A note on approximating the nearest stable discrete-time descriptor system with fixed rank

Nicolas Gillis∗  Michael Karow†  Punit Sharma‡

Abstract

Consider a discrete-time linear time-invariant descriptor system $Ex(k + 1) = Ax(k)$ for $k \in \mathbb{Z}_+$. In this paper, we tackle for the first time the problem of stabilizing such systems by computing a nearby regular index one stable system $\hat{E}x(k + 1) = \hat{A}x(k)$ with $\text{rank}(\hat{E}) = r$. We reformulate this highly nonconvex problem into an equivalent optimization problem with a relatively simple feasible set onto which it is easy to project. This allows us to employ a block coordinate descent method to obtain a nearby regular index one stable system. We illustrate the effectiveness of the algorithm on several examples.

Keywords. stability radius, linear discrete-time descriptor system, stability

1 Introduction

In [17, 7], authors have tackled the problem of computing the nearest stable matrix in the discrete case, that is, given an unstable matrix $A$, find the smallest perturbation $\Delta A$ with respect to Frobenius norm such that $\hat{A} = A + \Delta A$ has all its eigenvalues inside the unit ball centred at the origin. In this paper, we aim to generalize the results in [7] for matrix pairs $(E,A)$, where $E,A \in \mathbb{R}^{n,n}$.

The matrix pair $(E,A)$ is called regular if $\text{det}(\lambda E - A) \neq 0$ for some $\lambda \in \mathbb{C}$, which we denote $\text{det}(\lambda E - A) \not\equiv 0$, otherwise it is called singular. For a regular matrix pair $(E,A)$, the roots of the polynomial $\text{det}(zE - A)$ are called finite eigenvalues of the pencil $zE - A$ or of the pair $(E,A)$. A regular pair $(E,A)$ has $\infty$ as an eigenvalue if $E$ is singular. A regular real matrix pair $(E,A)$ can be transformed to Weierstraß canonical form [8], that is, there exist nonsingular matrices $W,T \in \mathbb{C}^{n,n}$ such that

$$E = W \begin{bmatrix} I_q & 0 \\ 0 & N \end{bmatrix} T \quad \text{and} \quad A = W \begin{bmatrix} J & 0 \\ 0 & I_{n-q} \end{bmatrix} T,$$

where $J \in \mathbb{C}^{q,q}$ is a matrix in Jordan canonical form associated with the $q$ finite eigenvalues of the pencil $zE - A$ and $N \in \mathbb{C}^{n-q,n-q}$ is a nilpotent matrix in Jordan canonical form corresponding to $n-q$ times the eigenvalue $\infty$. If $q < n$ and $N$ has degree of nilpotency $\nu \in \{1,2,\ldots\}$, that is, $N^\nu = 0$ and $N^i \neq 0$ for $i = 1,\ldots,\nu - 1$, then $\nu$ is called the index of the pair $(E,A)$. If $E$ is nonsingular, then by convention the index is $\nu = 0$; see for example [15, 18]. The matrix pair $(E,A) \in (\mathbb{R}^{n,n})^2$ is said to be
**stable** (resp. *asymptotically stable*) if all the finite eigenvalues of \( zE - A \) are in the closed (resp. open) unit ball and those on the unit circle are semisimple. The matrix pair \((E, A)\) is said to be *admissible* if it is regular, of index at most one, and stable.

The various distance problems for linear control systems is an important research topic in the numerical linear algebra community; for example, the distance to bounded realness \([1]\), the robust stability problem \([19]\), the stability radius problem for standard systems \([2, 13]\) and for descriptor systems \([3, 5]\), the nearest stable matrix problem for continuous-time systems \([17, 9, 14, 11]\) and for discrete-time systems \([17, 16, 12, 7]\), the nearest continuous-time admissible descriptor system problem \([8]\), and the nearest positive real system problem \([10]\).

For a given unstable matrix pair \((E, A)\), the discrete-time nearest stable matrix pair problem is to solve the following optimization problem

\[
\inf_{(\hat{E}, \hat{A}) \in S_{n,n}} \| E - \hat{E} \|^2_F + \| A - \hat{A} \|^2_F,
\]

where \( S_{n,n} \) is the set of admissible pairs of size \( n \times n \). This problem is the converse of stability radius problem for descriptor systems \([3, 5]\) and the discrete-time counterpart of continuous-time nearest stable matrix pair problem \([8]\). Such problems arise in systems identification where one needs to identify a stable matrix pair depending on observations \([17, 9]\). This is a highly nonconvex optimization problem because the set \( S_{n,n} \) is unbounded, nonconvex and neither open nor closed. In fact, consider the matrix pair

\[
(E, A) = \left( \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1/2 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right).
\]

The pair \((E, A)\) is regular since \( \det(\lambda E - A) = \det(\lambda - 1/2) \neq 0 \), of index one, and stable with the only finite eigenvalue \( \lambda_1 = 1/2 \). Thus \((E, A) \in S^{3,3}\). Let

\[
(\Delta E, \Delta A) = \left( \begin{bmatrix} 0 & 0 & 0 \\ 0 & \epsilon_1 & \epsilon_2 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\delta \end{bmatrix} \right),
\]

and consider the perturbed pair \((E + \Delta E, A + \Delta A)\). If we let \( \delta = \epsilon_1 = 0 \) and \( \epsilon_2 > 0 \), then the perturbed pair is still regular and stable as the only finite eigenvalue \( \lambda_1 = 1/2 \) belongs to the unit ball, but it is of index two. For \( \epsilon_2 = \delta = 0 \) and \( 0 < \epsilon_1 < 1 \), the perturbed pair is regular, of index one but has two finite eigenvalues \( \lambda_1 = 1/2 \) and \( \lambda_2 = 1/\epsilon_1 > 1 \). This implies that the perturbed pair is unstable. This shows that \( S^{3,3} \) is not open. Similarly, if we let \( \epsilon_1 = \epsilon_2 = 0 \) and \( \delta > 0 \), then as \( \delta \to 1 \) the perturbed pair becomes non-regular. This shows that \( S^{3,3} \) is not closed. The nonconvexity of \( S^{n,n} \) follows by considering for example

\[
\Sigma_1 = \left( I_2, \begin{bmatrix} 0.5 & 2 \\ 0 & 1 \end{bmatrix} \right), \quad \Sigma_2 = \left( I_2, \begin{bmatrix} 0.5 & 0 \\ -2 & 1 \end{bmatrix} \right),
\]

where \( \Sigma_1, \Sigma_2 \in S^{2,2} \), while \( \gamma \Sigma_1 + (1 - \gamma) \Sigma_2 \notin S^{2,2} \) for \( \gamma = \frac{1}{2} \), since \( \frac{1}{2} \Sigma_1 + \frac{1}{2} \Sigma_2 \) has two eigenvalues \( 0.75 \pm 0.96i \) outside the unit ball. Therefore it is in general difficult to work directly with the set \( S^{n,n} \). We explain in Section 2 the difficulty in generalizing the results in \([7]\) for problem \((P)\).
In this paper, we consider instead a rank-constrained nearest stable matrix pair problem. For this, let \( r < n \in \mathbb{Z}_+ \) and let us define a subset \( S_{r,n}^{n,n} \) of \( S^{n,n} \) by

\[
S_{r,n}^{n,n} := \left\{ (\hat{E}, \hat{X}) \in S^{n,n} : \text{rank}(\hat{E}) = r \right\}.
\]

For a given unstable matrix pair \((E, A)\), the rank-constrained nearest stable matrix pair problem requires to compute the smallest perturbation \((\Delta E, \Delta A)\) with respect to Frobenius norm such that \((E + \Delta E, A + \Delta A)\) is admissible with \(\text{rank}(E + \Delta E) = r\), or equivalently, we aim to solve the following optimization problem

\[
\inf_{(E, A) \in S_{r,n}^{n,n}} \|E - \hat{E}\|_F^2 + \|A - \hat{A}\|_F^2.
\]

The problem \((P_r)\) is also nonconvex as the set \(S_{r,n}^{n,n}\) is nonconvex. To solve \((P_r)\), we provide a simple parametrization of \(S_{r,n}^{n,n}\) in terms of a matrix quadruple \((T, W, U, B)\), where \(T, W \in \mathbb{R}^{n,n}\) are invertible, \(U \in \mathbb{R}^{r,r}\) is orthogonal, and \(B \in \mathbb{R}^{r,r}\) is a positive semidefinite contraction, see Section 2. This parametrization results in an equivalent optimization problem with a feasible set onto which it is easy to project, and we derive a block coordinate descent method to tackle it; see Section 3. We illustrate the effectiveness of our algorithm over several numerical examples in Section 4.

**Notation** Throughout the paper, \(X^T\) and \(\|X\|\) stand for the transpose and the spectral norm of a real square matrix \(X\), respectively. We write \(X \succ 0\) and \(X \succeq 0\) (\(X \preceq 0\)) if \(X\) is symmetric and positive definite or positive semidefinite (symmetric negative semidefinite), respectively. By \(I_m\) we denote the identity matrix of size \(m \times m\).

## 2 Reformulation of problem \((P_r)\)

As mentioned earlier the set \(S_{r,n}^{n,n}\) is nonconvex. It is also an unbounded set which is neither open nor closed. Consider

\[
\tilde{\Sigma}_1 = \left( \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} A_1 & 0 \\ 0 & I_{n-r} \end{bmatrix} \right), \quad \tilde{\Sigma}_2 = \left( \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} B_1 & 0 \\ 0 & I_{n-r} \end{bmatrix} \right),
\]

where \(A_1 = \begin{bmatrix} A & 0 \\ 0 & I_{r-2} \end{bmatrix}, B_1 = \begin{bmatrix} B & 0 \\ 0 & I_{r-2} \end{bmatrix}\), and \(A\) and \(B\) are defined as in \((1.3)\). We have that \(\tilde{\Sigma}_1, \tilde{\Sigma}_2 \in S_{r,n}^{n,n}\) because \((I_r, A_1)\) and \((I_r, B_1)\) are stable. Moreover \(\frac{1}{2}\tilde{\Sigma}_1 + \frac{1}{2}\tilde{\Sigma}_2 \notin S_{r,n}^{n,n}\) as it has two eigenvalues \(0.75 \pm 0.96i\) outside the unit ball hence \(S_{r,n}^{n,n}\) is non-convex. To show that \(S_{r,n}^{n,n}\) is neither open nor closed, let \((E, A)\) and \((\Delta E, \Delta A)\) be as defined in \((1.1)\) and \((1.2)\), and consider

\[
(\hat{E}, \hat{A}) = \left( \begin{bmatrix} I_{r-1} & 0 \\ 0 & E \end{bmatrix}, \begin{bmatrix} I_{r-1} & 0 \\ 0 & A \end{bmatrix} \right)
\]

and the perturbation

\[
(\Delta \hat{E}, \Delta \hat{A}) = \left( \begin{bmatrix} I_{r-1} & 0 \\ 0 & \Delta E \end{bmatrix}, \begin{bmatrix} I_{r-1} & 0 \\ 0 & \Delta A \end{bmatrix} \right).
\]

By using similar arguments as in the case of \(S^{n,n}\) one can show that \(S_{r,n}^{n,n}\) is neither open nor closed. Therefore it is difficult to compute a global solution to problem \((P_r)\) and to work directly with the set \(S_{r,n}^{n,n}\). For this reason, we reformulate the rank-constrained nearest stable matrix pair problem
into an equivalent problem with a relatively simple feasible set. In order to do this, we derive a parametrization of admissible pairs into invertible, symmetric and orthogonal matrices. We first recall a result from \cite{7} that gives a characterization for stable matrices.

**Theorem 1.** \cite{7} Theorem 1] Let $A \in \mathbb{R}^{n \times n}$. Then $A$ is stable if and only if $A = S^{-1}UBS$ for some $S, U, B \in \mathbb{R}^{n \times n}$ such that $S > 0$, $U^T U = I_n$, $B \succeq 0$, and $\|B\| \leq 1$.

We note that, in the proof of Theorem 1, only the invertibility of matrix $S$ is needed and the condition of symmetry on $S$ can be relaxed. We found that this relaxation on matrix $S$ does not make any difference on the numerical results in \cite{7}. The only gain is that the projection of the set of positive definite matrices takes some time and that can be avoided. Therefore, we rephrase the definition of a SUB matrix in \cite{7} and the corresponding characterization of stable matrices as follows.

**Theorem 2.** Let $A \in \mathbb{R}^{n \times n}$. Then $A$ is stable if and only if $A$ admits a SUB form, that is, $A = S^{-1}UBS$ for some $S, U, B \in \mathbb{R}^{n \times n}$ such that $S$ is invertible, $U^T U = I_n$, $B \succeq 0$, and $\|B\| \leq 1$.

**Theorem 3.** Let $E, A \in \mathbb{R}^{n \times n}$ be such that $\text{rank}(E) = r$. Then $(E, A)$ is admissible if and only if there exist matrices $T, W \in \mathbb{R}^{n \times n}$, $S, U, B \in \mathbb{R}^{r \times r}$ such that the matrices $T, W, S$ are invertible, $U^T U = I_r$, $B \succeq 0$, $\|B\| \leq 1$ such that

$$E = W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} T,$$

and

$$A = W \begin{bmatrix} S^{-1}UBS & 0 \\ 0 & I_{n-r} \end{bmatrix} T.$$  

(2.1)

**Proof.** For a regular index one pair $(E, A)$, there exist invertible matrices $W, T \in \mathbb{R}^{n \times n}$ such that

$$E = W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} T \quad \text{and} \quad A = W \begin{bmatrix} \tilde{A} & 0 \\ 0 & I_{n-r} \end{bmatrix} T,$$  

(2.2)

see \cite{4}. Further, the finite eigenvalues of $(E, A)$ and $\tilde{A}$ are same because $\det(\lambda E - A) = 0$ if and only if $\det(\lambda I_r - \tilde{A}) = 0$. Thus by stability of $(E, A)$ and Theorem 2, it follows that $\tilde{A}$ admits a SUB form, that is, there exist $S, U, B \in \mathbb{R}^{r \times r}$ such that $S$ is invertible, $U^T U = I_r$, $B \succeq 0$, $\|B\| \leq 1$, and $\tilde{A} = S^{-1}UBS$.

Conversely, it is easy to see that any matrix pair $(E, A)$ in the form (2.1) is regular and of index one. The stability of $(E, A)$ follows from Theorem 2 as the matrix $S^{-1}UBS$ is stable. \hfill \square

If the matrix $E$ is nonsingular, then Theorem 3 can be further simplified as follows.

**Theorem 4.** Let $E, A \in \mathbb{R}^{n \times n}$, and let $E$ be nonsingular. Then $(E, A)$ is admissible if and only if there exist matrices $S, U, B \in \mathbb{R}^{n \times n}$ such that $A = S^{-1}UBSE$, where $S$ is invertible, $U^T U = I_n$, $B \succeq 0$, and $\|B\| \leq 1$.

**Proof.** Since $E$ is nonsingular, the matrix pair $(E, A)$ can be equivalently written as a standard pair $(I_n, AE^{-1})$, and then stability of $(E, A)$ can be determined by the eigenvalues of $AE^{-1}$. That means, $(E, A)$ is stable if and only if $AE^{-1}$ is stable. Thus from Theorem 2, $AE^{-1}$ is stable if and only if $AE^{-1}$ admits a SUB form, that is, $AE^{-1} = S^{-1}UBS$ for some $S, U, B \in \mathbb{R}^{n \times n}$ such that $S$ is invertible, $U^T U = I_n$, $B \succeq 0$, and $\|B\| \leq 1$. \hfill \square

We note that, for a standard pair $(I_n, A)$ (with $E = I_n$), Theorem 3 coincides with Theorem 2 as in this case $W$ and $T$ can be chosen to be the identity matrix which yields $A = S^{-1}UBS$. A similar result also holds for asymptotically stable matrix pairs which can be seen as a generalization of \cite{7} Theorem 2.
Theorem 5. Let \( E, A \in \mathbb{R}^{n,n} \) be such that \( \text{rank}(E) = r \). Then \( (E, A) \) is regular, of index one and asymptotically stable if and only if there exist matrices \( T, W \in \mathbb{R}^{n,n} \), \( S, U, B \in \mathbb{R}^{r,r} \) such that the matrices \( T, W, S \) are invertible, \( U^T U = I_r \), \( B \succeq 0 \), \( \|B\| < 1 \) such that

\[
E = W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} T, \quad \text{and} \quad A = W \begin{bmatrix} S^{-1}UBS & 0 \\ 0 & I_{n-r} \end{bmatrix} T. \tag{2.3}
\]

Proof. The proof follows is similar to that of Theorem 3 by using [7, Theorem 2] instead of Theorem 2.

Note that the matrix \( S \) is invertible in Theorem 3 and therefore it can be absorbed in \( W \) and \( T \). The advantage is that this reduces the number of variables in the corresponding optimization problem.

Corollary 1. Let \( E, A \in \mathbb{R}^{n,n} \) be such that \( \text{rank}(E) = r \). Then \( (E, A) \) is admissible if and only if there exist invertible matrices \( T, W \in \mathbb{R}^{n,n} \), \( U, B \in \mathbb{R}^{r,r} \) with \( U^T U = I_r \), \( B \succeq 0 \), \( \|B\| \leq 1 \) such that

\[
E = W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} T, \quad \text{and} \quad A = W \begin{bmatrix} UB & 0 \\ 0 & I_{n-r} \end{bmatrix} T. \tag{2.4}
\]

In view of Corollary 1, the set \( S_{n,n}^r \) of restricted rank admissible pairs can be characterized in terms of matrix pairs (2.4), that is,

\[
S_{n,n}^r = \left\{ (W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} T, W \begin{bmatrix} UB & 0 \\ 0 & I_{n-r} \end{bmatrix} T) : \text{invertible } T, W \in \mathbb{R}^{n,n}, U, B \in \mathbb{R}^{r,r}, U^T U = I_r, B \succeq 0, \|B\| \leq 1 \right\}.
\]

This parametrization changes the feasible set and the objective function in problem (P_r) as

\[
(P_r) = \inf_{W, T \in \mathbb{R}^{n,n}, U, B \in \mathbb{R}^{r,r}, U^T U = I_r, \|B\| \leq 1} f(W, T, U, B), \tag{2.5}
\]

where

\[
f(W, T, U, B) = \left\| E - W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} T \right\|_F^2 + \left\| A - W \begin{bmatrix} UB & 0 \\ 0 & I_{n-r} \end{bmatrix} T \right\|_F^2.
\]

An advantage of this reformulation over (P_r) is that it is relatively easy to project onto the feasible set of (2.5). This enables us to use standard optimization schemes to solve it, see Section 3.

As mentioned in [8], for the standard pair \( (I_n, A) \) making \( A \) stable without perturbing the identity matrix gives an upper bound to the solution of \( (P_n) \), because

\[
\inf_{(M, X) \in S_{n,n}^n} \|I_n - M\|_F^2 + \|A - X\|_F^2 \leq \inf_{(I_n, X) \in S_{n,n}^n} \|A - X\|_F^2 = \inf_{(I_n, S^{-1}UBS) \in S_{n,n}^n} \|A - S^{-1}UBS\|_F^2. \tag{2.6}
\]

Note that the right hand side infimum in (2.6) is the distance of \( A \) from the set of stable matrices [7]. It is demonstrated in our numerical experiments that (as expected) the inequality in (2.6) is strict. We also note that similar arguments do not extend to the solution of problem \( (P_r) \), when \( r < n \). In this case, the distance of \( A \) from the set of stable matrices is not an upper bound for the solution of \( (P_r) \), see Section 4. We close the section with a remark that emphasizes the difficulty in solving \( (P_r) \) over \( (P_r) \).
Remark 1. In view of Corollary 1, the set $S^{n,n}$ of admissible pairs can be written as

$$S^{n,n} = \bigcup_{r=1}^{n} S^{n,n}_r.$$ 

Hence we have that

$$\mathcal{P} = \min_{r=1,2,...,n} \mathcal{P}_r.$$ 

To compute a solution of $\mathcal{P}$, a possible way is therefore to solve $n$ rank-constrained problems $\mathcal{P}_r$. For $n$ large, this would be rather costly as it makes the corresponding algorithm for $\mathcal{P}_r$ $n$ times more expensive than for $\mathcal{P}_r$. However, in practice, the rank $r$ has to be chosen close to the (numerical) rank of $E$ so that it can be estimated from the input data. Also, as we will see in Section 4, the error tends to change monotonically with $r$ (first it decreases as $r$ increases –unless $r = 1$ is the best value– and then increases after having achieved the best value for $r$) which could also be used to avoid computing the solutions for all $r$.

3 Algorithmic solution for $\mathcal{P}_r$

To solve (2.5), we use a block coordinate descent method and optimize alternatively over $W$, $T$ and $(U, B)$. For $T$, $U$ and $B$ fixed, the optimal $W$ can be computed using least squares, and similarly for the optimal $T$. Note that the least squares problem in $W$ (resp. $T$) can be solved independently for each row (resp. each column). To update $(U, B)$ for $W$ and $T$ fixed, we use the fast gradient method from [7] (it can be easily adapted by fixing $S$ to the identity and modifying the gradients).

Algorithm 1 Block Coordinate Descent Method for (2.5)

Input: An initialization $W \in \mathbb{R}^{n \times n}, T \in \mathbb{R}^{n \times n}, U \in \mathbb{R}^{r \times r}, B \in \mathbb{R}^{r \times r}$.

Output: An approximate solution $(W, T, U, B)$ to (2.5).

1: for $k = 1, 2, \ldots$ do
2: \quad $W \leftarrow \arg\min_{Y} f(Y, T, U, B)$; \hspace{1em} % Least squares problem
3: \quad $T \leftarrow \arg\min_{X} f(W, X, U, B)$; \hspace{1em} % Least squares problem
4: \quad Apply a few steps of the fast gradient method from [7] on
\quad \quad $\min_{(U, B) \text{ s.t. } U^T U = I_r, \|B\| \leq 1} f(W, T, U, B)$
\quad \quad to update $(U, B)$.
5: end for

3.1 Initialization

For simplicity, we only consider one initialization scheme in this paper which is similar to the one that performed best in [7]. However, it is important to keep in mind that Algorithm 1 is sensitive to initialization and that coming up with good initialization schemes is a topic of further research.

We take $W = T = I_n$ and $(U, B)$ as the optimal solution of

$$\min_{(U, B) \text{ s.t. } U^T U = I_r, \|B\| \leq 1} \|A_{1:r,1:r} - UB\|_F^2.$$ 

In this particular case, it can be computed explicitly using the polar decomposition of $A_{1:r,1:r}$ [2].
4 Numerical experiments

In this section, we apply Algorithm 1 on several examples. As far as we know, there does not exist any other algorithm to stabilize matrix pairs (in the discrete case) hence we cannot compare it to another technique. However, when $E = I_n$, we will compare to the fast gradient method of [7] which provides a nearby stable matrix (but does not allow to modify $E$). Our code is available from https://sites.google.com/site/nicolasgillis/ and the numerical examples presented below can be directly run from this online code. All tests are performed using Matlab R2015a on a laptop Intel CORE i7-7500U CPU @2.7GHz 24Go RAM. Algorithm 1 runs in $O(n^3)$ operations per iteration, including projections onto the set of orthogonal matrices, the resolution of the least squares problem and all necessary matrix-matrix products. Hence Algorithm 1 can be applied on a standard laptop with $n$ up to a thousand (each iteration on the specified laptop takes about 10 seconds for $r = n$).

4.1 Grcar matrix

Let us first consider the pair $(I_n, A)$ where $A$ is the Grcar matrix of dimension $n$ and order $k$ [9]. For $n = 10$ and $k = 3$, the nearest stable matrix found in [7] has relative error $\|A - \hat{A}\|_F = 3.88$. Applying Algorithm 1 with $r = n$, we obtain a matrix pair $(\hat{E}, \hat{A})$ such that $\|A - \hat{A}\|_F + \|E - \hat{E}\|_F = 1.88$. Figure 4.1 displays the evolution of the error (left) and the eigenvalues of the solutions (right). We observe that allowing $\hat{E}$ to be different than the identify matrix allows the matrix pair $(\hat{E}, \hat{A})$ to be much closer to $(I_n, A)$ and have rather different eigenvalues.

Effect of the dimension $n$  Let us perform the same experiment as above except that we increase the value of $n$. Table 4.1 compares the error of the nearest stable matrix and of the nearest stable matrix pair. As $n$ increases, the nearest stable matrix pair allows to decrease the error of approximation.
Table 4.1: Comparison of the error for Grcar matrices $A$ of order $k = 3$ and $E = I_n$ for different values of $n$.

| $n$ | $n = 5$ (30 s.) | $n = 10$ (60 s.) | $n = 20$ (120 s.) | $n = 50$ (300 s.) | $n = 100$ (600 s.) |
|-----|----------------|----------------|------------------|------------------|-------------------|
| Stable matrix | 1.76 | 3.88 | 15.89 | 68.18 | 160.00 |
| Stable pair | 1.16 | 1.88 | 3.02 | 8.69 | 20.41 |

**Effect of $r = \text{rank}(\hat{E})$ and $\text{rank}(E)$** Let us now perform more extensive numerical experiments on the Grcar matrix of dimension $n = 10$ of order 3. Let us fix $0 \leq p \leq n - 1$ and define $E(i, i) = 1$ for $i > p$ otherwise $E(i, j) = 0$ (that is, $E$ is the identity matrix where $p$ diagonal entries have been set to zero) with $\text{rank}(E) = n - p$. Table 4.2 gives the error of the solution obtained by Algorithm 1 for $r = 1, 2, \ldots, n$.

Table 4.2: Comparison of the error for Grcar matrices $A$ with $k = 3$, and $E(i, i) = 1$ for $i > p$ otherwise $E(i, j) = 0$.

| $\text{rank}(E)$ | $r = 1$ | $r = 2$ | $r = 3$ | $r = 4$ | $r = 5$ | $r = 6$ | $r = 7$ | $r = 8$ | $r = 9$ | $r = 10$ |
|-----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 10              | 9.02   | 8.05   | 7.09   | 6.20   | 5.44   | 4.63   | 3.94   | 3.16   | 2.16   | 1.88   |
| 9               | 8.04   | 7.08   | 6.13   | 5.33   | 4.61   | 3.83   | 3.16   | 2.16   | 1.57   | 1.36   |
| 8               | 7.05   | 6.10   | 5.17   | 5.16   | 4.16   | 3.16   | 2.16   | 1.46   | 1.37   | 1.42   |
| 7               | 6.05   | 5.13   | 4.22   | 4.16   | 2.83   | 2.17   | 1.57   | 1.52   | 1.44   | 1.44   |
| 6               | 5.07   | 4.17   | 3.27   | 3.16   | 2.04   | 1.32   | 1.34   | 1.69   | 1.69   | 1.91   |
| 5               | 4.09   | 3.24   | 2.34   | 1.76   | 1.20   | 1.52   | 1.30   | 1.56   | 1.74   | 3.13   |
| 4               | 3.12   | 2.26   | 1.49   | 1.25   | 1.19   | 1.24   | 1.30   | 1.68   | 2.96   | 2.95   |
| 3               | 2.13   | 1.31   | 0.69   | 1.19   | 1.23   | 1.29   | 1.69   | 2.83   | 2.83   | 2.83   |
| 2               | 1.15   | 0.41   | 0.69   | 1.19   | 1.22   | 1.27   | 1.91   | 2.79   | 2.79   | 2.79   |
| 1               | 0.17   | 0.81   | 1.21   | 1.36   | 1.22   | 2.71   | 2.72   | 2.72   | 4.33   | 4.33   |

We observe that

- In 6 out of the 10 cases, using $r = \text{rank}(E)$ provides the best solution. In 3 out of the 10 cases, using $r = \text{rank}(E) + 1$ provides the best solution, and in one case $r = \text{rank}(E) + 2$ provides the best solution. This illustrates the fact that the best value for $r$ should be close to the (numerical) rank of $E$. (Of course, since we use a single initialization, there is no guarantee that the error in Table 4.2 is the smallest possible.)

- In all cases, the error behaves monotonically, that is, it increases as the value of $r$ goes away from the best value.

The two observations above could be used in practice to tune effectively the value of $r$: start from a value close to the numerical rank of $E$, then try nearby values until the error increases.

Table 4.3 gives the computational time for the different cases. We use the following stopping criterion:

$$e(i) - e(i + 1) < 10^{-8}e(i),$$

where $e(i)$ is the error obtained at the $i$th iteration, and a time limit of 60 seconds.
Table 4.3: Time in seconds to compute the solution obtained in Table 4.2. The time limit is 60 seconds.

| rank(E) | r = 1  | r = 2  | r = 3  | r = 4  | r = 5  | r = 6  | r = 7  | r = 8  | r = 9  | r = 10 |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 10      | 0.59   | 1.00   | 0.72   | 0.56   | 2.44   | 2.20   | 2.08   | 43.92  | 13.41  | 49.16  |
| 9       | 0.23   | 0.58   | 0.44   | 2.34   | 2.53   | 17.89  | 3.25   | 24.19  | 43.34  | 47.72  |
| 8       | 0.22   | 0.89   | 0.42   | 0.36   | 13.16  | 1.16   | 16.34  | 3.67   | 8.09   | 11.02  |
| 7       | 0.16   | 0.81   | 0.77   | 0.61   | 13.63  | 10.20  | 60     | 60     | 60     | 60     |
| 6       | 0.19   | 0.67   | 0.73   | 1.19   | 27.42  | 60     | 60     | 60     | 60     | 60     |
| 5       | 0.17   | 0.83   | 0.59   | 4.22   | 60     | 60     | 39.17  | 60     | 60     | 60     |
| 4       | 0.25   | 1.20   | 0.39   | 29.11  | 60     | 21.80  | 60     | 60     | 40.81  | 60     |
| 3       | 0.14   | 0.28   | 0.13   | 31.08  | 60     | 60     | 60     | 55.08  | 60     | 60     |
| 2       | 0.06   | 0.17   | 17.50  | 60     | 60     | 60     | 60     | 60     | 60     | 60     |
| 1       | 0.02   | 23.56  | 26.61  | 48.11  | 60     | 60     | 37.95  | 60     | 60     | 60     |

We observe that the algorithm converges much faster when \( r \) is small. This can be partly explained by the smaller number of variables, being \( 2n^2 + 2r^2 \).

4.2 Scaled all-one matrix

In this section, we perform a similar experiment than in the previous section with \( A = \alpha ee^T \) where \( e \) is the vector of all ones, which is an example from \([12]\). For \( \alpha > 1/n \), the matrix is unstable. For \( 1/n \leq \alpha \leq 2/n \), the nearest stable matrix is \( ee^T/n \).

Let us take \( n = 10 \) and \( \alpha = 2/n = 0.2 \) for which the nearest stable matrix is \( A = 0.1ee^T \) with error 1. The nearest stable matrix pair computed by Algorithm [1] is given by \( A = 0.15ee^T \) and \( E = I_n + 0.05ee^T \) with error \( \frac{1}{2} \). As for the Grcar matrix, allowing \( E \) to be different from the identity matrix allows to reduce the error in approximating \( (I_n, A) \) significantly (by a factor of two).

References

[1] R. Alam, S. Bora, M. Karow, V. Mehrmann, and J. Moro, Perturbation theory for Hamiltonian matrices and the distance to bounded-realness, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 484–514.
[2] R. Byers, A bisection method for measuring the distance of a stable to unstable matrices, SIAM J. on Scientific and Statistical Computing, 9 (1988), pp. 875–881.
[3] R. Byers and N.K. Nichols, On the stability radius of a generalized state-space system, Linear Algebra and its Applications, 188 (1993), pp. 113–134.
[4] L. Dai, Singular Control Systems, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1989.
[5] N.H. Du, V.H. Linh, and V. Mehrmann, Robust stability of differential-algebraic equations, in Surveys in Differential-Algebraic Equations I, Berlin: Springer, 2013, pp. 63–95.
[6] F.R. Gantmacher, The Theory of Matrices I, Chelsea Publishing Company, New York, NY, 1959.
[7] N. Gillis, M. Karow, and P. Sharma, Stabilizing discrete-time linear systems, arXiv preprint arXiv:1802.08033, (2018).
[8] N. Gillis, V. Mehrmann, and P. Sharma, Computing nearest stable matrix pairs, Numerical Linear Algebra with Applications, (2018), p. e2153. doi:10.1002/nla.2153.
[9] N. Gillis and P. Sharma, *On computing the distance to stability for matrices using linear dissipative Hamiltonian systems*, Automatica, 85 (2017), pp. 113–121.

[10] ———, *Finding the nearest positive-real system*, SIAM Journal on Numerical Analysis, 56 (2018), pp. 1022–1047.

[11] Nicola Guglielmi and Christian Lubich, *Matrix stabilization using differential equations*, SIAM Journal on Numerical Analysis, 55 (2017), pp. 3097–3119.

[12] N. Guglielmi and V.Yu. Protasov, *On the closest stable/unstable nonnegative matrix and related stability radii*, arXiv:1802.03054, (2018).

[13] D. Hinrichsen and A.J. Pritchard, *Stability radii of linear systems*, Systems Control Lett., 7 (1986), pp. 1–10.

[14] C. Mehl, V. Mehrmann, and P. Sharma, *Stability radii for real linear Hamiltonian systems with perturbed dissipation*, BIT Numerical Mathematics, (2017). to appear, doi:10.1007/s10543-017-0654-0.

[15] V.L. Mehrmann, *The Autonomous Linear Quadratic Control Problem: Theory and Numerical Solution*, Lecture Notes in Control and Information Sciences, Springer Berlin Heidelberg, 1991.

[16] Yu Nesterov and V Yu Protasov, *Computing closest stable non-negative matrices*, (2017).

[17] F.-X. Orbandexivry, Yu. Nesterov, and P. Van Dooren, *Nearest stable system using successive convex approximations*, Automatica, 49 (2013), pp. 1195–1203.

[18] A. Varga, *On stabilization methods of descriptor systems*, Systems & Control Letters, 24 (1995), pp. 133–138.

[19] T. Zhou, *On nonsingularity verification of uncertain matrices over a quadratically constrained set*, IEEE Trans. on Automatic Control, 56 (2011), pp. 2206–2212.