MANCHESTER
1824

*Optimal scaling of generalized and polynomial
eigenvalue problems*

Betcke, T.

2007

MIMS EPrint: **2007.120**

Manchester Institute for Mathematical Sciences

School of Mathematics

The University of Manchester

OPTIMAL SCALING OF GENERALIZED AND POLYNOMIAL EIGENVALUE PROBLEMS*

T. BETCKE†

Abstract. Scaling is a commonly used technique for standard eigenvalue problems to improve the sensitivity of the eigenvalues. In this paper we investigate scaling for generalized and polynomial eigenvalue problems (PEPs) of arbitrary degree. It is shown that an optimal diagonal scaling of a PEP with respect to an eigenvalue can be described by the ratio of its normwise and componentwise condition number. Furthermore, the effect of linearization on optimally scaled polynomials is investigated. We introduce a generalization of the diagonal scaling by Lemonnier and Van Dooren to PEPs that is especially effective if some information about the magnitude of the wanted eigenvalues is available and also discuss variable transformations of the type $\lambda = \alpha\mu$ for PEPs of arbitrary degree.

Key words. polynomial eigenvalue problem, balancing, scaling, condition number, backward error

AMS subject classifications. 65F15, 15A18

1. Introduction. Scaling of standard eigenvalue problems $Ax = \lambda x$ is a well established technique that is implemented in the LAPACK routine `xGEBAL`. It goes back to work by Osborne, Parlett and Reinsch [14, 15]. The idea is to find a diagonal matrix D that scales the rows and columns of $A \in \mathbb{C}^{n \times n}$ in a given norm such that

$$\|D^{-1}ADe_i\| = \|e_i^* D^{-1}AD\|, \quad i = 1, \ldots, n,$$

where e_i is the ith unit vector. This is known as balancing. LAPACK uses the 1-norm. Balancing matrix rows and columns can often reduce the effect of rounding errors on the computed eigenvalues. However, as Watkins demonstrated [19], there are also cases in which balancing can lead to a catastrophic increase of the errors in the computed eigenvalues.

For generalized eigenvalue problems (GEPs) $Ax = \lambda Bx$ a scaling technique proposed by Ward [18] is implemented in the LAPACK routine `xGGBAL`. Its aim is to find diagonal matrices D_1 and D_2 such that the elements of D_1AD_2 and D_1BD_2 are scaled as equal in magnitude as possible.

A different approach for the scaling of GEPs is proposed by Lemonnier and Van Dooren [11]. In Section 5 we will come back to this. It is interesting to note that the default behavior of LAPACK (and also of MATLAB) is to scale nonsymmetric standard eigenvalue problems but not to scale GEPs.

In this paper we discuss the scaling of polynomial eigenvalue problems (PEPs) of the form

$$P(\lambda)x := (\lambda^\ell A_\ell + \cdots + \lambda A_1 + A_0)x = 0, \quad A_k \in \mathbb{C}^{n \times n}, \ A_\ell \neq 0, \ \ell \geq 1. \quad (1.1)$$

Every $\lambda \in \mathbb{C}$ for which there exists a solution $x \in \mathbb{C}^n \backslash \{0\}$ of $P(\lambda)x = 0$ is called an eigenvalue of P with associated right eigenvector x. We will also need left eigenvectors $y \in \mathbb{C}^n \backslash \{0\}$ defined by $y^*P(\lambda) = 0$.

†School of Mathematics, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK (`timo.betcke@maths.man.ac.uk`, `http://www.ma.man.ac.uk/~tbetcke`).

In Section 2 we review the definition of condition numbers and backward errors for the PEP (1.1). Then in Section 3 we investigate diagonal scalings of (1.1) of the form $D_1 P(\lambda) D_2$, where D_1 and D_2 are diagonal matrices in the set

$$\mathcal{D}_n := \{D : \ D \in \mathbb{C}^{n \times n} \text{ is diagonal and } \det(D) \neq 0\}.$$

We show that the minimal achievable normwise condition number of an eigenvalue by diagonal scaling of $P(\lambda)$ can be bounded by its componentwise condition number. This gives easily computable conditions on whether the condition number of eigenvalues can be improved by scaling. The results of that section can be applied to generalized linear and higher degree polynomial problems.

The most widely used technique to solve PEPs of degree $\ell \geq 2$ is to convert the associated matrix polynomial into a linear pencil, the process of linearization, and then solve the corresponding GEP. In Section 4 we investigate the difference between scaling before or after linearizing the matrix polynomial. Then in Section 5 we introduce a heuristic scaling strategy for PEPs that generalizes the idea of Lemonnier and Van Dooren. It is applicable to arbitrary polynomials of degree $\ell \geq 1$ and includes a weighting factor that, given some information about the magnitude of the wanted eigenvalues, can crucially improve the normwise condition numbers of eigenvalues after scaling.

Fan, Lin and Van Dooren [3] propose a transformation of variables of the form $\lambda = \alpha \mu$ for some parameter α for quadratic polynomials whose aim is to improve the backward stability of numerical methods for quadratic eigenvalue problems (QEPs) that are based on linearization. In Section 6 we extend this variable transformation to matrix polynomials of arbitrary degree $\ell \geq 2$.

Numerical examples illustrating our scaling algorithms are presented in Section 7. We conclude with practical remarks on how to put the results of this paper into practice.

Scaling routines for standard and generalized eigenvalue problems often include a preprocessing step that attempts to remove isolated eigenvalues by permutation of the matrices. This is for example implemented in the LAPACK routines `xGBAL` and `xGGBAL`. Since the permutation algorithm described in [18] can be easily adapted for matrix polynomials we will not discuss this further in this paper. But nevertheless, it is advisable to use this preprocessing step also for PEPs to reduce the problem dimension if possible.

All notation is standard. For a matrix A, we denote by $|A|$ the matrix of absolute values of the entries of A. Similarly, $|x|$ for a vector x denotes the absolute values of the entries of x. The vector of all ones is denoted by e, that is $e = \begin{bmatrix} 1, 1, \ldots, 1 \end{bmatrix}^T \in \mathbb{R}^n$.

2. Normwise and componentwise error bounds. An important tool to measure the quality of an approximate eigenpair $(\tilde{x}, \tilde{\lambda})$ of the PEP $P(\lambda)x = 0$ is its normwise backward error. With $\Delta P(\lambda) = \sum_{k=0}^{\ell} \lambda^k \Delta A_k$ it is defined for the 2-norm by

$$\eta_P(\tilde{x}, \tilde{\lambda}) := \min\{\epsilon : \ (P(\tilde{\lambda}) + \Delta P(\tilde{\lambda}))\tilde{x} = 0, \ \|\Delta A_k\|_2 \leq \epsilon \|A_k\|_2, \ k = 0 : \ell\}.$$

Tisseur [16] shows that

$$\eta_P(\tilde{x}, \tilde{\lambda}) = \frac{\|r\|_2}{\tilde{\alpha}\|\tilde{x}\|_2},$$

where $r = P(\tilde\lambda)\tilde x$ and $\tilde\alpha = \sum_{k=0}^{\ell} |\tilde\lambda|^k \|A_k\|_2$. The normwise backward error $\eta(\tilde\lambda)$ of a computed eigenvalue $\tilde\lambda$ is defined as

$$\eta_P(\tilde\lambda) = \min_{\substack{x\in\mathbb{C}^n \\ x\neq 0}} \eta_P(x,\tilde\lambda).$$

It follows immediately [16, Lemma 3] that $\eta_P(\tilde\lambda) = (\tilde\alpha\|P(\tilde\lambda)^{-1}\|_2)^{-1}$.

The sensitivity of an eigenvalue is measured by the condition number. It relates the forward error, that is the error in the computed eigenvalue $\tilde\lambda$, and the backward error $\eta_P(\tilde\lambda)$. To first order (meaning up to higher terms in the backward error) one has

$$\text{forward error} \leq \text{backward error} \times \text{condition number}. \tag{2.1}$$

The condition number of a simple, finite, nonzero eigenvalue $\lambda \neq 0$ is defined by

$$\kappa_P(\lambda) := \lim_{\epsilon\to 0} \sup\Big\{ \frac{|\Delta\lambda|}{\epsilon|\lambda|} \; : \; \big(P(\lambda+\Delta\lambda) + \Delta P(\lambda+\Delta\lambda)\big)(x+\Delta x) = 0,$$

$$\|\Delta A_k\|_2 \leq \epsilon\|A_k\|_2, \; k = 0:\ell\Big\}.$$

Let x be a right eigenvector and y be a left eigenvector associated with the eigenvalue λ of P. Then $\kappa_P(\lambda)$ is given by [16, Thm. 5]

$$\kappa_P(\lambda) = \frac{\|y\|_2\|x\|_2\alpha}{|y^*P'(\lambda)x||\lambda|}, \quad \alpha = \sum_{k=0}^{\ell} |\lambda|^k \|A_k\|_2. \tag{2.2}$$

Backward error and condition number can also be defined in a componentwise sense. The componentwise backward error of an eigenpair $(\tilde x, \tilde\lambda)$ is

$$\omega_P(\tilde x, \tilde\lambda) := \min\Big\{\epsilon : \; \big(P(\tilde\lambda) + \Delta P(\tilde\lambda)\big)\tilde x = 0; |\Delta A_k| \leq \epsilon|A_k|, \; k = 0:\ell\Big\}. \tag{2.3}$$

The componentwise condition number of a simple, finite, nonzero eigenvalue λ is defined as

$$\text{cond}_P(\lambda) := \lim_{\epsilon\to 0} \sup\Big\{\frac{|\Delta\lambda|}{\epsilon|\lambda|} \; : \; \big(P(\lambda+\Delta\lambda) + \Delta P(\lambda+\Delta\lambda)\big)(x+\Delta x) = 0,$$

$$|\Delta A_k| \leq \epsilon|A_k|, \; k = 0:\ell\Big\}. \tag{2.4}$$

The following theorem gives explicit expressions for these quantitites.

THEOREM 2.1. *The componentwise backward error of an approximate eigenpair* $(\tilde x, \tilde\lambda)$ *is given by*

$$\omega_P(\tilde x, \tilde\lambda) = \max_i \frac{|r_i|}{(\widetilde{A}|\tilde x|)_i}, \quad \widetilde{A} := \sum_{k=0}^{\ell} |\tilde\lambda|^k |A_k|, \tag{2.5}$$

where r_i denotes the ith component of the vector $P(\tilde\lambda)\tilde x$. The componentwise condition number of a simple, finite, nonzero eigenvalue λ with associated left and right eigenvectors y and x is given by

$$\text{cond}_P(\lambda) = \frac{|y|^*A|x|}{|\lambda||y^*P'(\lambda)x|}, \quad A := \sum_{k=0}^{\ell} |\lambda|^k |A_k|. \tag{2.6}$$

Proof. The proof is a slight modification of the proofs of [5, Thm. 3.1 and 3.2] along the lines of the proof of [16, Thm. 1]. □

Surveys of componentwise error analysis are contained in [6, 7]. The componentwise backward error and componentwise condition number are invariant under multiplication of $P(\lambda)$ from the left and the right with nonsingular diagonal matrices. In the next section we will use this property to characterize optimally scaled eigenvalue problems.

3. Optimal scalings. In this section we introduce the notion of an optimal scaling with respect to a certain eigenvalue and give characterizations of it.

Ultimately, we are interested in computing eigenvalues to as many digits as possible. Hence, we would like to find a scaling that leads to small forward errors. If we assume that we use a backward stable algorithm, that is the backward error is only a small multiple of the machine precision, then it follows from (2.1) that we can hope to compute an eigenvalue to many digits of accuracy by finding a scaling that minimizes the condition number.

In the following we define what we mean by a scaling of a matrix polynomial $P(\lambda)$.

DEFINITION 3.1. *Let $P(\lambda) \in \mathbb{C}^{n \times n}$ be a matrix polynomial. A scaling of $P(\lambda)$ is the matrix polynomial $D_1 P(\lambda) D_2$, where $D_1, D_2 \in \mathcal{D}_n$.*

It is immediately clear that the eigenvalues of a matrix polynomial $P(\lambda)$ are invariant under scaling. Furthermore, if (y, x, λ) is an eigentriplet of $P(\lambda)$ with eigenvalue λ and left and right eigenvector y and x, respectively then an eigentriplet of the scaling $D_1 P(\lambda) D_2$ is $(D_1^{-*}y, D_2^{-1}x, \lambda)$.

The following definition defines an optimal scaling of $P(\lambda)$ with respect to a given eigenvalue λ in terms of minimizing the condition number of λ.

DEFINITION 3.2. *Let λ be a simple, finite, nonzero eigenvalue of the matrix polynomial $P(\lambda)$. We call $P(\lambda)$ optimally scaled with respect to λ if*

$$\kappa_P(\lambda) = \inf_{D_1, D_2 \in \mathcal{D}_n} \kappa_{D_1 P D_2}(\lambda).$$

This definition of optimal scaling depends on the eigenvalue λ. We cannot expect that an optimal scaling for one eigenvalue also gives an optimal scaling for another eigenvalue. The following theorem states that a PEP is almost optimally scaled with respect to an eigenvalue λ, if the componentwise and normwise condition numbers of λ are close to each other. Furthermore, it gives explicit expressions for scaling matrices $D_1, D_2 \in \mathcal{D}_n$ that achieve an almost optimal scaling.

THEOREM 3.3. *Let λ be a simple, finite, nonzero eigenvalue of an $n \times n$ matrix polynomial $P(\lambda)$ with associated left and right eigenvectors y and x, respectively. Then*

$$\frac{1}{\sqrt{n}} \text{cond}_P(\lambda) \leq \inf_{D_1, D_2 \in \mathcal{D}_n} \kappa_{D_1 P D_2}(\lambda) \leq n \, \text{cond}_P(\lambda). \tag{3.1}$$

Moreover, if all the entries of y and x are nonzero, then for

$$D_1 = \text{diag}(|y|), \quad D_2 = \text{diag}(|x|).$$

we have

$$\kappa_{D_1 P D_2}(\lambda) \leq n \, \text{cond}_P(\lambda). \tag{3.2}$$

Proof. Let $A := \sum_{k=0}^{\ell} |\lambda|^k |A_k|$ and $\alpha := \sum_{k=0}^{\ell} |\lambda|^k \|A_k\|_2$. Using $\| |B| \|_2 \leq \sqrt{n} \|B\|_2$ [7, Lemma 6.6] for any matrix $B \in \mathbb{C}^{n \times n}$ the lower bound follows from

$$\text{cond}_P(\lambda) = \frac{|y|^* A |x|}{|\lambda| |y^* P'(\lambda) x|} \leq \frac{\|y\|_2 \|x\|_2 \|A\|_2}{|\lambda| |y^* P'(\lambda) x|} \leq \frac{\sqrt{n}\alpha \|y\|_2 \|x\|_2}{|\lambda| |y^* P'(\lambda) x|} = \sqrt{n}\kappa_P(\lambda)$$

and the fact that the componentwise condition number is invariant under diagonal scaling. For $\epsilon > 0$ define the vectors $\tilde{y}$ and $\tilde{x}$ by

$$\tilde{y}_i = \left\{ \begin{array}{ll} y_i, & y_i \neq 0 \\ \epsilon, & y_i = 0 \end{array} \right. \qquad \tilde{x}_i = \left\{ \begin{array}{ll} x_i, & x_i \neq 0 \\ \epsilon, & x_i = 0 \end{array} \right.$$

and consider the diagonal matrices

$$D_1 = \text{diag}(|\tilde{y}|), \quad D_2 = \text{diag}(|\tilde{x}|).$$

Using $\|B\|_2 \leq e^* |B| e$ for any matrix $B \in \mathbb{C}^{n \times n}$ [7, Table 6.2] we have

$$\kappa_{D_1 P D_2}(\lambda) = \frac{\|D_1^{-1} y\|_2 \|D_2^{-1} x\|_2 (\sum_{k=0}^{\ell} |\lambda|^k \|D_1 A_k D_2\|_2)}{|\lambda| |y^* P'(\lambda) x|} \leq \frac{n(\sum_{k=0}^{\ell} |\lambda|^k e^* |D_1 A_k D_2| e)}{|\lambda| |y^* P'(\lambda) x|}$$

$$= \frac{n(\sum_{k=0}^{\ell} |\lambda|^k \cdot |\tilde{y}|^* \cdot |A_k| \cdot |\tilde{x}|)}{|\lambda| |y^* P'(\lambda) x|} \longrightarrow n \, \text{cond}_P(\lambda) \text{ as } \epsilon \to 0. \tag{3.3}$$

The upper bounds in (3.1) and (3.2) follow immediately. $\square$

Theorem 3.3 is restricted to finite and nonzero eigenvalues. Assume that $\lambda = 0$ is an eigenvalue. Then we have to replace relative componentwise and normwise condition numbers by the absolute condition numbers

$$\kappa_P^{(a)}(\lambda) = \frac{\|y\|_2 \|x\|_2 \alpha}{|y^* P'(\lambda) x|}, \quad \text{cond}_P^{(a)}(\lambda) = \frac{|y|^* A |x|}{|y^* P'(\lambda) x|}.$$

With these condition numbers Theorem 3.3 is also valid for zero eigenvalues. If $P(\lambda)$ has an infinite eigenvalue the reversal rev $P(\lambda) := \lambda^\ell P(1/\lambda)$ has a zero eigenvalue and we can apply Theorem 3.3 using absolute condition numbers to rev $P(\lambda)$.

While Theorem 3.3 applies to generalized linear and polynomial problems it does not immediately apply to standard problems of the form $Ax = \lambda x$. The crucial difference is that for standard eigenvalue problems we assume the right-hand side identity matrix to be fixed and only allow scalings of the form $D^{-1} A D$ that leave the identity unchanged. However, if λ is an eigenvalue of A with associated left and right eigenvectors y and x that have nonzero entries we can still define $D_1 = \text{diag}(|y|)$ and $D_2 = \text{diag}(|x|)$ to obtain the generalized eigenvalue problem

$$D_1 A D_2 v = \lambda D_1 D_2 v, \tag{3.4}$$

where $x = D_2 v$. Since $D_1 D_2$ has positive diagonal entries there exists $\hat{D}$ such that $\hat{D}^2 = D_1 D_2$. We then obtain from (3.4) the standard eigenvalue problem

$$\hat{D}^{-1} D_1 A D_2 \hat{D}^{-1} \tilde{x} = \lambda \tilde{x}, \tag{3.5}$$

where $\tilde{x} = \hat{D} D_2^{-1} x$ and $|\tilde{x}| = \left[\sqrt{|y_1||x_1|}, \ldots, \sqrt{|y_n||x_n|} \right]^T$. For the scaled left eigenvector $\tilde{y}$ we have $\tilde{y} = \hat{D} D_1^{-1} y$ and $|\tilde{y}| = |\tilde{x}|$. If we define the normwise condition

number $k_A(\lambda)$ and the componentwise condition number $c_A(\lambda)$ for the eigenvalue λ of a standard eigenvalue problem by[1]

$$k_A(\lambda) = \frac{\|y\|_2 \|x\|_2}{|\lambda||y^T x|}, \quad c_A(\lambda) = \frac{|y|^T |x|}{|\lambda||y^T x|}$$

it follows for the scaling $D^{-1}AD$, where $D = D_2\hat{D}^{-1} = D_1^{-1}\hat{D}$ that

$$c_A(\lambda) = k_{D^{-1}AD}(\lambda).$$

But this scaling is not always useful as $\|D^{-1}AD\|_2$ can become large if x or y contain tiny entries.

There is a special case in which the scaling (3.5) also minimizes $\|D^{-1}AD\|_2$. If λ is the Perron root of an irreducible and non-negative matrix A the corresponding left and right eigenvectors y and x can be chosen to have positive entries. After scaling by D as described above we have $k_{D^{-1}AD}(\lambda) = 1$ and $\|D^{-1}AD\|_2 = \lambda$. This was investigated by Chen and Demmel in [2] who proposed a weighted balancing which is identical to the scaling described above for non-negative and irreducible A.

Theorem 3.3 gives us an easy way to check whether a matrix polynomial P is nearly optimally scaled with respect to an eigenvalue λ. We only need to compute the ratio

$$\frac{\kappa_P(\lambda)}{\mathrm{cond}_P(\lambda)} = \frac{\|y\|_2 \|x\|_2 \sum_{k=0}^{\ell} |\lambda|^k \|A_k\|_2}{|y|^* \left(\sum_{k=0}^{\ell} |\lambda|^k |A_k| \right) |x|} \tag{3.6}$$

after computing the eigenvalues and eigenvectors. If an eigensolver already returns normwise condition numbers this is only little extra effort. If $\frac{\kappa_P(\lambda)}{\mathrm{cond}_P(\lambda)} \gg n$ the eigensolver can give a warning to the user that the problem is badly scaled and that the error in the computed eigenvalue λ is likely to become smaller by rescaling P. Furthermore, from Theorem 3.3 it follows that a polynomial is nearly optimally scaled if the entries of the left and right eigenvectors have equal magnitude. This motivates a heuristic scaling algorithm, which is discussed in Section 5.

At the end we would like to emphasize that a diagonal scaling which improves the condition numbers of the eigenvalues needs not necessarily be a good scaling for eigenvectors. An example is the generalized linear eigenvalue problem $L(\lambda)x = 0$, where

$$L(\lambda) = \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} + \begin{bmatrix} 0 & 1 + 2 \cdot 10^{-8} & 2 \\ 2 & 10^{-8} & 1 \\ 1 & 1 + 10^{-8} & -1 \end{bmatrix}$$

One eigenvalue is $\lambda = 1$ with associated right eigenvector $x = \begin{bmatrix} 1 & -1 & 10^{-8} \end{bmatrix}^T$ and left eigenvector $y = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & -1 \end{bmatrix}^T$. The condition number[2] of the eigenvector x before scaling is approximately 33.1. After scaling with $D_1 = \mathrm{diag}(|y|)$ and $D_2 = \mathrm{diag}(|x|)$ it increases to $1.70 \cdot 10^9$. For the corresponding eigenvalue $\lambda = 1$ we have $\kappa_L(1) \approx 21.8$ and after scaling $\kappa_{D_1 L D_2}(1) \approx 19.6$. However, in most of our experiments we could not observe an increase of the eigenvector condition number after scaling.

---

[1] Choose $E = I$, $F = 0$ and $B = I$ in the Theorems 2.5 and 3.2 of [5].

[2] The condition number of the eigenvector was computed using Theorem 2.7 from [5] with the normalization vector $g = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$.

4. Scalings and linearizations. The standard way to solve the PEP (1.1) of degree $\ell \geq 2$ is to convert $P(\lambda)$ into a linear pencil

$$L(\lambda) = \lambda X + Y$$

having the same spectrum as $P(\lambda)$ and then solve the eigenproblem for L. Formally, $L(\lambda)$ is a linearization if

$$E(\lambda)L(\lambda)F(\lambda) = \begin{bmatrix} P(\lambda) & 0 \\ 0 & I_{(\ell-1)n} \end{bmatrix}$$

for some unimodular $E(\lambda)$ and $F(\lambda)$ [4, sect. 7.2]. For example,

$$C_1(\lambda) = \lambda \begin{bmatrix} A_\ell & 0 & \cdots & 0 \\ 0 & I_n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_n \end{bmatrix} + \begin{bmatrix} A_{\ell-1} & A_{\ell-2} & \cdots & A_0 \\ -I_n & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -I_n & 0 \end{bmatrix} \tag{4.1}$$

is a linearization of $P(\lambda)$, called the first companion form. In [12] Mackey, Mackey, Mehl and Mehrmann identified two vector spaces of pencils that are potential linearizations of $P(\lambda)$. Let $\Lambda := [\lambda^{\ell-1}, \lambda^{\ell-2}, \ldots, 1]^T$. Then these spaces are defined by

$$\mathbb{L}_1(P) = \left\{ L(\lambda) : L(\lambda)(\Lambda \otimes I_n) = v \otimes P(\lambda), \ v \in \mathbb{C}^\ell \right\},$$
$$\mathbb{L}_2(P) = \left\{ L(\lambda) : (\Lambda^T \otimes I_n)L(\lambda) = \tilde{v}^T \otimes P(\lambda), \ \tilde{v} \in \mathbb{C}^\ell \right\}.$$

The first companion linearization belongs to $\mathbb{L}_1(P)$ with $v = e_1$. Furthermore, the pencils in $\mathbb{L}_1(P)$ and $\mathbb{L}_2(P)$ that are not linearizations form a closed nowhere dense subset of measure zero in these spaces [12, Thm. 4.7].

Another important space of potential linearizations is given by

$$\mathbb{DL}(P) := \mathbb{L}_1(P) \cap \mathbb{L}_2(P).$$

In [12, Thm. 5.3] it is shown that each pencil $L(\lambda) \in \mathbb{DL}(P)$ is uniquely defined by a vector $v \in \mathbb{C}^\ell$ such that

$$L(\lambda)(\Lambda \otimes I_n) = v \otimes P(\lambda), \quad (\Lambda^T \otimes I_n)L(\lambda) = v^T \otimes P(\lambda).$$

There is a well defined relationship between the eigenvectors of linearizations $L(\lambda) \in \mathbb{DL}(P)$ and eigenvectors of $P(\lambda)$, namely for finite eigenvalues λ x is a right eigenvector of $P(\lambda)$ if and only if $\Lambda \otimes x$ is a right eigenvector of $L(\lambda)$ and y is a left eigenvector of $P(\lambda)$ if and only if $\overline{\Lambda} \otimes y$ is a left eigenvector of $L(\lambda)$ [12, Thm. 3.8 and Thm. 3.14].

A simple observation is that scaling $P(\lambda)$ leads to a scaling of $L(\lambda)$ within the same space of potential linearizations.

LEMMA 4.1. *Let $L(\lambda) \in \mathbb{S}(P)$ with vector v, where $\mathbb{S}(P) = \mathbb{L}_1(P)$, $\mathbb{L}_2(P)$, or $\mathbb{DL}(P)$. Then $(I_n \otimes D_1)L(\lambda)(I_n \otimes D_2)$ is in $\mathbb{S}(D_1PD_2)$ with the same vector v, where $D_1, D_2 \in \mathbb{C}^{n\times n}$ are nonsingular matrices.*

Proof. The statements follow immediately from the identities

$$(I \otimes D_1)L(\lambda)(I \otimes D_2)(\Lambda \otimes I_n) = v \otimes D_1 P(\lambda) D_2$$

and

$$(\Lambda^T \otimes I_n)(I \otimes D_1)L(\lambda)(I \otimes D_2) = \tilde{v}^T \otimes D_1 P(\lambda) D_2$$

for matrices $D_1, D_2 \in \mathbb{C}^{n \times n}$. $\square$

Hence, if we solve a PEP by a linearization in $\mathbb{L}_1(P)$, $\mathbb{L}_2(P)$ or $\mathbb{DL}(P)$ scaling of the original polynomial $P(\lambda)$ with matrices D_1 and D_2 is just a special scaling of the linearization $L(\lambda)$ with scaling matrices $(I \otimes D_1)$ and $(I \otimes D_2)$. If preserving structure of the linearization is not an issue we can scale the linearization $L(\lambda)$ directly with diagonal scaling matrices $\widetilde{D}_1$ and $\widetilde{D}_2$ that have $2\ell n$ free parameters compared to the $2n$ free parameters in D_1 and D_2. The following theorem gives a worst case bound on the ratio between the optimal condition numbers with the two different scaling strategies (i.e. scaling and then linearizing or linearizing and then scaling).

THEOREM 4.2. *Let λ be a simple finite eigenvalue of P and let $L(\lambda) \in \mathbb{DL}(P)$ with vector v. Then*

$$\inf_{D_1, D_2 \in \mathcal{D}_n} \kappa_{\widetilde{L}}(\lambda; v; D_1 P D_2) \leq \begin{cases} \ell^{1/2} n^{3/2} \left(\frac{|\lambda|^{2\ell} - 1}{|\lambda|^2 - 1} \right) \displaystyle\inf_{\tilde{D}_1, \tilde{D}_2 \in \mathcal{D}_{\ell n}} \kappa_{\tilde{D}_1 L \tilde{D}_2}(\lambda) & \text{for } |\lambda| \geq 1, \\[2ex] \frac{\ell^{1/2} n^{3/2}}{|\lambda|^{2(\ell-1)}} \left(\frac{|\lambda|^{2\ell} - 1}{|\lambda|^2 - 1} \right) \displaystyle\inf_{\tilde{D}_1, \tilde{D}_2 \in \mathcal{D}_{\ell n}} \kappa_{\tilde{D}_1 L \tilde{D}_2}(\lambda) & \text{for } |\lambda| < 1, \end{cases}$$

where $\kappa_{\widetilde{L}}(\lambda; v; D_1 P D_2)$ is the condition number of λ for the linearization $\widetilde{L}(\lambda) \in \mathbb{DL}(D_1 P D_2)$ with vector v.

Proof. Let y and x be left and right eigenvectors of $P(\lambda)$ associated with the eigenvalue λ. Since $L(\lambda) = \lambda X + Y \in \mathbb{DL}(P)$, its left and right eigenvectors associated with λ are $\overline{\Lambda} \otimes y$ and $\Lambda \otimes x$. Assume that y and x have no zero entries. The case of zero entries follows by a limit process similar to that in the proof of Theorem 3.3. Define $D_1 = \text{diag}(|y|)$ and $D_2 = \text{diag}(|x|)$. Since $\|\overline{\Lambda} \otimes (D_1^{-1} y)\|_2 = \|\Lambda\|_2 \|D_1^{-1} y\|_2$ and $\|\Lambda \otimes (D_2^{-1} x)\|_2 = \|\Lambda\|_2 \|D_2^{-1} x\|_2$ we have

$$\kappa_{\widetilde{L}}(\lambda; v; D_1 P D_2) =$$
$$\frac{\|\Lambda\|_2^2 \|D_1^{-1} y\|_2 \|D_2^{-1} x\|_2 (|\lambda| \|(I \otimes D_1) X (I \otimes D_2)\|_2 + \|(I \otimes D_1) Y (I \otimes D_2)\|_2)}{|\lambda| |(\overline{\Lambda} \otimes y)^* X (\Lambda \otimes x)|}$$

and therefore by using $\|B\|_2 \leq e^* |B| e$ for any $B \in \mathbb{C}^{n \times n}$

$$\kappa_{\widetilde{L}}(\lambda; v; D_1 P D_2) \leq \frac{\|\Lambda\|_2^2 n \hat{e}^* (|\lambda| |(I \otimes D_1) X (I \otimes D_2)| + |(I \otimes D_1) Y (I \otimes D_2)|) \hat{e}}{|\lambda| |(\overline{\Lambda} \otimes y)^* X (\Lambda \otimes x)|}$$

for $\hat{e} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^T \in \mathbb{R}^{\ell n}$. Assume that $|\lambda| \geq 1$. Since componentwise $\hat{e} \leq |\Lambda| \otimes e = |\overline{\Lambda}| \otimes e$ and

$$(|\overline{\Lambda}| \otimes e)^* (I \otimes D_1) = |\overline{\Lambda} \otimes y|^*, \quad (I \otimes D_2)(|\Lambda| \otimes e) = |\Lambda \otimes x|,$$

we obtain

$$\kappa_{\widetilde{L}}(\lambda; v; D_1 P D_2) \leq n \frac{\|\Lambda\|_2^2 |\overline{\Lambda} \otimes y|^* (|\lambda| |X| + |Y|) |\Lambda \otimes x|}{|\lambda| |(\overline{\Lambda} \otimes y)^* X (\Lambda \otimes x)|} = n \|\Lambda\|_2^2 \text{cond}_L(\lambda). \quad (4.2)$$

It holds that

$$\|\Lambda\|_2^2 = \left(\frac{|\lambda|^{2\ell} - 1}{|\lambda|^2 - 1} \right). \quad (4.3)$$

Furthermore, from Theorem 3.3 we know that

$$\frac{1}{\sqrt{\ell n}} \text{cond}_L(\lambda) \leq \inf_{\tilde{D}_1, \tilde{D}_2 \in \mathcal{D}_{\ell n}} \kappa_{\tilde{D}_1 L \tilde{D}_2}(\lambda). \tag{4.4}$$

Combining (4.2), (4.3) and (4.4) the proof for the case $|\lambda| \geq 1$ follows. The proof for $|\lambda| < 1$ is similar. The only essential difference is that now componentwise $\hat{e} \leq \frac{|\Lambda|}{|\lambda|^{\ell-1}} \otimes e$. $\square$

Theorem 4.2 suggests that for eigenvalues that are large or small in magnitude first linearizing and then scaling can in the worst case be significantly better than first scaling and then linearizing. However, if we first linearize and then scale the special structure of the linearization is lost.

How sharp are these bounds? In the following we discuss the case $|\lambda| \geq 1$. For the case $|\lambda| < 1$ analogous arguments can be used. Consider the QEP $Q(\lambda) = \lambda^2 A + \lambda B + C$, where

$$A = \begin{bmatrix} -0.6 & -0.1 \\ 2 & 0.1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -0.1 \\ 0.6 & -0.8 \end{bmatrix} \quad C = \begin{bmatrix} 3 \cdot 10^7 & 7 \cdot 10^7 \\ -1 \cdot 10^8 & 1.6 \cdot 10^8 \end{bmatrix} \tag{4.5}$$

and its linearization in $\mathbb{DL}(Q)$

$$L(\lambda) = \lambda X + Y := \lambda \begin{bmatrix} A & 0 \\ 0 & -C \end{bmatrix} + \begin{bmatrix} B & C \\ C & 0 \end{bmatrix},$$

which corresponds to the vector $v = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$. One eigenvalue of this pencil is $\lambda \approx 4.105 \cdot 10^4$. If we first scale Q using the left and right eigenvector associated with λ and then linearize, this eigenvalue has the condition number $1.2 \cdot 10^9$ for the linearization. If we first linearize the QEP and then scale the pencil $L(\lambda)$ using the left and right eigenvector of λ for the linearization this eigenvalue has the condition number 5.2. The ratio between the condition numbers is in magnitude what we would expect from applying Theorem 4.2.

However, Theorem 4.2 can be a large overestimate. Assume that $P(\lambda)$ is already almost optimally scaled in the sense of Theorem 3.3, that is $|y| = |x| = e$ for the left and right eigenvectors y and x associated with the simple finite eigenvalue λ of P. Let $L(\lambda) = \lambda X + Y$ be a linearization of P and let D_1 and D_2 be scaling matrices for L such that $|D_1^{-*}\tilde{y}| = |D_2^{-1}\tilde{x}| = e$ for the left and right eigenvectors $\tilde{y}$ and $\tilde{x}$ of L associated with the eigenvalue λ. The ratio of the condition numbers of the eigenvalue λ for the two pencils L and $D_1 L D_2$ is given by

$$\frac{\kappa_L(\lambda)}{\kappa_{D_1 L D_2}(\lambda)} = \frac{\|\tilde{x}\|_2 \|\tilde{y}\|_2}{\|D_1^{-*}\tilde{y}\|_2 \|D_2^{-1}\tilde{x}\|_2} \frac{|\lambda| \|X\|_2 + \|Y\|_2}{|\lambda| \|D_1 X D_2\|_2 + \|D_1 Y D_2\|_2}. \tag{4.6}$$

If $L(\lambda) \in \mathbb{DL}(P)$ (4.6) simplifies to

$$\frac{\kappa_L(\lambda)}{\kappa_{D_1 L D_2}(\lambda)} = \frac{1}{\ell} \left(\frac{|\lambda|^{2\ell} - 1}{|\lambda|^2 - 1} \right) \frac{|\lambda| \|X\|_2 + \|Y\|_2}{|\lambda| \|D_1 X D_2\|_2 + \|D_1 Y D_2\|_2}$$

since $|\tilde{x}| = |\tilde{y}| = |\Lambda \otimes e|$. This shows that for $|\lambda| > 1$ the upper bound in Theorem 4.2 can only be attained if

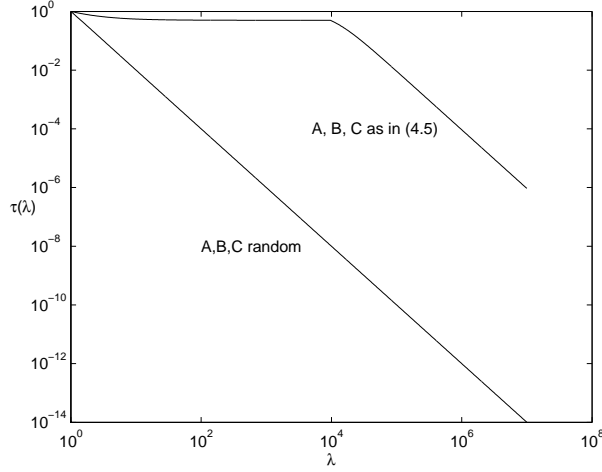$$\frac{|\lambda| \|X\|_2 + \|Y\|_2}{|\lambda| \|D_1 X D_2\|_2 + \|D_1 Y D_2\|_2} =: \tau(\lambda) \tag{4.7}$$

FIG. 4.1. *The function $\tau(\lambda)$ for a large range of values in the case of a random 2×2 QEP and the QEP from (4.5).*

is approximately constant in the range of the eigenvalues that we are interested in. For $L(\lambda) \in \mathbb{DL}(P)$ the matrices D_1 and D_2 are given as

$$D_1 = D_2 = \begin{bmatrix} |\lambda|^{\ell-1}I & & \\ & \ddots & \\ & & I \end{bmatrix} = \operatorname{diag}(|\Lambda|) \otimes I.$$

It follows that for $|\lambda|$ large enough

$$\tau(\lambda) \sim \gamma|\lambda|^{2-2\ell} \tag{4.8}$$

for some constant $\gamma > 0$ and therefore

$$\frac{\kappa_L(\lambda)}{\kappa_{D_1 L D_2}(\lambda)} \sim \frac{\gamma}{\ell}$$

in that case.

Especially, if the upper left $n \times n$ block of X is in norm comparable or larger than the other $n \times n$ subblocks of X we expect a good agreement of the asymptotic in (4.8) for all $|\lambda| > 1$, where γ is not much larger than 1. Only if the $n \times n$ subblocks of X and Y are of widely varying norm it is possible that $\tau(\lambda)$ is approximately constant for a large range of values of λ leading to the worst case bound in (4.2) being attained.

The situation is demonstrated in Figure 4.1. For a random 2×2 QEP $\tau(\lambda)$ decays like $\gamma|\lambda|^{-2}$, where $\gamma \approx 1$. For the QEP from (4.5) the function $\tau(\lambda)$ is almost constant for a long time leading to the worst case bound of Theorem 4.2 being attained in these range of values. Then at about 10^4 it starts decaying like $\gamma|\lambda|^{-2}$, where this time γ is in the order of 10^8.

One of the most frequently used linearizations for unstructured problems is the companion form (4.1). Unfortunately, we cannot immediately apply the previous results to it since the companion form is not in $\mathbb{DL}(P)$ but only in $\mathbb{L}_1(P)$. However, we can still compare the ratio in (4.6). Consider the QEP $Q(\lambda) = \lambda^2 A + \lambda B + C$.

The companion linearization takes the form

$$C_1(\lambda) = \lambda \begin{bmatrix} A & \\ & I \end{bmatrix} + \begin{bmatrix} B & C \\ -I & 0 \end{bmatrix}.$$

We assume that for the left and right eigenvectors y and x associated with the eigenvalue λ of Q we have $|y| = |x| = e$. Furthermore, let D_1 and D_2 again be chosen such that $|D_1^{-*}\tilde{y}| = |D_2^{-1}\tilde{x}| = e$, where $\tilde{y}$ and $\tilde{x}$ are the corresponding left and right eigenvectors for the eigenvalue λ of the companion linearization $C_1(\lambda) = \lambda X + Y$. The relationship between the eigenvectors of C_1 and the eigenvectors of P associated with a finite nonzero eigenvalue λ is given by

$$\tilde{x} = \Lambda \otimes x, \quad \tilde{y} = \begin{bmatrix} y \\ -\frac{1}{\lambda}C^*y \end{bmatrix}.$$

The formula for the left eigenvector is a consequence of [8, Theorem 3.2]. It follows that

$$\frac{\kappa_{C_1}(\lambda)}{\kappa_{\tilde{D}_1 C_1 \tilde{D}_2}(\lambda)} = \frac{1}{2n^{1/2}} \left(\frac{|\lambda|^4 - 1}{|\lambda|^2 - 1}\right)^{1/2} \left(n + \frac{1}{|\lambda|^2}\|C^*y\|_2^2\right)^{1/2} \tau(\lambda).$$

If $|\lambda| \gg 1$ this simplifies to

$$\frac{\kappa_{C_1}(\lambda)}{\kappa_{\tilde{D}_1 C_1 \tilde{D}_2}(\lambda)} \approx \frac{1}{2}|\lambda|\tau(\lambda),$$

which differs by a factor of $|\lambda|$ from the corresponding case using a $\mathbb{DL}(P)$ linearization. Asymptotically, we have

$$\tau(\lambda) \sim \gamma|\lambda|^{-1}, \quad |\lambda| \gg 1$$

for some factor γ and therefore $\frac{\kappa_{C_1}(\lambda)}{\kappa_{\tilde{D}_1 C_1 \tilde{D}_2}(\lambda)} \sim \frac{\gamma}{2}$, where again we expect this asymptotic to hold approximately for all $|\lambda| > 1$ with a value of γ that is not much larger than 1 if the $n \times n$ subblocks of X and Y do not differ too widely in norm.

5. A heuristic scaling strategy. For standard eigenvalue problems the motivation of scaling algorithms is based on the observation that in floating point arithmetic computed eigenvalues of a matrix A are at least perturbed by an amount of the order of $\epsilon_{mach}\|A\|$. Hence, by reducing $\|A\|$ one hopes to reduce the inaccuracies in the computed eigenvalues.

One way of minimizing $\|A\|$ is to find a nonsingular diagonal matrix D such that the rows and columns of A are balanced in the sense that

$$\|D^{-1}ADe_i\| = \|e_i^* D^{-1}AD\|, \quad i = 1, \ldots, n. \tag{5.1}$$

Osborne [14] shows that if A is irreducible and $\|\cdot\|$ is the 2-norm in (5.1) then for this D it holds that

$$\|D^{-1}AD\|_F = \inf_{\hat{D}\in\mathcal{D}_n} \|\hat{D}^{-1}A\hat{D}\|_F.$$

A routine that attempts to find a matrix D that balances the row and column norms of A is built into LAPACK under the name xGEBAL. It uses the 1-norm in the balancing condition (5.1). A description of the underlying algorithm is contained in [15].

For generalized eigenvalue problems $Ax = \lambda Bx$ Ward [18] proposes to find non-singular diagonal scaling matrices D_1 and D_2 such that the elements of the scaled matrices D_1AD_2 and D_1BD_2 would have absolute values close to unity. Then the relative perturbations in the matrix elements caused by computational errors would be of similar magnitude. To achieve this Ward proposes to minimize the function

$$\sum_{i,j=1}^{n} (r_i + c_j + \log|A_{ij}|)^2 + (r_i + c_j + \log|B_{ij}|)^2,$$

where the r_i and c_j are the logarithms of the absolute values of the diagonal entries of D_1 and D_2. The scaling by Ward can fail if the matrices A and B contain tiny entries that are not due to bad scaling [10, Example 2.16].

A different strategy for generalized eigenvalue problems is proposed by Lemonnier and Van Dooren [11]. By introducing the notion of generalized normal pencils they motivate a scaling strategy that aims to find nonsingular diagonal matrices D_1 and D_2 such that

$$\|D_1AD_2e_j\|_2^2 + \|D_1BD_2e_j\|_2^2 = \|e_i^*D_1AD_2\|_2^2 + \|e_i^*D_1BD_2\|_2^2 = 1, \quad i,j = 1,\ldots,n. \tag{5.2}$$

The scaling condition (5.2) can be generalized in a straight forward way to matrix polynomials of higher degree by

$$\sum_{k=0}^{\ell} \omega^{2k} \|D_1A_kD_2e_i\|_2^2 = 1, \quad \sum_{k=0}^{\ell} \omega^{2k} \|e_j^*D_1A_kD_2\|_2^2 = 1, \quad i,j = 1,\ldots,n \tag{5.3}$$

for some $\omega > 0$ that is chosen in magnitude close to the wanted eigenvalues. The intuitive idea behind (5.3) is to balance rows and columns of the coefficient matrices A_k while taking into account the weighting of the coefficient matrices induced by the eigenvalue paramter, that is for very large eigenvalues the rows and columns of A_ℓ dominate and for very small eigenvalues the rows and columns of A_0 dominate. This also reflects the result of Theorem (3.3) that the optimal scaling matrices are dependent on the wanted eigenvalue. In Section 7 we show that including the estimate ω can greatly improve the results of scaling.

In [11] Lemonnier and Van Dooren introduced a linearly convergent iteration to obtain matrices D_1 and D_2 consisting of powers of 2 that approximately satisfy (5.2). The idea in their code is to alternatively update D_1 and D_2 by first normalizing all rows of $\begin{bmatrix} A & B \end{bmatrix}$ and then all columns of $\begin{bmatrix} A \\ B \end{bmatrix}$. The algorithm repeats this operation until (5.2) is approximately satisfied. This iteration can easily be extended to weighted scaling of matrix polynomials. This is done in Alg. 1. The main difference to the Matlab code in [11] is the definition of the variable M in line 6 that now accommodates matrix polynomials and the weighting parameter ω.

If we do not have any estimate for the magnitude of the wanted eigenvalues a possible choice is to set $\omega = 1$ in (5.3). In that case all coefficient matrices have the same weight in that condition.

6. Transformations of the eigenvalue parameter. In the previous sections we investigated how diagonal scaling of $P(\lambda)$ by multiplication of $P(\lambda)$ with left and right scaling matrices $D_1, D_2 \in \mathcal{D}_n$ can improve the condition number of the eigenvalues. In this section we consider scaling a PEP by transforming the eigenvalue

---

Algorithm 1 Diagonal scaling of $P(\lambda) = \lambda^\ell A_\ell + \cdots + \lambda A_1 + A_0$.

Require: $A_0, \ldots, A_\ell \in \mathbb{C}^{n \times n}$, $\omega > 0$.
1: $M \leftarrow \sum_{k=0}^{\ell} |\lambda|^{2k} |A_k|^{.2}$, $D_1 \leftarrow I$, $D_2 \leftarrow I$ ($|A_k|^{.2}$ is entry-wise square)
2: $maxiter \leftarrow 5$
3: **for** $iter = 1$ to $maxiter$ **do**
4:    $emax \leftarrow 0$, $emin \leftarrow 0$
5:    **for** $i = 1$ to n **do**
6:       $d \leftarrow \sum_{j=0}^{n} M(i,j)$, $e \leftarrow -\text{round}(\frac{1}{2}\log_2 d)$
7:       $M(i,:) \leftarrow 2^{2e} \cdot M(i,:)$, $D_1(i,i) \leftarrow 2^e \cdot D_1(i,i)$
8:       $emax \leftarrow \max(emax, e)$, $emin \leftarrow \min(emin, e)$
9:    **end for**
10:   **for** $i = 1$ to n **do**
11:      $d \leftarrow \sum_{j=0}^{n} M(j,i)$, $e \leftarrow -\text{round}(\frac{1}{2}\log_2 d)$
12:      $M(:,i) \leftarrow 2^{2e} \cdot M(:,i)$, $D_2(i,i) \leftarrow 2^e \cdot D_2(i,i)$
13:      $emax \leftarrow \max(emax, e)$, $emin \leftarrow \min(emin, e)$
14:   **end for**
15:   **if** $emax \leq emin + 2$ **then**
16:      **BREAK**
17:   **end if**
18: **end for**
19: **return** D_1, D_2

---

parameter λ. This was proposed by Fan, Lin and Van Dooren for quadratics in [3] (see also [8]). Let $Q(\lambda) := \lambda^2 A_2 + \lambda A_1 + A_0$. Define the quadratic polynomial $\widetilde{Q}(\mu) = \mu^2 \widetilde{A}_2 + \mu \widetilde{A}_1 + \widetilde{A}_0$ as

$$\widetilde{Q}(\mu) := \beta Q(\alpha\mu) = \beta\mu^2\alpha^2 A_2 + \beta\mu\alpha A_1 + \beta A_0.$$

The parameters $\beta > 0$ and $\alpha > 0$ are chosen such that the 2-norms of the new coefficient matrices $\widetilde{A}_2 := \beta\alpha^2 A_2$, $\widetilde{A}_1 := \beta\alpha A_1$ and $\widetilde{A}_0 := \beta A_0$ are as close to 1 as possible, that is, we need to solve

$$\min_{\alpha>0,\beta>0} \max\{|\beta\alpha^2\|A_2\|_2 - 1|, |\beta\alpha\|A_1\|_2 - 1|, |\beta\|A_0\|_2 - 1|\}. \tag{6.1}$$

It is shown in [3] that the unique solution of (6.1) is given by

$$\alpha = \left(\frac{\|A_0\|_2}{\|A_2\|_2}\right)^{\frac{1}{2}}, \quad \beta = \frac{2}{\|A_0\|_2 + \|A_1\|_2\alpha}.$$

Hence, after scaling we have $\|\widetilde{A}_0\|_2 = \|\widetilde{A}_2\|_2$. The motivation behind this scaling is that solving a QEP by applying a backward stable algorithm to solve (4.1) is backward stable if $\|A_0\|_2 = \|A_1\|_2 = \|A_2\|_2$ [16, Thm. 7]. For matrix polynomials of arbitrary degree ℓ it is shown in [9] that with

$$\rho := \frac{\max_i \|A_i\|_2}{\min(\|A_0\|_2, \|A_\ell\|_2)} \geq 1$$

one has

$$\frac{2\sqrt{\ell}}{\ell+1}\frac{1}{\rho} \leq \frac{\inf_v \kappa_L(\lambda; v; P)}{\kappa_P(\lambda)} \leq \ell^2\rho,$$

where $\kappa_L(\lambda; v; P)$ is the condition number of the eigenvalue λ for the linearization $L(\lambda) \in \mathbb{DL}(P)$ with vector v. Hence, if $\rho \approx 1$ then there is $L(\lambda) \in \mathbb{DL}(P)$ such that $\kappa_L(\lambda; v; P) \approx \kappa_P(\lambda)$. For backward errors analogous results were shown in [8]. The aim is therefore to find a transformation of λ such that ρ is minimized. For the transformation $\lambda = \alpha\mu$ the solution is given in the following theorem.

THEOREM 6.1. *Let $P(\lambda)$ be a matrix polynomial of degree ℓ and define*

$$\rho(\alpha) := \frac{\max_{0 \le i \le \ell} \alpha^i \|A_i\|_2}{\min(\|A_0\|_2, \alpha^\ell \|A_\ell\|_2)}$$

for $\alpha > 0$. The unique minimizer of $\rho(\alpha)$ is $\alpha_{opt} = (\|A_0\|_2/\|A_\ell\|_2)^{\frac{1}{\ell}}$.

Proof. The function $\rho(\alpha)$ is continuous. Furthermore, for $\alpha \to 0$ and $\alpha \to \infty$ we have $\rho(\alpha) \to \infty$. Hence, there must be at least one minimium in $(0, \infty)$. Let $\tilde{\alpha}$ be a local minimizer. Now assume that $\|A_0\|_2 < \tilde{\alpha}^\ell \|A_\ell\|_2$. Then

$$\rho(\alpha) = \frac{1}{\|A_0\|_2} \max(\alpha\|A_1\|_2, \dots, \alpha^\ell \|A_\ell\|_2)$$

in a neighborhood of $\tilde{\alpha}$. But this function is strictly increasing in this neighborhood. Hence, $\tilde{\alpha}$ cannot be a minimizer. Similarly, the assumption $\|A_0\|_2 > \tilde{\alpha}^\ell \|A_\ell\|_2$ at the minimum leads to

$$\rho(\alpha) = \frac{1}{\alpha^\ell \|A_\ell\|_2} \max(\|A_0\|_2, \dots, \alpha^{\ell-1} \|A_{\ell-1}\|_2),$$

in a neighborhood of this minimum, which is strictly decreasing. A necessary condition for a minimizer is therefore given as $\|A_0\|_2 = \alpha^\ell \|A_\ell\|_2$, which has the unique solution $\alpha_{opt} = (\|A_0\|_2/\|A_\ell\|_2)^{\frac{1}{\ell}}$ in $(0, \infty)$. Since there must be at least one minimum of $\rho(\alpha)$ in $(0, \infty)$ it follows that α_{opt} is the unique minimizer there. $\qquad\square$

We emphasize that the variable transformation $\lambda = \alpha\mu$ does not change condition numbers or backward errors of the original polynomial problem. It only affects these quantities for the linearization $L(\lambda)$.

For the special case $\ell = 2$ this leads to the same scaling as proposed by Fan, Lin and Van Dooren. If $\|A_0\|_2 = \|A_\ell\|_2$ then $\alpha_{opt} = 1$ and we cannot improve ρ with the transformation $\lambda = \alpha\mu$. In that case one might consider more general Möbius transformations of the type

$$\widetilde{P}(\mu) := (c\mu + d)^\ell P\left(\frac{a\mu + b}{c\mu + d}\right), \quad a, b, c, d \in \mathbb{C}.$$

However, it is still unclear how to choose the parameters a, b, c, d in order to improve ρ for a specific matrix polynomial.

7. Numerical examples. We first present numerical experiments on sets of randomly generated PEPs. The test problems are created by defining $A_k := F_1^{(k)} \tilde{A}_k F_2^{(k)}$, where the entries of $\tilde{A}_k$ are $N(0, 1)$ distributed random numbers and the entries of $F_1^{(k)}$ and $F_2^{(k)}$ are jth powers of $N(0, 1)$ distributed random numbers obtained from the `randn` function in MATLAB. As j increases these matrices become more badly scaled and ill-conditioned. This is a similar strategy to create test matrices as was used in [11]. In our experiments we choose the parameter $j = 6$.
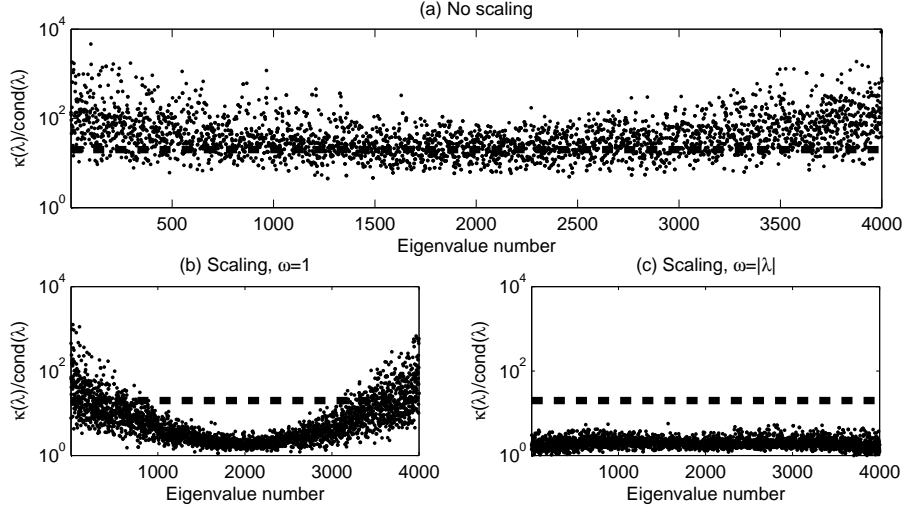
FIG. 7.1. *(a) The ratio of the normwise and componentwise condition numbers for the eigenvalues of* 100 *randomly created quadratic test problems of dimension* $n = 20$ *before scaling. (b) The same test set but now after scaling with* $\omega = 1$*. (c) Eigenvalue dependent scaling with* $\omega = |\lambda|$*.*
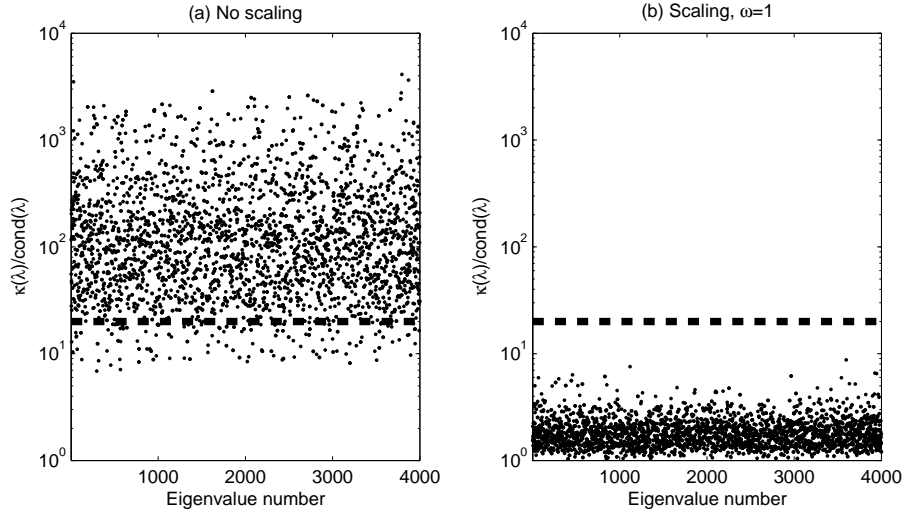

FIG. 7.2. *In this test all coefficient matrices of an eigenvalue problem are badly scaled in a similar way. (a) Ratio of normwise and componentwise condition condition numbers before scaling. (b) The same ratio after scaling with* $\omega = 1$*.*

In Figure 7.1(a) we show the ratio of the normwise and componentwise eigenvalue condition numbers of the eigenvalues for 100 quadratic eigenvalue problems of dimension $n = 20$. The eigenvalues range in magnitude from 10^{-8} to 10^8 and are sorted in ascending magnitude. According to Theorem 3.3 the ratio of normwise and componentwise condition number is smaller than n (shown by the dotted line) if the problem is almost optimally scaled for the corresponding eigenvalue. But only few eigenvalues satisfy this condition. Hence, we expect that scaling will improve the normwise

condition numbers of the eigenvalues in these test problems. In Figure 7.1(b) the test problems are scaled using Alg. 1 with the fixed parameter $\omega = 1$. Apart from the extreme ones all eigenvalues are now almost optimally scaled. In Figure 7.1(c) an eigenvalue dependent scaling is used, that is $\omega = |\lambda|$ for each eigenvalue λ. Now all eigenvalues are almost optimally scaled. This demonstrates that having some information about the magnitude of the wanted eigenvalues can greatly improve the results of scaling.

The source of badly scaled eigenvalue problems often lies in a nonoptimal choice of units in the modelling process, which can lead to all coefficient matrices A_k being badly scaled in a similar way. In that case it is not necessary to provide any kind of weighting. This is demonstrated by the example in Figure 7.2. The left plot in that figure shows the ratio of the normwise and componentwise condition numbers of the eigenvalues of another set of eigenvalue problems. Again, we choose $n = 20$ and $\ell = 2$. However, this time the matrices $F_1^{(k)}$ and $F_2^{(k)}$ in the definition $A_k := F_1^{(k)} \tilde{A}_k F_2^{(k)}$ are kept constant for all $k = 0, \ldots, \ell$. They only vary between different eigenvalue test problems. The right plot in Figure 7.2 shows the ratio of normwise and componentwise condition number after scaling using $\omega = 1$. Now all eigenvalue condition numbers are almost optimal.

Let us now consider the example of a 4th order PEP $(\lambda^4 A_4 + \lambda^3 A_3 + \lambda^2 A_2 + \lambda A_1 + A_0)x = 0$ derived from the Orr-Sommerfeld equation [17]. The matrices are created with the NLEVP benchmark collection [1]. To improve the scaling factor ρ we substitute $\lambda = \mu \alpha_{opt}$, where $\alpha_{opt} \approx 8.42 \cdot 10^{-4}$. This reduces ρ from $1.99 \cdot 10^{12}$ to 4.86. The ratio $\kappa_P(\mu)/\text{cond}_P(\mu)$ for the unscaled problem is shown in Figure 7.3(a). The x-axis denotes the absolute value $|\mu|$ of an eigenvalue μ. The horizontal line shows the dimension $n = 64$ of the problem. The large eigenvalues in this problem are far away from being optimally scaled. In Figure 7.3(b) we use Alg. 1 with the weighting parameter $\omega = 1$. This has almost no effect on the normwise condition numbers of the eigenvalues. In Figure 7.3(c) we use $\omega = 10^3$. Now the larger eigenvalues are almost optimally scaled while the normwise condition numbers of some of the smaller eigenvalues have become worse. Hence, in this example the right choice of the weighting parameter ω is crucial. If we want to improve the scaling of the large eigenvalue we need to choose ω as approximately the magnitude of these values to obtain good results. By diagonal scaling with D_1 and D_2 the scaling factor ρ might increase again. In this example, after diagonal scaling using the weight $\omega = 10^3$ ρ increases to $1.8 \cdot 10^5$. However, we can reduce this again by another variable transformation of the form $\mu = \tilde{\alpha}_{opt} \tilde{\mu}$. From Theorem 6.1 it follows that $\tilde{\alpha}_{opt} \approx 13.9$ and after this variable transformation ρ reduces to 67.6. Hence, at the end the condition numbers of the largest eigenvalues have decreased by a factor of about 10^5, while the scaling factor ρ has only increased by a factor of about 10.

Not only for polynomial problems can a weighted scaling significantly improve the condition numbers compared to unweighted scaling. In Figure 7.4 we show the results of scaling for the GEP $Kx = \lambda Mx$, where K and M are the matrices BCSSTK03 and BCSSTM03 from Matrix Market [13]. The dimension of the GEP is 112. While unweighted scaling improves the condition number of the smaller eigenvalues, the best result is obtained by using the weighting parameter $\omega = 10^7$. Then the condition number of all eigenvalues is improved considerably.

8. Some remarks about scaling in practice. In this concluding section we want to give based on the results of this paper some suggestions for practical scaling algorithms.
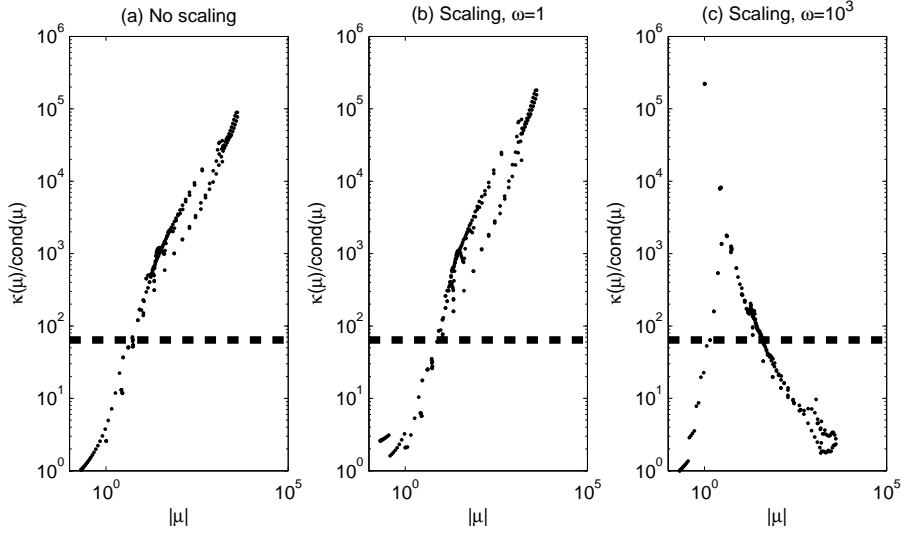
FIG. 7.3. *Scaling of a 4th order PEP. (a) $\kappa(\mu)/\mathrm{cond}(\mu)$ for the unscaled PEP. (b) The same ratio after scaling with $\omega = 1$. (c) Scaling with $\omega = 10^3$. The horizontal lines denote the dimension $n = 64$ of the PEP.*
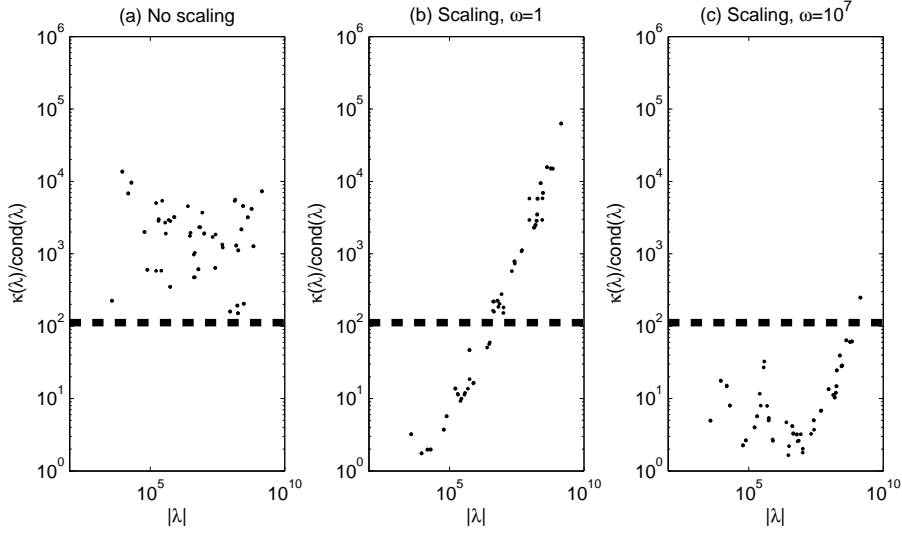

FIG. 7.4. *Scaling of the GEP $Kx = \lambda Mx$, where K and M are the matrices BCSSTK03 and BCSSTM03 from Matrix Market [13]. The best overall results are obtained with $\omega = 10^7$ (see right graph). The horizontal lines denote the dimension $n = 112$ of the GEP.*

1. *Compute $\kappa(\lambda)$ and $\mathrm{cond}(\lambda)$ for each eigenvalue.* At the moment eigensolvers often return a normwise condition number if desired by the user. It is only little more effort to additonally compute the ratio $\kappa(\lambda)/\mathrm{cond}(\lambda)$. From Theorem 3.3 it follows that a polynomial is almost optimally scaled for a certain eigenvalue if $\kappa(\lambda)/\mathrm{cond}(\lambda) \leq n$. If this condition is violated the user may decide to rescale the eigenvalue problem and then to recompute the eigenvalues in order to improve their accuracy.

2. *Use a weighted scaling.* The numerical examples in Section 7 show that the results of scaling can be greatly improved if ω is chosen to be of the magnitude of the wanted eigenvalues. In many applications this information is available from other considerations. If no information about the eigenvalues is available a reasonable choice is to set $\omega = 1$.

3. *First linearize and then scale if no special structure of the linearization is used.* The results in Section 4 show that one can obtain a smaller condition number if one scales after linearizing the polynomial $P(\lambda)$. If the eigenvalues of the linearization $L(\lambda)$ are computed without taking any special structure of $L(\lambda)$ into account this is therefore the preferable way. However, if the eigensolver uses the special structure of the linearization $L(\lambda)$ then one should scale the original polynomial $P(\lambda)$ and then linearize in order not to destroy this structure.

4. *Use a variable substitution of the type $\lambda = \alpha\mu$ to reduce the scaling factor ρ.* This technique, which was introduced by Fan, Lin and Van Dooren for quadratics and generalized in Theorem 6.1 often reduces the ratio of the condition number of an eigenvalue λ between the linearization and the original polynomial. In practice we would compute α using the Frobenius or an other cheaply computable norm.

The first two suggestions also apply to generalized linear eigenvalue problems and can be easily implemented to current standard solvers for them. Further research is needed for the effect of scaling on the backward error. Bounds on the backward error after scaling are difficult to obtain since the computed eigenvalues change after scaling and this change depends on the eigensolver.

REFERENCES

[1] T. Betcke, N. J. Higham, V. Mehrmann, C. Schröder, and F. Tisseur, *NLEVP: A collection of nonlinear eigenvalue problems*, tech. report, University of Manchester, to appear.

[2] T.-Y. Chen and J. W. Demmel, *Balancing sparse matrices for computing eigenvalues*, Linear Algebra Appl., 309 (2000), pp. 261–287.

[3] H.-Y. Fan, W.-W. Lin, and P. Van Dooren, *Normwise scaling of second order polynomial matrices*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 252–256.

[4] I. Gohberg, P. Lancaster, and L. Rodman, *Matrix Polynomials*, Academic Press, New York, 1982.

[5] D. J. Higham and N. J. Higham, *Structured backward error and condition of generalized eigenvalue problems*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 493–512.

[6] N. J. Higham, *A survey of componentwise perturbation theory in numerical linear algebra*, in Mathematics of Computation 1943–1993: A Half Century of Computational Mathematics, W. Gautschi, ed., vol. 48 of Proc. Sympos. Appl. Math., Providence, RI, 1994, AMS, pp. 49–77.

[7] ———, *Accuracy and Stability of Numerical Algorithms*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, second ed., 2002.

[8] N. J. Higham, R.-C. Li, and F. Tisseur, *Backward error of polynomial eigenproblems solved by linearization*, MIMS EPrint 2006.137, Manchester Institute for Mathematical Sciences, University of Manchester, Manchester, UK, June 2007. To appear in SIAM J. Matrix Anal. Appl.

[9] N. J. Higham, D. S. Mackey, and F. Tisseur, *The conditioning of linearizations of matrix polynomials*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1005–1028.

[10] D. Kressner, *Numerical methods for general and structured eigenvalue problems*, vol. 46 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, Berlin, 2005.

[11] D. Lemonnier and P. Van Dooren, *Balancing regular matrix pencils*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 253–263.

[12] D. S. MACKEY, N. MACKEY, C. MEHL, AND V. MEHRMANN, *Vector spaces of linearizations for matrix polynomials*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 971–1004.
[13] *Matrix Market.* http://math.nist.gov/MatrixMarket/.
[14] E. E. OSBORNE, *On pre-conditioning of matrices*, J. ACM, 7 (60), pp. 338–345.
[15] B. N. PARLETT AND C. REINSCH, *Balancing a matrix for calculation of eigenvalues and eigenvectors*, Numer. Math., 13 (1969), pp. 293–304.
[16] F. TISSEUR, *Backward error and condition of polynomial eigenvalue problems*, Linear Algebra Appl., 309 (2000), pp. 339–361.
[17] F. TISSEUR AND N. J. HIGHAM, *Structured pseudospectra for polynomial eigenvalue problems, with applications*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 187–208.
[18] R. C. WARD, *Balancing the generalized eigenvalue problem*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 141–152.
[19] D. S. WATKINS, *A case where balancing is harmful*, Electron. Trans. Numer. Anal., 23 (2006), pp. 1–4.