GENERALIZED EIGENVALUES OF NONSQUARE PENCILS WITH STRUCTURE

PABLO LECUMBERRI†, MARISOL GÓMEZ†, AND ALFONSO CARLOSENA‡

Abstract. This work deals with the generalized eigenvalue problem for nonsquare matrix pencils $A - \lambda B$ such that matrices $A, B \in \mathbb{M}_C(m \times n)$ show a given structure. More precisely, we assume they result from removing the first row of some matrix $G \in \mathbb{M}_C((m+1) \times n)$ in the case of $A$, and its last row in the case of $B$. This structured generalized eigenvalue problem can be found in signal processing methods and in the numerical computation of the greatest common divisor (GCD) of polynomials. Traditional methods for solving the problem $(A - \lambda B)v = 0$ do not yield valid solutions when the data are not exact, as is often the case in real applications. In this work we adopt a minimal perturbation approach. Taking into account the structure of the matrices involved, we develop a simple algorithm for the computation of the generalized eigenvalues.

Key words. nonsquare pencils, generalized eigenvalue, pseudospectra

AMS subject classifications. 15A18, 15A22, 65F10, 65F15

DOI. 10.1137/060669267

1. Introduction.

1.1. Nonsquare matrix pencils with structure. Given two matrices $A, B$ of the same dimensions, the set $\{A - \lambda B; \lambda \in \mathbb{C}\}$ constitutes a pencil. The generalized eigenvalues of the matrix pencil are those values of $\lambda$ for which there exist vectors $v$ different from zero, called generalized eigenvectors, such that the pair $(\lambda, v)$ satisfies

$$(A - \lambda B)v = 0.$$ 

The computation of the generalized eigenvalues has been regarded as a tool of great importance in engineering problems for decades. See [6] for applications in linear systems theory.

Matrices $A$ and $B$ form a regular matrix pencil $(A - \lambda B)$ if they are square and the characteristic polynomial $p(\lambda) = \det(A - \lambda B)$ is only zero for a finite number of values of $\lambda$. The generalized eigenvalue problem for this case has been extensively studied. In [12] there can be found some methods to solve it. The problem is more difficult when the matrices are rectangular or $(A - \lambda B)$ is square and singular for all values of $\lambda$. A matrix pencil built with this kind of matrices is known as a singular matrix pencil and its set of eigenvalues can be finite, empty, or infinite.

It is common to find applications, especially in the field of signal processing, that involve rectangular matrices $A, B \in \mathbb{M}_C(m, n), m > n$. Additional measurements, which should allow more accurate estimations, add more rows to matrices $A$ and $B$. Besides, matrices $A$ and $B$ often show some structure that must be taken into account for a correct treatment of the problem, or that may ease the complexity of...
the computation of the generalized eigenvalues. In this work we consider the following structure: We assume that matrix $A$ is the result of removing the first row of some matrix $G \in \mathcal{M}_C(m+1,n)$, while matrix $B$ results from removing the last row of $G$. These structured nonsquare matrix pencils appear, for example, in some important signal processing applications \cite{13, 9, 17} (in some of these examples, matrices $A, B$ also show some other structure, namely the Hankel structure of matrix $G$; this is not considered in this paper), and in the computation of the greatest common divisor of polynomials \cite{14, 15}, which in turn has application in diverse fields, such as network theory, automatic control, and computer-aided geometric design.

1.2. The use of the Kronecker canonical form (KCF). The Jordan canonical form of a square matrix, which describes its eigenvalues and invariant subspaces, can be extended to matrix pencils. A pencil $(A - \lambda B)$ is similar to its KCF \cite{11}, which is a pseudodiagonal matrix whose diagonal building blocks are Jordan blocks, with finite and infinite eigenvalues, and, in the case of singular pencils, singular blocks. The development of algorithms for the computation of the KCF has been a matter of research for decades. Nowadays, numerically stable routines are available (see \cite{1} and the references therein). However, the computation of the KCF is an ill-posed problem, in the sense that small perturbations to the pencil may yield a different KCF, probably a generic one, which only contains singular blocks. This feature hampers the application of the KCF to singular pencils derived from imperfectly known data, which is common in fields such as signal processing.

Previous study of perturbations in singular matrix pencils has been carried out under the light of the KCF. Stewart \cite{19} bounds the perturbation on generalized eigenvalues and eigenvectors implicitly assuming that the perturbation on the pencil does not change the structure of the KCF. A similar assumption is used in \cite{5} to bound the change in deflating subspaces of perturbed singular matrix pencils. In that paper, Demmel and Kågström also consider the regular case. Boley in \cite{2} gives bounds for the perturbation that must be applied to $A$ for the pencil $(A - \lambda B)$ to have a regular part in its KCF. Edelman, Elmroth, and Kågström in \cite{8} give bounds on the perturbation that makes a pencil less generic. These works give bounds to perturbations, which may be regarded as important information for some linear systems applications, but do not address the computation of the eigenvalues of the perturbed pencils, which in turn would be useful for many other real-world applications.

1.3. Contribution and structure of the paper. The goal of this paper is to propose a numerical method for finding generalized eigenvalues of nonsquare, and thus singular, matrix pencils. Aiming at its application to problems with real data, the method must be robust with respect to perturbations on the matrices of the pencil. This requirement leads us to discard matrix decompositions or transformations that reveal the structure of the pencil. We adopt the minimal perturbation approach (MPA) for the development of the numerical method. This paper can be seen as a follow-up to \cite{3} and \cite{4} which considers a particular structure for the matrix pencil. In those papers, the MPA for the eigenproblem of singular matrix pencils is presented. Roughly speaking, it consists of minimizing the norm of the perturbation that must be applied to a pencil so that it has a fixed number of generalized eigenvalues. The main contribution of this paper, apart from adapting the results in \cite{4} to the considered structure, is the transformation of the constrained optimization problem of the MPA into an unconstrained one with a low-dimensional parameter vector when only one eigenvalue is assumed to exist. Well-known optimization techniques \cite{16} can be applied to solve it, outperforming the numerical algorithm proposed in \cite{3} in terms of...
convergence rate and speed.

The paper is organized as follows: Section 2 reviews the MPA presented in [3] and [4] and its relation to pseudospectra. In section 3 the structure considered in this paper for the matrices that form the matrix pencil is introduced and incorporated to the constrained optimization problem that lies at the heart of the MPA. Then, it is shown to be equivalent to a simpler optimization problem and, in the case of seeking a single eigenvalue, it is further transformed into an unconstrained optimization problem that allows the use of simple numerical algorithms. Section 4 includes an example to show the differences between considering the structure of the matrix pencil and not doing so. Finally, we conclude the paper highlighting the main contribution and future research lines.

1.4. Notation. Matrices are denoted by boldface upper-case letters, such as $A$, whereas boldface lower-case letters, such as $v$, stand for column vectors. Scalars and polynomials are in roman typography, as $\lambda$ and $p(x)$, respectively. $A^T$ stands for the transpose of matrix $A$, $A^H$ stands for the conjugate transpose, and $\lambda^*$ denotes the complex conjugate of scalar $\lambda$. The 2-norm and the Frobenius norm of matrices and vectors are denoted by $\|\cdot\|_2$ and $\|\cdot\|_F$, respectively.

The inner product $\langle u, v \rangle$ is defined as $\langle u, v \rangle = u^H \cdot v$. Either notation will be used throughout the text. The projection of vector $u$ onto the range of $A$ is denoted by $P_A(u) = P_A \cdot u$, where $P_A$ is the corresponding matrix projection operator. Matrix $P^*_A$ is the projection operator over the complement subspace of the range of $A$. $0$ stands for a row or column vector whose elements are all equal to zero. Its actual dimensions will become clear from the context.

The $i$th singular value of a matrix $M \in M_C(m, n)$ is denoted by $\sigma_i(M)$, with $1 \leq i \leq \min(m, n)$. The singular values are assumed to be in decreasing order, i.e., $\sigma_1(A) \leq \sigma_2(A) \leq \cdots \leq \sigma_{\min(m,n)}(A)$.

2. Pseudospectra and the minimal perturbation approach. In some problems, particularly in singular matrix pencils, the eigenvalues do not change continuously with perturbations of the matrices involved. In these situations, the pseudospectra have proved to be a more useful tool to gain the insight usually provided by eigenvalues.

The spectra of a square matrix $A \in M_C(m, m)$ is defined as its set of eigenvalues:

$$\Lambda(A) = \{z \in \mathbb{C} : \exists u \neq 0, A \cdot u = z \cdot u\}.$$ 

The $\epsilon$-pseudospectra of $A$, $\Lambda_\epsilon(A)$, includes those elements of $\mathbb{C}$ that are eigenvalues of a matrix obtained from $A$ by a perturbation of Frobenius norm at most $\epsilon$ [20]:

$$\Lambda_\epsilon(A) = \{z \in \mathbb{C} : z \in \Lambda(A+E), \|E\|_F \leq \epsilon\}.$$ 

The spectra of $A$ is obtained as the $\epsilon$-pseudospectra when $\epsilon = 0$.

The extension to regular matrix pencils is treated in several papers; [7, 10] are some of them. Given two matrices $A, B \in M_C(m, m)$, the pseudospectra of the pair (allowing perturbation on both matrices) is defined as

$$\Lambda_\epsilon(A, B) = \left\{z \in \mathbb{C} : \exists u \neq 0, (A+E) \cdot u = z \cdot (B+F) \cdot u, \|E\|_F^2 + \|F\|_F^2 \leq \epsilon^2\right\}.$$ 

As in the case of a square matrix, when we are dealing with regular matrix pencils, there exists a nonempty spectra $\Lambda(A, B)$ and any connected region of the pseudospectra has elements of $\Lambda(A, B)$ inside it [10]. This does not happen with

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
singular matrix pencils. The definition of pseudospectra is generalized for rectangular matrices and general matrix pencils in [21], considering matrices in $\mathcal{M}_C(m, n)$ in (2.1).

The spectra of a singular matrix pencil may be empty, as well as its pseudospectra for values of $\epsilon$ small enough. As the value of $\epsilon$ is increased, the pseudospectra will eventually become nonempty. The MPA for the generalized eigenvalue problem for nonsquare matrix pencils [3] aims to find the smallest $\epsilon$ that allows a number of eigenvalues to be included in the pseudospectra of the pair $(A, B)$. It yields not only the eigenvalues, but also the exact perturbation matrices. The formulation of the MPA, also referred to as inverse pseudospectra analysis of the pencil, is as follows:

Given $A, B \in \mathcal{M}_C(m, n)$, find:

$$\min_{\{A_0, B_0, \{\lambda_k, v_k\}_{k=1}^p\}} \|A_0 - A\|_F^2 + \|B_0 - B\|_F^2$$

subject to:

$$\begin{cases}
(A_0 - \lambda_k B_0) v_k = 0, \\
\|v_k\|_2 = 1, \quad k = 1, \ldots, p
\end{cases}$$

\{v_1, \ldots, v_p\} \text{ is an independent set.}

This difficult optimization problem is shown in [4] to be equivalent to an optimization over the compact set of $p \times p$ unitary matrices. The result in [4] provides a feasible way to compute the least-squares (LS) estimation (or maximum likelihood estimation provided the errors in the elements of the matrices are independent and follow a Gaussian distribution) of the $p$ finite eigenvalues of a perturbed rectangular pencil. From a practical point of view, the number $p$ of finite eigenvalues (number of finite elementary divisors in the KCF) of the pair $(A_0, B_0)$ must be known beforehand and the optimization process can be complicated due to the large number of parameters.

Two particular cases that allow for a simpler formulation of the optimization problem are studied in [3]. The eigenproblem for matrices $A, B \in \mathcal{M}_C(m, 1)$ is studied in the first place, showing that it is equivalent to a total least squares problem. Then, the problem for matrices in $\mathcal{M}_C(m, n)$, with $n > 1$, is considered, assuming that a single finite eigenvalue is known to exist ($p = 1$). The simplified formulation of the minimal perturbation approach for this particular case is as follows:

Given $A, B \in \mathcal{M}_C(m, n)$, find:

$$\min_{\{A_0, B_0, \lambda, v\}} \|A_0 - A\|_F^2 + \|B_0 - B\|_F^2$$

subject to:

$$(A_0 - \lambda B_0) v = 0,$$

$$\|v\|_2 = 1.$$

A method for solving (2.3) is proposed. The optimization problem

$$\min_{\{\lambda, v\}} \frac{\|(A - \lambda B) v\|_2^2}{1 + |\lambda|^2}$$

subject to:

$$\|v\|_2 = 1$$

A method for solving (2.3) is proposed. The optimization problem

$$\min_{\{\lambda, v\}} \frac{\|(A - \lambda B) v\|_2^2}{1 + |\lambda|^2}$$

subject to:

$$\|v\|_2 = 1$$
is shown to be equivalent to (2.3) (both objective functions attain their conditioned minimum values at the same points), and a numerical algorithm guaranteed to converge to the local minima of \( f(\lambda, v) = \frac{\| (A - \lambda B)v \|^2_2}{1 + |\lambda|^2} \) is given. It is also noted that if several eigenvalues exist, \( f(\lambda, v) \) has minima at values of \( \lambda \) next to the eigenvalues. Different initial values may result in the optimization process converging at different local minima, and therefore, estimations of different eigenvalues. Note that, simple as this approach may be, the estimation of \( p \) finite eigenvalues (\( p > 1 \)) with this method is based on heuristics, unlike the LS estimation presented in [4].

When analyzing pseudospectra, we study a function that gives the norm of the smallest perturbation that must be applied to make \( \lambda \) an eigenvalue. The pseudospectra of a square matrix \( A \) is studied in [20] and the proposed function is the smallest singular value of \( (A - \lambda I) \), \( \sigma_{\min}(A - \lambda I) \), as only perturbations on \( A \) are allowed. In [10], the pseudospectra of regular matrix pencils is studied and another function is considered, as perturbation may occur on both matrices \( A \) and \( B \). In fact, the function that appears in [10] is equivalent to \( f(\lambda, v) \), which provides a measurement of the magnitude of the perturbation that must be applied to the pair \( (A, B) \) for \( \lambda \) to be its finite eigenvalue.

The position of the global minimum of \( f(\lambda, v) \) yields a LS estimation of the eigenvalue (provided the assumptions \( n > 1, p = 1 \) are valid and matrices \( A, B \) are unstructured). However, when using the proposed method for finding several eigenvalues, the minimization of \( f(\lambda, v) \) is performed for each eigenvalue on its own, without considering how the perturbation that would make a value of \( \lambda \) an eigenvalue would affect the other eigenvalues in the set. It is not a LS estimation of the set of eigenvalues, but nevertheless it may be accurate enough for many applications.

3. Minimal perturbation approach for structured nonsquare matrix pencils.

3.1. Introduction. We adopt the MPA for nonsquare matrix pencils \( (A - \lambda B) \) with a particular structure, namely matrix \( A \in \mathcal{M}_C(m, n) \) is the bottom \( m \times n \) submatrix of some matrix \( G \in \mathcal{M}_C(m+1, n) \), while \( B \in \mathcal{M}_C(m, n) \) is the top \( m \times n \) submatrix of \( G \). This structure may be written in a compact way as follows:

\[
\begin{align*}
A &= \begin{bmatrix} 0 & I_m \end{bmatrix} \cdot G, \\
B &= \begin{bmatrix} I_m & 0 \end{bmatrix} \cdot G,
\end{align*}
\]

where \( I_m \) is an \( m \times m \) identity matrix.

Structure (3.1) usually stems from the special characteristics of a particular problem or application, and matrix \( G \) comprises the data that is subject to perturbations. In this framework, as errors appear in the elements of \( G \), it is more natural to take the objective function of the MPA as the norm of the perturbation applied to matrix \( G \), \( \| G_o - G \|^2_F \), instead of the sum of the norms of the perturbation suffered by matrices \( A \) and \( B \), \( \| A_o - A \|^2_F + \| B_o - B \|^2_F \). These functions are slightly different, as it can be seen from (3.1) that the elements of the first and last row of the perturbation matrix \( G_o - G \) contribute once to the latter objective function, whereas the rest of the elements contribute twice. Taking this observation into account, the MPA for a nonsquare matrix pencil that shows structure (3.1) adopts the following conditioned
optimization formulation:

\[
\min \{ \| G_0 - G \|_F^2 \}
\]

subject to:

\[
\begin{align*}
\left\{ \begin{array}{l}
( \begin{bmatrix} 0 & I_m \end{bmatrix} - \lambda_k \begin{bmatrix} I_m & 0 \end{bmatrix} ) G_0 v_k &= 0 \\
\| v_k \|_2 &= 1
\end{array} \right\}, \; k = 1, \ldots, p
\end{align*}
\]

\{ v_1, \ldots, v_p \} \text{ independent set.}

As it is pointed out in [3] and section 2, an easy-to-solve optimization formulation may be obtained if the scope of the problem is limited to simple cases. We assume that matrices \((A, B)\) are such that there exists one single finite eigenvalue. As has been stated previously, if the matrix pencil comes from the perturbation of another pencil with several finite eigenvalues, these will appear as local minima of the MPA function. Problem (3.2) is simplified to

\[
\begin{align*}
\min_{\{ G_0, \lambda, v \}} & \| G_0 - G \|_F^2 \\
\text{subject to:} & \\
\left\{ \begin{array}{l}
( \begin{bmatrix} 0 & I_m \end{bmatrix} - \lambda \begin{bmatrix} I_m & 0 \end{bmatrix} ) G_0 v = 0 \\
\| v \|_2 &= 1
\end{array} \right\}
\end{align*}
\]

In this section we give a solution for problem (3.2) as a minimization over the set \(\{ \{ \lambda_k \}_{k=1}^p \} \). This solution is then particularized to the case \(p = 1\) (only a single eigenvalue is assumed to exist) and it is shown that the corresponding constrained optimization problem (3.3) can be transformed into an unconstrained minimization of a function of \(\lambda\). The objective function of this unconstrained optimization problem can be seen as a scalar function of \(R^2\). Nonlinear optimization algorithms like gradient-descent, conjugate-gradient or Gauss–Newton method can be used to find a local minimum, with the advantage that the reduced number of parameters lowers the complexity of the function and the number of local minima. This procedure is simpler and faster than the iterative algorithm proposed in [3], or the minimization of \(f(\lambda, v)\) through standard optimization algorithms.

### 3.2. Definition of vectors and matrices.

Before performing the simplification of the optimization problem, we introduce some preliminary results that will be used. Given a set of complex numbers \(\{ \{ \lambda_k \in \mathbb{C} \}_{k=1}^p \} \), matrix \(W(\lambda_1, \lambda_2, \ldots, \lambda_p) \in \mathcal{M}_\mathbb{C}(m+1, p)\) is defined as the Vandermonde matrix:

\[
W(\lambda_1, \lambda_2, \ldots, \lambda_p) = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
\lambda_1 & \lambda_2 & \cdots & \lambda_p \\
\lambda_1^2 & \lambda_2^2 & \cdots & \lambda_p^2 \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_1^m & \lambda_2^m & \cdots & \lambda_p^m
\end{bmatrix}.
\]

In the following, when the list of arguments is clear from the context, it will be omitted for the sake of clarity. A well-known property of Vandermonde matrices is that they have full column rank as long as \(\lambda_i \neq \lambda_j, \; 1 \leq i \neq j \leq p\), and \((m+1) \geq p\).
We also define matrix $D(\lambda) \in \mathcal{M}_C(m+1, m)$ as

\begin{equation}
D(\lambda) = \begin{bmatrix} 0 & I_m \end{bmatrix} - \lambda^* \begin{bmatrix} I_m & 0 \end{bmatrix}.
\end{equation}

Matrix $D(\lambda)$ has some relevant properties that are easy to prove:

- $D(\lambda)$ has full column rank $m$.
- $\text{Range}\{D(\lambda)\}$ is an $m$-dimensional subspace in $\mathbb{C}^{m+1}$.
- The left null space of $D(\lambda)$ is a subspace of dimension 1 spanned by the Vandermonde vector $W(\lambda)$. The corresponding unit-norm vector is denoted by $d_\perp \in \mathbb{C}^{m+1}$:

\begin{equation}
d_\perp = \frac{1}{\sqrt{\sum_{k=0}^{m} |\lambda|^2^k}} \begin{bmatrix} 1 \\ \lambda \\ \vdots \\ \lambda^m \end{bmatrix}.
\end{equation}

As $D(\lambda)$ has full column rank, the hermitian matrix $D(\lambda)^H \cdot D(\lambda)$ is regular and the projection operator $P_D \in \mathcal{M}_C(m, m)$ over $\text{Range}\{D(\lambda)\}$ can be computed as

\begin{equation}
P_D = D(\lambda) \cdot \left(D(\lambda)^H \cdot D(\lambda)\right)^{-1} \cdot D(\lambda)^H = (I - d_\perp \cdot d_\perp^H).
\end{equation}

### 3.3. Equivalent optimization problems.

The following lines give a solution for problem (3.2) as a minimization over the set $\{\{\lambda_k\}_{k=1}^p\}$, and then the constrained optimization problem (3.3) is transformed into an unconstrained minimization of a function that only depends on $\lambda$. These transformations are made in three steps that we present as theorems.

**Theorem 1.** The optimization problem posed in (3.2) is equivalent to the optimization problem

\begin{equation}
\min_{\{\{\lambda_k\}_{k=1}^p\}} \sum_{i=1}^{p} \sigma_{(n-p+i)}(G_\perp \cdot \mathbf{G}),
\end{equation}

where $G_\perp$ is the projection matrix over the complement subspace of the range of the Vandermonde matrix $W$, which is defined in (3.4). The optimal solution for $G_\perp$ is obtained by a rank-$p$ perturbation to the original matrix $\mathbf{G}$:

\begin{equation}
G_\perp = \mathbf{G} - P_\perp \cdot \mathbf{G} \cdot Q \cdot Q^H,
\end{equation}

with $Q \in \mathcal{M}_C(n, p)$ a matrix composed of the singular vectors of $P_\perp \cdot \mathbf{G}$ corresponding to its smallest singular values.

**Proof.** We will begin the transformation of the optimization problem (3.2) by noting that conditions

\begin{equation}
\begin{bmatrix} 0 & I_m \end{bmatrix} \cdot G_\perp \cdot v_k = \lambda_k \cdot \begin{bmatrix} I_m & 0 \end{bmatrix} \cdot G_\perp v_k, \quad k = 1, \ldots, p
\end{equation}

are equivalent to

\begin{align*}
\left[ G_\perp \cdot v_k \right]_1 &= h_k, \\
\left[ G_\perp \cdot v_k \right]_i &= \lambda_{i-1} \cdot \left[ G_\perp \cdot v_k \right]_{i-1}, \quad i = 2, \ldots, m+1, \quad k = 1, \ldots, p
\end{align*}
where $|x|_i$ denotes the $i$th element of vector $x$, and $h_k \in \mathbb{C}$, $k = 1, \ldots, p$. Therefore, (3.10) will hold if and only if

$$G_o \cdot V = W \cdot H$$

(3.11)

for some diagonal matrix $H \in \mathcal{M}_\mathbb{C}(p, p)$ built with $h_1, h_2, \ldots, h_p$ as its diagonal elements. Since the Vandermonde matrix $W \in \mathcal{M}_\mathbb{C}(m+1, p)$ has full column rank (3.4), so does $V$, which amounts to the set $\{v_k\}_{k=1}^p$ being independent, as required by the last condition of (3.2).

Let $T \in \mathcal{M}_\mathbb{C}(p, p)$ be an invertible matrix such that $V \cdot T = Q$, with $Q \in \mathcal{M}_\mathbb{C}(n, p)$ a matrix with orthonormal columns. Then, multiplying (3.11) by $T$ we get

$$G_o \cdot V \cdot T = W \cdot H \cdot T,$$

or, equivalently,

$$G_o \cdot Q = W \cdot R.$$  

(3.12)

Equation (3.12) is equivalent to equation (3.10) and $R \in \mathcal{M}_\mathbb{C}(p, p)$ is a regular matrix.

Let $U \in \mathcal{M}_\mathbb{C}(n, n-p)$ denote a matrix with orthonormal columns such that $S = [Q \ U]$ is a $n \times n$ unitary matrix. Taking $\Delta = G_o - G$, the objective function of the optimization problem can be written as follows:

$$\|G - G_o\|_F = \|\Delta\|_F = \|\Delta \cdot S\|_F = \|([\Delta \cdot Q \ \Delta \cdot U])\|_F$$

(3.13)

In the last equality, the decomposition of the perturbation matrix $\Delta = \Delta^{(Q)} + \Delta^{(U)}$ has been used:

$$\Delta = \Delta \cdot S \cdot S^H = \Delta \cdot (Q \cdot Q^H + U \cdot U^H) = \Delta \cdot Q \cdot Q^H + \Delta \cdot U \cdot U^H = \Delta^{(Q)} + \Delta^{(U)}.$$  

The rows of $\Delta^{(Q)}$ and $\Delta^{(U)}$ are the projections of the rows of $\Delta$ onto the range of $Q$ and $U$, respectively, so $\Delta^{(Q)} \cdot U = 0$ and $\Delta^{(U)} \cdot Q = 0$.

Then, condition (3.12) can be written in terms of the perturbation matrix as

$$\left(G + \Delta^{(Q)}\right) \cdot Q = W \cdot R.$$  

(3.14)

It is clear that the minimum of the objective function (3.13) subject to condition (3.14) will be attained for a perturbation matrix such that $\Delta^{(U)} = 0$. The objective function can be replaced by $\|\Delta^{(Q)} \cdot Q\|_F$ and, consequently, problem (3.3) is equivalent to

$$\min_{\{G_o \cdot (\lambda^k)_{k=1}^p, Q\}} \|G \cdot Q - G_o \cdot Q\|_F$$

subject to:

$$G_o \cdot Q = W \cdot R$$

(3.15a)

for some invertible matrix $R \in \mathcal{M}_\mathbb{C}(p, p)$ and with $Q \in \mathcal{M}_\mathbb{C}(n, p)$ a matrix with orthonormal columns. The matrix $G_o$ that solves the constrained optimization problem satisfies

$$G_o = G + \Delta = G + \Delta^{(Q)} = G + \Delta^{(Q)} \cdot Q \cdot Q^H.$$
From condition (3.14), we have

\[(3.16) \quad G_o = G + W \cdot R \cdot Q^H - G \cdot Q \cdot Q^H.\]

Condition (3.15b) can be incorporated to the objective function (3.15a), yielding the following optimization problem:

\[(3.17) \quad \min_{\{R, \{\lambda_k\}_{k=1}^p, Q\}} \|G \cdot Q - W \cdot R\|_F.\]

Given matrix Q and a set \{\lambda_k\}_{k=1}^p, the value of R that minimizes the objective function (3.17) is the one that makes the columns of W \cdot R equal to the projection of the corresponding columns of G \cdot Q over the range of W:

\[(3.18) \quad W \cdot R = P_W \cdot G \cdot Q.\]

Hence, the objective function can be written as:

\[\|G \cdot Q - P_W \cdot G \cdot Q\|_F = \|(I - P_W) \cdot G \cdot Q\|_F = \|P_W^+ \cdot G \cdot Q\|_F.\]

The minimum of the objective function will be attained for a matrix Q such that its columns are the singular vectors of \(P_W^+ \cdot G\) corresponding to its smallest singular values:

\[\min_{\{\{\lambda_k\}_{k=1}^p, Q\}} \|P_W^+ \cdot G \cdot Q\|_F = \min_{\{\{\lambda_k\}_{k=1}^p\}} \sum_{i=1}^p \sigma_{n-p+i} (P_W^+ \cdot G).\]

Finally, an expression for the optimal solution \(G_o\) in terms of matrix G and a perturbation will be given. From (3.16) and (3.18), we have

\[G_o = G + P_W \cdot G \cdot Q \cdot Q^H - G \cdot Q \cdot Q^H = G - (I - P_W) \cdot G \cdot Q \cdot Q^H = G - P_W^+ \cdot G \cdot Q \cdot Q^H.\]

**Corollary 1.** The optimization problem posed in (3.3) is equivalent to the optimization problem

\[(3.19a) \quad \min_{\{\lambda, v\}} v^H \cdot G^H \cdot P_D \cdot G \cdot v
\]

subject to:

\[(3.19b) \quad \|v\|_2 = 1\]

with \(P_D\) defined in (3.7). The optimal solution for \(G_o\) is obtained by a rank-one perturbation to the original matrix G:

\[G_o = G - P_D \cdot G \cdot v \cdot v^H.\]

**Proof.** This corollary follows directly from Theorem 1, taking \(p = 1\), and the definitions in section 3.2. Note that in this case \(W = h_1 \cdot d_\perp\), for some \(h_1 \in \mathbb{C}\), and \(Q = v\). □

Corollary 1 provides a simple formulation for the MPA when only one eigenvalue is assumed to exist. Still, the minimization problem it poses is complicated due to the number of parameters involved, which discourages the use of standard optimization techniques. The following theorems show that if matrix G has orthonormal columns,
then problem (3.8) is equivalent to the unconstrained minimization of a function of \( \lambda \).

The assumption of \( G \) having orthonormal columns is valid in many practical applications. For example, in the case of the greatest common divisor (GCD) computation, the columns of matrix \( G \) span the null-space of a matrix \( P \) defined in terms of the coefficients of the polynomials \([14]\), so \( G \) can be computed with orthonormal columns from the singular value decomposition of \( P \).

On the other hand, the generalized eigenvalues of a matrix pencil that shows structure (3.1) are the roots of the Vandermonde vectors \([1 \ \lambda^* \ \ldots \ \lambda^* \ m]\)^T that are included in Range(\( G \)) \([18]\). Therefore, a matrix \( \hat{G} \) with orthonormal columns that has the same range as \( G \) gives rise to a matrix pencil \((\hat{A} - \lambda \hat{B})\), with \( \hat{A} = [0 \ \ I_m] \hat{G} \) and \( \hat{B} = [I_m \ 0] \cdot \hat{G} \), that has the same eigenvalues.

Although the MPA yields different results for \( G \) and \( \hat{G} \), the position of the minima of the objective function when \( G \) is considered may still be a good estimate for the approximate eigenvalues of \((A - \lambda B)\).

**THEOREM 2.** The optimization problem posed in (3.19) is equivalent to the optimization problem

\[
\begin{align*}
(3.20a) \quad & \max_{\{\lambda, v\}} |d^H \cdot G \cdot v|^2 \\
subject \ to: \quad & e^H \cdot G \cdot v = 0. \\
(3.20c) \quad & ||v||_2 = 1.
\end{align*}
\]

**Proof.** We will make use of the inner-product and projection function notation for transforming the objective function (3.19a):

\[
\begin{align*}
(3.21) \quad & v^H \cdot G^H \cdot P_D \cdot G \cdot v = \langle G \cdot v, P_{D(\lambda)}(G \cdot v) \rangle \\
& \quad = \langle P_{D(\lambda)}(G \cdot v) + P_d(G \cdot v), P_{D(\lambda)}(G \cdot v) \rangle \\
& \quad = \langle P_{D(\lambda)}(G \cdot v), P_{D(\lambda)}(G \cdot v) \rangle,
\end{align*}
\]

where the orthogonality of projections onto the range of \( D(\lambda) \) and \( d_\perp \) has been taken into account.

Under the assumption that the columns of \( G \) are orthonormal, the norm of \( G \cdot v \) will be the same as the norm of \( v \), which must be equal to 1 (3.19b). This property fixes a relationship between the inner-products of the projections of \( G \cdot v \):

\[
\begin{align*}
\langle G \cdot v, G \cdot v \rangle = & \langle P_{D(\lambda)}(G \cdot v) + P_d(G \cdot v), P_{D(\lambda)}(G \cdot v) + P_d(G \cdot v) \rangle \\
= & \langle P_{D(\lambda)}(G \cdot v), P_{D(\lambda)}(G \cdot v) \rangle + \langle P_d(G \cdot v), P_d(G \cdot v) \rangle = 1.
\end{align*}
\]

From (3.21) and (3.22) it is clear that our objective function is

\[
\|G - G_0\|_F^2 = ||G \cdot v||_2^2 - ||P_d(G \cdot v)||_2^2 = 1 - ||P_d(G \cdot v)||_2^2.
\]

Its minimization is equivalent to the maximization of \( ||P_d(G \cdot v)||_2^2 \). Consequently, the objective of our optimization problem is the maximization of

\[
\|P_d(G \cdot v)||_2^2 = \|d_\perp \cdot d^H \cdot G \cdot v||_2^2 = \|d_\perp||_2^2 \cdot |d^H \cdot G \cdot v|^2 = |d^H \cdot G \cdot v|^2.
\]
Theorem 3. The optimization problem posed in (3.20) is equivalent to the optimization problem

\[
\min_{\lambda} \frac{1}{f(\lambda)}
\]

with

\[
f(\lambda) = d^H_L \cdot G \cdot G^H \cdot d_L.
\]

Proof. We define vector \( x \) as the vector in the direction of \( v \) that makes \( d^H_L \cdot G \cdot x = 1 \). Since \( v \) is a unit-norm vector (3.20c), \( v = \frac{x}{\|x\|_2} \), and the objective function (3.20a) can be written in terms of \( x \) as

\[
\left| d^H_L \cdot G \cdot v \right|^2 = \frac{|d^H_L \cdot G \cdot x|^2}{\|x\|^2_2} = \frac{1}{\|x\|^2_2}.
\]

Thus, the optimization problem (3.20) is equivalent to

\[
\min_{\{\lambda, x\}} \|x\|^2_2
\]

subject to:

\[
d^H_L \cdot G \cdot x = 1.
\]

Condition (3.24b) poses an undetermined linear system. The minimum norm solution for \( x \) will minimize function (3.24a). For every \( \lambda \), the minimum norm solution of the respective undetermined linear system can be computed as the product of the Moore–Penrose pseudoinverse of the coefficients matrix and the independent term. The pseudoinverse of \( d^H_L \cdot G \) can be constructed from its singular value decomposition [12], \( d^H_L \cdot G = u \cdot \sigma \cdot v^H \), with

\[
u = 1 \quad \sigma = \|G^H \cdot d_L\|_2 \quad v = \frac{G^H \cdot d_L}{\|G^H \cdot d_L\|_2^2}.
\]

Then, the pseudoinverse of \( d^H_L \cdot G \) is defined as

\[
(d^H_L \cdot G)^+ = v \cdot \frac{1}{\sigma} \cdot u^H = \frac{G^H \cdot d_L \cdot d_L}{\|G^H \cdot d_L\|_2^2}.
\]

And the minimum norm solution of the undetermined linear system is

\[
x = (d^H_L \cdot G)^+ \cdot 1 = (d^H_L \cdot G)^+.
\]

With this result, the objective function we are trying to minimize is equivalent to

\[
\|x\|^2_2 = x^H \cdot x = \frac{d^H_L \cdot G \cdot G^H \cdot d_L}{\|G^H \cdot d_L\|^2_2} \cdot \frac{d^H_L \cdot G \cdot G^H \cdot d_L}{\|G^H \cdot d_L\|^2_2} = \frac{d^H_L \cdot G \cdot G^H \cdot d_L}{\|G^H \cdot d_L\|^2_2}
\]

\[
= \frac{\|G^H \cdot d_L\|^2_2 \cdot 1}{\|G^H \cdot d_L\|^2_2} = \frac{1}{\|G^H \cdot d_L\|^2_2}, \quad \Box
\]
4. Example. In order to obtain a matrix pencil with the structure considered in this paper, we pose a numerical GCD computation problem. First, we take a polynomial \( p(x) \) with roots \(-0.5, -1.25, 0.2 \pm 1.4i \) and 0.625:

\[
p(x) = (x + 0.5) \cdot (x + 1.25) \cdot (x^2 - 0.4x + 2) \cdot (x - 0.625).
\]

Then we randomly generate 20 polynomials of degree 44 with their coefficients following a normal distribution with zero mean and variance equal to one. After multiplying these polynomials with \( p(x) \), we have 20 polynomials whose GCD is \( p(x) \). Matrix \( X \in M_{50 \times 20} \) is built with the coefficients of these 20 polynomials as columns. In [14, 17] it is proved that the only generalized eigenvalues of the matrix pencil \( A - \lambda B \) with \( A = [0 \ I] \cdot G, \ B = [I \ 0] \cdot G, \) and \( G \in M_{50 \times 30} \) the orthogonal complement of \( X \), are the roots of the GCD, \( p(x) \). Figure 4.1(a) shows a plot of \( \epsilon(\lambda) = \log(1 - f(\lambda)) \). This function has been chosen so that its minima can be clearly spotted in Figure 4.1. Five minima, located at values of \( \lambda \) equal to the roots of \( p(x) \), can be seen.

The MPA is an alternative to traditional methods (computation of the KCF or other structure-revealing forms) when the available data are contaminated by noise and the latter yield no valid results. Figures 4.1(b) and 4.1(c) show plots of function \( \epsilon(\lambda) \) for matrix \( X \) perturbed with matrices whose elements are taken from zero-mean normal distributions of variance \( \sigma = 0.5 \) and \( \sigma = 1 \), respectively. It can be observed that the minima are not as deep as in the noiseless case (Figure 4.1(a)) and that they are slightly deviated from their correct position. Some minima may vanish for strong perturbations, and therefore, in these situations, the MPA may fail to provide an estimation of the generalized eigenvalues.

Having established the suitability of the MPA for low and moderate noise levels in our example, we proceed to compare the optimization method proposed in [3] and the minimization of function \( \epsilon(\lambda) \) (3.23). To this end, we must choose a method for the unconstrained minimization of \( \epsilon(\lambda) \). Due to its simplicity, we make use of the steepest descent method, although more sophisticated strategies, such as trust region methods, show better convergence properties [16].

Figure 4.1(d) shows a plot of \( \epsilon(\lambda) \) for a perturbation of level \( \sigma = 0.5 \) with the points reached at every iteration for both minimization methods, starting from a hand-picked initial point. The steepest descent method over \( \epsilon(\lambda) \) (triangles) converges to the minimum faster than the optimization algorithm proposed in [3] (circles). This can also be seen in Figure 4.2. It shows the values of \( \epsilon(\lambda) \) at every iteration for two different hand-picked initial points for both optimization methods. Lines of the same style (solid or dashed) denote the same initialization, whereas triangles correspond to points found with the steepest descent method and circles correspond to the method proposed in [3]. Note that this hand-picked starting points are considerably further away from the location of the minima than the initial values of \( \lambda \) obtained by the squaring method of initialization [3].

Apart from the rate of convergence, another issue that should be taken into account is the numerical complexity of each iteration. The method of minimization for unstructured matrix pencils requires the computation of the singular vector related to the smallest singular value of \( (A - \lambda B) \) at every iteration. The SVD for a \( m \times n \) matrix is an operation of complexity \( O(mn^2) \) (this value can be lowered by suitable factorization or partial decomposition [12]). On the other hand, the gradient and Hessian of \( f(\lambda) \) can be computed with a complexity of \( O(m) \).
5. Conclusions and future work. This work has addressed the MPA for non-square matrix pencils \((A - \lambda B)\). We have considered the following structure for matrices \(A, B \in M_C(m, n)\): 

\[
A = [0 \ I] \cdot G, \quad B = [I \ 0] \cdot G,
\]

Fig. 4.1. Log of function \((1 - f(\lambda))\) with exact data (a), with \(\sigma = 0.5\) (b) and with \(\sigma = 1\) (c). Iterations of minimization algorithms (d): Method proposed in [3] (circles, solid line) and steepest descent over \(f(\lambda)\) defined in (3.23) (triangles, dashed line).

Fig. 4.2. Values of \(\log(1 - f(\lambda))\) at every iteration for different starting points for the method proposed in [3] (circles) and steepest descent over \(f(\lambda)\) defined in (3.23) (triangles).
with $G \in M_C(m+1,n)$. The matrices of the pencil are derived from matrix $G$. This observation leads to a slightly different MPA formulation, as the norm of the perturbation in $G$ is considered, instead of the norm of the perturbations in $A$ and $B$. The MPA can be formulated as a minimization over the set $\{\{\lambda_k\}_{k=1}^p\}$. Besides, when $p = 1$, this structure allows the transformation of the constrained optimization problem that must be solved to compute the solution of the MPA into an unconstrained one. Then, well-known methods for unconstrained minimization can be used, improving the rate of convergence and simplicity of previous algorithm for the computation of the solution.

The structure considered in this paper appears in some signal processing methods and, more notably, in the numerical computation of the GCD of polynomials. However, once the GCD computation problem is formulated as an eigenvalue problem for a rectangular matrix pencil, previous proposed solutions consist in matrix transformations that reveal the generalized eigenvalues. This way of proceeding is not suitable for perturbed data, due to the ill-posedness of the transformations, while the MPA may still yield valid results in this situation. Our work in the future will focus in this application, comparing this method to other numerical GCD computation methods.

Acknowledgments. The authors would like to thank professor I. Lizasoain for her very helpful comments and suggestions while this article was being written. The authors would also like to thank the referees for their valuable comments and suggestion greatly improving this paper.

REFERENCES


