Condition numbers for inversion of Fiedler companion matrices☆

Fernando De Terán[a], Froilán M. Dopico[b,*], Javier Pérez[a]

[a] *Departamento de Matemáticas, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Spain.*
[b] *Instituto de Ciencias Matemáticas CSIC-UAM-UC3M-UCM and Departamento de Matemáticas, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Spain.*

Abstract

The Fiedler matrices of a monic polynomial $p(z)$ of degree n are $n \times n$ matrices with characteristic polynomial equal to $p(z)$ and whose nonzero entries are either 1 or minus the coefficients of $p(z)$. Fiedler matrices include as particular cases the classical Frobenius companion forms of $p(z)$. Frobenius companion matrices appear frequently in the literature on control and signal processing, but it is well known that they posses many properties that are undesirable numerically, which limit their use in applications. In particular, as n increases, Frobenius companion matrices are often nearly singular, i.e., their condition numbers for inversion are very large. Therefore, it is natural to investigate whether other Fiedler matrices are better conditioned than the Frobenius companion matrices or not. In this paper, we present explicit expressions for the condition numbers for inversion of all Fiedler matrices with respect the Frobenius norm, i.e., $\|A\|_F = \sqrt{\sum_{ij} |a_{ij}|^2}$. This allows us to get a very simple criterion for ordering all Fiedler matrices according to increasing condition numbers and to provide lower and upper bounds on the ratio of the condition numbers of any pair of Fiedler matrices. These results establish that if $|p(0)| \leq 1$, then the Frobenius companion matrices have the largest condition number among all Fiedler matrices of $p(z)$, and that if $|p(0)| > 1$, then the Frobenius companion matrices have the smallest condition number. We also provide families of polynomials where the ratio of the condition numbers of pairs of Fiedler matrices can be arbitrarily large and prove that this can only happen when both Fiedler matrices are very ill-conditioned. We finally study some properties of the singular values of Fiedler matrices and determine how many of the singular values of a Fiedler matrix are equal to one.

Keywords: condition numbers, Fiedler companion matrices, Frobenius companion matrices, inverses of Fiedler companion matrices, polynomials, singular values, staircase matrices
2010 MSC: 65F35, 15A12, 15A18, 15B99

1. Introduction

Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ be a monic polynomial of degee n with $a_i \in \mathbb{C}$, for $i = 0, 1, \ldots, n-1$. The first Frobenius companion matrix of $p(z)$ is defined as

$$C_1 = \begin{bmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_1 & -a_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \tag{1}$$

☆This work was partially supported by the Ministerio de Economía y Competitividad of Spain through grant MTM-2009-09281.

*Corresponding author
Email addresses: `fteran@math.uc3m.es` (Fernando De Terán), `dopico@math.uc3m.es` (Froilán M. Dopico), `jpalvaro@math.uc3m.es` (Javier Pérez)

and it has the property that $p(z) = \det(zI - C_1)$. Other similar Frobenius companion matrices that appear in the literature can be obtained by transposition and/or by reversing the order of rows and columns of C_1 [7, pp. 146–149], [9, p. 105].

Frobenius companion matrices are important in theory, numerical computations, and applications. For instance, they are the building blocks of the rational canonical form of a matrix [7, 9], MATLAB computes all the roots of a polynomial by applying the QR algorithm to a balanced Frobenius companion matrix [10], and they appear frequently in the control and signal processing literature (see [8] and [9, Section 10.4] and the references therein).

Frobenius companion matrices arise in control theory because any single-input controllable system can be transformed into a companion form system and, also, because the structure of companion systems greatly simplifies theoretical considerations such as feedback analysis [8]. However, as n increases, Frobenius matrices are known to posses many properties that are undesirable numerically. For instance, stable ones are nearly unstable, controllable ones are nearly uncontrollable, and nonsingular ones are nearly singular, that is, they have large condition numbers for inversion. These properties were studied in detail in [8]. In particular, the study of the behavior of the spectral condition number $\kappa_2(C_1) = \|C_1\|_2 \|C_1^{-1}\|_2$ presented in [8] is based on the following remarkable property of C_1: it is possible to derive explicit expressions for its singular values and at least $n-2$ of the singular values of C_1 are equal to 1 (see also [9, Section 10.4]).

In 2003, Fiedler expanded significantly the family of companion matrices associated with the monic polynomial $p(z)$ [6]. These matrices were named *Fiedler matrices* in [3]. Every Fiedler matrix shares with C_1 two key properties: (i) its characteristic polynomial is $p(z)$, and (ii) $(n-1)$ of its nonzero entries are equal to 1 and the remaining nonzero entries are equal to $-a_i$, for $i = 0, \ldots, n-1$, with exactly one copy of each. Fiedler matrices have attracted considerable attention very recently in the area of nonlinear eigenvalue problems, since they can be used for constructing linearizations of regular, singular, and rectangular matrix polynomials [1, 3, 5]. In addition, some matrix pencils constructed by simple transformations of Fiedler matrices have been used to design structure preserving linearizations of different classes of structured matrix polynomials [1, 2, 4, 11].

It is natural to investigate whether some Fiedler matrices have properties that are more convenient numerically than Frobenius companion matrices or not. In this context, we study in this paper the condition numbers for inversion of Fiedler matrices of scalar polynomials, with the purpose of comparing them and to provide a simple criterion that allows us to determine in advance which Fiedler matrices of a fixed polynomial $p(z)$ have the smallest condition number. The first point to be remarked is that there are not explicit expressions for the singular values of those Fiedler matrices that are different from the Frobenius companion matrices (see Section 6), which prevents the use of the spectral norm in our developments. We have used instead the Frobenius norm of an $n \times n$ matrix, that is, $\|A\|_F = \sqrt{\sum_{ij} |a_{ij}|^2}$, which satisfies $\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\,\|A\|_2$, see [7]. Therefore, the condition number $\kappa_F(A) = \|A\|_F \|A^{-1}\|_F$ in the Frobenius norm satisfies $\kappa_2(A) \leq \kappa_F(A) \leq n\,\kappa_2(A)$ and $n \leq \kappa_F(A)$, in contrast with $1 \leq \kappa_2(A)$.

We have obtained a simple explicit expression in terms of the coefficients of $p(z)$ for the condition number of any Fiedler matrix in the Frobenius norm. This allows us to get a simple criterion for ordering all Fiedler matrices of $p(z)$ according to increasing condition numbers. This ordering establishes in particular that: (i) if $|p(0)| = 1$, then all Fiedler matrices have the same condition number; (ii) if $|p(0)| < 1$, then the Frobenius companion matrices have the largest condition number among all Fiedler matrices of $p(z)$; and (iii) if $|p(0)| > 1$, then the Frobenius companion matrices have the smallest condition number. The important fact in applications is to know whether one matrix has a condition number *much* smaller than another or not. With this goal in mind, we provide simple lower and upper bounds for the ratio of the condition numbers of every pair of Fiedler matrices and we prove that there exist families of polynomials $p(z)$ for which these ratios can be arbitrarily large or small. However, we also prove that this only happens when both Fiedler matrices are very ill-conditioned. Loosely speaking, this means that there is no any polynomial $p(z)$ for which one Fiedler matrix has a small condition number (close to n) while others have very large condition numbers. We finally study some properties of the singular values of Fiedler matrices. More precisely, we determine how many of their singular values are equal to one and, for those that are not, we show that they can be obtained from the square roots of the eigenvalues of certain matrices that may have a size much smaller than n and that are easily constructible from the coefficients of $p(z)$.

The paper is organized as follows. In Section 2, we recall the definition of Fiedler matrices and establish the notation and their most basic properties. Section 3 shows how to construct the inverse of a Fiedler matrix and studies some of its properties. Section 4 contains all the results concerning condition numbers of Fiedler matrices in the Frobenius norm. In Section 5, we introduce the concept of *staircase matrices* and determine their rank. These results are then used in Section 6 to prove some properties of the singular values of Fiedler matrices. Finally, Section 7 gives some conclusions and describes possible future work on Fiedler companion matrices.

2. Definition and basic properties of Fiedler Matrices

Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ be a polynomial with $a_i \in \mathbb{C}$, for $i = 0, 1, \ldots, n-1$. From $p(z)$, we define the $n \times n$ matrices

$$M_0 := \begin{bmatrix} I_{n-1} & 0 \\ 0 & -a_0 \end{bmatrix} \quad \text{and} \quad M_k := \begin{bmatrix} I_{n-k-1} & & & \\ & -a_k & 1 & \\ & 1 & 0 & \\ & & & I_{k-1} \end{bmatrix}, \quad k = 1, \ldots, n-1, \tag{2}$$

which are the basic factors used to build all the Fiedler matrices. Here and in the rest of the paper I_j denotes the $j \times j$ identity matrix. In [6] Fiedler matrices are constructed as

$$M_{i_1} M_{i_2} \cdots M_{i_n},$$

where $(i_1, i_2, \ldots, i_n)$ is any possible permutation of the n-tuple $(0, 1, \ldots, n-1)$. In order to better express certain key properties of this permutation and the resulting Fiedler matrix, in [3] the authors index the product of the M_i-factors in a slightly different way, as it is described in the following definition.

Definition 2.1. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$ and let M_i, for $i = 0, 1, \ldots, n-1$, be the matrices defined in* (2)*. Given any bijection $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$, the Fiedler matrix of $p(z)$ associated with σ is the $n \times n$ matrix*

$$M_\sigma := M_{\sigma^{-1}(1)} \cdots M_{\sigma^{-1}(n)}. \tag{3}$$

Note that $\sigma(i)$ describes the position of the factor M_i in the product $M_{\sigma^{-1}(1)} \cdots M_{\sigma^{-1}(n)}$, i.e., $\sigma(i) = j$ means that M_i is the jth factor in the product.

When necessary, we will explicitly indicate the dependence of the matrices (2) and (3) on a certain polynomial $p(z)$ by writing $M_i(p)$ and $M_\sigma(p)$. This family of matrices $\{M_k\}_{k=0}^{n-1}$ satisfies the following commutativity relations

$$M_i M_j = M_j M_i \quad \text{for } |i - j| \neq 1. \tag{4}$$

It is proved in [6] that all Fiedler matrices of $p(z)$ are similar, and so have $p(z)$ as characteristic polynomial. The set of Fiedler matrices includes the *first and second Frobenius companion forms* of $p(z)$, that is, the matrix C_1 defined in (1) and $C_2 := C_1^T$. More precisely,

$$C_1 = M_{n-1} M_{n-2} \cdots M_1 M_0 \quad \text{and} \quad C_2 = M_0 M_1 \cdots M_{n-2} M_{n-1}\,. \tag{5}$$

Observe that the matrices M_i are symmetric, therefore the transpose of any Fiedler matrix is another Fiedler matrix which corresponds to reverse the order of the M_i factors in (3).

The set of Fiedler matrices also includes four pentadiagonal matrices which have a much smaller bandwidth than C_1 and C_2 for large n. This property is of interest in fast numerical methods for computing roots of polynomials. These pentadiagonal matrices are constructed in Example 2.2.

Example 2.2. *Let $A = M_1 M_3 \cdots$ be the product of all factors M_i with odd index, and let $B = M_2 M_4 \cdots$ be the product of all factors M_i with even index, excluding M_0. Clearly A and B are tridiagonal matrices, and M_0 is diagonal, so the product of A, B, and M_0 in any order yields a pentadiagonal matrix. There are four Fiedler pentadiagonal matrices, because M_0 commutes with B. These are*

$$P_1 = M_0 BA, \quad P_2 = BAM_0, \quad P_3 = AM_0 B, \quad P_4 = M_0 AB.$$

Notice that, due to the commutativity relations (4) *and the fact that the matrices* M_i *are symmetric we have* $P_3 = P_1^T$ *and* $P_4 = P_2^T$*. Therefore there are essentially only two different pentadiagonal Fiedler matrices. For a polynomial* $p(z) = z^8 + \sum_{k=0}^{7} a_k z^k$ *with degree* 8*, they are explicitly:*

$$P_1 = \begin{bmatrix} -a_7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -a_6 & 0 & -a_5 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -a_4 & 0 & -a_3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -a_2 & 0 & -a_1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -a_0 & 0 \end{bmatrix}, \quad P_2 = \begin{bmatrix} -a_7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -a_6 & 0 & -a_5 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -a_4 & 0 & -a_3 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -a_2 & 0 & -a_1 & -a_0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

The relations (4) imply that some Fiedler matrices associated with different bijections σ are equal. For example, for $n = 3$, the Fiedler matrices $M_0M_2M_1$ and $M_2M_0M_1$ are equal. These relations suggest that the relative positions of the matrices M_i and M_{i+1} in the product M_σ are of fundamental interest in studying Fiedler matrices. This motivates Definition 2.3, partially introduced in [3].

Definition 2.3. *Let* $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ *be a bijection.*

(a) *For* $i = 0, \ldots, n-2$*, we say that* σ *has a* consecution *at* i *if* $\sigma(i) < \sigma(i+1)$ *and that* σ *has an* inversion *at* i *if* $\sigma(i) > \sigma(i+1)$*.*

(b) *The* consecution-inversion structure sequence of σ*, denoted by* $\mathrm{CISS}(\sigma)$*, is the tuple* $(\mathfrak{c}_0, \mathfrak{i}_0, \mathfrak{c}_1, \mathfrak{i}_1, \ldots, \mathfrak{c}_\ell, \mathfrak{i}_\ell)$*, where* σ *has* $\mathfrak{c}_0$ *consecutive consecutions at* $0, 1, \ldots, \mathfrak{c}_0 - 1$*;* $\mathfrak{i}_0$ *consecutive inversions at* $\mathfrak{c}_0, \mathfrak{c}_0 + 1, \ldots, \mathfrak{c}_0 + \mathfrak{i}_0 - 1$ *and so on, up to* $\mathfrak{i}_\ell$ *inversions at* $n - 1 - \mathfrak{i}_\ell, \ldots, n-2$*.*

(c) *The* reduced consecution-inversion structure sequence of σ*, denoted by* $\mathrm{RCISS}(\sigma)$*, is the sequence obtained from* $\mathrm{CISS}(\sigma)$ *after removing the zero entries.*

(d) *The* number of initial consecutions or inversions of σ*, denoted by* t_σ*, is*

$$t_\sigma = \begin{cases} \mathfrak{c}_0 & \text{if } \mathfrak{c}_0 \neq 0, \\ \mathfrak{i}_0 & \text{if } \mathfrak{c}_0 = 0. \end{cases}$$

Remark 2.4. The following simple observations on Definition 2.3 will be used freely.

1. σ has a consecution at i if and only if M_i is to the left of M_{i+1} in the Fiedler matrix M_σ, while σ has an inversion at i if and only if M_i is to the right of M_{i+1} in M_σ.
2. Note that $\mathfrak{c}_0$ and $\mathfrak{i}_\ell$ in $\mathrm{CISS}(\sigma)$ may be zero (in the first case, σ has an inversion at 0 and in the second one it has a consecution at $n-2$) but $\mathfrak{i}_0, \mathfrak{c}_1, \mathfrak{i}_1, \ldots, \mathfrak{i}_{\ell-1}, \mathfrak{c}_\ell$ are all strictly positive. These conditions uniquely determine $\mathrm{CISS}(\sigma)$ and, in particular, the parameter ℓ.
3. According to the previous comment, $\mathrm{RCISS}(\sigma) = \mathrm{CISS}(\sigma)$ if and only if $\mathfrak{c}_0 \neq 0$ and $\mathfrak{i}_\ell \neq 0$.
4. $1 \leq t_\sigma \leq n-1$.

Example 2.5. *For the pentadiagonal matrix* P_1 *in Example* 2.2 *with degree* $n = 8$*, we have* $\mathrm{CISS}(\sigma_1) = (1, 1, 1, 1, 1, 1, 1, 0)$, $\mathrm{RCISS}(\sigma_1) = (1, 1, 1, 1, 1, 1, 1)$*, and* $t_{\sigma_1} = 1$*.*

For the pentadiagonal matrix P_2 *with degree* $n = 8$*, we have* $\mathrm{CISS}(\sigma_2) = (0, 2, 1, 1, 1, 1, 1, 0)$, $\mathrm{RCISS}(\sigma_2) = (2, 1, 1, 1, 1, 1)$*, and* $t_{\sigma_2} = 2$*.*

For the first Frobenius companion matrix C_1 *for arbitrary degree* n *in* (5)*, we have* $\mathrm{CISS}(\mu_1) = (0, n-1)$, $\mathrm{RCISS}(\mu_1) = (n-1)$*, and* $t_{\mu_1} = n-1$*.*

For the second Frobenius companion matrix C_2 *for arbitrary degree* n *in* (5)*, we have* $\mathrm{CISS}(\mu_2) = (n-1, 0)$, $\mathrm{RCISS}(\mu_2) = (n-1)$*, and* $t_{\mu_2} = n-1$*.*

2.1. A multiplication free algorithm to construct Fiedler matrices

To construct a Fiedler matrix M_σ, the obvious way is to perform the multiplication of all the M_i factors directly, but in [5, Algorithm 1], the authors give an algorithm which constructs Fiedler matrices without performing any arithmetic operation. Algorithm 1 in [5] considers the general case of Fiedler linearization of nonmonic matrix polynomials. In Theorem 2.6 we recall this algorithm only for monic scalar polynomials. Here, and in the rest of the paper, we use *MATLAB notation for submatrices*, that is, $A(i:j,:)$ indicates the submatrix of A consisting of rows i through j and $A(:,k:l)$ indicates the submatrix of A consisting of columns k through l.

Theorem 2.6. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$, let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection, and let M_σ be the Fiedler matrix of $p(z)$ associated with σ. Then* `Algorithm 1` *constructs M_σ.*

> `Algorithm 1.` *Given $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ and a bijection σ, the following algorithm constructs M_σ.*
>
> if σ has a consecution at 0 then
> $$W_0 = \begin{bmatrix} -a_1 & 1 \\ -a_0 & 0 \end{bmatrix}$$
> else
> $$W_0 = \begin{bmatrix} -a_1 & -a_0 \\ 1 & 0 \end{bmatrix}$$
> endif
> for $i = 1 : n-2$
> if σ has a consecution at i then
> $$W_i = \begin{bmatrix} -a_{i+1} & 1 & 0 \\ W_{i-1}(:,1) & 0 & W_{i-1}(:,2:i+1) \end{bmatrix}$$
> else
> $$W_i = \begin{bmatrix} -a_{i+1} & W_{i-1}(1,:) \\ 1 & 0 \\ 0 & W_{i-1}(2:i+1,:) \end{bmatrix}$$
> endif
> endfor
> $M_\sigma = W_{n-2}$

The interest of this algorithm, apart from constructing Fiedler matrices without performing any arithmetic operation, is that it allows to prove easily some elementary properties of Fiedler matrices. For instance, since `Algorithm 1` performs $n-1$ "if" decisions, we get that there are at most 2^{n-1} different Fiedler matrices associated with any $p(z)$ of degree $n \geq 2$. In fact, with a little bit of extra effort, the reader may prove by induction on W_i that if $a_0 \neq -1$, then all these 2^{n-1} Fiedler matrices are really different, i.e., different for any set of specific values of the coefficients $a_0, a_1, \ldots, a_{n-1}$. However, if $a_0 = -1$, then `Algorithm 1` produces the same W_0 for σ having either a consecution or an inversion at 0, and there are only 2^{n-2} different Fiedler matrices. We summarize these results without proof in Corollary 2.7.

Corollary 2.7. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$.*

(a) *If $a_0 \neq -1$, then there are 2^{n-1} different Fiedler matrices associated with $p(z)$.*

(b) *If $a_0 = -1$, then there are 2^{n-2} different Fiedler matrices associated with $p(z)$.*

From `Algorithm 1`, it is also very easy to prove Theorem 2.8 via an straightforward induction on the matrices W_i. We omit the proof, since Theorem 2.8 is a particular case of the much more general result [5, Theorem 3.10], which proves several structural properties of Fiedler linearizations of rectangular matrix polynomials.

Theorem 2.8. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$, let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection, and let M_σ be the Fiedler matrix of $p(z)$ associated with σ. Then:*

(a) *M_σ has n entries equal to $-a_0, -a_1, \ldots, -a_{n-1}$, with exactly one copy of each.*

(b) *M_σ has $n-1$ entries equal to 1.*

(c) *The rest of the entries of M_σ are equal to 0.*

(d) *If an entry equal to 1 of those in part* (b) *is at position (i, j), then either the rest of the entries in the ith row of M_σ are equal to 0 or the rest of the entries in the jth column of M_σ are equal to 0.*

In plain words, Theorem 2.8 establishes the fact that any Fiedler matrix has the same entries as the first and second Frobenius companion forms, although placed on different positions. So all the Fiedler matrices associated with a given polynomial have the same Frobenius norm.

Corollary 2.9. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$, let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection, and let M_σ be the Fiedler matrix of $p(z)$ associated with σ. Then:*

$$||M_\sigma||_F = \sqrt{(n-1) + |a_0|^2 + |a_1|^2 + \cdots + |a_{n-1}|^2}, \tag{6}$$

which is independent on σ and depends only on $p(z)$.

3. The inverse of a Fiedler Matrix

For $k = 1, \ldots, n-1$, the matrices M_k defined in (2) are nonsingular for any value of the coefficients a_k, while the matrix M_0 is nonsingular if and only if $a_0 \neq 0$. In this case, the inverses of these matrices are

$$M_0^{-1} = \begin{bmatrix} I_{n-1} & 0 \\ 0 & -1/a_0 \end{bmatrix}, \qquad M_k^{-1} = \begin{bmatrix} I_{n-k-1} & & & \\ & 0 & 1 & \\ & 1 & a_k & \\ & & & I_{k-1} \end{bmatrix}, \quad k = 1, 2, \ldots, n-1. \tag{7}$$

For any bijection σ, the Fiedler matrix M_σ in (3) is nonsingular if and only if $a_0 \neq 0$, that is, if $z = 0$ is not a root of $p(z)$, and equation (7) allows us to obtain a factorized expression of M_σ^{-1} given by

$$M_\sigma^{-1} = (M_{\sigma^{-1}(1)} \cdots M_{\sigma^{-1}(n)})^{-1} = M_{\sigma^{-1}(n)}^{-1} \cdots M_{\sigma^{-1}(1)}^{-1}.$$

However, as we did in `Algorithm 1` for M_σ, it is possible to construct the inverse of any Fiedler matrix via the simple `Algorithm 2`. This algorithm allows us to prove easily some key properties of M_σ^{-1} in Theorem 3.2. Note that `Algorithm 2` is not operation free, although the only arithmetic operations involved are multiplications of certain coefficients of $p(z)$ by $1/a_0$ (see Theorem 3.2).

Theorem 3.1. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$ and $a_0 \neq 0$, let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection, and let M_σ be the Fiedler matrix of $p(z)$ associated with σ. Then* `Algorithm 2` *constructs M_σ^{-1}.*

`Algorithm 2`. *Given $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$, with $a_0 \neq 0$, and a bijection σ, the following algorithm constructs M_σ^{-1}.*

if σ has a consecution at 0 then

$$B_0 = \begin{bmatrix} 0 & -1/a_0 \\ 1 & -a_1/a_0 \end{bmatrix}$$

else

$$B_0 = \begin{bmatrix} 0 & 1 \\ -1/a_0 & -a_1/a_0 \end{bmatrix}$$

endif

for $i = 1 : n-2$

if σ has a consecution at i then

$$B_i = \begin{bmatrix} 0 & B_{i-1}(1,:) \\ 1 & a_{i+1} B_{i-1}(1,:) \\ 0 & B_{i-1}(2:i+1,:) \end{bmatrix}$$

else

$$B_i = \begin{bmatrix} 0 & 1 & 0 \\ B_{i-1}(:,1) & a_{i+1}B_{i-1}(:,1) & B_{i-1}(:,2:i+1) \end{bmatrix}$$

endif

endfor

$M_\sigma^{-1} = B_{n-2}$

Proof. Let $\{W_0, W_1, \ldots, W_{n-2}\}$ be the sequence of matrices constructed by `Algorithm 1` and $\{B_0, B_1, \ldots, B_{n-2}\}$ be the sequence of matrices constructed by `Algorithm 2`. The proof consists of proving by induction that $W_iB_i = I_{i+2}$, i.e., that $B_i = W_i^{-1}$, which implies the theorem just by taking $i = n-2$.

If σ has a consecution at 0, then a direct multiplication of 2×2 matrices leads to $W_0B_0 = I_2$. The same happens if σ has an inversion at 0. Let us assume that $W_{i-1}B_{i-1} = I_{i+1}$ for some $i-1 \geq 0$ and let us prove $W_iB_i = I_{i+2}$. If σ has a consecution at i, then, from `Algorithms 1` and `2`, we get

$$W_iB_i = \begin{bmatrix} 1 & -a_{i+1}B_{i-1}(1,:) + a_{i+1}B_{i-1}(1,:) \\ 0 & W_{i-1}(:,1)B_{i-1}(1,:) + W_{i-1}(:,2:i+1)B_{i-1}(2:i+1,:) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & W_{i-1}B_{i-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & I_{i+1} \end{bmatrix}.$$

If σ has an inversion at i, then the proof is similar and is omitted. □

`Algorithm 2` allows us to easily get information on the entries of M_σ^{-1} in Theorem 3.2.

Theorem 3.2. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$ and $a_0 \neq 0$, let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection, let M_σ be the Fiedler matrix of $p(z)$ associated with σ, and let t_σ be the number of initial consecutions or inversions of σ. Then:*

(a) *M_σ^{-1} has $t_\sigma + 1$ entries equal to $-\frac{1}{a_0}, -\frac{a_1}{a_0}, \ldots, -\frac{a_{t_\sigma}}{a_0}$, with exactly one copy of each.*

(b) *M_σ^{-1} has $n - 1 - t_\sigma$ entries equal to $a_{t_\sigma+1}, a_{t_\sigma+2}, \ldots, a_{n-1}$, with exactly one copy of each.*

(c) *M_σ^{-1} has $n-1$ entries equal to 1.*

(d) *The rest of the entries of M_σ^{-1} are equal to 0.*

Proof. Recall that, according to Definition 2.3, $t_\sigma = \mathfrak{c}_0$ if $\mathfrak{c}_0 \neq 0$, and $t_\sigma = \mathfrak{i}_0$ if $\mathfrak{c}_0 = 0$. The case $\mathfrak{c}_0 = 0$ follows from the case $\mathfrak{c}_0 \neq 0$ by applying the result to the Fiedler matrix M_σ^T, which corresponds to a bijection with $\mathfrak{i}_0$ initial consecutions.

Therefore, we prove only the result for $t_\sigma = \mathfrak{c}_0 \neq 0$. In this case, the bijection σ has consecutions at $0, 1, 2, \ldots, \mathfrak{c}_0 - 1$ and an inversion at $\mathfrak{c}_0$. Therefore, a direct application of `Algorithm 2` leads to

$$B_{\mathfrak{c}_0-1} = \begin{bmatrix} 0 & 0 & \ldots & & 0 & -1/a_0 \\ 1 & 0 & \ddots & & 0 & -a_{\mathfrak{c}_0}/a_0 \\ & 1 & \ddots & & & \vdots \\ & & \ddots & \ddots & & \vdots \\ & & & 1 & 0 & -a_2/a_0 \\ & & & & 1 & -a_1/a_0 \end{bmatrix}, \quad B_{\mathfrak{c}_0} = \left[\begin{array}{c|cccccc} 0 & 1 & 0 & \ldots & & 0 & 0 \\ \hline 0 & 0 & 0 & \ldots & & 0 & -1/a_0 \\ 1 & a_{\mathfrak{c}_0+1} & 0 & \ddots & & 0 & -a_{\mathfrak{c}_0}/a_0 \\ 0 & 0 & 1 & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & \ddots & & \vdots \\ \vdots & \vdots & & & 1 & 0 & -a_2/a_0 \\ 0 & 0 & & & & 1 & -a_1/a_0 \end{array}\right]. \quad (8)$$

Observe that the nonzero entries of $B_{\mathfrak{c}_0}$ are: $\mathfrak{c}_0 + 1$ entries equal to 1, $-\frac{1}{a_0}, -\frac{a_1}{a_0}, \ldots, -\frac{a_{\mathfrak{c}_0}}{a_0}$, and $a_{\mathfrak{c}_0+1}$. In addition, both the first row and the first column of $B_{\mathfrak{c}_0}$ satisfy that they have only one nonzero entry and that this entry is equal to 1.

From `Algorithm 2`, one obtains by inspection the following property: if the first row and the first column of B_{i-1} satisfy that they have only one nonzero entry and that this entry is equal to 1, then (a) the nonzero entries of B_i are those of B_{i-1} together with an additional 1 and a_{i+1}, and (b) the first row and the first column of B_i have also only one nonzero entry and this entry is equal to 1.

This property and (8) imply that the nonzero entries of $M_\sigma^{-1} = B_{n-2}$ are those of $B_{\mathfrak{c}_0}$ together with $n - 2 - \mathfrak{c}_0$ entries equal to 1 and $a_{\mathfrak{c}_0+2}, a_{\mathfrak{c}_0+3}, \ldots, a_{n-1}$. This completes the proof. □

Theorem 3.2 allows us to give an explicit expression for the Frobenius norm of the inverse of any Fiedler matrix.

Corollary 3.3. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$ and $a_0 \neq 0$, let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection, let M_σ be the Fiedler matrix of $p(z)$ associated with σ, and let t_σ be the number of initial consecutions or inversions of σ. Then:*

$$||M_\sigma^{-1}||_F^2 = (n-1) + \frac{1 + |a_1|^2 + \cdots + |a_{t_\sigma}|^2}{|a_0|^2} + |a_{t_\sigma+1}|^2 + \cdots + |a_{n-1}|^2.$$

In contrast with $||M_\sigma||_F$, the quantity $||M_\sigma^{-1}||_F$ depends on σ, although only through the number of its initial consecutions or inversions t_σ. This implies that very different Fiedler matrices can have inverses with the same Frobenius norm and, therefore, can have the same condition numbers in the Frobenius norm.

4. Condition numbers for inversion in the Frobenius norm

In this section, we start by presenting in Theorem 4.1 an explicit expression for the condition number of any Fiedler matrix in the Frobenius norm, as an immediate consequence of Corollaries 2.9 and 3.3. This expression will allow us to easily establish several relevant properties of these condition numbers.

Theorem 4.1. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$ and $a_0 \neq 0$, let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection, let M_σ be the Fiedler matrix of $p(z)$ associated with σ, and let t_σ be the number of initial consecutions or inversions of σ. Define*

$$N(p)^2 := (n-1) + |a_0|^2 + |a_1|^2 + \cdots + |a_{n-1}|^2.$$

Then,

$$\kappa_F^2(M_\sigma) = N(p)^2 \left((n-1) + \frac{1 + |a_1|^2 + \cdots + |a_{t_\sigma}|^2}{|a_0|^2} + |a_{t_\sigma+1}|^2 + \cdots + |a_{n-1}|^2 \right). \tag{9}$$

Corollary 4.2 gives crude lower and upper bounds on $\kappa_F(M_\sigma)$ that are independent on σ and show that, for any σ, $\kappa_F(M_\sigma)$ is large if and only if $|a_0|$ is small or $|a_i|$ is large for some $i = 0, 1, \ldots, n-1$ (or both).

Corollary 4.2. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$ and $a_0 \neq 0$, let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection, and let M_σ be the Fiedler matrix of $p(z)$ associated with σ.*

(a) *If $|a_0| \leq 1$, then*

$$\frac{\sqrt{n-1+|a_1|^2 + \cdots + |a_{n-1}|^2}}{|a_0|} \leq \kappa_F(M_\sigma) \leq \frac{n + |a_1|^2 + \cdots + |a_{n-1}|^2}{|a_0|}.$$

(b) *If $|a_0| > 1$, then*

$$\sqrt{(n-1) + |a_0|^2 + |a_1|^2 + \cdots + |a_{n-1}|^2} \leq \kappa_F(M_\sigma) \leq (n-1) + |a_0|^2 + |a_1|^2 + \cdots + |a_{n-1}|^2.$$

The proof of Corollary 4.2 is omitted since it follows trivially from Theorem 4.1. We would like to remark that it is natural that $\kappa_F(M_\sigma)$ is large if $|a_0|$ is small, because M_σ is singular when $a_0 = 0$. However, it might not be so clear why $\kappa_F(M_\sigma)$ is large, i.e., M_σ is close in relative distance to a singular matrix, if $|a_i|$ is large for some $i = 0, 1, \ldots, n-1$. The reason resides in Theorem 2.8-(d), because if some $|a_i| \gg 1$, then a tiny relative normwise perturbation can turn one of the entries equal to 1 in M_σ into 0 and can make the matrix singular. This property shows that "representing" a polynomial $p(z)$ via a Fiedler companion matrix is not convenient if some $|a_i| \gg 1$ because the "structural" entries equal to one are fragile under non-structured tiny perturbations.

Another direct consequence of Theorem 4.1 is Corollary 4.3, which gives a necessary and sufficient condition for two Fiedler matrices to have the same condition numbers for any monic polynomial $p(z)$.

Corollary 4.3. *Let $\mathbb{P}_n = \{z^n + \sum_{k=0}^{n-1} a_k z^k \, : \, a_0 \neq 0\}$ be the set of monic polynomials of degree $n \geq 2$ without roots equal to 0. Let $\sigma_1, \sigma_2 : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be two bijections and let t_{σ_1} and t_{σ_2} be, respectively, the numbers of initial consecutions or inversions of σ_1 and σ_2. Let $M_{\sigma_1}(p)$ and $M_{\sigma_2}(p)$ be, respectively, the Fiedler matrices of $p \in \mathbb{P}_n$ associated with σ_1 and σ_2. Then, $t_{\sigma_1} = t_{\sigma_2}$ if and only if $\kappa_F(M_{\sigma_1}(p)) = \kappa_F(M_{\sigma_2}(p))$ for all $p \in \mathbb{P}_n$.*

Proof. It is obvious that $t_{\sigma_1} = t_{\sigma_2}$ implies $\kappa_F(M_{\sigma_1}(p)) = \kappa_F(M_{\sigma_2}(p))$ for all $p \in \mathbb{P}_n$ by (9). To prove the converse, assume that $\kappa_F(M_{\sigma_1}(p)) = \kappa_F(M_{\sigma_2}(p))$ for all $p \in \mathbb{P}_n$ and proceed by contradiction, i.e., assume $t_{\sigma_1} \neq t_{\sigma_2}$. More precisely assume without loss of generality that $t_{\sigma_1} < t_{\sigma_2}$. Take $p(z)$ such that $a_0 = 2$, $a_{t_{\sigma_2}} = 1$, and $a_i = 0$ for $i \neq 0, t_{\sigma_2}$. Then $\kappa_F(M_{\sigma_1}(p)) = \kappa_F(M_{\sigma_2}(p))$ and (9) imply $1/4 + 1 = (1+1)/4$, which is a contradiction. □

Example 4.4. *In this example all considered Fiedler matrices correspond to the same polynomial $p(z)$. According to Example 2.5, the condition numbers in Frobenius norm of the classical Frobenius companion matrices C_1 and C_2 in (5) are equal. This is obvious because $C_2 = C_1^T$. It is however somewhat surprising that the condition numbers of the two pentadiagonal matrices P_1 and P_2 introduced in Example 2.2 are, in general, different. This follows from Corollary 4.3 and the fact $t_{\sigma_1} = 1$ for P_1 and $t_{\sigma_2} = 2$ for P_2 (see Example 2.5). In fact, we will see in Theorem 4.10 that these condition numbers can be arbitrarily different for properly chosen polynomials.*

4.1. Ordering Fiedler matrices according to condition numbers in the Frobenius norm

Given a polynomial $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$, with $n \geq 2$ and $a_0 \neq 0$, and a number t such that $1 \leq t \leq n-1$, Corollary 4.3 establishes that all Fiedler matrices of $p(z)$ in the set

$$\mathcal{S}_t(p) := \{M_\sigma(p) \, : \, t_\sigma = t\}$$

have the same condition number $\kappa_F(M_\sigma(p))$. In the generic case $a_0 \neq -1$ (recall Corollary 2.7) the cardinality of $\mathcal{S}_t(p)$ is given by

$$|\mathcal{S}_t(p)| = \begin{cases} 2^{n-1-t}, & \text{if } t < n-1, \\ 2, & \text{if } t = n-1. \end{cases} \tag{10}$$

This can be seen as follows. If $t_\sigma = n-1$, then σ has $n-1$ consecutions and no inversions, or vice versa. This corresponds to the two classical Frobenius companion matrices. If $t_\sigma = t < n-1$, then σ has consecutions at $0, 1, \ldots, t-1$ and an inversion at t, or vice versa. For each of these two cases, we can select freely the consecutions/inversions at $t+1, \ldots, n-2$. This can be done in 2^{n-2-t} different ways, that according to `Algorithm 1` give each of them a different Fiedler matrix. The value of t in $\mathcal{S}_t(p)$ and expression (9) allow us to order all Fiedler matrices of $p(z)$ by increasing/decreasing condition numbers in Corollary 4.5. *Observe that there are only three possible different orders of this type, since the order via increasing/decreasing condition numbers is the same for all polynomials with $|p(0)| < 1$, it is also the same for all polynomials with $|p(0)| > 1$, and also the same for all polynomials with $|p(0)| = 1$.*

Corollary 4.5. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$ and $a_0 \neq 0$, and let t be a number such that $1 \leq t \leq n-1$. Let $\mathcal{S}_t(p) = \{M_\sigma(p) \, : \, t_\sigma = t\}$ be the set of Fiedler matrices of $p(z)$ associated with bijections σ whose number of initial consecutions or inversions is equal to t. Define*

$$\kappa(t) := \kappa_F(M_\sigma(p)), \quad \text{for } M_\sigma(p) \in \mathcal{S}_t(p), \tag{11}$$

which does not depend on the specific bijection σ as long as $t_\sigma = t$. Then the following results hold.

(a) *If $|a_0| < 1$, then $\kappa(1) \leq \kappa(2) \leq \cdots \leq \kappa(n-1)$.*

(b) *If $|a_0| = 1$, then $\kappa(1) = \kappa(2) = \cdots = \kappa(n-1)$.*

(c) *If $|a_0| > 1$, then $\kappa(1) \geq \kappa(2) \geq \cdots \geq \kappa(n-1)$.*

Proof. The result follows from (9), since this expression makes obvious that if $|a_0| < 1$, then $\kappa_F(M_\sigma(p))$ increases as the number t_σ of coefficients $|a_i|^2$ divided by $|a_0|^2$ increases. The other cases are proved in a similar way. □

Remark 4.6. From Corollary 4.5 we see that if $|a_0| < 1$, then the two Frobenius companion matrices have the largest condition number among all Fiedler matrices of $p(z)$, since the set $\mathcal{S}_{n-1}(p) = \{M_\sigma(p) : t_\sigma = n-1\}$ contains only the two Frobenius companion matrices. On the contrary, the Fiedler matrices in $\mathcal{S}_1(p) = \{M_\sigma(p) : t_\sigma = 1\}$ have the smallest condition number among all Fiedler matrices of $p(z)$ if $|a_0| < 1$. If n is large, then there are many Fiedler matrices with smallest condition number, since according to (10), $\mathcal{S}_1(p)$ has 2^{n-2} elements. In particular, $\mathcal{S}_1(p)$ contains the pentadiagonal Fiedler matrices P_1 and $P_3 = P_1^T$ (see Examples 2.2 and 2.5), but not the pentadiagonal matrices P_2 and $P_4 = P_2^T$, which have a larger condition number if $|a_0| < 1$.

If $|a_0| > 1$, then similar remarks hold but with reverse order for the magnitudes of the condition numbers. In this case, the Frobenius companion matrices have the smallest condition number among all Fiedler matrices of $p(z)$.

The clear and simple ordering of Fiedler matrices according to condition numbers in the Frobenius norm presented in Corollary 4.5 does not hold for condition numbers in other matrix norms often used in the literature as, for instance, the $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ [7]. This is one of the reasons why we have chosen to use the Frobenius norm in this paper. Of course, the equivalence of all these norms via constants smaller than or equal to n implies that the order in Corollary 4.5 between $\kappa(t)$ and $\kappa(t+1)$ can be broken in other norms only if $\kappa(t)$ and $\kappa(t+1)$ are not very different. We illustrate these points in Example 4.7, which also shows that an ordering based on the number of initial consecutions or inversions of σ is not possible for these other norms.

Example 4.7. *In Figure* 1 *we consider the polynomials* $p_1(z) = z^5 + 0.01z^4 + 0.01z^3 + 0.01z^2 + 0.01z + 0.01$ *and* $p_2(z) = z^5 + 10z^4 + z^3 + z^2 + 10z + 0.01$, *both of degree* 5. *We have constructed in MATLAB the* 8 *Fiedler matrices for each of these polynomials associated with bijections that have an inversion at* 0. *The matrices associated with bijections that have a consecution at* 0 *are the transposes of the previous ones and have not been considered for simplicity. Each of these Fiedler matrices has been labeled with an index from* 1 *to* 8, *according to the following table.*

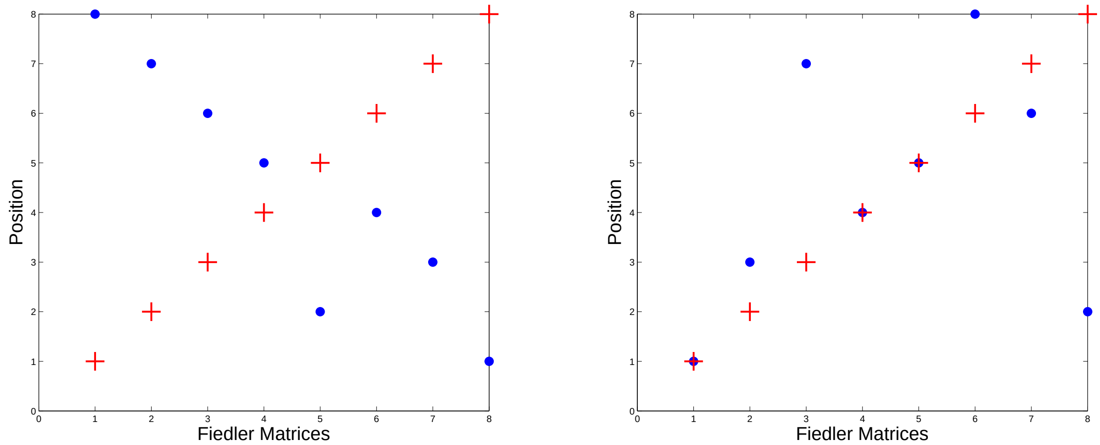
index	CISS(σ)	t_σ
1	$(0,4)$	4
2	$(0,3,1,0)$	3
3	$(0,2,1,1)$	2
4	$(0,2,2,0)$	2

index	CISS(σ)	t_σ
5	$(0,1,1,2)$	1
6	$(0,1,1,1,1,0)$	1
7	$(0,1,2,1)$	1
8	$(0,1,3,0)$	1

These indices are represented in the horizontal axes of the plots in Figure 1. *For these* 8 *Fiedler matrices of each polynomial* $p_1(z)$ *and* $p_2(z)$, *we have computed their condition numbers in the* 2*-norm (that is, the ratio between the largest and smallest singular values) and we have ordered the matrices by decreasing magnitudes of these condition numbers, i.e., the matrix with the largest condition number is in the first position. The positions of the Fiedler matrices with respect this ordering are represented in the vertical axes of the plots in Figure* 1 *by using the symbol "•". In addition, the positions of the same Fiedler matrices with respect the ordering corresponding to decreasing Frobenius condition numbers are represented in the vertical axes by using the symbol "+". We observe that the ordering with respect the* 2*-norm condition number differs completely from* $p_1(z)$ *to* $p_2(z)$, *and in both cases is very different from the one corresponding to Frobenius condition numbers. Other interesting point to be remarked is that, both for* $p_1(z)$ *and* $p_2(z)$, *the condition numbers in the* 2*-norm of the* 8 *considered Fiedler matrices are all different each other, and so the same value of* t_σ *does not imply the same condition number in the* 2*-norm, in contrast with the behaviour in the Frobenius norm. This is not seen in Figure* 1, *but it may be easily checked by the reader with MATLAB. Finally, we mention that the condition numbers in the Frobenius norm for the Fiedler matrices of* $p_1(z)$ *range from* 200.063 *to* 200.093, *while the corresponding to* $p_2(z)$ *range from* $1.443 \cdot 10^4$ *to* $2.045 \cdot 10^4$.

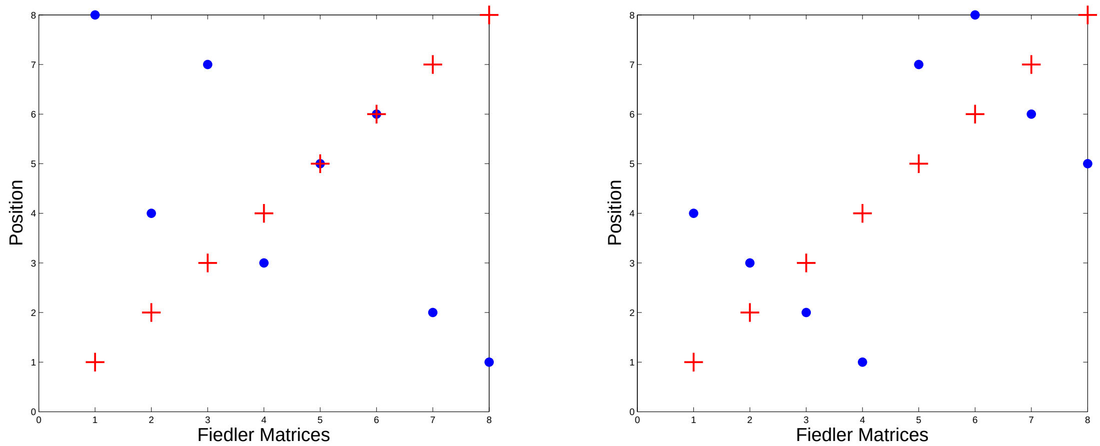
In Figure 2, *we repeat the same experiment for the* 1*-norm instead of the* 2*-norm and for the polynomials* $p_3(z) = z^5 + 10z^4 + 100z^3 + 10z^2 + 100z + 0.01$ *and* $p_4(z) = z^5 + 100z^4 + 10z^3 + 100z^2 + 10z + 0.01$. *The results obtained are similar to those in the* 2*-norm. In this case, the condition numbers in the Frobenius norm for the Fiedler matrices of* $p_3(z)$ *range from* $1.421 \cdot 10^6$ *to* $2.020 \cdot 10^6$, *while the corresponding to* $p_4(z)$ *range from* $2.020 \cdot 10^6$ *to* $0.144 \cdot 10^6$.

We do not show experiments in the ∞-norm, because the ∞-norm condition number of a matrix is the 1-norm condition number of its transpose, and the transpose of any Fiedler matrix is another Fiedler matrix with the same number of initial consecutions or inversions.



(a) $p_1(z) = z^5 + 0.01z^4 + 0.01z^3 + 0.01z^2 + 0.01z + 0.01$ (b) $p_2(z) = z^5 + 10z^4 + z^3 + z^2 + 10z + 0.01$

Figure 1: Ordering Fiedler matrices of a fixed polynomial according to decreasing condition numbers in the 2-norm (•) and in the Frobenius norm (+).



(a) $p_3(z) = z^5 + 10z^4 + 100z^3 + 10z^2 + 100z + 0.01$ (b) $p_4(z) = z^5 + 100z^4 + 10z^3 + 100z^2 + 10z + 0.01$

Figure 2: Ordering Fiedler matrices of a fixed polynomial according to decreasing condition numbers in the 1-norm (•) and in the Frobenius norm (+).

4.2. The ratio of the condition numbers of two Fiedler matrices

The important fact in applications is not whether one matrix is better conditioned than another or not. The really important fact is to know whether the condition number of one matrix is much smaller than the condition number of another or not. Therefore, we study in this section the ratio between the condition numbers in the Frobenius norm of any pair of Fiedler matrices of a fixed polynomial $p(z)$ that have different numbers of initial consecutions or inversions.

Lemma 4.8 states a simple technical result that will be used in the rest of this section.

Lemma 4.8. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$ and $a_0 \neq 0$, let $\sigma, \mu : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be two bijections, let M_σ and M_μ be the Fiedler matrices of $p(z)$ associated with σ and μ, and let t_σ and t_μ be the numbers of initial consecutions or inversions of σ and μ. Assume that $t_\sigma < t_\mu$ and define*

$$g_{\sigma,\mu} := (n-1) + \frac{1 + |a_1|^2 + \cdots + |a_{t_\sigma}|^2}{|a_0|^2} + |a_{t_\mu+1}|^2 + \cdots + |a_{n-1}|^2, \tag{12}$$

where if $t_\mu = n-1$, then $|a_{t_\mu+1}|^2 + \cdots + |a_{n-1}|^2$ is not present. Then

$$\left(\frac{\kappa_F(M_\mu)}{\kappa_F(M_\sigma)}\right)^2 = \frac{g_{\sigma,\mu} + \dfrac{|a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2}{|a_0|^2}}{g_{\sigma,\mu} + |a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2} \quad \text{and} \quad \left(\frac{\kappa_F(M_\sigma)}{\kappa_F(M_\mu)}\right)^2 = \frac{g_{\sigma,\mu} + |a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2}{g_{\sigma,\mu} + \dfrac{|a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2}{|a_0|^2}}.$$

Proof. It is another corollary of (9). Simply note that $\kappa_F^2(M_\mu) = N(p)^2 \left(g_{\sigma,\mu} + \frac{|a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2}{|a_0|^2}\right)$ and $\kappa_F^2(M_\sigma) = N(p)^2 \left(g_{\sigma,\mu} + |a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2\right)$. □

The ratios of condition numbers in Lemma 4.8 are complicated functions of the coefficients of the polynomial $p(z)$. Theorem 4.9 provides simple upper bounds for these ratios, which show that distinct Fiedler matrices of the same polynomial $p(z)$ may have very different condition numbers only if some of the coefficients $a_2, a_3, \ldots, a_{n-1}$ of the polynomial is very large, and a_0 is very small or very large.

Theorem 4.9. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$ and $a_0 \neq 0$, let $\sigma, \mu : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be two bijections, let M_σ and M_μ be the Fiedler matrices of $p(z)$ associated with σ and μ, and let t_σ and t_μ be the numbers of initial consecutions or inversions of σ and μ. Assume that $t_\sigma < t_\mu$ and define*

$$S_{\sigma,\mu} := \sum_{i=t_\sigma+1}^{t_\mu} |a_i|^2 \quad \text{and} \quad A = \max_{2 \leq i \leq n-1} |a_i|. \tag{13}$$

Then, the following statements hold.

(a) *If $|a_0| < 1$, then*

$$1 \leq \frac{\kappa_F(M_\mu)}{\kappa_F(M_\sigma)} \leq \min\left\{\sqrt{1 + S_{\sigma,\mu}}\,,\, \frac{1}{|a_0|}\right\} \leq \min\left\{\sqrt{1 + (n-2)A^2}\,,\, \frac{1}{|a_0|}\right\}. \tag{14}$$

(b) *If $|a_0| > 1$, then*

$$1 \leq \frac{\kappa_F(M_\sigma)}{\kappa_F(M_\mu)} \leq \min\left\{\sqrt{1 + \frac{S_{\sigma,\mu}}{n-1}}\,,\, |a_0|\right\} \leq \min\left\{\sqrt{1 + \frac{n-2}{n-1}A^2}\,,\, |a_0|\right\}. \tag{15}$$

Observe that the rightmost upper bounds in parts (a) *and* (b) *are both independent of σ and μ.*

Proof. Part (a). From Corollary 4.5 and Lemma 4.8, we have

$$\begin{aligned} 1 \leq \left(\frac{\kappa_F(M_\mu)}{\kappa_F(M_\sigma)}\right)^2 &\leq \frac{g_{\sigma,\mu} + \dfrac{|a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2}{|a_0|^2}}{g_{\sigma,\mu}} = 1 + \frac{|a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2}{g_{\sigma,\mu}\,|a_0|^2} \\ &\leq 1 + |a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2 = 1 + S_{\sigma,\mu}\,, \end{aligned}$$

where in the last inequality we have used that $1 < g_{\sigma,\mu}\,|a_0|^2$. To get the rightmost bound in Part (a), recall that $1 \leq t_\sigma\,,\, t_\mu \leq (n-1)$. So $S_{\sigma,\mu} \leq (n-2)A^2$. Next, we bound the ratio of condition numbers by $1/|a_0|$. To this purpose define $y := S_{\sigma,\mu}/g_{\sigma,\mu} \geq 0$ and $\alpha := 1/|a_0|^2 > 1$. Therefore Lemma 4.8 implies

$$\left(\frac{\kappa_F(M_\mu)}{\kappa_F(M_\sigma)}\right)^2 = \frac{1 + \alpha\, y}{1 + y}. \tag{16}$$

Observe that the function $g(y) = (1 + \alpha\, y)/(1 + y)$ satisfies: (i) $g(0) = 1$; (ii) $\lim_{y\to\infty} g(y) = \alpha$; and (iii) $g'(y) = (\alpha - 1)/(1+y)^2 > 0$. Therefore, $1 \leq g(y) < \alpha$, if $y \geq 0$, and (16) implies that $\kappa_F(M_\mu)/\kappa_F(M_\sigma) < 1/|a_0|$.

Part (b). From Corollary 4.5 and Lemma 4.8, we have

$$1 \leq \left(\frac{\kappa_F(M_\sigma)}{\kappa_F(M_\mu)}\right)^2 \leq \frac{g_{\sigma,\mu} + |a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2}{g_{\sigma,\mu}} = 1 + \frac{|a_{t_\sigma+1}|^2 + \cdots + |a_{t_\mu}|^2}{g_{\sigma,\mu}} \leq 1 + \frac{S_{\sigma,\mu}}{n-1},$$

where in the last inequality we have used that $n - 1 < g_{\sigma,\mu}$. To get the rightmost bound in Part (b), we use again that $S_{\sigma,\mu} \leq (n-2)A^2$. Next, we bound the ratio of condition numbers by $|a_0|$. To this purpose define $y := S_{\sigma,\mu}/g_{\sigma,\mu} \geq 0$ and $\alpha := 1/|a_0|^2 < 1$. Therefore Lemma 4.8 implies

$$\left(\frac{\kappa_F(M_\sigma)}{\kappa_F(M_\mu)}\right)^2 = \frac{1 + y}{1 + \alpha\, y}. \tag{17}$$

Observe that the function $h(y) = (1 + y)/(1 + \alpha\, y)$ satisfies: (i) $h(0) = 1$; (ii) $\lim_{y\to\infty} h(y) = 1/\alpha$; and (iii) $h'(y) = (1 - \alpha)/(1 + \alpha y)^2 > 0$. Therefore, $1 \leq h(y) < 1/\alpha$, if $y \geq 0$, and (17) implies that $\kappa_F(M_\sigma)/\kappa_F(M_\mu) < |a_0|$. □

It is obvious that there exist polynomials $p(z)$ for which the upper bound $\min\left\{\sqrt{1 + S_{\sigma,\mu}}\,,\ 1/|a_0|\right\}$ in (14) (resp. $\min\left\{\sqrt{1 + S_{\sigma,\mu}/(n-1)}\,,\ |a_0|\right\}$ in (15)) can be as large as desired, but this does not mean necessarily that the ratio $\kappa_F(M_\mu)/\kappa_F(M_\sigma)$ (resp. $\kappa_F(M_\sigma)/\kappa_F(M_\mu)$) for these polynomials is large. In fact, note that $\min\left\{\sqrt{1 + S_{\sigma,\mu}}\,,\ 1/|a_0|\right\}$ (resp. $\min\left\{\sqrt{1 + S_{\sigma,\mu}/(n-1)}\,,\ |a_0|\right\}$) does not depend of the coefficients of the polynomial that define the magnitude $g_{\sigma,\mu}$ in (12), with the exception of a_0. Therefore, according to Lemma 4.8, the upper bounds in (14) or (15) cannot determine the actual values of the ratios $\kappa_F(M_\mu)/\kappa_F(M_\sigma)$ or $\kappa_F(M_\sigma)/\kappa_F(M_\mu)$. Theorem 4.10 shows that if we fix a priori an arbitrary value of the upper bound in (14) or in (15), then there exist polynomials for which this upper bound is almost attained and another polynomials for which the ratios of the condition numbers of Fiedler matrices are arbitrarily close to 1. Note that, in particular, Theorem 4.10 shows that there are polynomials for which the ratios of the Frobenius condition numbers of two distinct Fiedler matrices can be arbitrarily large or arbitrarily small.

Theorem 4.10. *Let $\sigma, \mu : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be two bijections and let t_σ and t_μ be the numbers of initial consecutions or inversions of σ and μ. Assume that $t_\sigma < t_\mu$. For any polynomial $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$, with $n \geq 2$, let $g_{\sigma,\mu}(p)$ be the expression in* (12)*, let $S_{\sigma,\mu}(p)$ be the first expression in* (13)*, and let $M_\sigma(p)$ and $M_\mu(p)$ be the Fiedler matrices of $p(z)$ associated with σ and μ. Let $\boldsymbol{b} > 1$ be a given real number and define the sets of polynomials*

$$\mathcal{L}_{\boldsymbol{b}} := \left\{ p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k \ :\ \min\left\{ \sqrt{1 + S_{\sigma,\mu}(p)}\,,\ \frac{1}{|a_0|} \right\} = \boldsymbol{b},\ 0 \neq |a_0| < 1 \right\},$$

$$\mathcal{M}_{\boldsymbol{b}} := \left\{ p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k \ :\ \min\left\{ \sqrt{1 + \frac{S_{\sigma,\mu}(p)}{n-1}}\,,\ |a_0| \right\} = \boldsymbol{b},\ 1 < |a_0| \right\}.$$

Then the following statements hold.

(a) *For all $\epsilon > 0$, there exists $p(z) \in \mathcal{L}_{\boldsymbol{b}}$ such that*

$$1 \leq \left(\frac{\kappa_F(M_\mu(p))}{\kappa_F(M_\sigma(p))}\right)^2 \leq 1 + \epsilon.$$

In particular, this happens for any $p(z) \in \mathcal{L}_{\boldsymbol{b}}$ whose coefficient a_1 satisfies $S_{\sigma,\mu}(p)/\epsilon \leq |a_1|^2$.

(b) *For all* $\epsilon > 0$, *there exists* $p(z) \in \mathcal{L}_{\boldsymbol{b}}$ *such that*

$$\frac{\boldsymbol{b}^2}{1+\epsilon} \le \left(\frac{\kappa_F(M_\mu(p))}{\kappa_F(M_\sigma(p))}\right)^2 \le \boldsymbol{b}^2.$$

In particular, this happens for any $p(z) \in \mathcal{L}_{\boldsymbol{b}}$ *such that* $S_{\sigma,\mu}(p)$ *satisfies* $\max\{1/|a_0|^2, g_{\sigma,\mu}(p)/\epsilon\} \le S_{\sigma,\mu}(p)$. *Note that in this case* $1/|a_0| = \boldsymbol{b}$.

(c) *For all* $\epsilon > 0$, *there exists* $p(z) \in \mathcal{M}_{\boldsymbol{b}}$ *such that*

$$1 \le \left(\frac{\kappa_F(M_\sigma(p))}{\kappa_F(M_\mu(p))}\right)^2 \le 1+\epsilon.$$

In particular, this happens for any $p(z) \in \mathcal{M}_{\boldsymbol{b}}$ *whose coefficient* a_1 *satisfies* $(|a_0|^2\, S_{\sigma,\mu}(p)/\epsilon) \le |a_1|^2$.

(d) *For all* $\epsilon > 0$, *there exists* $p(z) \in \mathcal{M}_{\boldsymbol{b}}$ *such that*

$$\frac{\boldsymbol{b}^2}{1+\epsilon} \le \left(\frac{\kappa_F(M_\sigma(p))}{\kappa_F(M_\mu(p))}\right)^2 \le \boldsymbol{b}^2.$$

In particular, this happens for any $p(z) \in \mathcal{M}_{\boldsymbol{b}}$ *such that* $|a_0| = \boldsymbol{b}$ *and* $S_{\sigma,\mu}(p)$ *satisfies* $(|a_0|^2\, g_{\sigma,\mu}(p)/\epsilon) \le S_{\sigma,\mu}(p)$.

Proof. Part (a). Let $p(z) \in \mathcal{L}_{\boldsymbol{b}}$ be such that its coefficient a_1 satisfies $S_{\sigma,\mu}(p)/\epsilon \le |a_1|^2$. In the following developments all magnitudes refer to $p(z)$, but the dependence on $p(z)$ is dropped for simplicity. From Corollary 4.5 and Lemma 4.8, we get

$$1 \le \left(\frac{\kappa_F(M_\mu)}{\kappa_F(M_\sigma)}\right)^2 = \frac{1+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}\,|a_0|^2}}{1+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}}} \le 1+\frac{S_{\sigma,\mu}}{g_{\sigma,\mu}\,|a_0|^2} \le 1+\frac{S_{\sigma,\mu}}{|a_1|^2} \le 1+\epsilon\,.$$

Part (b). Let $p(z) \in \mathcal{L}_{\boldsymbol{b}}$ be such that $S_{\sigma,\mu}(p)$ satisfies $\max\{1/|a_0|^2, g_{\sigma,\mu}(p)/\epsilon\} \le S_{\sigma,\mu}(p)$. In the following developments all magnitudes refer to $p(z)$, but the dependence on $p(z)$ is dropped for simplicity. From Lemma 4.8 and Theorem 4.9, we get

$$\boldsymbol{b}^2 \ge \left(\frac{\kappa_F(M_\mu)}{\kappa_F(M_\sigma)}\right)^2 = \frac{1+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}\,|a_0|^2}}{1+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}}} \ge \frac{\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}\,|a_0|^2}}{1+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}}} \ge \frac{\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}\,|a_0|^2}}{\epsilon\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}}+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}}} = \frac{\dfrac{1}{|a_0|^2}}{1+\epsilon} = \frac{\boldsymbol{b}^2}{1+\epsilon}\,.$$

Part (c). Let $p(z) \in \mathcal{M}_{\boldsymbol{b}}$ be such that its coefficient a_1 satisfies $(|a_0|^2\, S_{\sigma,\mu}(p)/\epsilon) \le |a_1|^2$. In the following developments all magnitudes refer to $p(z)$, but the dependence on $p(z)$ is dropped for simplicity. From Corollary 4.5 and Lemma 4.8, we get

$$1 \le \left(\frac{\kappa_F(M_\sigma)}{\kappa_F(M_\mu)}\right)^2 = \frac{1+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}}}{1+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}\,|a_0|^2}} \le 1+\frac{S_{\sigma,\mu}}{g_{\sigma,\mu}} \le 1+\frac{S_{\sigma,\mu}}{\dfrac{|a_1|^2}{|a_0|^2}} \le 1+\epsilon\,.$$

Part (d). Let $p(z) \in \mathcal{M}_{\boldsymbol{b}}$ be such that $|a_0| = \boldsymbol{b}$ and $S_{\sigma,\mu}(p)$ satisfies $(|a_0|^2\, g_{\sigma,\mu}(p)/\epsilon) \le S_{\sigma,\mu}(p)$. In the following developments all magnitudes refer to $p(z)$, but the dependence on $p(z)$ is dropped for simplicity. From Lemma 4.8 and Theorem 4.9, we get

$$\boldsymbol{b}^2 \ge \left(\frac{\kappa_F(M_\sigma)}{\kappa_F(M_\mu)}\right)^2 = \frac{1+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}}}{1+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}\,|a_0|^2}} \ge \frac{\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}}}{\epsilon\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}\,|a_0|^2}+\dfrac{S_{\sigma,\mu}}{g_{\sigma,\mu}\,|a_0|^2}} = \frac{|a_0|^2}{1+\epsilon} = \frac{\boldsymbol{b}^2}{1+\epsilon}\,.$$

□

Remark 4.11. We have shown in Theorem 4.10 how to easily construct families of polynomials where the upper bounds in Theorem 4.9 for the ratios of condition numbers of different Fiedler matrices of the same polynomial are essentially attained, and other families where they are far from being attained. The reader should keep in mind that there are other families of polynomials satisfying the same properties.

Corollary 4.5 and Theorem 4.10 suggest that for polynomials with $|p(0)| < 1$ one should avoid the use of the classical Frobenius companion matrices and to use, instead, Fiedler matrices with a number of initial consecutions or inversions equal to one, as for instance P_1 in Example 2.2. This would lead to use matrices with the smallest possible condition number that, in addition, for certain polynomials may be arbitrarily smaller than the condition numbers of other Fiedler matrices. For polynomials with $|p(0)| > 1$ the situation is the opposite, and Frobenius companion matrices are the best choice from the point of view of condition numbers for inversion. However, Theorem 4.12 tells us that, given a monic polynomial $p(z)$, if there are two distinct Fiedler matrices with very different condition numbers, then both matrices are very ill-conditioned. Therefore, different Fiedler matrices may have very different condition numbers but only in cases where these matrices are nearly singular. The reciprocal is not true, because there may be two different Fiedler matrices nearly singular but having exactly the same condition number, as it is shown by Corollaries 4.2 and 4.5-(b).

Theorem 4.12. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$ and $a_0 \neq 0$, let $\sigma, \mu : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be two bijections, let M_σ and M_μ be the Fiedler matrices of $p(z)$ associated with σ and μ, and let t_σ and t_μ be the numbers of initial consecutions or inversions of σ and μ. Assume that $t_\sigma < t_\mu$. Then the following results hold.*

(a) *If $|a_0| < 1$, then*

$$1 \leq \left(\frac{\kappa_F(M_\mu)}{\kappa_F(M_\sigma)} \right)^2 \leq \kappa_F(M_\sigma) \leq \kappa_F(M_\mu) \, .$$

(b) *If $|a_0| > 1$, then*

$$1 \leq \frac{\kappa_F(M_\sigma)}{\kappa_F(M_\mu)} \leq \kappa_F(M_\mu) \leq \kappa_F(M_\sigma) \, .$$

Proof. Part (a). From Theorem 4.9 and with the notation used there, we get

$$1 \leq \left(\frac{\kappa_F(M_\mu)}{\kappa_F(M_\sigma)} \right)^2 = \frac{\kappa_F(M_\mu)}{\kappa_F(M_\sigma)} \frac{\kappa_F(M_\mu)}{\kappa_F(M_\sigma)} \leq \sqrt{1 + S_{\sigma,\mu}} \, \frac{1}{|a_0|} \leq \|M_\sigma\|_F \, \|M_\sigma^{-1}\|_F \, ,$$

where the last inequality follows from Corollaries 2.9 and 3.3.

Part (b). From Theorem 4.9 and with the notation used there, we get

$$1 \leq \frac{\kappa_F(M_\sigma)}{\kappa_F(M_\mu)} \leq |a_0| \leq \|M_\sigma\|_F \, \|M_\sigma^{-1}\|_F \, ,$$

where the last inequality follows again from Corollaries 2.9 and 3.3. □

Remark 4.13. The difference between the statements of parts (a) and (b) in Theorem 4.12 is striking, but the next example shows that $(\kappa_F(M_\sigma)/\kappa_F(M_\mu))^2$ cannot be used in part (b). Consider the monic polynomial

$$p(z) = 10^4 + 2\,z + 2\,z^2 + 2 \cdot 10^5\, z^3 + 2 \cdot 10^5\, z^4 + 2\,z^5 + 2\,z^6 + 3\,z^7 + z^8$$

and the bijections σ and μ with consecution-inversion structure sequences $\mathrm{CISS}(\sigma) = (2, 1, 1, 2, 1, 0)$ and $\mathrm{CISS}(\mu) = (4, 2, 1, 0)$. In this case, $\kappa_F(M_\mu) = 8.124 \cdot 10^6$, $\kappa_F(M_\sigma) = 8.005 \cdot 10^{10}$, and $(\kappa_F(M_\sigma)/\kappa_F(M_\mu))^2 = 9.709 \cdot 10^7$. However, we see in this example that $(\kappa_F(M_\sigma)/\kappa_F(M_\mu)) << \kappa_F(M_\mu)$. We have observed the same behavior in all the examples that we have tested with large values of $\kappa_F(M_\sigma)/\kappa_F(M_\mu)$. Therefore, we think that the result in part (b) of Theorem 4.12 can be considerably improved.

5. Staircase matrices

We have mentioned in the Introduction that there exist simple explicit expressions for the singular values of the Frobenius companion matrices of a monic polynomial of degree n in terms of the coefficients of the polynomial and that at least $n-2$ singular values are equal to one [8, 9]. These nice properties do not extend to other Fiedler matrices, but the singular values of Fiedler matrices still retain some interesting properties that will be analyzed in Section 6. This analysis is based on the notion of *staircase matrix* and the determination of its rank in Theorem 5.15. These questions are the subject of the present section.

Staircase matrices are matrices whose nonzero entries follow a very special pattern. We assume throughout this section that these matrices have more than one row or more than one column to avoid the trivial 1×1 case that may complicate the definition.

Definition 5.1. *Given a matrix $A = [a_{ij}] \in \mathbb{C}^{m \times p}$, we say that A is a* staircase matrix *if A satisfies the following properties:*

1. *If $a_{i,j_1} \neq 0$ and $a_{i,j_2} \neq 0$, for some $1 \leq i \leq m$ and $1 \leq j_1 \leq j_2 \leq p$, then $a_{ij} \neq 0$ for all $j_1 \leq j \leq j_2$.*
2. *If $a_{i1} = a_{i2} = \cdots = a_{i,j-1} = 0$ and $a_{ij} \neq 0$, for some $1 < i \leq m$, $1 \leq j \leq p$, then $a_{i-1,j} \neq 0$ and $a_{i-1,j+1} = 0$, whenever $j+1 \leq p$.*
3. *$a_{11} \neq 0$ and $a_{mp} \neq 0$.*

A matrix $B = [b_{ij}] \in \mathbb{C}^{m \times p}$ is said to be a generalized staircase matrix *if it is obtained from a staircase matrix by turning some nonzero entries into zero entries.*

The first condition in Definition 5.1 means that all nonzero entries in a given row of A are placed in consecutive columns. The second condition means that the first nonzero entry in a given row of A is placed in the same column as the last nonzero entry of the immediate upper row.

Example 5.2. *The following matrices are staircase matrices:*

$$A = \begin{bmatrix} \times & \times & \times & & & & & \\ & & \times & \times & & & & \\ & & & \times & \times & \times & & \\ & & & & & \times & & \\ & & & & & \times & & \\ & & & & & \times & \times & \times \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} \times & \times & \times & & & \\ & & \times & \times & & \\ & & & \times & \times & \times \\ & & & & & \times \\ & & & & & \times \\ & & & & & \times \end{bmatrix},$$

where the symbol $\times$ denotes the nonzero entries (here and in all the examples of this section). A generalized staircase matrix compatible with A is obtained by replacing some of the $\times$ entries of A by 0. In plain words, one can say that generalized staircase matrices may have "holes" in the steps.

Notice that, as a consequence of the second and third conditions in Definition 5.1, every row and every column in a staircase matrix has at least one nonzero entry.

Definition 5.3. *Let $A \in \mathbb{C}^{m \times p}$ be a staircase matrix. We say that a nonzero entry a_{ij} is a* corner *entry of A if one (or both) of the following conditions holds:*

1. *$i = j = 1$ or $i = m$ and $j = p$.*
2. *a_{ij} is the first or the last nonzero entry in the ith row and there are more than one nonzero entries in the ith row.*

Example 5.4. *Let A and C be the staircase matrices in Example 5.2. Then*

$$A = \begin{bmatrix} \otimes & \times & \otimes & & & & & \\ & & \otimes & \otimes & & & & \\ & & & \otimes & \times & \otimes & & \\ & & & & & \times & & \\ & & & & & \times & & \\ & & & & & \otimes & \times & \otimes \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} \otimes & \times & \otimes & & & \\ & & \otimes & \otimes & & \\ & & & \otimes & \times & \otimes \\ & & & & & \times \\ & & & & & \times \\ & & & & & \otimes \end{bmatrix},$$

and the entries with the symbols $\otimes$ are the corner entries.

Definition 5.5. *Let $A \in \mathbb{C}^{m\times p}$ be a staircase matrix. We define the* ordered list of corner entries *of A as the ordered list $(a_{i_1,j_1}, a_{i_2,j_2}, \ldots, a_{i_{t+1},j_{t+1}})$ of all corner entries of A, where the corner entry a_{i_r,j_r} precedes the corner entry a_{i_s,j_s} if $i_r < i_s$ or $i_r = i_s$ and $j_r < j_s$.*

Example 5.6. *For the staircase matrices in Example* 5.4*, the ordered lists of corner entries of A and C are, respectively, $\{a_{11}, a_{13}, a_{23}, a_{24}, a_{34}, a_{36}, a_{66}, a_{68}\}$ and $\{c_{11}, c_{13}, c_{23}, c_{24}, c_{34}, c_{36}, c_{66}\}$.*

Notice that for two consecutive entries in the ordered list of corner entries of A, say a_{i_k,j_k} and $a_{i_{k+1},j_{k+1}}$, we always have $i_k = i_{k+1}$ or $j_k = j_{k+1}$ (but not both). This motivates the following definition.

Definition 5.7. *Let $A \in \mathbb{C}^{m\times p}$ be a staircase matrix and $(a_{i_1,j_1}, \ldots, a_{i_{t+1},j_{t+1}})$ be the ordered list of corner entries of A. Then, for $1 \le k \le t$, the* kth flight *of A is the set of entries*

- *$a_{i_k,j_k}, a_{i_k,j_k+1}, \ldots, a_{i_k,j_{k+1}}$ if $i_k = i_{k+1}$, or*
- *$a_{i_k,j_k}, a_{i_k+1,j_k}, \ldots, a_{i_{k+1},j_k}$ if $j_k = j_{k+1}$.*

Notice that the number of flights of a staircase matrix A is equal to the number of corner entries of A minus one. We are particularly interested in the lengths of the flights of A. This notion is made precise in Definition 5.8.

Definition 5.8. *Let $A \in \mathbb{C}^{m\times p}$ be a staircase matrix and $(a_{i_1,j_1}, \ldots, a_{i_{t+1},j_{t+1}})$ be the ordered list of corner entries of A. The* flight-length sequence *of A is the sequence*

$$\mathcal{F}(A) := (f_1, f_2, \ldots, f_t),$$

where $f_k = \max\{i_{k+1} - i_k, j_{k+1} - j_k\}$, for $k = 1, \ldots, t$.

We note that the kth term f_k in the flight-length sequence of A is equal to the number of entries in the kth flight of A minus one.

Definition 5.9. *Let $\mathbf{s} = (s_1, s_2, \ldots, s_t)$ be an ordered list of nonnegative integers. For each $j = 1, 2, \ldots, t$, we define the* length of the string of ones at the jth position *of $\mathbf{s}$, denoted by l_j, as*

- *a positive integer $l_j > 0$, if the following three conditions are satisfied:*
  - (i) $s_j = s_{j+1} = \cdots = s_{j+l_j-1} = 1$,
  - (ii) $s_{j-1} \neq 1$ *or* $j = 1$, *and*
  - (iii) $s_{j+l_j} \neq 1$ *or* $j + l_j - 1 = t$.
- *$l_j = 0$, otherwise.*

Let $(l_1, l_2, \ldots, l_t)$ be the ordered list of the lengths of the strings of ones of $\mathbf{s}$. Then, the list of positive lengths of the strings of ones of $\mathbf{s}$*, denoted by $\mathcal{L}(\mathbf{s})$, is the ordered list obtained from $(l_1, l_2, \ldots, l_t)$ after removing all zero entries. If $\mathbf{s}$ is a list containing no ones, then we set $\mathcal{L}(\mathbf{s}) := (0)$.*

Example 5.10. *For the list $\mathbf{s} = (2, 1, 1, 1, 3, 1, 1, 2)$, the list of the lengths of the strings of ones is $(0, 3, 0, 0, 0, 2, 0, 0)$, so we have $\mathcal{L}(\mathbf{s}) = (3, 2)$.*

Until now, we have not established any relationship between staircase matrices and Fiedler matrices. However, both types of matrices are closely connected in a way that will be shown in Section 6. In order to introduce this connection, we show in Theorem 5.11 that every staircase matrix with n nonzero entries can be constructed from the consecutions and inversions of a bijection $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$. Moreover, we show that the reduced consecution-inversion structure sequence of σ introduced in Definition 2.3, RCISS(σ), is the flight-length sequence of the matrix *in reversed order.* The reader is invited to focus on the similarities between `Algorithm 3` in Theorem 5.11 and `Algorithm 1` in Theorem 2.6, which will be exploited in depth in Section 6. However, note that in `Algorithm 3` we use the MATLAB notation $V(:, j : \texttt{end})$ to indicate the submatrix of V consisting of columns j through the last column (a similar notation is used for rows), because the sizes of the constructed matrices are not fixed. They depend on the number of consecutions and inversions of σ. In addition, if expressions like $V(:, 2 : 1)$ appear in `Algorithm 3`, then they should be understood as empty matrices. We warn also the reader that in `Algorithm 3` the staircase matrix is constructed starting from the lower-right entry, which may seem unnatural, but it is convenient for establishing the connection with Fiedler matrices and `Algorithm 1`.

Theorem 5.11. *Let $x_0, x_1, \ldots, x_{n-1}$ be $n \geq 2$ complex nonzero numbers, let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection, and consider the following algorithm:*

`Algorithm 3.` *Given $x_0, x_1, \ldots, x_{n-1}$ nonzero numbers and a bijection σ, the following algorithm constructs a matrix $\widetilde{V}_\sigma$ whose nonzero entries are precisely $x_0, x_1, \ldots, x_{n-1}$.*

if σ has a consecution at 0 then

$$\widetilde{V}_0 = \begin{bmatrix} x_1 \\ x_0 \end{bmatrix}$$

else

$$\widetilde{V}_0 = \begin{bmatrix} x_1 & x_0 \end{bmatrix}$$

endif

for $i = 1 : n-2$

if σ has a consecution at i then

$$\widetilde{V}_i = \begin{bmatrix} x_{i+1} & 0 \\ \widetilde{V}_{i-1}(:,1) & \widetilde{V}_{i-1}(:,2:\texttt{end}) \end{bmatrix}$$

else

$$\widetilde{V}_i = \begin{bmatrix} x_{i+1} & \widetilde{V}_{i-1}(1,:) \\ 0 & \widetilde{V}_{i-1}(2:\texttt{end},:) \end{bmatrix}$$

endif

endfor

$\widetilde{V}_\sigma = \widetilde{V}_{n-2}$.

Then the matrix $\widetilde{V}_\sigma$ is a staircase matrix. Moreover, if $\mathrm{RCISS}(\sigma) = (p_1, p_2, \ldots, p_t)$ *is the reduced consecution-inversion structure sequence of σ, then the flight-length sequence of $\widetilde{V}_\sigma$ is $\mathcal{F}(\widetilde{V}_\sigma) = (p_t, p_{t-1}, \ldots, p_2, p_1)$.*

Conversely, given a staircase matrix A with n nonzero entries and flight-length sequence $\mathcal{F}(A) = (f_1, f_2, \ldots, f_t)$, there exists a bijection $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ such that $\mathrm{RCISS}(\sigma) = (f_t, \ldots, f_2, f_1)$ *and $A = \widetilde{V}_\sigma$, where $\widetilde{V}_\sigma$ is the output of* `Algorithm 3` *with the inputs σ and the list of the nonzero entries of A ordered from the lower-right to the upper-left entry*[1].

Proof. The proof is easy, so we only sketch the main points. In the proof we use a family of bijections $\sigma_i : \{0, 1, \ldots, i+1\} \to \{1, \ldots, i+2\}$, for $i = 0, 1, \ldots, n-2$, such that σ_i has a consecution (resp. inversion) at j, $0 \leq j \leq i$, if and only if σ has a consecution (resp. inversion) at j. Observe that $\widetilde{V}_i$ is constructed by applying `Algorithm 3` to the numbers $x_0, x_1, \ldots, x_{i+1}$ and the bijection σ_i. The bijection σ_{n-2} may be taken to be equal to σ.

Let us prove first the properties of $\widetilde{V}_\sigma$. It is obvious that $\widetilde{V}_0$ is a staircase matrix that has $\mathcal{F}(\widetilde{V}_0) = (1)$, and that $\mathrm{RCISS}(\sigma_0) = (1)$. Next, we proceed by induction. Assume that $\widetilde{V}_{i-1}$, for some $i - 1 \geq 0$, is a staircase matrix that has $\mathcal{F}(\widetilde{V}_{i-1}) = (p'_s, p_{s-1}, \ldots, p_2, p_1)$, where $(p_1, p_2, \ldots, p_{s-1}, p'_s) = \mathrm{RCISS}(\sigma_{i-1})$. Then the structure of `Algorithm 3` makes obvious that $\widetilde{V}_i$ is also a staircase matrix. Also if σ has two consecutions or two inversions at $i-1$ and i, then $\mathcal{F}(\widetilde{V}_i) = (p'_s + 1, p_{s-1}, \ldots, p_2, p_1)$ and also $(p_1, p_2, \ldots, p_{s-1}, p'_s + 1) = \mathrm{RCISS}(\sigma_i)$. If σ has a consecution at $i-1$ and an inversion at i, or viceversa, then $\mathcal{F}(\widetilde{V}_i) = (1, p'_s, p_{s-1}, \ldots, p_2, p_1)$ and also $(p_1, p_2, \ldots, p_{s-1}, p'_s, 1) = \mathrm{RCISS}(\sigma_i)$ (note that in this case $p'_s = p_s$). Therefore, the result is true for $\widetilde{V}_i$. The result in the statement follows by taking $i = n-2$.

The "converse statement" is also immediate just by looking carefully at `Algorithm 3` and the reader is invited to complete the details. The only point to be remarked is that σ is not determined only by $\mathrm{RCISS}(\sigma)$. It is needed to also know whether σ has a consecution or an inversion at 0. Note that if the last flight of A, with length f_t, is an horizontal flight, i.e., it corresponds to entries in the same row, then σ has inversions at $0, 1, \ldots, f_t - 1$. On the contrary, if the last flight of A is a vertical flight, i.e., it corresponds to entries in the same column, then σ has consecutions at $0, 1, \ldots, f_t - 1$. ☐

[1]More precisely, this order corresponds to list all the flights of A from 1 to t, to remove repeated entries, and to reverse the order of the obtained list.

Given n ordered nonzero numbers $x_0, x_1, \ldots, x_{n-1}$, Theorem 5.11 establishes a correspondence between bijections $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ and staircase matrices A that have as nonzero entries $x_0, x_1, \ldots, x_{n-1}$, with x_0 being the lower-right entry and x_{n-1} being the upper-left. Taking into account the relationship between $\mathrm{RCISS}(\sigma)$ and $\mathcal{F}(A)$ in Theorem 5.11, Definition 5.12 introduces, from two different but equivalent perspectives, the list that will allow us to determine the rank of a staircase matrix in Theorem 5.15.

Definition 5.12. (a) *Let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, 2, \ldots, n\}$ with $n \geq 2$ be a bijection and let $\mathrm{RCISS}(\sigma) = (p_1, p_2, \ldots, p_t)$ be the reduced consecution-inversion structure sequence of σ. Let $\mathbf{s} := (p_{t-1}, \ldots, p_2)$. Then, the* rank-determining list *of σ, denoted by $\mathcal{L}(\sigma)$, is $\mathcal{L}(\mathbf{s})$, that is, the list of positive lengths of the strings of ones of $\mathbf{s}$ introduced in Definition 5.9. If $\mathrm{RCISS}(\sigma)$ has 1 or 2 entries, then $\mathbf{s}$ is empty and we set $\mathcal{L}(\sigma) := (0)$.*

(b) *Let A be a staircase matrix and let $\mathcal{F}(A) = (f_1, f_2, \ldots, f_t)$ be its flight-length sequence. Let $\mathbf{s} := (f_2, \ldots, f_{t-1})$. Then, the* rank-determining list *of A, denoted by $\mathcal{L}(A)$, is $\mathcal{L}(\mathbf{s})$, that is, the list of positive lengths of the strings of ones of $\mathbf{s}$ introduced in Definition 5.9. If $\mathcal{F}(A)$ has 1 or 2 entries, then $\mathbf{s}$ is empty and we set $\mathcal{L}(A) := (0)$.*

Observe that $\mathcal{L}(\sigma)$ has been defined without any reference to staircase matrices. It depends only on the bijection σ. However, if $\widetilde{V}_\sigma$ is *any* staircase matrix constructed by `Algorithm 3` for this bijection, then $\mathcal{L}(\sigma) = \mathcal{L}(\widetilde{V}_\sigma)$, as a consequence of Theorem 5.11. Also, given any staircase matrix A, Theorem 5.11 guarantees that there exists a bijection σ such that $\mathcal{L}(A) = \mathcal{L}(\sigma)$. Therefore we will use the notation $\mathcal{L}(\sigma)$ or $\mathcal{L}(A)$ depending on which is more convenient for the specific result we are considering. We illustrate these concepts in Example 5.13.

Example 5.13. *The staircase matrix A in Example 5.2 has $\mathcal{F}(A) = (2, 1, 1, 1, 2, 3, 2)$ as flight-length sequence. Then A can be constructed by `Algorithm 3` via a bijection $\sigma : \{0, 1, \ldots, 12\} \to \{1, \ldots, 13\}$ with $\mathrm{RCISS}(\sigma) = (2, 3, 2, 1, 1, 1, 2)$. Note that the first two entries of $\mathrm{CISS}(\sigma)$ are $\mathfrak{c}_0 = 0$ and $\mathfrak{i}_0 = 2$, since the last flight of A corresponds to entries in the same row. Moreover, we have $\mathcal{L}(A) = \mathcal{L}(\sigma) = (3)$.*

Theorem 5.15 is the key result of this section. It gives the simple formula (18) for the rank of any staircase matrix in terms of its flight-length sequence. The formula (18) shows that to determine the rank of a staircase matrix is not completely trivial. The idea of the proof is to perform *Gaussian Elimination by rows and columns* starting from the upper-left corner. We illustrate this procedure in a simple case in Example 5.14, and then we state Theorem 5.15.

Example 5.14. *By Gaussian elimination, it is easy to determine the rank of any staircase matrix. Consider the matrix A in Example 5.2. Using elementary row and column replacement operations starting from the upper-left entry, we can transform the matrix A into*

$$A = \begin{bmatrix} \times & \times & \times & & & & & \\ & & \times & \times & & & & \\ & & & \times & \times & \times & & \\ & & & & & \times & & \\ & & & & & \times & & \\ & & & & & \times & \times & \times \end{bmatrix} \sim \begin{bmatrix} \times & 0 & 0 & & & & & \\ & & \times & 0 & & & & \\ & & & \times & 0 & 0 & & \\ & & & & & \times & & \\ & & & & & 0 & & \\ & & & & & 0 & \times & 0 \end{bmatrix}.$$

Hence $\operatorname{rank} A = 5$. *As can be seen in Theorem 5.15, the rank of a staircase matrix A can be obtained from the number of flights of A and the sequence $\mathcal{L}(A)$. It is important to notice the role played by those flights of length 1 different from the first and the last flights.*

In the rest of the paper, given a real number x, we will use the standard notation $\lceil x \rceil$ for the smallest integer which is greater than or equal to x.

Theorem 5.15. *Let $A \in \mathbb{C}^{m \times p}$ be a staircase matrix, let $\mathcal{F}(A) = (f_1, f_2, \ldots, f_t)$ be the flight-length sequence of A, and let $\mathcal{L}(A) = (l_1, l_2, \ldots, l_q)$ be the rank-determining list of A. Then the rank of A is equal to*

$$\operatorname{rank} A = t - \sum_{j=1}^{q} \left\lceil \frac{l_j}{2} \right\rceil. \tag{18}$$

Proof. The proof proceeds by induction on the number of flights t. For $t = 1$, the result is obviously true because all staircase matrices with only one flight have $\operatorname{rank} A = 1$ and $\mathcal{L}(A) = (0)$, so

$$t - \sum_{j=1}^{q} \left\lceil \frac{l_j}{2} \right\rceil = 1 = \operatorname{rank} A.$$

By a similar argument the result is also true for $t = 2$, since in this case $\operatorname{rank} A = 2$ and $\mathcal{L}(A) = (0)$. Now, let us assume that the result is true for any staircase matrix with $t - 1 \geq 2$ flights. Let A and $\widehat{A}$ be staircase matrices with $\mathcal{F}(A) = (f_1, f_2, \ldots, f_t)$ and $\mathcal{F}(\widehat{A}) = (f_1, f_2, \ldots, f_{t-1})$ and such that A is obtained from $\widehat{A}$ by adding one flight with length f_t. Note that A and $\widehat{A}$ have different sizes. We distinguish two cases.

Case 1: $f_{t-1} \neq 1$. In fact, according to Definition 5.8 this means $f_{t-1} > 1$. In this case, $\mathcal{L}(A) = \mathcal{L}(\widehat{A}) = (l_1, l_2, \ldots, l_q)$. The reason is that $\mathcal{L}(A)$ is determined by the strings of ones in $(f_2, \ldots, f_{t-1})$, while $\mathcal{L}(\widehat{A})$ is determined by the strings of ones in $(f_2, \ldots, f_{t-2})$, and in both cases the strings of ones are the same.

In addition, $\operatorname{rank} A = 1 + \operatorname{rank} \widehat{A}$. To see this, assume without loss of generality that the last flight of A has all its entries in the same row (otherwise we transpose the matrix, which preserves the rank and the flight-length sequence). Therefore, A has more columns than $\widehat{A}$ and the same number of rows, i.e., $A \in \mathbb{C}^{m \times p}$ and $\widehat{A} \in \mathbb{C}^{m \times \ell}$ with $\ell < p$, and the last flight of $\widehat{A}$ has all its entries in the same column. This and the fact $f_{t-1} > 1$ mean that the last two rows of A are

$$A(m-1:m,:) = \left[\begin{array}{cccc|ccc} 0 & \cdots & 0 & \times & 0 & \ldots & 0 \\ 0 & \cdots & 0 & \times & \times & \ldots & \times \end{array}\right] \sim \left[\begin{array}{cccc|ccc} 0 & \cdots & 0 & \times & 0 & \ldots & 0 \\ 0 & \cdots & 0 & 0 & \times & \ldots & \times \end{array}\right] = A'(m-1:m,:), \tag{19}$$

where the symbol $\times$ denotes nonzero entries, the vertical line separates $\widehat{A}$ from those columns of A that are not columns of $\widehat{A}$, and we have performed an elementary row replacement operation to get A'. Since $A(1:m-1, \ell+1:p) = 0$, (19) implies that $\operatorname{rank} A = \operatorname{rank} A' = 1 + \operatorname{rank} \widehat{A}' = 1 + \operatorname{rank} \widehat{A}$.

The above equalities $\mathcal{L}(A) = \mathcal{L}(\widehat{A})$ and $\operatorname{rank} A = 1 + \operatorname{rank} \widehat{A}$, and the induction hypothesis imply

$$\operatorname{rank} A = \operatorname{rank} \widehat{A} + 1 = t - 1 - \sum_{j=1}^{q} \left\lceil \frac{l_j}{2} \right\rceil + 1 = t - \sum_{j=1}^{q} \left\lceil \frac{l_j}{2} \right\rceil,$$

which proves the result for A in **Case 1**.

Case 2: $f_{t-1} = 1$. In this case $\mathcal{L}(A) = (l_1, l_2, \ldots, l_q)$ with $l_q \geq 1$, and we need to distinguish two subcases: l_q even and l_q odd.

Case 2.1: $l_q = 2k$, with $k > 0$ an integer. In this case $l_q > 1$ and $\mathcal{L}(\widehat{A}) = (l_1, l_2, \ldots, l_{q-1}, l_q - 1)$. Definition 5.12 and $l_q = 2k$ imply

$$f_{t-2k} = f_{t-(2k-1)} = \cdots = f_{t-1} = 1, \quad \text{with } t - 2k \geq 2, \text{ and } f_{t-2k-1} > 1 \text{ if } t - 2k > 2. \tag{20}$$

Assume, as in **Case 1**, that the last flight of A has all its entries in the same row, which implies that $A \in \mathbb{C}^{m \times p}$ and $\widehat{A} \in \mathbb{C}^{m \times \ell}$ with $\ell < p$, and also that the last flight of $\widehat{A}$ has its two entries in the same column. This and (20) imply that if $t - 2k > 2$

$$A(m-k-1:m,:) = \left[\begin{array}{ccccccccc|ccc} 0 & \cdots & 0 & \times & & & & & & 0 & \ldots & 0 \\ \vdots & & \vdots & \times & \times & & & & & \vdots & & \vdots \\ \vdots & & \vdots & & \times & \times & & & & \vdots & & \vdots \\ \vdots & & \vdots & & & \ddots & \ddots & & & \vdots & & \vdots \\ \vdots & & \vdots & & & & & \times & \times & 0 & \ldots & 0 \\ 0 & \cdots & 0 & & & & & & \times & \times & \ldots & \times \end{array}\right], \tag{21}$$

where the vertical line separates $\widehat{A}$ from those columns of A that are not columns of $\widehat{A}$. If we perform elementary row replacement operations in $A(m-k-1:m,:)$ starting from the top we get

$$A(m-k-1:m,:) \sim \left[\begin{array}{ccccccc|ccc} 0 & \cdots & 0 & \times & & & & 0 & \ldots & 0 \\ \vdots & & \vdots & 0 & \times & & & \vdots & & \vdots \\ \vdots & & \vdots & & 0 & \times & & \vdots & & \vdots \\ \vdots & & \vdots & & & \ddots & \ddots & \vdots & & \vdots \\ \vdots & & \vdots & & & 0 & \times & 0 & \ldots & 0 \\ 0 & \cdots & 0 & & & & 0 & \times & \ldots & \times \end{array}\right], \tag{22}$$

which implies

$$\operatorname{rank} A = 1 + \operatorname{rank} \widehat{A}. \tag{23}$$

If $t - 2k = 2$, then $A(m-k-1:m,:)$ is as in (21) but removing all the left-most columns of zeros. So we also get (23).

The equalities $\mathcal{L}(\widehat{A}) = (l_1, l_2, \ldots, l_{q-1}, l_q - 1)$, (23), and the induction hypothesis imply

$$\operatorname{rank} A = \operatorname{rank} \widehat{A} + 1 = t - 1 - \sum_{j=1}^{q-1} \left\lceil \frac{l_j}{2} \right\rceil - \left\lceil \frac{2k-1}{2} \right\rceil + 1 = t - \sum_{j=1}^{q-1} \left\lceil \frac{l_j}{2} \right\rceil - \left\lceil \frac{2k}{2} \right\rceil = t - \sum_{j=1}^{q} \left\lceil \frac{l_j}{2} \right\rceil,$$

which proves the result for A in *Case 2.1.*

Case 2.2: $l_q = 2k+1$, with $k \geq 0$ an integer. The proof is similar to the one of *Case 2.1*, so we only emphasize the main differences and omit the details. To begin with, in this case

$$\mathcal{L}(\widehat{A}) = \begin{cases} (l_1, l_2, \ldots, l_{q-1}, l_q - 1), & \text{if } l_q > 1, \\ (l_1, l_2, \ldots, l_{q-1}), & \text{if } l_q = 1. \end{cases},$$

and one has to distinguish the cases $l_q > 1$ and $l_q = 1$. In both of them, it is satisfied that

$$\operatorname{rank} A = \operatorname{rank} \widehat{A}. \tag{24}$$

This follows because in this case the structure of A is

$$A(m-k-2:m,:) = \left[\begin{array}{cccccccc|ccc} \cdots & \cdots & 0 & & & & & & 0 & \ldots & 0 \\ \cdots & \cdots & \times & \times & & & & & 0 & \ldots & 0 \\ 0 & & 0 & \times & \times & & & & \vdots & & \vdots \\ \vdots & & \vdots & & \times & \times & & & \vdots & & \vdots \\ \vdots & & \vdots & & & \ddots & \ddots & & \vdots & & \vdots \\ \vdots & & \vdots & & & & \times & \times & 0 & \ldots & 0 \\ 0 & \cdots & 0 & & & & & \times & \times & \ldots & \times \end{array}\right], \tag{25}$$

and elementary column replacement operations starting from the left-most $\times$ entry shown in (25) allows us to make zeros all the entries to the right of the vertical line.

The equalities $\mathcal{L}(\widehat{A}) = (l_1, l_2, \ldots, l_{q-1}, l_q - 1)$ (if $l_q > 1$), (24), and the induction hypothesis imply

$$\operatorname{rank} A = \operatorname{rank} \widehat{A} = t - 1 - \sum_{j=1}^{q-1} \left\lceil \frac{l_j}{2} \right\rceil - \left\lceil \frac{2k}{2} \right\rceil = t - \sum_{j=1}^{q-1} \left\lceil \frac{l_j}{2} \right\rceil - (k+1) = t - \sum_{j=1}^{q-1} \left\lceil \frac{l_j}{2} \right\rceil - \left\lceil \frac{2k+1}{2} \right\rceil = t - \sum_{j=1}^{q} \left\lceil \frac{l_j}{2} \right\rceil,$$

which proves the result for A in *Case 2.2.* Observe that the case $l_q = 1$ follows by taking $k = 0$ in the equation above. ☐

Theorem 5.15 shows, in particular, that the rank of a staircase matrix is not an increasing function of the number of flights, as it might be thought at a first glance, since intermediate flights of length 1 also affects the rank. Example 5.16 illustrates this fact.

Example 5.16. *Consider the following staircase matrices A and B*

$$A = \begin{bmatrix} \times & \times & \times \\ & & \times \\ & & \times \\ & & \times \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \times & \times & \times & & \\ & & \times & \times & \times \end{bmatrix}.$$

A and B have both 6 nonzero entries, A has 2 flights, B has 3 flights, and $\operatorname{rank} A = \operatorname{rank} B = 2$.

Next consider the staircase matrices

$$C = \begin{bmatrix} \times & & & & \\ \times & \times & & & \\ & \times & \times & & \\ & & \times & & \\ & & \times & \times & \times \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} \times & & & & \\ \times & \times & \times & & \\ & & \times & & \\ & & \times & \times & \times \\ & & & & \times \end{bmatrix}.$$

C and D have both 9 nonzero entries, C has 6 flights, and D has 5 flights. In addition, $\operatorname{rank} C = 4$ *and* $\operatorname{rank} D = 5$, *that is, the matrix with less flights have larger rank.*

Next, we bound the rank of a generalized staircase matrix B. Since B is constructed by turning some nonzero entries of a staircase matrix A into zero entries, it seems that $\operatorname{rank} B$ has to be smaller than or equal to $\operatorname{rank} A$. This is true, as we will see in Theorem 5.18, but a rigorous proof of this fact requires some work, since for general matrices the operation of turning a nonzero entry into zero may increase the rank. For instance, in MATLAB notation, $\operatorname{rank}[1,1;1,1] = 1$ and $\operatorname{rank}[1,0;1,1] = 2$. The proof of Theorem 5.18 relies on Lemma 5.17.

Lemma 5.17. *If $B \in \mathbb{C}^{m\times p}$ is a generalized staircase matrix, $1 \le i_1 < i_2 < \cdots < i_d \le m$, and $1 \le j_1 < j_2 < \cdots < j_d \le p$, then*

$$\det B(\{i_1, i_2, \dots, i_d\}, \{j_1, j_2, \dots, j_d\}) = b_{i_1,j_1} b_{i_2,j_2} \cdots b_{i_d,j_d},$$

where $B(\{i_1, i_2, \dots, i_d\}, \{j_1, j_2, \dots, j_d\})$ is the submatrix of B that lies in the rows indexed by $\{i_1, i_2, \dots, i_d\}$ and in the columns indexed by $\{j_1, j_2, \dots, j_d\}$.

Proof. The proof is by induction on d. For $d = 1$ the result is trivial. Let us assume that the result is true for $d - 1 \ge 1$, and let us prove it for d. Consider that the matrix B is constructed from a staircase matrix A such that (1) $a_{i_1,k} \neq 0$, if $c_1 \le k \le c_1'$, and (2) $a_{i_1,k} = 0$, if $1 \le k \le c_1 - 1$ or $c_1' + 1 \le k \le p$. We split the proof in two cases.

Case 1: $1 \le j_1 \le c_1' - 1$. In this case the definition of generalized staircase matrix implies that all entries in the column j_1 of B below the row i_1 are equal to zero. Then the Laplace expansion of $\det B(\{i_1, i_2, \dots, i_d\}, \{j_1, j_2, \dots, j_d\})$ along the first column gives

$$\det B(\{i_1, i_2, \dots, i_d\}, \{j_1, j_2, \dots, j_d\}) = b_{i_1,j_1} \det B(\{i_2, \dots, i_d\}, \{j_2, \dots, j_d\}) = b_{i_1,j_1} b_{i_2,j_2} \cdots b_{i_d,j_d}, \tag{26}$$

where the last equality follows from the induction hypothesis.

Case 2: $c_1' \le j_1 \le p$. In this case the definition of generalized staircase matrix implies that all entries in the row i_1 of B to the right of the column j_1 are equal to zero. Then the Laplace expansion of $\det B(\{i_1, i_2, \dots, i_d\}, \{j_1, j_2, \dots, j_d\})$ along the first row gives again (26). □

Theorem 5.18. *Let $A \in \mathbb{C}^{m\times p}$ be a staircase matrix, let $\mathcal{F}(A) = (f_1, f_2, \dots, f_t)$ be the flight-length sequence of A, and let $\mathcal{L}(A) = (l_1, l_2, \dots, l_q)$ be the rank-determining list of A. If $B \in \mathbb{C}^{m\times p}$ is any generalized staircase matrix that is obtained by turning some nonzero entries of A into zero entries, then*

$$\operatorname{rank} B \le t - \sum_{j=1}^{q} \left\lceil \frac{l_j}{2} \right\rceil. \tag{27}$$

Proof. Lemma 5.17 and the definition of generalized staircase matrix imply that if a minor of B is nonzero, then the same minor of A is nonzero. So $\operatorname{rank} B \le \operatorname{rank} A$ and the result follows from Theorem 5.15. □

Recall that we used `Algorithm 3` in Theorem 5.11 to construct a staircase matrix $\widetilde{V}_\sigma$ via a bijection σ. It is clear that if we allow zero numbers among the inputs $x_0, x_1, \ldots, x_{n-1}$, then `Algorithm 3` constructs a generalized staircase matrix coming from turning some nonzero entries of $\widetilde{V}_\sigma$ into zero. In addition, according to Definition 5.12 and the discussion in the paragraph just after it, $\mathcal{L}(\sigma) = \mathcal{L}(\widetilde{V}_\sigma)$. Therefore, Corollary 5.19 follows immediately from Theorem 5.18. Here we associate to a bijection σ the magnitude r_σ that will be often used in Section 6.

Corollary 5.19. *Let $x_0, x_1, \ldots, x_{n-1}$ be $n \geq 2$ complex numbers not necessarily different from zero, and let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection. Let $\mathcal{L}(\sigma) = (l_1, l_2, \ldots, l_q)$ be the rank-determining list of σ introduced in Definition 5.12, and let $\widetilde{V}_\sigma$ be the matrix constructed by `Algorithm 3`. Let t be the number of entries of* RCISS(σ)*. Then*

$$\operatorname{rank} \widetilde{V}_\sigma \leq r_\sigma \,, \quad \textit{where} \quad r_\sigma := t - \sum_{j=1}^{q} \left\lceil \frac{l_j}{2} \right\rceil . \tag{28}$$

Moreover, if $x_k \neq 0$ for all $k = 0, 1, \ldots, n-1$, then $\operatorname{rank} \widetilde{V}_\sigma = r_\sigma$.

5.1. Maximal rank of staircase matrices with a fixed number of nonzero entries

This section can be skipped in a first reading, although it will be referred to in some parts of Section 6. Theorem 5.15 provides a formula for the rank of a staircase matrix A depending on the number of flights and the rank-determining list of A. In this section, fixed the number of nonzero entries, we consider the problem of identifying those staircase matrices that have maximal rank. This problem is solved in Theorem 5.21. To get this result, we first prove Lemma 5.20, where we give an upper bound for the rank depending only on the number of nonzero entries, and we provide a necessary condition and a (different) sufficient condition for this bound to be attained. For a given real number x we use the standard notation $\lfloor x \rfloor$ to denote the largest integer which is smaller than or equal to x.

Lemma 5.20. *Let A be a staircase matrix with $n \geq 2$ nonzero entries, and let $\mathcal{F}(A) = (f_1, \ldots, f_t)$ be the flight-length sequence of A. Then*

(a) $\operatorname{rank} A \leq \left\lfloor \dfrac{n+1}{2} \right\rfloor$.

(b) *If $f_i = 1$ for all $i = 1, \ldots, t$, then* $\operatorname{rank} A = \left\lfloor \dfrac{n+1}{2} \right\rfloor$.

(c) *If* $\operatorname{rank} A = \left\lfloor \dfrac{n+1}{2} \right\rfloor$ *, then $f_i \leq 2$ for all $i = 1, \ldots, t$.*

Proof. (a) Let $d_c(A)$ and $d_r(A)$ denote the number of columns and rows of A, respectively. Since A is a staircase matrix, we have $d_c(A) + d_r(A) = n + 1$ (this follows easily from `Algorithm 3`). Hence, the result follows from the inequalities $\operatorname{rank} A \leq \min\{d_c(A), d_r(A)\} \leq \frac{n+1}{2}$.

(b) Following the notation of Theorem 5.15, for a staircase matrix A in the conditions of the statement we have $t = n - 1$ and $\mathcal{L}(A) = (n - 3)$, so (18) gives $\operatorname{rank} A = n - 1 - \left\lceil \frac{n-3}{2} \right\rceil = \left\lfloor \frac{n+1}{2} \right\rfloor$.

(c) We proceed by contradiction. Let $1 \leq i_0 \leq t$ be such that $f_{i_0} \geq 3$. We will construct a staircase matrix $\widehat{A}$ with exactly n nonzero entries and with $\operatorname{rank} \widehat{A} = \operatorname{rank} A + 1$. Using (a) this immediately implies that $\operatorname{rank} A < \left\lfloor \frac{n+1}{2} \right\rfloor$, which contradicts the hypothesis. Let $\widehat{A}$ be a staircase matrix such that

$$\mathcal{F}(\widehat{A}) = (f_1, \ldots, f_{i_0-1}, s_{i_0}, u_{i_0}, v_{i_0}, f_{i_0+1}, \ldots, f_t),$$

where $u_{i_0} = 1$ and s_{i_0}, v_{i_0} are positive integers such that $s_{i_0} + u_{i_0} + v_{i_0} = f_{i_0}$, and $\widehat{A}$ is constructed by creating 3 flights from the i_0th flight of A. This matrix $\widehat{A}$ always exists, since $f_{i_0} \geq 3$. It is obvious that $\widehat{A}$ has n nonzero entries. Now, let us prove that $\operatorname{rank} \widehat{A} = \operatorname{rank} A + 1$. For this, we assume without loss of generality that the i_0th flight of A has all its entries in the same row, we use Gaussian elimination by rows and columns starting from the $(1, 1)$ entry on A and $\widehat{A}$, and consider the following two cases:

- If the first (leftmost) entry of the i_0th flight of A (equivalently of $\widehat{A}$) is a pivot, then the i_0th, the (i_0+1)th, and the (i_0+2)th flights of $\widehat{A}$ follow the pattern

$$\begin{array}{ccccccc} \boxed{\times} & \cdots & \times & \times & & & \\ & & & \boxed{\times} & \times & \cdots & \times \end{array},$$

  where $\boxed{\times}$ denote pivot entries. The remaining flights of $\widehat{A}$ have exactly the same structure as the flights of A (all but the i_0th one). As a consequence, $\widehat{A}$ has one more pivot than A.

- If the first (leftmost) entry of the i_0th flight of A is not a pivot, then the i_0th, the (i_0+1)th, and the (i_0+2)th flights of $\widehat{A}$ follow the pattern

$$\begin{array}{ccccccc} \times & \boxed{\times} & \cdots & \times & & & \\ & & & \boxed{\times} & \times & \cdots & \times \end{array},$$

  if $s_{i_0} \geq 2$, or

$$\begin{array}{cccccc} \times & \boxed{\times} & & & & \\ & \times & \boxed{\times} & \times & \cdots & \times \end{array},$$

  if $s_{i_0} = 1$. Again, the remaining flights of $\widehat{A}$ have the same structure as the ones of A (all but the i_0th one), so $\widehat{A}$ has one more pivot than A.

□

Part (b) of Lemma 5.20 provides a particular type of staircase matrices where the maximum rank, given in part (a), is attained. This type corresponds to staircase matrices having only flights of length 1. It is natural to ask whether or not there are other staircase matrices for which this maximum rank is attained. The answer is given in Theorem 5.21, where we provide necessary and sufficient conditions for a staircase matrix A to be of maximal rank, and we prove that this may happen for matrices with flights of lengths larger than 1.

Theorem 5.21. *Let A be a staircase matrix with $n \geq 2$ nonzero entries. Let $\mathcal{F}(A) = (f_1, \ldots, f_t)$ be the flight-length sequence of A, and let $\mathcal{L}(A) = (l_1, \ldots, l_q)$ be the rank-determining list of A. Let α be the number of ones in $\{f_1, f_t\}$. Then* $\operatorname{rank} A = \lfloor \frac{n+1}{2} \rfloor$ *if and only if $f_i \leq 2$ for all $i = 1, \ldots, t$ and one of the following sets of conditions hold:*

(a) *n is odd, $\alpha = 2$, and l_i is even for all $i = 1, \ldots, q$;*

(b1) *n is even, $\alpha = 1$, and l_i is even for all $i = 1, \ldots, q$; or*

(b2) *n is even, $\alpha = 2$, and there is exactly one odd element among the elements of $\mathcal{L}(A)$.*

Proof. We will assume from the beginning that $f_i \leq 2$ for all i as a consequence of Lemma 5.20-(c). With this assumption, set n_1 (resp. n_2) for the number of flights of length 1 (resp. 2) of A. Then, following the notation in the statement, we have

$$n_1 = \sum_{i=1}^{q} l_i + \alpha\,, \qquad t = n_1 + n_2\,, \qquad \text{and} \qquad n = n_1 + 2n_2 + 1\,.$$

Hence,

$$n = 2t - \left(\sum_{i=1}^{q} l_i + \alpha \right) + 1. \tag{29}$$

Now, we distinguish the cases n odd and n even.

(a) Let n be odd. Then $\operatorname{rank} A = \lfloor (n+1)/2 \rfloor = (n+1)/2$ if and only if

$$\frac{n+1}{2} = t - \sum_{i=1}^{q} \left\lceil \frac{l_i}{2} \right\rceil ,$$

by Theorem 5.15. By (29), this is equivalent to

$$\sum_{i=1}^{q} \frac{l_i}{2} = \sum_{i=1}^{q} \left\lceil \frac{l_i}{2} \right\rceil + 1 - \frac{\alpha}{2}. \tag{30}$$

If $\alpha = 0$ or $\alpha = 1$, then (30) is not possible, since

$$\sum_{i=1}^{q} \frac{l_i}{2} < \sum_{i=1}^{q} \left\lceil \frac{l_i}{2} \right\rceil + \frac{1}{2}.$$

Then $\alpha = 2$, and (30) holds if and only if l_i is even for all $i = 1, \dots, q$.

(b) Let n be even. Then $\operatorname{rank} A = \lfloor (n+1)/2 \rfloor = n/2$ if and only if

$$\frac{n}{2} = t - \sum_{i=1}^{q} \left\lceil \frac{l_i}{2} \right\rceil ,$$

by Theorem 5.15. Using (29) again, this is equivalent to

$$\sum_{i=1}^{q} \frac{l_i}{2} = \sum_{i=1}^{q} \left\lceil \frac{l_i}{2} \right\rceil + \frac{1}{2} - \frac{\alpha}{2}. \tag{31}$$

Again, if $\alpha = 0$, then (31) does not hold. If $\alpha = 1$, then (31) holds if and only if l_i is even for all $i = 1, \dots, q$. Finally, if $\alpha = 2$, then (31) holds if and only if one, and exactly one, among the numbers $l_1, \dots, l_q$ is odd.

□

Example 5.22 illustrates with staircase matrices having some flights with lengths larger than 1 the three situations presented in Theorem 5.21, where the maximal rank is attained.

Example 5.22. (a) *Let A be the following staircase matrix with 11 nonzero entries, which is equivalent to B through elementary row and column replacement operations.*

$$A = \begin{bmatrix} \times & \times & & & & \\ & \times & & & & \\ & \times & \times & \times & & \\ & & & \times & & \\ & & & \times & \times & \times \\ & & & & & \times \end{bmatrix} \sim \begin{bmatrix} \times & 0 & & & & \\ & \times & & & & \\ & 0 & \times & 0 & & \\ & & & \times & & \\ & & & 0 & \times & 0 \\ & & & & & \times \end{bmatrix} = B.$$

The matrix A has $\operatorname{rank} A = 6 = (11+1)/2$, $\mathcal{F}(A) = (1,2,2,2,2,1)$, $\alpha = 2$, *and* $\mathcal{L}(A) = (0)$, *so we are in case* (a) *of Theorem* 5.21.

(b1) *Now, let A be the following staircase matrix with* 12 *nonzero entries*

$$A = \begin{bmatrix} \times & \times & & & & \\ & \times & & & & \\ & \times & \times & \times & & \\ & & & \times & & \\ & & & \times & \times & \times \\ & & & & & \times \\ & & & & & \times \end{bmatrix} \sim \begin{bmatrix} \times & 0 & & & & \\ & \times & & & & \\ & 0 & \times & 0 & & \\ & & & \times & & \\ & & & 0 & \times & 0 \\ & & & & & \times \\ & & & & & 0 \end{bmatrix} = B.$$

Then $\operatorname{rank}(A) = 6 = 12/2 = \lfloor (12+1)/2 \rfloor$, $\mathcal{F}(A) = (1,2,2,2,2,2)$, $\alpha = 1$, *and* $\mathcal{L}(A) = (0)$, *so we are in case* (b1) *of Theorem* 5.21.

(b2) *In the last example, the staircase matrix A has 8 nonzero entries.*

$$A = \begin{bmatrix} \times & & & & & \\ \times & \times & \times & & & \\ & & \times & \times & \times & \\ & & & & \times & \end{bmatrix} \sim \begin{bmatrix} \times & & & & & \\ 0 & \times & 0 & & & \\ & & \times & 0 & 0 & \\ & & & & \times & \end{bmatrix} = B.$$

We have rank $A = 4 = 8/2 = \lfloor (8+1)/2 \rfloor$, $\mathcal{F}(A) = (1,2,1,2,1)$, $\alpha = 2$, *and* $\mathcal{L}(A) = (1)$, *so we are in case* (b2) *of Theorem* 5.21.

Notice that if the staircase matrix A is of maximal rank equal to $\lfloor \frac{n+1}{2} \rfloor$, then $\alpha = 0$ cannot occur. The maximum rank $\lfloor \frac{n+1}{2} \rfloor$ considered in Theorem 5.21 is related to Theorem 6.5 in next section. We will explain there this relationship.

6. Singular values of Fiedler matrices

We have commented in the Introduction that in [8] (see also [9]), the authors prove that the Frobenius companion matrices associated with the monic polynomial $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ (that is (1) and its transpose) have $n-2$ singular values equal to 1 and that the largest and the smallest singular values are the square roots of the following explicit expressions

$$\frac{1 + \sum_{k=0}^{n-1} |a_k|^2 \pm \sqrt{\left(1 + \sum_{k=0}^{n-1} |a_k|^2\right)^2 - 4|a_0|^2}}{2}. \tag{32}$$

The deep reason behind these properties is that C_1 can be written as a sum of a unitary matrix plus a matrix with rank one as follows

$$C_1 = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_1 & -a_0 - 1 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix}. \tag{33}$$

This expression immediately allows us to prove that C_1 has at least $n-2$ singular values equal to 1 and that the squares of the remaining two singular values can be obtained as the eigenvalues of a simple 2×2 matrix (we will present in Lemma 6.3 a general version of this result). In fact, the unitary matrix in the sum (33) has an additional property: it is a *permutation matrix*, i.e., a matrix obtained by permuting the rows (or columns) of the identity matrix. Fiedler matrices different from the Frobenius companion matrices cannot be expressed as "unitary plus rank-one matrices", but we will see in this section that every Fiedler matrix of $p(z)$ can be expressed as a sum of a permutation matrix plus a matrix whose rank varies from 1 to $\lfloor (n+1)/2 \rfloor$. In addition, we will show how to construct these two summands via simple algorithms and how to determine the rank of the second summand. In plain words, this will imply that many Fiedler matrices admit expressions as "unitary plus low-rank matrices" and so have a certain number of singular values equal to 1. We will also determine this number. Before proving these results, we illustrate in Example 6.1 these ideas.

Example 6.1. *We consider monic polynomials with degree 8, i.e., $p(z) = z^8 + \sum_{k=0}^{7} a_k z^k$.*

1. *We consider first the pentadiagonal matrix P_1 in Example* 2.2. *It can be written as*

$$P_1 = \begin{bmatrix} 0&1&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&1&0 \end{bmatrix} + \begin{bmatrix} -a_7&0&0&0&0&0&0&0 \\ -a_6&0&-a_5&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&-a_4&0&-a_3&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&-a_2&0&-a_1&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&-a_0-1&0 \end{bmatrix}. \tag{34}$$

The first summand is a permutation matrix and the second one has rank at most $4 = \lfloor (n+1)/2 \rfloor$ *(to see this, perform Gaussian elimination by rows). In fact, if* $a_i \neq 0$ *for* $i = 1, \ldots, 7$*, then the rank is exactly* 4.

2. *The second example corresponds to a Fiedler matrix with* $\mathrm{CISS}(\sigma) = (4, 3)$*. It can be expressed as*

$$M_\sigma = \begin{bmatrix} 0&0&0&0&1&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&1 \\ 0&0&0&1&0&0&0&0 \end{bmatrix} + \begin{bmatrix} -a_7 & -a_6 & -a_5 & -a_4 & 0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0 \\ 0&0&0&-a_3&0&0&0&0 \\ 0&0&0&-a_2&0&0&0&0 \\ 0&0&0&-a_1&0&0&0&0 \\ 0&0&0&-a_0-1&0&0&0&0 \end{bmatrix}. \tag{35}$$

The first summand is again a permutation matrix and the second one has rank at most 2*. In fact, if* $a_i \neq 0$ *for* $i = 1, \ldots, 7$*, then the rank is exactly* 2*.* `Algorithm 1` *allows the reader to easily check that these properties hold for Fiedler matrices of polynomials of arbitrary degree* $n \geq 2$ *associated with bijections that have all their consecutions in consecutive indices and all their inversions in consecutive indices, that is, those with* $\mathrm{CISS}(\sigma) = (\mathfrak{c}_0, \mathfrak{i}_0)$*,* $\mathfrak{c}_0 \neq 0$ *and* $\mathfrak{i}_0 \neq 0$*, or those with* $\mathrm{CISS}(\sigma) = (0, \mathfrak{i}_0, \mathfrak{c}_1, 0)$*,* $\mathfrak{i}_0 \neq 0$ *and* $\mathfrak{c}_1 \neq 0$*.*

Observe that, if all zero rows and columns are removed in the second summands in (34) *and* (35)*, then, in both cases, generalized staircase matrices are obtained. This property holds for every Fiedler matrix and is the key point in this section.*

Our first result in this section is Theorem 6.2, which proves rigorously that any Fiedler matrix M_σ can be written as $U_\sigma + V_\sigma$, where U_σ is a permutation matrix (and so unitary) and V_σ is a matrix such that after removing all its zero rows and columns becomes a staircase matrix. This property will allow us to bound the rank of V_σ via Corollary 5.19.

Theorem 6.2. *Let* $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ *with* $n \geq 2$*, let* $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ *be a bijection, and consider the following algorithm:*

`Algorithm 4.` *Given* $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ *and a bijection* σ*, the following algorithm constructs a pair of* $n \times n$ *matrices* U_σ *and* V_σ*.*

If σ has a consecution at 0 then

$$U_0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \qquad V_0 = \begin{bmatrix} -a_1 & 0 \\ -a_0 - 1 & 0 \end{bmatrix}$$

else

$$U_0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \qquad V_0 = \begin{bmatrix} -a_1 & -a_0 - 1 \\ 0 & 0 \end{bmatrix}$$

endif

for $i = 1 : n-2$

if σ has a consecution at i then

$$U_i = \begin{bmatrix} 0 & 1 & 0 \\ U_{i-1}(:,1) & 0 & U_{i-1}(:,2:i+1) \end{bmatrix}; \qquad V_i = \begin{bmatrix} -a_{i+1} & 0 & 0 \\ V_{i-1}(:,1) & 0 & V_{i-1}(:,2:i+1) \end{bmatrix}$$

else

$$U_i = \begin{bmatrix} 0 & U_{i-1}(1,:) \\ 1 & 0 \\ 0 & U_{i-1}(2:i+1,:) \end{bmatrix}; \qquad V_i = \begin{bmatrix} -a_{i+1} & V_{i-1}(1,:) \\ 0 & 0 \\ 0 & V_{i-1}(2:i+1,:) \end{bmatrix}$$

endif

endfor

$U_\sigma = U_{n-2}$

$V_\sigma = V_{n-2}$

Then the following statements hold.

(a) *If M_σ is the Fiedler matrix of $p(z)$ associated with σ, then,*

$$M_\sigma = U_\sigma + V_\sigma\,. \tag{36}$$

(b) *U_σ is a permutation matrix and, therefore, it is a unitary matrix.*

(c) *If all the zero rows and columns of V_σ are removed, then the resulting matrix is the generalized staircase matrix $\widetilde{V}_\sigma$ constructed by* `Algorithm 3` *for the inputs $x_0 = -a_0 - 1, x_1 = -a_1, \ldots, x_{n-1} = -a_{n-1}$ and σ.*

(d) *Let $\mathcal{L}(\sigma) = (l_1, l_2, \ldots, l_q)$ be the rank-determining list of σ introduced in Definition 5.12, and let t be the number of entries of* RCISS(σ)*. Then*

$$\operatorname{rank} V_\sigma \le r_\sigma \le \left\lfloor \frac{n+1}{2} \right\rfloor, \quad \textit{where} \quad r_\sigma := t - \sum_{j=1}^{q} \left\lceil \frac{l_j}{2} \right\rceil. \tag{37}$$

Moreover, if $a_0 + 1 \neq 0$ and $a_i \neq 0$ for all $i = 1, \ldots, n-1$, then $\operatorname{rank} V_\sigma = r_\sigma$.

Proof. Part (a). If we compare `Algorithms 1` and `4`, then we see that $W_0 = U_0 + V_0$. The proof is an induction on W_i, U_i, and V_i. Assume that $W_{i-1} = U_{i-1} + V_{i-1}$ for some $i - 1 \ge 0$. Then the structures of `Algorithms 1` and `4` make obvious that $W_i = U_i + V_i$. The result follows by taking $i = n - 2$.

Part (b). Again the proof is by induction on U_i. By definition U_0 is a 2×2 permutation matrix. Assume that U_{i-1} for some $i-1 \ge 0$ is a $(i+1) \times (i+1)$ permutation matrix. Then, `Algorithm 4` implies that U_i is a $(i+2) \times (i+2)$ permutation matrix. The result follows by taking $i = n - 2$.

Part (c). We perform an induction on the matrices V_i and $\widetilde{V}_i$ constructed by `Algorithms 4` and `3`, respectively. It is trivial to see that if we remove all zero rows and columns of V_0, then we obtain $\widetilde{V}_0$. Let us assume that the same is true for V_{i-1} and $\widetilde{V}_{i-1}$ for some $i - 1 \ge 0$, and let us prove the result for V_i and $\widetilde{V}_i$. For this purpose note that the first row and the first column of all matrices in the sequence $\{V_0, V_1, \ldots, V_{n-2}\}$ are not identically zero. Therefore, neither the first row nor the first column of V_{i-1} are removed to get $\widetilde{V}_{i-1}$. With this property in mind, it is clear from `Algorithms 4` and `3` that if we remove all zero rows and columns of V_i, then we get $\widetilde{V}_i$. The result follows by taking $i = n - 2$.

Part (d). Since removing zero rows and columns does not change the rank, we get $\operatorname{rank} V_\sigma = \operatorname{rank} \widetilde{V}_\sigma$, and the result is a direct consequence of Corollary 5.19 and Lemma 5.20. □

Parts (a) and (d) of Theorem 6.2 imply, in particular, that any Fiedler matrix M_σ associated with a bijection σ having a low number (compared to n) of entries in RCISS(σ) *can be decomposed as a sum of a unitary matrix U_σ plus a low-rank matrix V_σ.* The relationship between the rank of V_σ and the number of singular values of M_σ equal to 1 is established in Lemma 6.3, *which is valid for matrices much more general than Fiedler matrices.* In addition, Lemma 6.3 will allow us to reduce the computation of those singular values of M_σ that are not equal to 1, to the computation of the eigenvalues of a matrix whose size may be much smaller than n.

Lemma 6.3. *Let $A = U + LR \in \mathbb{C}^{n\times n}$, where $U \in \mathbb{C}^{n\times n}$ is a unitary matrix, $L \in \mathbb{C}^{n\times r}$, and $R \in \mathbb{C}^{r\times n}$. If $2r < n$, then A has at least $n - 2r$ singular values equal to 1, and the other $2r$ singular values are the square roots of the eigenvalues of the matrix*

$$H = I + \begin{bmatrix} R \\ L^*U \end{bmatrix} \begin{bmatrix} U^*L + R^*L^*L & R^* \end{bmatrix} \in \mathbb{C}^{2r\times 2r}. \tag{38}$$

Proof. The singular values of $A = U + LR$ are the square roots of the eigenvalues of A^*A. In the conditions of the statement,

$$\begin{aligned} A^*A &= (U + LR)^*(U + LR) = U^*U + R^*L^*U + U^*LR + R^*L^*LR \\ &= I + \begin{bmatrix} U^*L + R^*L^*L & R^* \end{bmatrix} \begin{bmatrix} R \\ L^*U \end{bmatrix} =: I + \widetilde{L}\widetilde{R}, \end{aligned}$$

where $\widetilde{L} \in \mathbb{C}^{n\times 2r}$ and $\widetilde{R} \in \mathbb{C}^{2r\times n}$. Therefore $\operatorname{rank}(\widetilde{L}\widetilde{R}) \le 2r$. Now, recall that the eigenvalues of $\widetilde{R}\widetilde{L} \in \mathbb{C}^{2r\times 2r}$, together with an additional $n-2r$ eigenvalues equal to 0, are the eigenvalues of $\widetilde{L}\widetilde{R} \in \mathbb{C}^{n\times n}$

[7, Theorem 1.3.20]. Hence, the eigenvalues of $H = I + \widetilde{R}\widetilde{L} \in \mathbb{C}^{2r\times 2r}$ together with an additional $n - 2r$ eigenvalues equal to 1 are the eigenvalues of $A^*A = I + \widetilde{L}\widetilde{R} \in \mathbb{C}^{n\times n}$. These are, precisely, the squares of the singular values of A. □

The application of Lemma 6.3 to a Fiedler matrix M_σ requires to factorize the matrix V_σ in (36) as $V_\sigma = L_\sigma R_\sigma$, where $L_\sigma \in \mathbb{C}^{n\times r_\sigma}$, $R_\sigma \in \mathbb{C}^{r_\sigma\times n}$, and r_σ was defined in (37). This is done in Lemma 6.4 via `Algorithm 5`. In this algorithm, submatrices like $L(:, 2 : 1)$ or $R(2 : 1, :)$ indicate empty matrices.

Lemma 6.4. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$, let $\sigma : \{0, 1, \ldots , n-1\} \to \{1, \ldots , n\}$ be a bijection, let V_σ be the matrix constructed by* `Algorithm 4`*, and let r_σ be the number defined in* (37)*. Consider the following algorithm:*

`Algorithm 5.` *Given $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ and a bijection σ, the following algorithm constructs a pair of matrices L_σ and R_σ.*

if σ has a consecution at 0 then

$$L_{-1} = \begin{bmatrix}-a_0 - 1\end{bmatrix}; \qquad R_{-1} = \begin{bmatrix}1\end{bmatrix}; \qquad L_0 = \begin{bmatrix}-a_1 \\ -a_0 - 1\end{bmatrix}; \qquad R_0 = \begin{bmatrix}1 & 0\end{bmatrix}$$

else

$$L_{-1} = \begin{bmatrix}1\end{bmatrix}; \qquad R_{-1} = \begin{bmatrix}-a_0 - 1\end{bmatrix}; \qquad L_0 = \begin{bmatrix}1 \\ 0\end{bmatrix}; \qquad R_0 = \begin{bmatrix}-a_1 & -a_0 - 1\end{bmatrix}$$

endif

for $i = 1 : n - 2$

if σ has an inversion at $i - 1$ and a consecution at i then

$$L_i = \begin{bmatrix}-a_{i+1} & 0 \\ -a_i & L_{i-2}(1, :) \\ 0 & 0 \\ 0 & L_{i-2}(2 : i, :)\end{bmatrix}; \qquad R_i = \begin{bmatrix}1 & 0 & 0 \\ 0 & 0 & R_{i-2}\end{bmatrix}$$

elseif σ has a consecution at $i - 1$ and an inversion at i then

$$L_i = \begin{bmatrix}1 & 0 \\ 0 & 0 \\ 0 & L_{i-2}\end{bmatrix}; \qquad R_i = \begin{bmatrix}-a_{i+1} & -a_i & 0 & 0 \\ 0 & R_{i-2}(:, 1) & 0 & R_{i-2}(:, 2 : i)\end{bmatrix}$$

elseif σ has consecutions at $i - 1$ and i then

$$L_i = \begin{bmatrix}-a_{i+1} & 0 \\ L_{i-1}(:, 1) & L_{i-1}(:, 2 : \texttt{end})\end{bmatrix}; \qquad R_i = \begin{bmatrix}R_{i-1}(:, 1) & 0 & R_{i-1}(:, 2 : i + 1)\end{bmatrix}$$

elseif σ has inversions at $i - 1$ and i then

$$L_i = \begin{bmatrix}L_{i-1}(1, :) \\ 0 \\ L_{i-1}(2 : i + 1, :)\end{bmatrix}; \qquad R_i = \begin{bmatrix}-a_{i+1} & R_{i-1}(1, :) \\ 0 & R_{i-1}(2 : \texttt{end}, :)\end{bmatrix}$$

endif

endfor

$L_\sigma = L_{n-2}$

$R_\sigma = R_{n-2}$

Then $V_\sigma = L_\sigma R_\sigma$, with $L_\sigma \in \mathbb{C}^{n\times r_\sigma}$ and $R_\sigma \in \mathbb{C}^{r_\sigma\times n}$. In addition, if $a_0 + 1 \neq 0$ and $a_i \neq 0$ for all $i = 1, \ldots , n - 1$, then $\operatorname{rank} V_\sigma = \operatorname{rank} L_\sigma = \operatorname{rank} R_\sigma = r_\sigma$.

Proof. We prove first $V_\sigma = L_\sigma R_\sigma$. To this purpose, let $\{V_0, V_1, \ldots , V_{n-2}\}$ (recall $V_{n-2} = V_\sigma$) be the sequence of matrices constructed by `Algorithm 4`. In addition, we define $V_{-1} := -a_0 - 1$. We will also consider the sequences $\{L_{-1}, L_0, L_1, \ldots , L_{n-2}\}$ and $\{R_{-1}, R_0, R_1, \ldots , R_{n-2}\}$ of matrices constructed by `Algorithm 5`. The proof consists of proving by induction that $V_i = L_i R_i$ for $i = -1, 0, 1, \ldots , n - 2$ (so, the result follows by taking $i = n - 2$). It is obvious that $V_{-1} = L_{-1}R_{-1}$ and $V_0 = L_0 R_0$. With a little bit more of effort, it is also straightforward to show via a direct computation that $V_1 = L_1 R_1$ holds. Let us assume that $V_j = L_j R_j$ for all $j = -1, 0, 1, \ldots , i - 1$, with $i - 1 \geq 1$, and let us prove $V_i = L_i R_i$. In the first place, it follows immediately from `Algorithm 5` and the induction hypothesis that the sizes of L_i and R_i allow us to multiply them. Next, we have two distinguish the four cases that appear in `Algorithm 5`:

(a) If σ has an inversion at $i-1$ and a consecution at i, then `Algorithm 5` implies that

$$L_iR_i = \begin{bmatrix} -a_{i+1} & 0 \\ -a_i & L_{i-2}(1,:) \\ 0 & 0 \\ 0 & L_{i-2}(2:i,:) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & R_{i-2} \end{bmatrix} = \begin{bmatrix} -a_{i+1} & 0 & 0 \\ -a_i & 0 & L_{i-2}(1,:)R_{i-2} \\ 0 & 0 & 0 \\ 0 & 0 & L_{i-2}(2:i,:)R_{i-2} \end{bmatrix}.$$

By the induction hypothesis $L_{i-2}R_{i-2} = V_{i-2}$, so

$$L_iR_i = \begin{bmatrix} -a_{i+1} & 0 & 0 \\ -a_i & 0 & V_{i-2}(1,:) \\ 0 & 0 & 0 \\ 0 & 0 & V_{i-2}(2:i,:) \end{bmatrix}. \tag{39}$$

On the other hand, if σ has an inversion at $i-1$ and a consecution at i, then `Algorithm 4` implies

$$V_{i-1} = \begin{bmatrix} -a_i & V_{i-2}(1,:) \\ 0 & 0 \\ 0 & V_{i-2}(2:i,:) \end{bmatrix} \quad \text{and} \quad V_i = \begin{bmatrix} -a_{i+1} & 0 & 0 \\ -a_i & 0 & V_{i-2}(1,:) \\ 0 & 0 & 0 \\ 0 & 0 & V_{i-2}(2:i,:) \end{bmatrix}. \tag{40}$$

Therefore, (39) and (40) imply that $V_i = L_iR_i$.

(b) If σ has a consecution at $i-1$ and an inversion at i, then the proof is similar to the one of Case (a) and is omitted.

(c) If σ has consecutions at $i-1$ and i, then `Algorithm 4` implies that

$$V_i = \begin{bmatrix} -a_{i+1} & 0 & 0 \\ V_{i-1}(:,1) & 0 & V_{i-1}(:,2:i+1) \end{bmatrix}. \tag{41}$$

Before completing the proof, it is needed to prove the following auxiliary result: *if σ has a consecution at k, for some $k = 0,1,\dots,n-2$, then the matrix R_k constructed by* `Algorithm 5` *satisfies* $R_k(1,:) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$. By definition, $R_0 = [1,0]$, so the result is true for $k = 0$. We follow by induction. Assume that $R_{k-1}(1,:) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$ if σ has a consecution at $k-1$ for some $k-1 \geq 0$, and let us prove the result for k. If σ has a consecution at k, then we need to consider only two out of the four cases in `Algorithm 5`: (1) σ has an inversion at $k-1$ and a consecution at k; and (2) σ has a consecution at $k-1$ and a consecution at k. In Case (1), $R_k(1,:) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$ by construction. In Case (2), $R_k(1,:) = \begin{bmatrix} R_{k-1}(1,1) & 0 & R_{k-1}(1,2:k+1) \end{bmatrix}$ and the result follows from the induction assumption.

Next we continue with the proof. If σ has consecutions at $i-1$ and i, then `Algorithm 5` and the auxiliary result imply that

$$\begin{aligned} L_iR_i &= \begin{bmatrix} -a_{i+1} & 0 \\ L_{i-1}(:,1) & L_{i-1}(:,2:\texttt{end}) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ R_{i-1}(2:\texttt{end},1) & 0 & R_{i-1}(2:\texttt{end},2:i+1) \end{bmatrix} \\ &= \begin{bmatrix} -a_{i+1} & 0 & 0 \\ L_{i-1}(:,1) + L_{i-1}(:,2:\texttt{end})R_{i-1}(2:\texttt{end},1) & 0 & L_{i-1}(:,2:\texttt{end})R_{i-1}(2:\texttt{end},2:i+1) \end{bmatrix} \\ &= \begin{bmatrix} -a_{i+1} & 0 & 0 \\ V_{i-1}(:,1) & 0 & V_{i-1}(:,2:i+1) \end{bmatrix}, \end{aligned} \tag{42}$$

where the last equality follows from the induction hypothesis $L_{i-1}R_{i-1} = V_{i-1}$ and the auxiliary result, which implies $R_{i-1}(1,:) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$. Equations (41) and (42) imply $V_i = L_iR_i$.

(d) If σ has inversions at $i-1$ and i, then the proof is similar to the one of Case (c) and is omitted. We only remark that in this case it is needed to prove the following auxiliary result: *if σ has an inversion at k, for some $k = 0,1,\dots,n-2$, then the matrix L_k constructed by* `Algorithm 5` *satisfies* $L_k(:,1) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$.

Next, we prove that if $a_0 + 1 \neq 0$ and $a_i \neq 0$ for all $i = 1, \ldots, n-1$, then $L_\sigma \in \mathbb{C}^{n \times r_\sigma}$, $R_\sigma \in \mathbb{C}^{r_\sigma \times n}$, and $\operatorname{rank} V_\sigma = \operatorname{rank} L_\sigma = \operatorname{rank} R_\sigma = r_\sigma$. It is very easy to see by induction that if $a_0 + 1 \neq 0$ and $a_i \neq 0$ for all $i = 1, \ldots, n-1$, then the structure of `Algorithm 5` implies that, for $i = 0, 1, \ldots, n-2$, all matrices L_i have full column rank and all matrices R_i have full row rank. In particular, $L_\sigma = L_{n-2} \in \mathbb{C}^{n \times r}$ has full column rank and $R_\sigma = R_{n-2} \in \mathbb{C}^{r \times n}$ has full row rank. Since $V_\sigma = L_\sigma R_\sigma$ and $\operatorname{rank} V_\sigma = r_\sigma$ by Theorem 6.2-(d), we get that $r = r_\sigma$ and $\operatorname{rank} L_\sigma = \operatorname{rank} R_\sigma = r_\sigma$.

Finally, observe that the sizes of the matrices $L_\sigma \in \mathbb{C}^{n \times r}$ and $R_\sigma \in \mathbb{C}^{r \times n}$ depend only on σ and n and not on the specific values of the coefficients $a_0, a_1, \ldots, a_{n-1}$ of $p(z)$. Therefore the sizes of L_σ and R_σ are always $L_\sigma \in \mathbb{C}^{n \times r_\sigma}$ and $R_\sigma \in \mathbb{C}^{r_\sigma \times n}$. □

Finally, as a direct corollary of Theorem 6.2, Lemma 6.3, and Lemma 6.4, we state Theorem 6.5, which is our concluding result on singular values of Fiedler matrices. For completeness and for making easy future references, we include again in the statement all notions involved.

Theorem 6.5. *Let $p(z) = z^n + \sum_{k=0}^{n-1} a_k z^k$ with $n \geq 2$, let $\sigma : \{0, 1, \ldots, n-1\} \to \{1, \ldots, n\}$ be a bijection, let M_σ be the Fiedler matrix of $p(z)$ associated with σ, let $\mathcal{L}(\sigma) = (l_1, l_2, \ldots, l_q)$ be the rank-determining list of σ introduced in Definition 5.12, and let t be the number of entries of $\operatorname{RCISS}(\sigma)$. Let us define*

$$r_\sigma := t - \sum_{j=1}^{q} \left\lceil \frac{l_j}{2} \right\rceil ,$$

which depends only on σ and not on $p(z)$. If $2r_\sigma < n$, then the following statements hold.

(a) *M_σ has at least $n - 2r_\sigma$ singular values equal to 1.*

(b) *The remaining $2r_\sigma$ singular values of M_σ are the square roots of the eigenvalues of the following $2r_\sigma \times 2r_\sigma$ matrix*

$$H_\sigma(p) = I + \begin{bmatrix} R_\sigma \\ L_\sigma^* U_\sigma \end{bmatrix} \begin{bmatrix} U_\sigma^* L_\sigma + R_\sigma^* L_\sigma^* L_\sigma & R_\sigma^* \end{bmatrix} \in \mathbb{C}^{2r_\sigma \times 2r_\sigma} , \tag{43}$$

where $U_\sigma \in \mathbb{C}^{n \times n}$ is the permutation matrix constructed by `Algorithm 4`, and $L_\sigma \in \mathbb{C}^{n \times r_\sigma}$ and $R_\sigma \in \mathbb{C}^{r_\sigma \times n}$ are the matrices constructed by `Algorithm 5`.

Proof. We combine equation (36) with $V_\sigma = L_\sigma R_\sigma$, from Lemma 6.4, to obtain $M_\sigma = U_\sigma + L_\sigma R_\sigma$. Then, apply Lemma 6.3 and get the result. □

Note that if the parameter t is small ($t \ll n$), then r_σ is also small, since $r_\sigma \leq t$, which implies that M_σ has many singular values equal to 1 and that the matrix $H_\sigma(p)$ has a small size. For almost all polynomials, $H_\sigma(p)$ has not eigenvalues equal to 1 and so M_σ has exactly $n - 2r_\sigma$ singular values equal to 1. Unfortunately, the potential small size of $H_\sigma(p)$ does not allow us to find explicit formulas for its eigenvalues (as it is illustrated in Example 6.7). This is only possible for Frobenius companion matrices because in this case $H_\sigma(p)$ is 2×2. We use in Example 6.6 the approach of Theorem 6.5 to recover the formulas (32) of the singular values of Frobenius companion matrices.

From Theorem 6.2-(d), we have $r_\sigma \leq \lfloor (n+1)/2 \rfloor$. In addition, observe that all Fiedler matrices for which $r_\sigma < \lfloor (n+1)/2 \rfloor$, satisfy $2r_\sigma < n$ and, so, have at least one singular value equal to 1. For those Fiedler matrices with $r_\sigma = \lfloor (n+1)/2 \rfloor$ Theorem 6.5 does not apply and they do not have any *guaranteed* singular value equal to 1. These matrices are characterized as those such that the staircase matrix $\widetilde{V}_\sigma$ in Theorem 6.2-(c) satisfies Theorem 5.21 (recall that $\mathcal{L}(\sigma) = \mathcal{L}(\widetilde{V}_\sigma)$ and that the number of entries of $\operatorname{RCISS}(\sigma)$ and $\mathcal{F}(\widetilde{V}_\sigma)$ are equal). In particular, Theorem 6.5 does not apply to some (but not all) of the Fiedler pentadiagonal matrices introduced in Example 2.2. We will illustrate this fact in Example 6.8.

Example 6.6. *We apply here Theorem 6.5 to C_1 in (1), that is, to the first Frobenius companion matrix. From Example 2.5, we know that C_1 corresponds to a bijection μ_1 with only inversions and with $\operatorname{RCISS}(\mu_1) = (n-1)$. Therefore, in this case, $t = 1$ and $\mathcal{L}(\mu_1) = (0)$, which implies $r_{\mu_1} = 1$ and that C_1 has at least $n-2$ singular values equal to 1. To determine the remaining 2 singular values, we use `Algorithm 4` to construct U_{μ_1} and `Algorithm 5` to construct L_{μ_1} and R_{μ_1} and we get that U_{μ_1} is the first summand in the right-hand side of (33), $L_{\mu_1} = [1, 0, \ldots, 0]^T \in \mathbb{C}^{n \times 1}$, and $R_{\mu_1} =$*

$[-a_{n-1}, -a_{n-2}, \ldots, -a_1, -a_0 - 1] \in \mathbb{C}^{1\times n}$ *(of course, this can be also seen by simple inspection of* (33)*). With these matrices, we easily obtain*

$$H_{\mu_1}(p) = \begin{bmatrix} |a_{n-1}|^2 + \cdots + |a_1|^2 + |a_0|^2 + \overline{a}_0 + 1 & |a_{n-1}|^2 + \cdots + |a_1|^2 + |a_0|^2 + a_0 + \overline{a}_0 + 1 \\ -\overline{a}_0 & -\overline{a}_0 \end{bmatrix}.$$

It is immediate to show that the eigenvalues of this matrix are given by (32)*, whose square roots are the* 2 *remaining singular values of* C_1*.*

Example 6.7. *Here, we apply Theorem* 6.5 *to the Fiedler matrix* M_σ *corresponding to bijections* σ *having a consecution at* 0 *and inversions at* $1, 2, \ldots, n-2$*. Explicitly, this matrix and its decomposition* (36) *are*

$$M_\sigma = \begin{bmatrix} -a_{n-1} & \cdots & \cdots & -a_1 & 1 \\ 1 & & & 0 & 0 \\ & \ddots & & \vdots & \vdots \\ & & 1 & 0 & 0 \\ & & & -a_0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 \\ 1 & & & & 0 \\ & \ddots & & & \vdots \\ & & 1 & & \vdots \\ & & & 1 & 0 \end{bmatrix} + \begin{bmatrix} -a_{n-1} & \cdots & -a_1 & 0 \\ & & \vdots & \vdots \\ & & \vdots & \vdots \\ & & \vdots & \vdots \\ & & -a_0 - 1 & 0 \end{bmatrix}. \quad (44)$$

In this case $\mathrm{CISS}(\sigma) = (1, n-2)$*,* $\mathrm{RCISS}(\sigma) = (1, n-2)$*,* $t = 2$*, and* $\mathcal{L}(\sigma) = (0)$*. Therefore,* $r_\sigma = 2$ *and* M_σ *has at least* $n-4$ *singular values equal to* 1*. To determine the remaining* 4 *singular values, we use again* `Algorithms 4` *and* `5` *to construct* U_σ*,* L_σ*, and* R_σ*. The matrix* U_σ *is the first summand of the right-hand side of* (44)*, which is the same matrix as in Example* 6.6*. For* L_σ *and* U_σ*, we obtain*

$$L_\sigma = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & -a_0 - 1 \end{bmatrix} \in \mathbb{C}^{n\times 2} \quad \text{and} \quad R_\sigma = \begin{bmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_2 & -a_1 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \in \mathbb{C}^{2\times n}.$$

With these matrices, we obtain after some algebra

$$H_\sigma(p) = \begin{bmatrix} 1 + \sum_{k=1}^{n-1} |a_k|^2 & -\overline{a}_0 a_1 (a_0 + 1) & \sum_{k=1}^{n-1} |a_k|^2 & -a_1 \\ -\overline{a}_1 & -a_0 + |a_0 + 1|^2 & -\overline{a}_1 & 1 \\ 1 & 0 & 1 & 0 \\ \overline{a}_1(\overline{a}_0 + 1) & -\overline{a}_0 |a_0 + 1|^2 & \overline{a}_1(\overline{a}_0 + 1) & -\overline{a}_0 \end{bmatrix}.$$

The square roots of the eigenvalues of $H_\sigma(p)$ *are the* 4 *remaining singular values of* M_σ*. However, it is not easy to obtain (if possible) explicit expressions for them, although we remark the fact that we have reduced an* $n \times n$ *(with* n *arbitrary) singular value problem to a* 4×4 *eigenvalue problem.*

Example 6.8. *Our last example illustrates that, except in one case, Theorem* 6.5 *does not apply to the pentadiagonal matrices introduced in Example* 2.2 *and, so, these matrices do not have, in general, any singular value equal to* 1*. Since* $P_3 = P_1^T$ *and* $P_4 = P_2^T$*, we consider only* P_1 *and* P_4*.*

For P_1*, it is easy to show that* $\mathrm{RCISS}(\sigma_1) = (1, 1, \ldots, 1)$ *with* $n-1$ *entries. Therefore* $\mathcal{L}(\sigma_1) = (n-3)$*, which gives* $r_{\sigma_1} = \lfloor (n+1)/2 \rfloor$ *both if* n *is even or odd.*

For P_4 *a surprise arises. It can be seen that* $\mathrm{RCISS}(\sigma_4) = (2, 1, \ldots, 1)$ *with* $n-2$ *entries, which implies* $\mathcal{L}(\sigma_2) = (n-4)$*. This implies* $r_{\sigma_4} = \lfloor (n+1)/2 \rfloor$ *if* n *is even, but* $r_{\sigma_4} = (n-1)/2 < \lfloor (n+1)/2 \rfloor$ *if* n *is odd. Therefore if* n *is odd, the pentadiagonal matrices* P_4 *and* P_2 *have, in general, only one singular value equal to* 1*.*

7. Conclusions and future work

We have performed a very detailed study of the condition numbers for inversion of Fiedler companion matrices of monic polynomials $p(z)$ in the Frobenius norm. This study is based on new properties for the inverses of Fiedler companion matrices. Among many other results, we have established that, from the point of view of condition numbers for inversion, the classical Frobenius companion matrices should not be used if $|p(0)| < 1$, since they have the largest condition number among all the Fiedler matrices of $p(z)$ and one should use, instead, any Fiedler matrix having a number of initial consecutions or inversions equal to 1. On the contrary, if $|p(0)| > 1$, then the Frobenius companion matrices are the ones to be used, since they have the smallest condition number among all the Fiedler matrices of $p(z)$. In the border case $|p(0)| = 1$ all Fiedler matrices of $p(z)$ have the same condition number. We have also seen that the singular values of Frobenius companion matrices have very simple properties that are not shared by any other Fiedler matrix. Nonetheless, the singular values of Fiedler matrices still retain some interesting properties that we have carefully studied. This study is based on the developments that we have presented on staircase matrices. As far as we know, this paper is the first work on perturbation properties of Fiedler matrices. Probably, the most interesting problem in this area is the study of eigenvalue condition numbers of Fiedler matrices. This will be the subject of future research.

References

[1] E. N. Antoniou, S. Vologiannidis, A new family of companion forms of polynomial matrices, Electron. J. Linear Algebra 11 (2004) 78–87.

[2] M. I. Bueno, F. De Terán, and F. M. Dopico, Recovery of eigenvectors and minimal bases of matrix polynomials from generalized Fiedler linearizations, SIAM J. Matrix Anal. Appl. 32 (2011) 463–483.

[3] F. De Terán, F. M. Dopico, and D. S. Mackey, Fiedler companion linearizations and the recovery of minimal indices, SIAM J. Matrix Anal. Appl. 31 (2010) 2181–2204.

[4] F. De Terán, F. M. Dopico, and D. S. Mackey, Palindromic companion forms for matrix polynomials of odd degree, J. Comput. Appl. Math. 236 (2011) 1464–1480.

[5] F. De Terán, F. M. Dopico, D. S. Mackey, Fiedler companion linearizations for rectangular matrix polynomials, to appear in Linear Algebra Appl. Also available as MIMS-EPrint 2011.110, Manchester Institute for Mathematical Sciences, Manchester, UK, 2011.

[6] M. Fiedler, A note on companion matrices, Linear Algebra Appl., 372 (2003) 325–331.

[7] R. A. Horn, C. R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.

[8] C. Kenney, A. J. Laub, Controllability and stability radii for companion form systems, Math. Control Signals Systems 1 (1988) 239–256.

[9] A. J. Laub, Matrix Analysis for Scientists & Engineers, SIAM, Philadelphia, 2005.

[10] The MathWorks, Inc., MATLAB Reference Guide, Natick, MA, 1992.

[11] S. Vologiannidis and E. N. Antoniou, A permuted factors approach for the linearization of polynomial matrices, Math. Control Signals Syst. 22 (2011) 317–342.