Applied Numerical Mathematics ••• (••••) •••–•••

Contents lists available at ScienceDirect

Applied Numerical Mathematics

www.elsevier.com/locate/apnum

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

Nicolas Gillis [a,*,1], Michael Karow [b], Punit Sharma [c,2]

[a] Department of Mathematics and Operational Research, Faculté Polytechnique, Université de Mons, Rue de Houdain 9, 7000 Mons, Belgium
[b] TU Berlin, Institut für Mathematik, Straße des 17. Juni 136, 10623 Berlin, Germany
[c] Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi-110016, India

ARTICLE INFO

Article history:
Received 12 July 2018
Received in revised form 22 January 2019
Accepted 3 September 2019
Available online xxxx

Keywords:
Stability radius
Linear discrete-time systems
Stability
Convex optimization

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 $\operatorname{rank}(\hat{E}) = r$. 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.

1. Introduction

In [17,10], 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 Δ_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 [10] for matrix pairs (E, A), where $E, A \in \mathbb{R}^{n,n}$.

The matrix pair (E, A) is called *regular* if $\det(\lambda E - A) \neq 0$ for some $\lambda \in \mathbb{C}$, which we denote $\det(\lambda E - A) \not\equiv 0$, otherwise it is called *singular*. For a regular matrix pair (E, A), the roots of the polynomial $\det(zE - A)$ are called *finite eigenvalues* of the pencil $zE - A$ or of the pair (E, A). A regular pair (E, A) has ∞ *as an eigenvalue* if E is singular. A regular real matrix pair (E, A) can be transformed to *Weierstraß canonical form* [6], 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 ∞. If $q < n$ and N

* Corresponding author.
E-mail addresses: nicolas.gillis@umons.ac.be (N. Gillis), karow@math.tu-berlin.de (M. Karow), punit.sharma@maths.iitd.ac.in (P. Sharma).

[1] N. Gillis acknowledges the support of the European Research Council (ERC starting grant no 679515), and the Fonds de la Recherche Scientifique - FNRS (incentive grant for scientific research no F.4501.16).
[2] P. Sharma acknowledges the support of the DST-Inspire Faculty Award (MI01807-G) by Government of India and Institute SEED Grant (NPN5R) by IIT Delhi.

https://doi.org/10.1016/j.apnum.2019.09.004
0168-9274/

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 ν is called the *index of the pair* (E, A). If E is nonsingular, then by convention the index is $\nu = 0$; see for example [15,19]. 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 [20], the stability radius problem for standard systems [2,13] and for descriptor systems [3,5], the nearest stable matrix problem for continuous-time systems [17,7,14,11] and for discrete-time systems [17,16,12,10], the nearest continuous-time admissible descriptor system problem [9], and the nearest positive real system problem [8].

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

$$\inf_{(\hat{E},\hat{A})\in\mathcal{S}^{n,n}} \|E - \hat{E}\|_F^2 + \|A - \hat{A}\|_F^2, \tag{$\mathcal{P}$}$$

where $\mathcal{S}^{n,n}$ is the set of admissible pairs of size $n \times n$. This problem is the converse of the stability radius problem for descriptor systems [3,5] and the discrete-time counter part of continuous-time nearest stable matrix pair problem [9]. Such problems arise in systems identification where one needs to identify a stable matrix pair depending on observations [17, 7]. This is a highly nonconvex optimization problem because the set $\mathcal{S}^{n,n}$ is unbounded, nonconvex and neither open nor closed. In fact, consider the matrix pair

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

The pair (E, A) is regular since $\det(\lambda E - A) = \det(\lambda - 1/2) \not\equiv 0$, of index one, and stable with the only finite eigenvalue $\lambda_1 = 1/2$. Thus $(E, A) \in \mathcal{S}^{3,3}$. Let

$$(\Delta_E, \Delta_A) = \left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & \epsilon_1 & \epsilon_2 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\delta \end{bmatrix} \right), \tag{2}$$

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 $\mathcal{S}^{3,3}$ 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 $\mathcal{S}^{3,3}$ is not closed. The nonconvexity of $\mathcal{S}^{n,n}$ follows by considering for example

$$\Sigma_1 = \Big(I_2, \underbrace{\begin{bmatrix} 0.5 & 2 \\ 0 & 1 \end{bmatrix}}_{A}\Big), \quad \Sigma_2 = \Big(I_2, \underbrace{\begin{bmatrix} 0.5 & 0 \\ -2 & 1 \end{bmatrix}}_{B}\Big), \tag{3}$$

where $\Sigma_1, \Sigma_2 \in \mathcal{S}^{2,2}$, while $\gamma\Sigma_1 + (1-\gamma)\Sigma_2 \notin \mathcal{S}^{2,2}$ 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 $\mathcal{S}^{n,n}$. We explain in Section 2 the difficulty in generalizing the results in [10] for problem ($\mathcal{P}$).

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 $\mathcal{S}_r^{n,n}$ of $\mathcal{S}^{n,n}$ by

$$\mathcal{S}_r^{n,n} := \left\{ (\hat{E}, \hat{X}) \in \mathcal{S}^{n,n} : \operatorname{rank}(\hat{E}) = r \right\}.$$

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

$$\inf_{(\hat{E},\hat{A})\in\mathcal{S}_r^{n,n}} \|E - \hat{E}\|_F^2 + \|A - \hat{A}\|_F^2. \tag{$\mathcal{P}_r$}$$

The main reason for considering the rank-constrained problem ($\mathcal{P}_r$) over ($\mathcal{P}$) is of not being able to reformulate the set $\mathcal{S}^{n,n}$. A necessary and sufficient condition for a descriptor system (E, A) to be admissible is the existence of a matrix $X = X^T$ such that LMI's

Please cite this article in press as: N. Gillis et al., A note on approximating the nearest stable discrete-time descriptor systems with fixed rank, Appl. Numer. Math. (2019), https://doi.org/10.1016/j.apnum.2019.09.004

$$EXE^T \succeq 0 \quad \text{and} \quad AXA^T \preceq EXE^T \tag{4}$$

hold true [18]. In contrast to the matrix case [10] where a parametrization of the set of stable matrices was possible because of a positive definite solution P of the Schur's equation $A^T PA - P \prec 0$, for a genuine descriptor system ($\text{rank}(E) < n$) a solution X to (4) always is indefinite. The indefiniteness of the matrix X governed by (4) turns out to be a major obstacle for the reformulation of the problem $(\mathcal{P})$. This obstacle can be avoided by imposing a rank constraint on E. Some of the applications of the rank-constraint admissible pair (E, A) could be seen as follows.

- It enables us to parameterize the set $\mathcal{S}_r^{n,n}$ in terms of the 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.
- The set $\mathcal{S}^{n,n}$ of admissible pairs can be written as

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

  This implies that

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

  Therefore, to compute a solution of $(\mathcal{P})$, a possible way is therefore to solve n rank-constrained problems $(\mathcal{P}_r)$, see Remark 1.
- Such situations may be useful when descriptor systems are directly generated from data where the constraints are added as a second step. One practical example is of a circuit or power net, where one discretizes the flow and adds the Kirchhoff laws afterward.

The problem $(\mathcal{P}_r)$ is also nonconvex as the set $\mathcal{S}_r^{n,n}$ is nonconvex. To solve $(\mathcal{P}_r)$, we provide a simple parametrization of $\mathcal{S}_r^{n,n}$ in 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 $(\mathcal{P}_r)$

As mentioned earlier the set $\mathcal{S}_r^{n,n}$ 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 (3). We have that $\tilde{\Sigma}_1, \tilde{\Sigma}_2 \in \mathcal{S}_r^{n,n}$ because (I_r, A_1) and (I_r, B_1) are stable. Moreover $\frac{1}{2}\tilde{\Sigma}_1 + \frac{1}{2}\tilde{\Sigma}_2 \notin \mathcal{S}_r^{n,n}$ as it has two eigenvalues $0.75 \pm 0.96i$ outside the unit ball hence $\mathcal{S}_r^{n,n}$ is non-convex. To show that $\mathcal{S}_r^{n,n}$ is neither open nor closed, let (E, A) and (Δ_E, Δ_A) be as defined in (1) and (2), and consider

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

and the perturbation

$$(\Delta_{\tilde{E}}, \Delta_{\tilde{A}}) = \left(\begin{bmatrix} I_{r-1} & 0 \\ 0 & \Delta_E \end{bmatrix}, \begin{bmatrix} I_{r-1} & 0 \\ 0 & \Delta_A \end{bmatrix} \right).$$

By using similar arguments as in the case of $\mathcal{S}^{n,n}$ one can show that $\mathcal{S}_r^{n,n}$ is neither open nor closed. Therefore it is difficult to compute a global solution to problem $(\mathcal{P}_r)$ and to work directly with the set $\mathcal{S}_r^{n,n}$. 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 [10] that gives a characterization for stable matrices.

Theorem 1. [10, Theorem 1] *Let $A \in \mathbb{R}^{n,n}$. Then A is stable if and only if $A = S^{-1}UBS$ for some $S, U, B \in \mathbb{R}^{n,n}$ such that $S \succ 0$, $U^T U = I_n$, $B \succeq 0$, and $\|B\| \leq 1$.*

Please cite this article in press as: N. Gillis et al., A note on approximating the nearest stable discrete-time descriptor systems with fixed rank, Appl. Numer. Math. (2019), https://doi.org/10.1016/j.apnum.2019.09.004

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 [10]. The only gain is that the projection of S on the set of positive definite matrices takes some time and that can be avoided. Therefore, we rephrase the definition of a SUB matrix in [10] and the corresponding characterization of stable matrices as follows.

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

Theorem 3. *Let $E, A \in \mathbb{R}^{n,n}$ be such that $\mathrm{rank}(E) = r$. Then (E, A) is admissible 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, $U^TU = I_r$, $B \succeq 0$, $\|B\| \leq 1$ such that*

$$E = W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} T, \quad \text{and} \quad A = W \begin{bmatrix} S^{-1}UBS & 0 \\ 0 & I_{n-r} \end{bmatrix} T. \tag{5}$$

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

$$E = W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} T \qquad \text{and} \qquad A = W \begin{bmatrix} \tilde{A} & 0 \\ 0 & I_{n-r} \end{bmatrix} T, \tag{6}$$

see [4]. Further, the finite eigenvalues of (E, A) and $\tilde{A}$ are the 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 $S, U, B \in \mathbb{R}^{r,r}$ such that S is invertible, $U^TU = I_r$, $B \succeq 0$, $\|B\| \leq 1$, and $\tilde{A} = S^{-1}UBS$.

Conversely, it is easy to see that any matrix pair (E, A) in the form (5) is regular and of index one. The stability of (E, A) follows from Theorem 2 as the matrix $S^{-1}UBS$ is stable. □

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

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

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 [10, Theorem 2].

Theorem 5. *Let $E, A \in \mathbb{R}^{n,n}$ be such that $\mathrm{rank}(E) = r$. 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, $U^TU = I_r$, $B \succeq 0$, $\|B\| < 1$ such that*

$$E = W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} T, \quad \text{and} \quad A = W \begin{bmatrix} S^{-1}UBS & 0 \\ 0 & I_{n-r} \end{bmatrix} T. \tag{7}$$

Proof. The proof is similar to that of Theorem 3 by using [10, 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 $\mathrm{rank}(E) = r$. Then (E, A) is admissible if and only if there exist invertible matrices $T, W \in \mathbb{R}^{n,n}$, and $U, B \in \mathbb{R}^{r,r}$ with $U^TU = I_r$, $B \succeq 0$ and $\|B\| \leq 1$ such that*

$$E = W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} T, \quad \text{and} \quad A = W \begin{bmatrix} UB & 0 \\ 0 & I_{n-r} \end{bmatrix} T. \tag{8}$$

In view of Corollary 1, the set $\mathcal{S}_r^{n,n}$ of restricted rank admissible pairs can be characterized in terms of matrix pairs (8), that is,

Please cite this article in press as: N. Gillis et al., A note on approximating the nearest stable discrete-time descriptor systems with fixed rank, Appl. Numer. Math. (2019), https://doi.org/10.1016/j.apnum.2019.09.004

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

This parametrization changes the feasible set and the objective function in problem $(\mathcal{P}_r)$ as

$$(\mathcal{P}_r) \quad = \quad \inf_{W,T \in \mathbb{R}^{n,n},\, UB \in \mathbb{R}^{r,r},\, U^T U = I_r,\, \|B\| \leq 1} f(W,T,U,B), \tag{9}$$

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 $(\mathcal{P}_r)$ is that it is relatively easy to project onto the feasible set of (9). This enables us to use standard optimization schemes to solve it, see Section 3.

As mentioned in [9], for the standard pair (I_n, A) making A stable without perturbing the identity matrix gives an upper bound to the solution of $(\mathcal{P}_n)$, because

$$\inf_{(M,X) \in \mathcal{S}_n^{n,n}} \|I_n - M\|_F^2 + \|A - X\|_F^2 \leq \inf_{(I_n,X) \in \mathcal{S}_n^{n,n}} \|A - X\|_F^2 = \inf_{(I_n, S^{-1}UBS) \in \mathcal{S}_n^{n,n}} \|A - S^{-1}UBS\|_F^2. \tag{10}$$

Note that the right hand side infimum in (10) is the distance of A from the set of stable matrices [10]. It is demonstrated in our numerical experiments that (as expected) the inequality in (10) is strict. We also note that similar arguments do not extend to the solution of problem $(\mathcal{P}_r)$, 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 $(\mathcal{P}_r)$, see Section 4. We close the section with a remark that emphasizes the difficulty in solving $(\mathcal{P})$ over $(\mathcal{P}_r)$.

Remark 1. As mentioned in the introduction, we have that

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

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 $(\mathcal{P})$ 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 solutions for $(\mathcal{P}_r)$

To solve (9), we propose two different methods. The first method is a projected fast gradient method and uses exactly the same steps as the method proposed in [10]. In a few words, it performs a standard projected gradient scheme with line search and uses extracted points to accelerate it. We will refer to this first method as FGM. The second method is a block coordinate descent method and optimizes 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 [10] (it can be easily adapted by fixing S to the identity and modifying the gradients). We will refer to this second method as BCD (see Algorithm 1).

Convergence FGM and BCD are both guaranteed to decrease the objective function at each step because we use a line-search step within the fast gradient method; see [10]. Hence, for every iterate (W, T, U, B), we have

$$\|\hat{E}(W,T) - E\|_F^2 + \|\hat{A}(W,T,U,B) - A\|_F^2 \leq f_0 \tag{11}$$

where $\hat{A}(W,T,U,B)$ and $\hat{E}(W,T)$ are the approximations of A and E, respectively, as done in (9), and f_0 is the initial objective function value. Since the objective function is bounded from below by zero, this implies that the objective function values converge to some value f^*. Moreover, the approximations $(\hat{E}, \hat{A})$ generated at each step of the algorithm are in a compact set because of (11). Therefore, there exists a subsequence of approximations $(\hat{E}, \hat{A})$ generated by FGM and BCD that converge to some limit point $(\hat{E}^*, \hat{A}^*)$ with objective function value f^*. However, it is more difficult to prove convergence of the iterates (T, W, U, B) as T and W are not bounded. It is possible to add an upper bound on the norm of T and W

Please cite this article in press as: N. Gillis et al., A note on approximating the nearest stable discrete-time descriptor systems with fixed rank, Appl. Numer. Math. (2019), https://doi.org/10.1016/j.apnum.2019.09.004

Algorithm 1 Block Coordinate Descent (BCD) Method for (9).

Require: An initialization $W \in \mathbb{R}^{n\times n}$, $T \in \mathbb{R}^{n\times n}$, $U \in \mathbb{R}^{r\times r}$, $B \in \mathbb{R}^{r\times r}$.
Ensure: An approximate solution (W, T, U, B) to (9).

1: **for** $k = 1, 2, \ldots$ **do**
2: $W \leftarrow \operatorname{argmin}_Y f(Y, T, U, B)$; *% Least squares problem*
3: $T \leftarrow \operatorname{argmin}_X f(W, X, U, B)$; *% Least squares problem*
4: Apply a few steps of the fast gradient method from [10] on

$$\min_{(U,B) \text{ s.t. } U^TU=I_r, \|B\|\leq 1} f(W, T, U, B)$$

to update (U, B).
5: **end for**

to guarantee a subsequence of iterates to converge, but we have not observed in practice that this was an issue. Providing a rigorous proof of convergence of the iterates of FGM and BCD to a stationary point of (9) is a difficult problem which we leave as a question for further research. Another challenging direction of research is to characterize the speed of convergence of such algorithms in the basin of attraction of local minima.

3.1. Initialization

For simplicity, we only consider one initialization scheme in this paper which is similar to the one that performed best in [10]. However, it is important to keep in mind that FGM and BCD are 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

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

In this particular case, it can be computed explicitly using the polar decomposition of $A_{1:r,1:r}$ [10].

4. Numerical experiments

In this section, we apply FGM and BCD 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 [10] 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 preformed using Matlab R2015a on a laptop Intel CORE i7-7500U CPU @2.7GHz 24Go RAM. BCD and FGM run in $O(n^3)$ operations per iteration, including projections onto the set of orthogonal matrices, the resolution of the least squares problem (for BCD) and all necessary matrix-matrix products. Hence BCD and FGM can be applied on a standard laptop with n up to a thousand (for $n = 1000$, 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 [7]. For $n = 10$ and $k = 3$, the nearest stable matrix found in [10] has relative error $\|A - \hat{A}\|_F^2 = 3.88$. Applying BCD with $r = n$, we obtain a matrix pair $(\hat{E}, \hat{A})$ such that $\|A - \hat{A}\|_F^2 + \|E - \hat{E}\|_F^2 = 1.88$. FGM converges towards a different solution (although initialized with the same point) with larger error 1.91. Fig. 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 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. Note that, as in the previous example, BCD generates better solutions than FGM; especially as n increases.

Effect of $r = \operatorname{rank}(\hat{E})$ *and* $\operatorname{rank}(E)$ 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 $\operatorname{rank}(E) = n - p$. We use the following stopping criterion:

Please cite this article in press as: N. Gillis et al., A note on approximating the nearest stable discrete-time descriptor systems with fixed rank, Appl. Numer. Math. (2019), https://doi.org/10.1016/j.apnum.2019.09.004

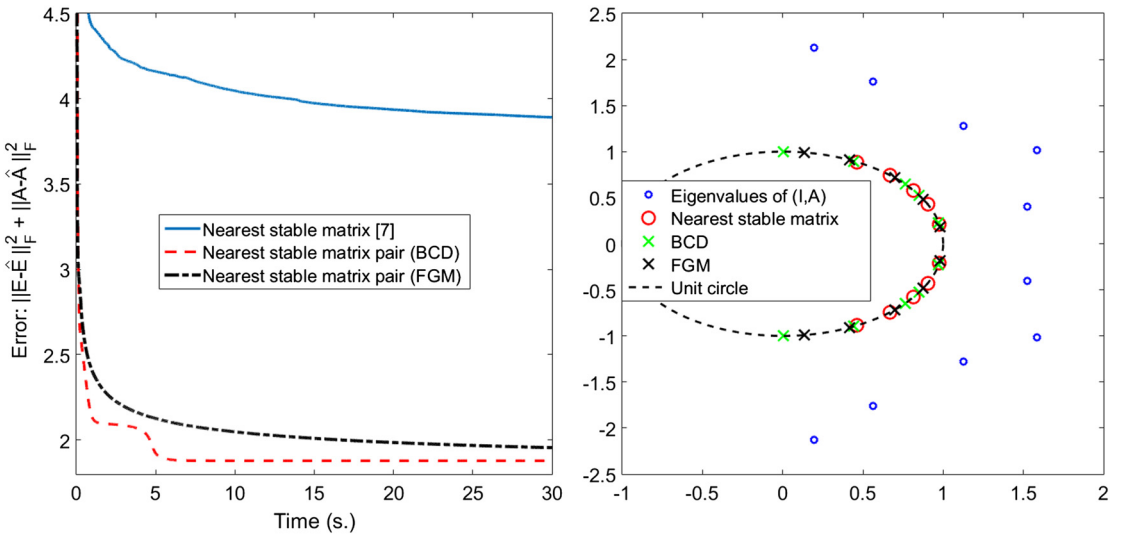

Fig. 1. (Left) Evolution of the error $\|E - \hat{E}\|_F^2 + \|A - \hat{A}\|_F^2$ for the Grcar matrix of dimension 10 and order 3 in the matrix and matrix pair cases (in the matrix case, $\hat{E} = I_n$). (Right) Location of the eigenvalues of A, and of the solutions in the matrix case and in the matrix pair case.

Table 1
Comparison of the error for Grcar matrices A of order $k = 3$ and $E = I_n$ for different values of 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	160.08
Stable pair - BCD	1.16	1.88	3.02	8.69	20.41
Stable pair - FGM	1.16	1.91	4.70	10.60	24.14

Table 2
Comparison of the error for Grcar matrices A with $k = 3$, and $E(i, i) = 1$ for $i > p$ otherwise $E(i, j) = 0$. For each value of rank(E), the first row is the error obtained by BCD, the second by FGM. The lowest error for each value of rank(E) is highlighted in bold.

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**
	9.02	8.62	7.74	6.20	5.44	4.63	3.94	3.37	2.17	1.91
9	8.04	7.08	6.13	5.33	4.61	3.83	3.16	2.16	1.57	**1.36**
	8.04	7.08	7.16	5.92	5.16	3.86	3.16	2.16	1.63	1.52
8	7.05	6.10	5.17	5.16	4.16	3.16	2.16	1.46	**1.37**	1.42
	8.16	6.10	6.16	5.16	4.16	3.16	2.16	1.59	1.66	1.88
7	6.05	5.13	4.22	4.16	2.83	2.17	1.57	1.52	1.44	1.44
	6.05	5.13	4.22	4.16	3.14	2.17	1.47	**1.35**	1.59	1.88
6	5.07	4.17	3.27	3.16	2.04	**1.32**	1.34	1.69	1.69	1.91
	5.07	4.17	3.27	3.16	2.14	**1.32**	1.33	1.47	1.86	2.12
5	4.09	3.24	2.34	1.76	**1.20**	1.52	1.30	1.56	1.74	3.13
	4.09	3.24	2.34	1.76	1.25	1.40	1.44	1.39	1.66	3.07
4	3.12	2.26	1.49	1.25	1.19	1.24	1.30	1.68	2.96	2.95
	3.12	2.26	1.49	**1.00**	1.27	1.22	1.35	1.51	2.97	3.03
3	2.13	1.31	**0.69**	1.19	1.23	1.29	1.69	2.83	2.83	2.83
	2.13	1.31	**0.69**	1.19	1.26	1.30	1.45	2.84	2.86	3.01
2	1.15	**0.41**	1.06	1.22	1.27	1.27	1.91	2.79	2.79	2.79
	1.15	**0.41**	1.18	1.22	1.29	1.44	1.80	2.83	3.00	4.42
1	**0.17**	0.81	1.21	1.21	1.36	1.22	2.71	2.72	2.72	4.33
	0.17	1.18	1.21	1.22	1.28	2.70	2.73	2.74	3.24	4.34

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

where $e(i)$ is the error obtained at the ith iteration, and a time limit of 60 seconds.

Table 2 gives the error of the solution obtained by BCD and FGM for $r = 1, 2, \ldots, n$. We observe that

Table 3
Time in seconds to compute the solution obtained in Table 2. The time limit is 60 seconds. For each value of rank(E), the first row is the error obtained by BCD, the second by FGM.

rank(E)	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	$r=8$	$r=9$	$r=10$
10	0.36	1.27	0.86	0.69	3.36	3.63	3.03	60	39.52	60
	0.08	0.19	0.14	0.06	0.09	0.13	0.42	1.27	60	60
9	0.31	0.80	0.61	2.64	4.23	19.27	60	29.67	60	50.59
	0.05	0.05	0.06	0.44	0.52	0.88	13.67	60	60	60
8	0.28	1.13	0.48	0.45	60	5.89	23.36	60	60	12.84
	0.03	0.06	0.08	0.19	1.92	6.53	21.06	60	60	60
7	0.23	1.05	0.89	0.75	13.98	60	60	60	60	60
	0.05	0.06	0.09	0.19	3.61	60	60	60	60	60
6	0.25	0.75	0.88	15.53	30.97	60	60	60	60	60
	0.05	0.06	0.33	0.39	26.77	15.58	48.08	60	60	60
5	0.23	1.14	0.77	4.56	60	60	60	60	60	60
	0.05	0.09	0.11	0.72	1.08	60	60	60	60	60
4	0.31	1.30	0.48	60	60	39.31	60	60	60	60
	0.05	0.09	0.11	0.50	1.44	60	60	60	60	60
3	0.20	0.33	0.19	60	60	60	60	60	60	60
	0.05	0.06	0.06	0.56	60	60	60	60	60	60
2	0.11	0.25	60	60	60	60	60	60	60	60
	0.02	0.08	0.38	3.67	60	60	37.48	60	60	60
1	0.03	60	60	60	60	23.41	60	60	60	60
	0.02	0.23	1.67	60	60	8.80	60	60	15.94	60

- In 7 out of the 10 cases, using $r = \text{rank}(E)$ provides the best solution. In the other 3 cases, using $r = \text{rank}(E) + 1$ 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 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.
- BCD performs in average better than FGM, obtaining in 82 out of the 100 cases the lowest error.

The first 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 3 gives the computational time for the different cases. We observe that the algorithms converge much faster when r is small. This can be partly explained by the smaller number of variables, being $2n^2 + 2r^2$. Although FGM generated in general worse solutions (see Table 2), it seems that it converges faster in most cases (in 65 out of the 100 cases).

4.2. Scaled all-one matrix

In this section, we perform an experiment similar to the previous one with $(E, A) = (I_n, \alpha ee^T)$ where e is the vector of all ones, which is an example from [12]. For $\alpha > 1/n$, the matrix $A = \alpha ee^T$ 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 $\hat{A} = 0.1ee^T$ with error 1. The nearest stable matrix pair computed by both BCD and FGM is given by $\hat{A} = 0.15ee^T$ 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).

Acknowledgements

The authors would like to thank the reviewer for his insightful comments which helped improve the paper significantly. The authors would also like to thank Volker Mehrmann for helpful discussion and advice.

References

[1] R. Alam, S. Bora, M. Karow, V. Mehrmann, J. Moro, Perturbation theory for Hamiltonian matrices and the distance to bounded-realness, SIAM J. Matrix Anal. Appl. 32 (2) (2011) 484–514.
[2] R. Byers, A bisection method for measuring the distance of a stable to unstable matrices, SIAM J. Sci. Stat. Comput. 9 (1988) 875–881.
[3] R. Byers, N. Nichols, On the stability radius of a generalized state-space system, Linear Algebra Appl. 188 (1993) 113–134.
[4] L. Dai, Singular Control Systems, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1989.
[5] N. Du, V. Linh, V. Mehrmann, Robust stability of differential-algebraic equations, in: Surveys in Differential-Algebraic Equations I, Springer, Berlin, 2013, pp. 63–95.
[6] F. Gantmacher, The Theory of Matrices I, Chelsea Publishing Company, New York, NY, 1959.
[7] N. Gillis, P. Sharma, On computing the distance to stability for matrices using linear dissipative Hamiltonian systems, Automatica 85 (2017) 113–121.

Please cite this article in press as: N. Gillis et al., A note on approximating the nearest stable discrete-time descriptor systems with fixed rank, Appl. Numer. Math. (2019), https://doi.org/10.1016/j.apnum.2019.09.004

[8] N. Gillis, P. Sharma, Finding the nearest positive-real system, SIAM J. Numer. Anal. 56 (2) (2018) 1022–1047.
[9] N. Gillis, V. Mehrmann, P. Sharma, Computing nearest stable matrix pairs, Numer. Linear Algebra Appl. (2018) e2153, https://doi.org/10.1002/nla.2153.
[10] N. Gillis, M. Karow, P. Sharma, Stabilizing discrete-time linear systems, arXiv preprint, arXiv:1802.08033.
[11] N. Guglielmi, C. Lubich, Matrix stabilization using differential equations, SIAM J. Numer. Anal. 55 (6) (2017) 3097–3119.
[12] N. Guglielmi, V. Protasov, On the closest stable/unstable nonnegative matrix and related stability radii, arXiv:1802.03054.
[13] D. Hinrichsen, A. Pritchard, Stability radii of linear systems, Syst. Control Lett. 7 (1986) 1–10.
[14] C. Mehl, V. Mehrmann, P. Sharma, Stability radii for real linear Hamiltonian systems with perturbed dissipation, BIT Numer. Math. 57 (3) (2017) 811–843.
[15] V. Mehrmann, The Autonomous Linear Quadratic Control Problem: Theory and Numerical Solution, Lecture Notes in Control and Information Sciences, Springer, Berlin, Heidelberg, 1991.
[16] Y. Nesterov, V.Y. Protasov, Computing closest stable non-negative matrices, http://www.optimization-online.org/DB_HTML/2017/08/6178.html.
[17] F.-X. Orbandexivry, Y. Nesterov, P. Van Dooren, Nearest stable system using successive convex approximations, Automatica 49 (5) (2013) 1195–1203.
[18] A. Rehm, F. Allögwer, An LMI approach towards stabilization of discrete-time descriptor systems, IFAC Proc. Vol. 35 (1) (2002) 77–82.
[19] A. Varga, On stabilization methods of descriptor systems, Syst. Control Lett. 24 (2) (1995) 133–138.
[20] T. Zhou, On nonsingularity verification of uncertain matrices over a quadratically constrained set, IEEE Trans. Autom. Control 56 (9) (2011) 2206–2212.

Please cite this article in press as: N. Gillis et al., A note on approximating the nearest stable discrete-time descriptor systems with fixed rank, Appl. Numer. Math. (2019), https://doi.org/10.1016/j.apnum.2019.09.004