unsatisfactory.

The biggest difficulty with pole placement is that it is not known how the eigenvalues should be chosen to guarantee that the feedback matrix has a

[1] Fakultät für Mathematik, TU Chemnitz-Zwickau, PSF 964, D-09009 Chemnitz. Research support by DFG-Grant La 767/3-2, Große unsymmetrische Gleichungssysteme in Forschergruppe *Scientific Parallel Computing*, Chemnitz

small norm. In multi-input systems, where there is freedom in the choice of the matrix F, this freedom can be used to minimize the norm of F [5]. Another approach in the multi-input case is to minimize the condition number of the eigenvector matrix of the closed loop system over all possible feedback matrices that assign the required eigenvalues [6]. This choice minimizes a bound for the distance to instability [6]. Also here it is not known how one should choose the eigenvalues so that the optimal condition number is small. *Despite these difficulties, for small dimensions pole placement is often used successfully in practice.*

Another stabilization method, that has often better numerical properties is the solution of an appropriately chosen linear quadratic optimal control problem:

$$J = \min_u \int_0^\infty (x^T Q x + u^T R u)\, dt \qquad (2)$$

subject to (1) with appropriately chosen nonnegative definite matrix $Q = C^T C$ and positive definite matrix R.

The standard theory for such optimal control problems, e.g. [7, 8], shows that if (A, B) is stabilizable and (A, C) is detectable, then the linear quadratic optimal problem (2), (1) has the unique solution

$$u = Fx = -R^{-1} B^T X x, \qquad (3)$$

where X is the unique nonnegative definite solution of the algebraic Riccati equation

$$A^T X + XA - XBR^{-1}B^T X + C^T C = 0 \qquad (4)$$

and the corresponding closed loop system

$$\dot{x} = (A + BF)x = (A - BR^{-1}B^T X)x \qquad (5)$$

is stable. (The pair of matrices (A, B) is said to be *stabilizable* if Rank $(\lambda I - A, B) = n$ for all $\lambda \in \mathcal{C}$ with nonnegative real part. (A, C) is said to be *detectable* if (A^T, C^T) is stabilizable.)

Thus, by finding the nonnegative definite solution of the Riccati equation, the system can be stabilized. But one still has the choice of the cost matrices $R, Q = C^T C$ and clearly these should be chosen, so that the closed loop system is insensitive to perturbations. At least it should be guaranteed that small perturbations do not make the system unstable again. Typically for this approach the cost matrix $Q = 0$ is chosen in which case the Riccati equation reduces to a Lyapunov equation for the inverse of X [9]. This choice of Q can be motivated from the fact, that this choice leads to a minimum norm feedback. We will discuss this in Section 4.

This approach was already used in the classical stabilization algorithms which were based on the reduction to Schur form [10, 11, 4]. These methods work efficiently for small and medium sized problems ($n \leq 500$).

For large scale control problems ($n > 500$) none of the approaches discussed previously is feasible. The pole placement problem for such systems is extremely ill-conditioned and for the Riccati approach we essentially have to compute an n-dimensional invariant subspace of a $2n \times 2n$ Hamiltonian matrix, plus a matrix inversion [12]. Furthermore even if the system matrices are sparse, the solution X will be a full matrix.

For such large problems, therefore, other methods have to be considered. One suggestion that has been made is to use partial pole placement [13], but the difficulty of this approach is the same as that for the standard pole placement problem. Other suggestion are the use of iterative methods in the solution of the algebraic Riccati equation, or the Lyapunov equation which occurs in Newton's method applied to the Riccati equation, [14, 15, 16, 17]. None of these approaches is satisfactory so far, since it is difficult to guarantee that the stabilizing solution of the Riccati equation is obtained and also to guarantee the convergence of the iterative method.

We can summarize our previous discussion as follows: Given the problem to stabilize system (1) via feedback, we can do this via pole placement or the solution of a linear quadratic control problem. In both cases there is a lot freedom in the design of the problem. In the pole placement approach the eigenvalues can be chosen freely, in the optimal control approach the cost function is still free of choice. Currently for both approaches an optimal method is not known. On the other hand several measures of optimality can be considered. Whichever measure we choose, we should head for a closed loop system which is insensitive to perturbations in the feedback matrix F in order to guarantee that the computed closed loop system is really stable.

In order to achieve this we should try to maximize the distance to instability [18], i.e. the smallest perturbation which makes the system unstable

$$\delta(A) := \min_\mu \sigma_n(A - \mu i I), \qquad (6)$$

where σ_n denotes the smallest singular value. If $A + BF$ is diagonalizable and $A + BF = W \Lambda W^{-1}$ is the spectral decomposition of the closed loop matrix, then a lower bound for the distance to instability for the closed loop system is given by

$$\frac{1}{\mathrm{cond}_2(W)} \delta(\Lambda) \leq \delta(A + BF). \qquad (7)$$

Thus minimizing $\mathrm{cond}_2(W)$ will maximize a lower bound for the distance to instability. A pole placement method that minimizes $\mathrm{cond}(W)$ among all possible choices of feedback that assign the correct poles was introduced in [6]. This method, however, is very costly and unfeasible for large control problems.

In view of all these difficulties we suggest a new stabilization approach which is feasible for large sized problems $n < 5000$, where A has only few (less than 100) unstable poles.

This new approach is very closely related to the classical method suggested by Varga [9]. But instead of computing the Schur form which is infeasible for large matrices we computed the subspace splitting via the matrix sign-function method, (see [19] and the references therein), which has recently received quite a lot of interest due to its inherent parallelizability. We use the sign function method to split the complete space $\mathcal{R}^n$ into two subspaces which are the real invariant subspaces of A with respect to the stable eigenvalues and unstable eigenvalues, respectively. Using a projection method similar to that suggested in [13], the problem is reduced to a subsystem problem for the stabilization in the invariant subspace corresponding to the unstable eigenvalues. Based on the Schur method such an approach was previously suggested in [9]. For the subproblem we show that the optimal choice of a cost functional in the Riccati approach is obtained with $Q = 0$ so that this stabilization problem can be obtained via the solution of a Lyapunov equation for $Y = X^{-1}$:

$$\tilde{A}Y + Y\tilde{A}^T - \tilde{R} = 0. \tag{8}$$

Lifting the solution of (8) into the complete space we obtain a stabilizing feedback for the original problem.

The paper is organized as follows: In Section 2, we discuss the relationship between the cost functionals for the complete problem and the stabilization problem.

In the Section 3 we discuss the choice of the cost functional that leads to minimal norm of the feedback gain matrix.

The distance to instability is discussed in Section 4 to show that reducing $\|F\|_2$ is a crucial point in the stabilization problem.

In Section 5 some numerical examples are given to support the theoretical results.

Throughout the paper $\lambda(A)$ denotes the spectrum of A, I the identity matrix of suitable size, and we write $A \leq B$ ($A < B$) if $B - A$ is nonnegative definite (positive definite). We also assume that (A, B) is stabilizable. We denote the open left half plane by $\mathcal{C}^-$.

2. Subspace Splitting and Stabilization

In this section we discuss the use of the matrix sign-function to split the complete space into two invariant subspaces of A with respect to its stable and unstable eigenvalues.

The matrix sign function was first introduced by Roberts in a technical report (which appeared only significantly later [20]) as

$$\operatorname{sign}(A) := \frac{1}{\pi i} \int_\gamma (zI - A)^{-1} dz - I, \tag{9}$$

where γ is any closed connected set in the complex plane containing all eigenvalues of A with positive real part. An alternative definition in matrix terminology was given by Beavers and Denman [21, 22] using the Jordan canonical form.

We will use the following definition via the Schur-form, see [23]:

Suppose that A has the real Schur form, e.g. [24],

$$W^T A W = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \tag{10}$$

where we assume that all eigenvalues of A_{11} are in the left half plane and all eigenvalues of A_{22} are in the right half plane. Let Y be a solution of the Sylvester equation $YA_{22} - A_{11}Y = 2A_{12}$, then the sign function is given by

$$\operatorname{sign}(A) := W \begin{bmatrix} -I_p & Y \\ 0 & I_{n-p} \end{bmatrix} W^T. \tag{11}$$

Soon after the introduction of the sign function it was recognized that it can be useful method for the computation of eigenvalues and invariant subspaces [21, 25, 22, 26, 19, 27, 28, 29, 20] and for the solution of Riccati and Sylvester equations [25, 30, 31].

The difficulty with the matrix sign function is that it is not defined for matrices with eigenvalues on the imaginary axis and that the evaluation of the sign function is an ill-conditioned problem for matrices with eigenvalues close to this axis, due to the discontinuouity of the sign-function in the matrix elements. This is one reason why for quite some time the matrix sign function method has been associated with being an unstable method. But as recent results show [27, 23] for the computation of invariant subspaces it can be considered as accurate as for the example the transformation to Schur form. For matrices with eigenvalues on or near the imaginary axis, however, special activities have to be devised.

As in the Schur form, the calculation of the unstable invariant subspace via the matrix sign function leads to an orthogonal matrix W such that the system matrix is transformed as in (10) and

$$W^T B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}. \qquad (12)$$

Let $W = (W_1, W_2)$ be partitioned analogous to the partitioning in $W^T AW$. Since A_{22} is unstable, the problem of stabilizing the system given by A, B is then reduced to the stabilization of the smaller system

$$\dot{x}_2 = A_{22} x_2 + B_2 u. \qquad (13)$$

Suppose that there is a feedback matrix F_2 such that $Re(\mu) < 0$ for all $\mu \in \lambda(A_{22} + B_2 F_2)$, then the closed loop system obtained with the feedback matrix $F = \begin{bmatrix} 0 & F_2 \end{bmatrix} W^T = F_2 W_2^T$ is stable. This is seen immediately from

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \begin{bmatrix} 0 & F_2 \end{bmatrix}$$
$$= \begin{bmatrix} A_{11} & A_{12} + B_1 F_2 \\ 0 & A_{22} + B_2 F_2 \end{bmatrix}.$$

The difficulty with this approach, as well as with most of the pole placement approaches, is that the closed loop system may be much more ill-conditioned than the original system. Here ill-conditioning means that the eigenvalues of the closed loop system can change drastically if they are subject to small perturbations. This could mean that small perturbations may move the eigenvalues back into the right half plane. And such a situation is certainly not acceptable in practice. Thus we must be careful in the choice of the feedback matrix F. But since there is a lot of freedom in the choice of F, we may use this freedom to make the closed loop system as well-conditioned as possible.

To analyze the freedom we use the following results on the stabilization properties of the nonnegative definite solution of the Riccati equation. The first result is essentially due to Willems [32].

Lemma 1 *Let X_i, $i = 1, 2$ be real symmetric nonnegative definite solutions of the algebraic Riccati equations*

$$A^T X + XA - XBR^{-1}B^T X + Q_i = 0, \quad i = 1, 2,$$

respectively, such that

$$\lambda(A - BR^{-1}B^T X_i) \subset \mathcal{C}^-, \; i = 1, 2.$$

Then $0 \le Q_2 \le Q_1$ implies $J(Q_2) \le J(Q_1)$, where the cost functional $J(Q)$ as a function of Q is defined as in (2).

Proof: Inserting the optimal feedback (3) for the equation with Q_1 into the cost functional with cost matrix Q_2 we obtain

$$\begin{aligned} J(Q_2) &= \min_u \textstyle\int_0^\infty (x^T Q_2 x + u^T Ru) dt \\ &\le \textstyle\int_0^\infty (x^T Q_2 x + x^T X_1 BR^{-1} B^T X_1 x) dt. \end{aligned}$$

Then, since $Q_2 \le Q_1$, we have

$$J(Q_2)$$
$$\le \int_0^\infty (x^T Q_1 x + x^T X_1 BR^{-1}B^T X_1 x) \; dt = J(Q_1),$$

which completes the proof. $\square$

Using this lemma, we can show how in some special cases the value of the cost functional $J(Q)$ is decreased.

It is clear that the optimal value of the cost functional decreases if the column dimension of the matrix B is increased by adding columns. This follows, since

$$\min_{u_1,u_2} \int_0^\infty (x^T Q x + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}^T \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}) dt$$

$$\le \min_{u_1} \int_0^\infty (x^T Q x + u_1^T R_{11} u_1) dt,$$

where the minima are taken subject to $\dot{x} = Ax + [B_1, B_2] \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ or $\dot{x} = Ax + B_1 u_1$, respectively. This is obtained directly by chosing $u_2 = 0$.

Furthermore we have the following theorem:

Theorem 2 *Let*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix},$$

with A in Schur form. Let $Q = \mathrm{diag}(Q_1, Q_2)$, where Q_1 is positive semidefinite of appropriate dimension and let $J(Q)$ be the minimum of the cost functional. Furthermore let that

$$J_2 := \int_0^\infty (x_2^T Q_2 x_2 + x_2^T X_2 B_2 R^{-1} B_2^T X_2 x_2) dt$$

be the value of the cost functional obtained by inserting the optimal control $u_2 = -R^{-1}B_2^T X_2 x_2$ obtained via the algebraic Riccati equation

$$A_{22}^T X_2 + X_2 A_{22} - X_2 B_2 R^{-1} B_2^T X_2 + Q_2 = 0$$

for the subsystem $\dot{x}_2 = A_{22} x_2 + B_2 u(t)$. Then $J_2 \le J(Q)$.

Proof: The appended matrix $\tilde{X} = \mathrm{diag}(0, X_2)$ satisfies the Riccati equation

$$A^T \tilde{X} + \tilde{X} A - \tilde{X} BR^{-1} B\tilde{X} + \mathrm{diag}(0, Q_2) = 0.$$

As $\operatorname{diag}(0, Q_2) \leq Q$, Lemma 1 implies that $J_2 = J(\operatorname{diag}(0, Q_2)) \leq J(\operatorname{diag}(Q_1, Q_2))$. □

From these results we see that by increasing the dimension of B and by appropriately decreasing the matrix Q in the cost functional, we can decrease the minimum of the cost functional. Increasing the column dimension m of B is usually not an option, but the cost functional can still be chosen freely. In the next section we discuss how to choose the cost functional in order to minimize the norm of the feedback matrix.

3. The Minimum Norm Feedback

In this section, we consider the stabilization problem for a problem where the system matrix is completely unstable as is the case for our projected subproblem (13).

The following Theorem is probably well-known, but we do not know a reference.

Theorem 3 *If we consider the cost functional as a function of Q then*

$$\min_{Q \geq 0} J(Q) = J(0).$$

Furthermore suppose that $Re(\lambda) > 0$ for all $\lambda \in \lambda(A)$ and suppose that

$$A^T X + XA - XBR^{-1}B^T X = 0, \qquad (14)$$

has a nonsingular solution X. If $F = -R^{-1}B^T X$ is the corresponding feedback, then the eigenvalues $A + BF$ are the negatives of the eigenvalues A.

Proof: The first part of the theorem follows trivially from Lemma 1. For the second part observe that we can rewrite the Riccati equation (14) as

$$X(A + BF) = -A^T X.$$

Since all eigenvalues of A have negative real part, X is assumed nonsingular, it follows that the eigenvalues of $A + BF$ are those of $-A^T$. □

As we are looking for a nonsingular solution of the degenerate Riccati equation we can equivalently solve the Lyapunov equation

$$AY + YA^T = BR^{-1}B^T, \qquad (15)$$

where $X = Y^{-1}$, [10].

For relatively small sized Lyapunov equations there are efficient algorithms available [33, 34]. For large problems there is recently a lot of interest in iterative methods, see [16] and the references therein.

Remark 1 We have assumed that $Re(\lambda) > 0$ for all eigenvalues of A. Without this restriction the minimum of the function $J(Q)$ still occurs for $Q = 0$. The nonnegative definite solution X of the Riccati equation (4) with $Q = 0$ has rank equal to the number of eigenvalues of A with positive real parts, see [35]. The eigenvalues of $A + BF$ are a combination of eigenvalues with negative real part of A and negatives of eigenvalues of A with positive real parts. However we still have to assume that A has no pure imaginary eigenvalues, since otherwise the sign-function approach is not feasible.

Although the value of the cost functional $J(Q)$ partly reflects the size of $\|F\|_2$, we are merely interested in minimal values for $\|F\|_2$. This is, however, still an open problem and we only present the result in the case that B is a nonsingular matrix. We begin with another Lemma of Willems [32].

Lemma 4 *Let X_i, $i = 1, 2$ be real symmetric solutions of the algebraic Riccati equations*

$$A^T X_i + X_i A - X_i BR^{-1}B^T X_i + Q_i = 0, \; i = 1, 2$$

respectively and assume that all eigenvalues of $A - BR^{-1}B^T X_1$ have negative real part. Then $0 \leq Q_2 \leq Q_1$ implies $X_2 \leq X_1$.

Using this Lemma we can prove the following theorem:

Theorem 5 *Suppose that all eigenvalues of A have positive real part. Let B be square nonsingular and let $R = (B^T B)^{1/2}$ be the positive square root of $B^T B$ (cf. [24]). Let X be the nonnegative definite solution of the algebraic Riccati equation*

$$A^T X + XA - XBR^{-1}B^T X + Q = 0,$$

so that all eigenvalues of $A - BR^{-1}B^T X$ have negative real part. Then the minimum norm feedback matrix F taken over all positive semidefinite matrices Q occurs for $Q = 0$. It is given by $F = -R^{-1}B^T X$, where X is the positive definite solution of the degenerate Riccati equation (14). Furthermore the eigenvalues of $A + BF$ are the negatives of those of A.

Proof: Let X_1 and X_2 be the nonnegative definite solutions of the Riccati equations

$$A^T X + XA - XBR^{-1}B^T X + Q_i = 0, \quad i = 1, 2$$

for $0 \leq Q_2 \leq Q_1$. Let $F_i = -R^{-1}B^T X_i, i = 1, 2$. Then Lemma 4 implies $X_2 \leq X_1$. Thus $\|X_2\|_2 \leq \|X_1\|_2$. Observe that $R^{-1}B^T$ is an orthogonal matrix and therefore $\|F_1\|_2 = \|X_1\|_2$ and $\|F_2\|_2 = \|X_2\|_2$. Thus $\|F_2\|_2 \leq \|F_1\|_2$ and the minimum of $\|F\|_2$ occurs at $Q = 0$. □

Remark 2 It is hard to prove that Theorem 5 is true for general B. In order to prove $||R^{-1}B^TX_2||_2 \le ||R^{-1}B^TX_1||_2$, we would need to prove that

$$\max_y(y^TR^{-1}B^TX_2^2BR^{-1}y) \\ \le \max_y(y^TR^{-1}B^TX_1^2BR^{-1}y).$$

This inequality, however, does not hold in general as the following example demostrates. Take

$$B = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ R = I,\ X_1 = \begin{bmatrix} 6 & -4 \\ -4 & 20 \end{bmatrix}, \\ X_2 = \begin{bmatrix} 5 & -6 \\ -6 & 8 \end{bmatrix}.$$

Then $X_1 - X_2 = \begin{bmatrix} 1 & 2 \\ 2 & 12 \end{bmatrix}$ is positive definite but $||R^{-1}B^TX_2||_2 = \sqrt{61} > \sqrt{52} = ||R^{-1}B^TX_1||_2$. On the other hand X_1, X_2 and B are not independent of each other so we may expect that Theorem 5 holds for a much wider class of problems with $m < n$.

On the other hand we apply Theorem 5 only for the (usually) small subsystem (13), for which we may even have more inputs than states. In this sense our result is appropriate in quite general situations.

It is natural to ask what happens when the minimization problem includes R. The answer is, that minimizing the norm of feedback matrices among all $0 \le Q$ is usually sufficient, since we can always scale the problem so that $||R||_2 = 1$ (see [36]). In fact, let $\alpha = ||R||_2$, then $\tilde{X} = X/\alpha$ satisfies the Riccati equation

$$A^T\frac{X}{\alpha} + \frac{X}{\alpha}A - \frac{X}{\alpha}B\tilde{R}^{-1}B^T\frac{X}{\alpha} + \frac{Q}{\alpha} = 0,$$

where $\tilde{R} = R/\alpha$. Observe that the feedback matrices produced by both Riccati equations are same, i.e.

$$\tilde{F} = -\tilde{R}^{-1}B^T\tilde{X} = -R^{-1}B^TX = F.$$

Example 3 Consider the system given by

$$A = \begin{bmatrix} 0.1 & 1 & 10 & 0 & 0 \\ -1 & 0.1 & 0 & 10 & 0 \\ & & 2 & 1 & 10 \\ & & -1 & 2 & 0 \\ & & & & 5 \end{bmatrix}, B = \begin{bmatrix} 5 & 4 & 3 \\ 4 & 5 & 4 \\ 3 & 4 & 5 \\ 1 & 3 & 4 \\ 1 & 1 & 3 \end{bmatrix},$$

and let $R = \alpha I$ and $Q = \beta I$. The following table shows the optimal stabilizing feedback $||F||_2$ as a function of α and β.

$\beta \setminus \alpha$	10^{-4}	10^{-2}	1	10^2	10^4
10^{-4}	9.80	6.41	6.01	5.98	5.98
10^{-2}	23.7	9.80	6.41	6.01	5.98
1	147	23.7	9.80	6.41	6.01
10^2	1397	147	23.7	9.80	6.41
10^4	10^4	1397	147	23.7	9.80

The Toeplitz structure of the above table is in accordance with our theoretical analysis, only one parameter plays a role. The minimum norm feedback matrix F with $||F||_2 = 5.9833$ occurs at $\beta = 0, \alpha = 1$ and $\lambda(A+BF) = \{-5.0000, -0.1000 \pm 1.0000i, -2.0000 \pm 1.0000i\}$.

In this section we have discussed the minimization of the feedback F in two different measures, the value of the cost functional $J(Q)$ and $||F||_2$. In the first case and in special situations also in the second case the optimal F is obtained for the choice $Q = 0$ in the cost functional.

In the next question we discuss as a different measure of optimality the distance to instabilty.

4. Distance to Instability

We have already discussed in the introduction that the goal in using the freedom in the choice of the stabilizing feedback matrix is to make the closed loop system insensitive to perturbations. In the previous section we have attempted to minimize the norm of the feedback matrix to achieve this goal. In this section we try to maximize the distance to instability. If we could solve this problem, we would certainly obtain the best choice in terms of robustness.

Let A be stable, then the distance of A to the nearest matrix in the set of unstable matrices is measured as follows [18]:

$$\delta(A) = \min_{\mu} \sigma_n(A - \mu iI),$$

where σ_n is the smallest singular value of $A - \mu iI$ and $i = \sqrt{-1}$. For small sized problems efficient algorithms are available for computing this distance [37, 38]. We wish to solve the optimization problem

$$\max_F \delta(A + BF),$$

among all feedback matrices F that stabilize the system. Suppose that $A+BF$ is diagonalizable and $A + BF = W\Lambda W^{-1}$ is the spectral decomposition. Then it is clear that a lower bound for $\delta(A + BF)$ is given by

$$\frac{1}{\mathrm{cond}_2(W)}\delta(\Lambda) \le \delta(A + BF).$$

In [6] a robust pole assignment algorithms is based on minimizing the lower bound $\mathrm{cond}_2(W)$ for $\delta(A+BF)$. However, if eigenvalues are close, then this bound can be arbitrary small, even though the $\delta(A + BF)$ is not [39]. Also the method that optimizes $\mathrm{cond}_2(W)$ is not feasible for large scale problems.

In the following theorem the above lower bound is improved. To do this, the matrix $A+BF$ is assumed to be in block Schur form (10). Then the *separation* of A_{11} and A_{22}, $\mathrm{sep}(A_{11}, A_{22})$, is defined as the smallest singular value of Y, where $A_{11}Y - YA_{22} = A_{12}$ [40].

Theorem 6 *Let F be a feedback matrix derived from the stabilization algorithm such that all eigenvalues of $A+BF$ have negative real part. Suppose that $A+BF$ has the Schur form*

$$A = W \begin{bmatrix} A_{11} & A_{12}+B_1F_2 \\ 0 & A_{22}+B_2F_2 \end{bmatrix} W^T.$$

Then

$$\frac{\eta^2}{(\eta + \|A_{12}+B_1F_2\|_2)^2} \min\{\delta(A_{11}), \delta(A_{22}+B_2F_2)\} \leq \delta(A+BF),$$

where

$$\eta := \min_{\mu} sep(A_{11}-\mu iI, A_{22}+B_2F_2-\mu iI).$$

Proof: The result is true for $\eta = 0$. So we may assume that $\eta \neq 0$ and hence there exists a matrix Y satisfying

$$\begin{aligned} &(A_{11}-\mu iI)Y - Y(A_{22}+B_2F_2-\mu iI) \\ &+(A_{12}+B_1F_2) = 0, \end{aligned} \tag{16}$$

such that

$$\begin{bmatrix} I & -Y \\ 0 & I \end{bmatrix} \begin{bmatrix} A_{11}-\mu iI & A_{12}+B_1F_2 \\ 0 & A_{22}+B_2F_2-\mu iI \end{bmatrix} * \begin{bmatrix} I & Y \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11}-\mu iI & 0 \\ 0 & A_{22}+B_2F_2-\mu iI \end{bmatrix}.$$

It follows that

$$\begin{aligned} &\sigma_n(\begin{bmatrix} A_{11}-\mu iI & 0 \\ 0 & A_{22}+B_2F_2-\mu iI \end{bmatrix}) \\ &\leq \mathrm{cond}(\begin{bmatrix} I & -Y \\ 0 & I \end{bmatrix})* \\ &\sigma_n(\begin{bmatrix} A_{11}-\mu iI & A_{12}+B_1F_2 \\ 0 & A_{22}+B_2F_2-\mu iI \end{bmatrix}) \\ &\leq (1+\max_\mu \|Y\|_2)^2 * \\ &\sigma_n(\begin{bmatrix} A_{11}-\mu iI & A_{12}+B_1F_2 \\ 0 & A_{22}+B_2F_2-\mu iI \end{bmatrix}). \end{aligned} \tag{17}$$

Then by (16) we obtain that

$$\|Y\|_2 \leq \frac{\|A_{12}+B_1F_2\|_2}{sep(A_{11}-\mu iI, A_{22}+B_2F_2-\mu iI)} \leq \frac{\|A_{12}+B_1F_2\|_2}{\eta}.$$

Substituting the upper bound into (17) completes the proof. □

If the norm of $A_{12}+B_1F_2$ is large or if there are close eigenvalues of A_{11} and $A_{22}+B_2F_2$ then the lower bound may be very small. It is often the case that close eigenvalues make only the lower bound small but not $\delta(A+BF)$ [39]. Since the norm of F is minimized via the solution of the Lyapunov equation (14), the stabilization method in general does not make $\delta(A+BF)$ much smaller than $\delta(A_{11})$. This observation is demonstrated in our numerical examples in the next section.

5. Numerical Examples

Based on the previous discussion we suggest the following stabilization method for a large control system with only few unstable poles:

Algorithm 1 Stabilization

Input: System matrices $A \in \mathbf{R}^{n\times n}$, $B \in \mathbf{R}^{n\times m}$.

Output: Feedback matrix $F \in \mathbf{R}^{m\times n}$ such that $A+BF$ is stable.

1. **Newton iteration for the matrix sign function**
   Iterate
   $$X_0 = A,\ X_{k+1} = (X_k + X_k^{-1})/2,\ k = 1,2,\ldots$$
   until $\|X_{k+1}-X_k\|_1 \leq 1000 * N\|X_{k+1}\|_1^2\ eps$, where eps is the machine precision. Set $\mathrm{sign}(A) := X_{k+1}$.
2. **Computation of an orthogonal basis for the unstable invariant subspace.**
   Compute the QR decomposition with column pivoting (see [24])
   $$W^T(\mathrm{sign}(A) - I)\Pi = \begin{bmatrix} S_{11} & S_{12} \\ 0 & 0 \end{bmatrix},$$
   and partition $W = (W_1, W_2)$ analogously.
3. Set $A_{22} = W_2^TAW_2,\ B_2 = W_2^TB$.
4. Solve the Lyapunov equation
   $$A_{22}Y + YA_{22}^T = B_2B_2^T$$
   using for example the Bartels/Stewart algorithm (cf. [24]).
5. Set $F_2 = -B_2^TY^{-1}$ and set $F = \begin{bmatrix} 0 & F_2 \end{bmatrix} W^T$.

Remark 3 We have the following comments on this algorithm:

1. A detailed analysis of the numerical properties of methods for the compution of the matrix sign function is given in [19, 27, 23]. In particular it is shown in [23] that the matrix sign function method

is as good as the Schur method for computing invariant subspaces.

2. Other methods for the computation of the matrix sign-function can be used to replace Newton's method in Step 1., see [19, 27]. Also other methods for the solution of the Lyapunov equation can be employed, see [24].

3. In our algorithm, we have chosen $R = I$ in the cost functional that leads to the Lyapunov equation. There are of course other choices, for example $R = (B^T B)^{1/2}$, but numerical examples indicate that the choice $R = I$ is better.

Example 4 For Example 1 Algorithm 1 yields a feedback with $\|F\|_2 = 463.2583$ and for Example 2 the norm of the feedback is $\|F\|_2 = 204.7319$. A comparison of the pole assignment methods and the stabilization method via the Riccati equation shows that the norm $\|F\|_2$ is drastically reduced.

Example 5 This example demonstrates that when the dimension of the input matrix becomes larger, the norm of the optimal feedback matrix F goes down. Let

$$A = \begin{bmatrix} 0.1 & 10 & & & \\ & 0.3 & 10 & & \\ & & 2 & 10 & \\ & & & 4 & 10 \\ & & & & 5 \end{bmatrix},$$

$$B = \begin{bmatrix} 5 & 4 & 3 & 1 & 1 \\ 4 & 5 & 4 & 3 & 1 \\ 3 & 4 & 5 & 4 & 3 \\ 1 & 3 & 4 & 5 & 4 \\ 1 & 1 & 3 & 4 & 5 \end{bmatrix}.$$

By adding in successively more columns in B, we have the following norms of optimal feedback matrices .

m	1	2	3	4	5
$\|F\|_2$	16.10	11.55	3.99	2.28	1.80

Example 6 This example is taken from [19]. It was originally suggested by Chatelin. It has the form

$$A = Q(D+N)Q^T,$$

where Q is a random orthogonal matrix and N is a matrix with $n_{ij} = \begin{cases} \alpha, & \text{if } j = i+30; \\ 0, & \text{otherwise} \end{cases}$ and $\alpha = 10$. D has the form $D = \mathrm{diag}(D_1, \ldots, D_{350}, 1, 2, 3, \ldots, 10)$ with $D_k = \begin{bmatrix} x_k & y_k \\ -y_k & x_k \end{bmatrix}$, where $x_k = -y_k^2/10$, $y_k = -0.1k$, $k = 1, 2, \ldots, 350$ and B is a random $n \times 15$ matrix. The stable eigenvalues of A lie on the curve $x = -y^2/10$.

Due to the large size of the matrix, we were not able to compute the distance to instability. The sign-function method produced a factorization of the form

$$\begin{bmatrix} A_{11} & A_{12} \\ E_{21} & A_{22} \end{bmatrix},$$

where the perturbation E_{21} is very small,

$$\frac{\|E_{21}\|_1}{\|A\|_1} = 3.027 * 10^{-14}.$$

The stabilization procedure produced a feedback matrix F with $\|F\|_2 = 43.662$.

If we take the 105×105 matrix constructed in a similar way with $\alpha = 10^3$, five unstable eigenvalues $1, 2, 3, 4, 5$ and a 105×3 matrix B with rows $[k, 200-k, \sqrt{k}]$, $k = 1, 2 \ldots, 105$ we obtained an analogous result. In this case we computed $\delta(A_{11}) = 6.3145 * 10^{-5}$, where the minimum occurs for the parameter $\mu = 1.0002$. Our computation yields $\delta(A + BF) = 9.4497 * 10^{-7}$ and $\mu = 0.9961$.

Example 7 We ran 100 tests with a random 100×100 system matrix A with less than 7 unstable eigenvalues and a random 100×6 matrix B. None of the norms of the feedback matrices was larger than 45.

6. Conclusions and Future Work

A new algorithm is presented to do stabilization for a large linear control system with only a few unstable eigenvalues. The method is based on the matrix sign function method and the solution of a small Lyapunov equation. Both theoretical results and numerical examples are presented to analyze the properties of this new algorithm.

An argument that is often used in favor of pole placement algorithms is that pole placemnt allows to place the poles in specified regions for example angular sectors in the left half plane. The approach that we discussed in this paper, i.e. using the sign-function method to split the stable from the unstable subspace, is not limited to this situation. Since the sign-function can be used to locate the eigenvalues in any rectangular or parallelogram domain in the complex plane [19], we can use the described method also to identify the poles which are not in the region we wish them to be in, and use a similar approach to move the poles that are not into the specified region. A detailed analysis of this method is currently under investigation.

We thank Dr. V. Sima for many helpful comments and discussions.

References

[1] H. Jarausch, *Zur numerischen Untersuchung von parabolischen Differentialgleichungen mit Hilfe einer adaptiven spektralen Zerlegung*, Habilitationsschrift, RWTH Aachen, Aachen, 1990.

[2] M. Arnold, *Algorithms and conditioning for eigenvalue assignment*, PhD thesis, Northern Illinois University, Department of Mathematical Sciences, DeKalb, Il., 1993.

[3] G. Miminis and C. C. Paige, "An algorithm for pole assignment of time invariant linear systems", *International Journal of Control*, vol. 35, pp. 341–354, 1982.

[4] A. Varga, "A Schur method for pole assignment", *IEEE Trans. Automat. Control*, vol. AC-27, pp. 517–519, 1981.

[5] L.H. Keel, J.A. Fleming, and S.P. Bhattacharyya, "Minimum norm pole assignment via Sylvester's equation", *Lin. Alg. and its Role in Systems theory, Cont. Math. AMS*, vol. 47, pp. 265–272, 1984.

[6] J. Kautsky, N. Nichols, and P. Van Dooren, "Robust pole assignment in linear state feedback", *Int. J. Control*, vol. 41, pp. 1129–1155, 1985.

[7] T. Kailath, *Systems Theory*, Prentice-Hall, Englewood Cliffs, New York, 1980.

[8] H.W. Knobloch and H. Kwakernaak, *Lineare Kontrolltheorie*, Springer-Verlag, Berlin, 1985.

[9] A. Varga, "On stabilization algorithms for linear time-invariant systems", *Rev. Roum. Sci. Techn.-Electrotechn. et Energ.*, vol. 26, pp. 115–124, 1981.

[10] E.S. Armstrong, "An extension of Bass' algorithm for stabilizing linear continuous constant systems", *IEEE Trans. Automat. Control*, vol. AC-20, pp. 826–829, 1975.

[11] V. Sima, "An efficient Schur method to solve the stabilization problem", *IEEE Trans. Automat. Control*, vol. AC-26, pp. 724–725, 1981.

[12] A.J. Laub, "Invariant subspace methods for the numerical solution of algebraic Riccati equations", in *The Riccati equation*, S. Bittanti, A. Laub, and J. Willems, Eds. 1991, pp. 163–196, Springer-Verlag.

[13] Y. Saad, "Projection and deflations methods for partial pole assignment in linear state feedback", *IEEE Trans. Automat. Control*, vol. AC-33, pp. 290–297, 1988.

[14] G. Starke, *Rationale Minimierungsprobleme in der komplexen Ebene im Zusammenhang mit der Bestimmung optimaler ADI-Parameter*, Dissertation, Universität Karlsruhe, 1989.

[15] A.S. Hodel and K. Poolla, "Heuristic approaches to the solution of very large sparse Lyapunov and algebraic Riccati equations", in *Proceedings of the 27th IEEE Conference on Decision and Control*, 1988, pp. 2217–2222.

[16] I.M. Jaimoukha and E.M. Kasenally, "Krylov subspace methods for solving large Lyapunov equations", *SIAM J. Numer. Anal.*, vol. 31, pp. 227–251, 1994.

[17] I.G. Rosen and C. Wang, "A multilevel technique for the approximate solution of operator Lyapunov and algebraic Riccati equations", Tech. Rep., Center for Applied Mathematical Sciences, University of California, Los Angeles, 1994.

[18] C. F. Van Loan, "How near is a stable matrix to an unstable matrix?", in *Linear Algebra and its Role in Systems Theory*, B. N. Datta, Ed. 1985, vol. 47 of *Contemporary Math.*, pp. 465–478, Amer. Math. Soc.

[19] Z. Bai and J. Demmel, "Design of a parallel nonsymmetric eigenroutine toolbox, Part I", in *Proceeding of the sixth SIAM conference on parallel processing for scientific computing*, 1993.

[20] J. Roberts, "Linear model reduction and solution of the algebraic Riccati equation by the use of the sign function", *Internat. J. Control*, vol. 32, pp. 667–687, 1980.

[21] A.N. Beavers and E.D. Denman, "A computational method for eigenvalues and eigenvectors of a matrix with real eigenvalues.", *Numer. Math.*, vol. 21, pp. 389–396, 1973.

[22] E. Denman and R. Beavers, "The matrix sign function and computations in systems", *Appl. Math. Comput.*, vol. 2, pp. 63–94, 1976.

[23] R. Byers, C. He, and V. Mehrmann, "The matrix sign function method and the computation of invariant subspaces", Tech. Rep., Fakultät f. Mathematik, TU Chemnitz-Zwickau, PSF 964, D-09009 Chemnitz, FRG, 1994.

[24] G.H. Golub and C.F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Md, 1989.

[25] R. Byers, "Solving the algebraic Riccati equation with the matrix sign function", *Linear Algebra Appl.*, vol. 85, pp. 267–279, 1987.

[26] E.D. Denman and J. Leyva-Ramos, "Spectral decomposition of a matrix using the generalized sign matrix", *Appl. Math. and Comput.*, vol. 8, pp. 237–250, 1981.

[27] Z. Bai, J. Demmel, and M. Gu, "Inverse free parallel spectral divide and conquer algorithms for nonsymmetric eigenproblem", Tech. Rep., Berkeley University, 1994.

[28] C. Kenney and A.J. Laub, "Rational iterative methods for the matrix sign function", Report 89–08, Scientific Computation Laboratory, University of California of Santa Barbara, Santa Barbara, 1989.

[29] A.N. Malyshev, "Parallel algorithms for solving some spectral problems of linear algebra", *Lin. Alg. Appl.*, vol. 188/189, pp. 489–520, 1993.

[30] J.D. Gardiner and A.J. Laub, "A generalization of the matrix sign function solution for algebraic Riccati equations", *Internat. J. Control*, vol. 44, pp. 823–832, 1986.

[31] J.D. Gardiner and A.J. Laub, "Parallel algorithms for algebraic Riccati equations", *Internat. J. Control*, vol. 54, pp. 1317–1333, 1991.

[32] J.C. Willems, "Least squares stationary optimal control and the algebraic Riccati equation", *IEEE Trans. Automat. Control*, vol. 16, pp. 621–634, 1971.

[33] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation AX+XB=C", *Algorithm 432 Comm.*, vol. 15, pp. 820–826, 1972.

[34] S.J. Hammarling, "Numerical solution of the stable, nonnegative definite Lyapunov equation", *IMA J. Numer. Anal.*, vol. 2, pp. 303–323, 1982.

[35] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, Academic Press, Orlando, 1985.

[36] R. Byers, "Numerical condition of the algebraic Riccati equation", *Lin. Alg. and its Role in Systems theory. Cont. Math. AMS*, vol. 47, pp. 35–51, 1984.

[37] R. Byers, "A bisection method for measuring the distance of a stable to unstable matrices", *SIAM J. Sci. Stat. Comput.*, vol. 9, pp. 875–881, 1988.

[38] L. Elsner and C. He, "An algorithm for computing the distance to uncontrollability", *Syst. Contr. Letters*, vol. 17, pp. 453–464, 1991.

[39] J. W. Demmel, "A counterexample for two conjectures about stability", *IEEE Trans. Automat. Control*, vol. AC-32, pp. 340–342, 1987.

[40] J. G. Sun and G. W. Stewart, *Matrix perturbation theory*, Academic Press, Orlando, 1990.