

# Numerical solution of a quadratic eigenvalue problem<sup>1</sup>

Chun-Hua Guo

*Department of Mathematics and Statistics, University of Regina, Regina,  
Saskatchewan, Canada S4S 0A2*

Dedicated to Peter Lancaster on the occasion of his 75th birthday

---

## Abstract

We consider the quadratic eigenvalue problem (QEP)  $(\lambda^2 M + \lambda G + K)x = 0$ , where  $M = M^T$  is positive definite,  $K = K^T$  is negative definite, and  $G = -G^T$ . The eigenvalues of the QEP occur in quadruplets  $(\lambda, \bar{\lambda}, -\lambda, -\bar{\lambda})$  or in real or purely imaginary pairs  $(\lambda, -\lambda)$ . We show that all eigenvalues of the QEP can be found efficiently and with the correct symmetry, by finding a proper solvent  $X$  of the matrix equation  $MX^2 + GX + K = 0$ , as long as the QEP has no eigenvalues on the imaginary axis. This solvent approach works well also for some cases where the QEP has eigenvalues on the imaginary axis.

*AMS classification:* 15A18; 15A24; 65F30

*Keywords:* quadratic eigenvalue problem; quadratic matrix equation; solvent; cyclic reduction

---

## 1 Introduction

Quadratic eigenvalue problems (QEPs) appear in many applications. One of the early works on this topic is the monograph [15] by Peter Lancaster. For a recent survey on this topic, see [24]. The QEP is to find scalars  $\lambda$  and nonzero vectors  $x$  satisfying

$$(\lambda^2 M + \lambda G + K)x = 0,$$

---

*Email address:* [chguo@math.uregina.ca](mailto:chguo@math.uregina.ca) (Chun-Hua Guo).

<sup>1</sup> This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

where  $M, G, K$  are  $n \times n$  real matrices and  $x$  is the eigenvector corresponding to the eigenvalue  $\lambda$ . In this paper we are only interested in finding the eigenvalues. When  $M$  is nonsingular, the QEP has  $2n$  finite eigenvalues. We assume that the problem size  $n$  is not too large and that all  $2n$  eigenvalues are needed.

The standard approach for finding all  $2n$  eigenvalues of the QEP is to use a proper linearization and solve a  $2n \times 2n$  generalized eigenvalue problem. Another approach is based on the factorization of the matrix polynomial

$$L(\lambda) = \lambda^2 M + \lambda G + K. \quad (1)$$

Indeed, it is well known that  $L(\lambda)$  admits the factorization

$$L(\lambda) = (\lambda M + MX + G)(\lambda I - X)$$

if and only if  $X$  is a solution (called a solvent) of the corresponding quadratic matrix equation

$$MX^2 + GX + K = 0 \quad (2)$$

(see [15, Theorem 3.3] or [6, Corollary 3.6]). Therefore, when (2) has a solution  $X$ , the  $2n$  eigenvalues of the QEP (also called the eigenvalues of  $L(\lambda)$ ) can be found by finding the eigenvalues of the matrix  $X$  and the matrix pencil  $\lambda M + MX + G$ . This approach has been explored by Davis [4] and more recently by Higham and Kim [12]. The difficulty associated with this solvent approach is obvious. It is possible that the matrix equation (2) does not have a solvent. Even if a solvent exists, the computation of a solvent can still be a difficult task. As a result, the solvent approach can outperform the linearization approach only for special types of QEPs. The main purpose of this paper is to identify a class of QEPs, for which the solvent approach is truly successful.

## 2 The quadratic eigenvalue problem

The special QEP that we will consider is

$$(\lambda^2 M + \lambda G + K)x = 0, \quad (3)$$

where  $M = M^T$  is positive definite,  $K = K^T$  is negative definite, and  $G = -G^T$ . The QEP (3) has been studied recently in [16] and [18]. In [18] the emphasis is on the case where the matrices  $M, G, K$  are large and sparse, and the QEP (3) has no eigenvalues on the imaginary axis. In [16] the emphasis

is on the QEP (3) corresponding to a “strongly stable” gyroscopic system (i.e., the eigenvalues of the QEP (3) and all nearby QEPs with the same symmetry structure are on the imaginary axis and are semisimple). It is well known that the eigenvalues of (3) have “Hamiltonian structure”, i.e., they occur in quadruplets  $(\lambda, \bar{\lambda}, -\lambda, -\bar{\lambda})$  or in real or purely imaginary pairs  $(\lambda, -\lambda)$  ([16]). In this paper we are mainly interested in the QEP (3) that has no eigenvalues on the imaginary axis. To get some insights about the QEP (3) having eigenvalues very close to the imaginary axis, we will also consider the limit case in which some eigenvalues of the QEP make “initial touch” on the imaginary axis (i.e., the QEP has some eigenvalues on the imaginary axis, but in every neighborhood of the QEP there is a QEP of the same type that does not have eigenvalues on the imaginary axis). However, QEPs corresponding to strongly stable gyroscopic systems are excluded from our study.

We first assume that the QEP (3) has no eigenvalues on the imaginary axis. If we can find a solvent  $X$  of the matrix equation (2) whose eigenvalues are the  $n$  eigenvalues of  $L(\lambda)$  in the right half plane, the remaining  $n$  eigenvalues of  $L(\lambda)$  are obtained by symmetry without any computation.

The success of this procedure is contingent on the existence of a solvent whose eigenvalues are in the right half plane.

**Lemma 1** *If the QEP (3) has no eigenvalues on the imaginary axis, then the matrix equation (2) has a solvent whose eigenvalues are in the right half plane.*

**PROOF.** The proof can be completed quite easily by using a result by Langer [17]. We will apply the result as presented in Theorem 11.2 of [6]. Let  $\mu = i\lambda$ . We rewrite  $L(\lambda)$  as

$$\hat{L}(\mu) = -M^{1/2}(\mu^2 I + \mu \hat{G} + \hat{K})M^{1/2},$$

where  $\hat{G} = M^{-1/2}(iG)M^{-1/2}$  and  $\hat{K} = -M^{-1/2}KM^{-1/2}$ . Since  $\hat{G}$  and  $\hat{K}$  are Hermitian,

$$\tilde{L}(\mu) = \mu^2 I + \mu \hat{G} + \hat{K}$$

is a self-adjoint monic matrix polynomial of degree 2. It is clear that  $\tilde{L}(\mu)$  has  $n$  eigenvalues above the real axis and  $n$  eigenvalues below the real axis. By Theorem 11.2 of [6],  $\tilde{L}(\mu)$  has a right divisor  $\mu I - R$  such that the  $n$  eigenvalues of  $R$  are the  $n$  eigenvalues of  $\tilde{L}(\mu)$  above the real axis. By Corollary 3.6 of [6], the matrix  $R$  satisfies the equation

$$R^2 + \hat{G}R + \hat{K} = 0.$$

Then it is easily verified that

$$X = -iM^{-1/2}RM^{1/2} \quad (4)$$

is a solvent of (2) whose eigenvalues are in the right half plane.  $\square$

Later on, we shall see that the solvent  $X$  given by (4) is actually a real matrix. So the computed complex eigenvalues of  $X$  (by the QR algorithm) will appear in conjugate pairs. Since the eigenvalues of the QEP in the left half plane are obtained directly by symmetry, the Hamiltonian structure for the eigenvalues of the QEP is preserved.

Our task is then to find the solvent  $X$  or more precisely the eigenvalues of  $X$ , with high efficiency.

We will not try to find the matrix  $X$  directly to find its  $n$  eigenvalues on the right half plane. Instead, we use the Cayley transform  $X = (I + Y)(I - Y)^{-1}$ . The matrix equation (2) now becomes

$$B_0 + B_1Y + B_2Y^2 = 0, \quad (5)$$

where

$$B_0 = M + K + G, \quad B_1 = 2(M - K), \quad B_2 = M + K - G. \quad (6)$$

We are now interested in the solution  $Y$  of (5) whose  $n$  eigenvalues are inside the unit circle. There is no need to recover the matrix  $X$  by the Cayley transform after the matrix  $Y$  is found. We can use the QR algorithm to find the eigenvalues  $\mu_i$  ( $i = 1, \dots, n$ ) of the matrix  $Y$ . The eigenvalues of  $X$  are obtained by  $\lambda_i = (1 + \mu_i)/(1 - \mu_i)$  ( $i = 1, \dots, n$ ).

A solution  $Y$  of (5) may be found by Newton's method with exact line searches [13], which is applicable to more general quadratic matrix equations. However, the solution  $Y$  obtained by that method does not necessarily have all eigenvalues inside the unit circle.

Since  $B_2 = B_0^T$  and  $B_1 = B_1^T$  is positive definite, the matrix equation (5) is closely related to the matrix equation

$$X + A^T X^{-1} A = Q, \quad (7)$$

where  $Q$  is symmetric positive definite. Indeed, letting  $Y = -X^{-1}A$ , we can rewrite (7) as

$$A + QY + A^T Y^2 = 0,$$

of which (5) is a special case. When the matrix  $Y$  is given in the form of  $-X^{-1}A$ , there is a need to compute the eigenvalues of  $Y$  by applying the QZ algorithm to the matrix pencil  $\lambda X + A$  if the matrix  $X$  is ill-conditioned. The method in [13] does not have this problem, but the solution computed by that method is not necessarily our desired solution.

### 3 Results on the matrix equation (7)

The matrix equation (7) has been studied by a number of authors (see [1], [5], [9], [10], [19], [23], [25]). For this equation, we are interested in symmetric positive definite solutions. For Hermitian matrices  $W_1$  and  $W_2$ , we write  $W_1 \geq W_2$  ( $W_1 > W_2$ ) if  $W_1 - W_2$  is positive semidefinite (definite). A real symmetric solution  $X_+$  of (7) is called maximal if  $X_+ \geq X$  for any real symmetric solution  $X$  of the matrix equation. The maximal solution is the solution we need for application to the QEP (3).

The following result is obtained in [5].

**Theorem 2** *Equation (7) has a positive definite solution if and only if the rational matrix function  $\psi(\lambda) = \lambda A + Q + \lambda^{-1}A^T$  is regular (i.e., the determinant of  $\psi(\lambda)$  is not identically zero) and  $\psi(\lambda) \geq 0$  for all  $\lambda$  on the unit circle. If (7) has a positive definite solution, then it has a maximal solution  $X_+$ . Moreover,  $\rho(X_+^{-1}A) \leq 1$  (where  $\rho(\cdot)$  denotes the spectral radius); for any other positive definite solution  $X$ ,  $\rho(X^{-1}A) > 1$ .*

It is also known that  $\rho(X_+^{-1}A) < 1$  if and only if  $\psi(\lambda) > 0$  for all  $\lambda$  on the unit circle (see [10]).

The maximal solution  $X_+$  of (7) can be found by the fixed point iteration.

#### Algorithm 1

$$\begin{aligned} Y_0 &= Q, \\ Y_{n+1} &= Q - A^T Y_n^{-1} A, \quad n = 0, 1, \dots \end{aligned}$$

For Algorithm 1, we have  $Y_0 \geq Y_1 \geq \dots$ , and  $\lim_{n \rightarrow \infty} Y_n = X_+$  (see, e.g., [5]).

Moreover,

$$\limsup_{n \rightarrow \infty} \sqrt[n]{\|Y_n - X_+\|} \leq (\rho(X_+^{-1}A))^2,$$

where  $\|\cdot\|$  is any matrix norm (see [10]).

When  $\rho(X_+^{-1}A)$  is close to 1, a much more efficient method for finding the solution  $X_+$  is based on the cyclic reduction method. The algorithm is as follows (see [19]).

**Algorithm 2**

$$\begin{aligned} X_0 &= Q_0 = Q, & A_0 &= A, \\ X_{n+1} &= X_n - A_n^T Q_n^{-1} A_n, \\ Q_{n+1} &= Q_n - A_n Q_n^{-1} A_n^T - A_n^T Q_n^{-1} A_n \\ A_{n+1} &= -A_n Q_n^{-1} A_n, & n &= 0, 1, \dots \end{aligned}$$

The next result is given in [19].

**Theorem 3** *For the matrices  $Q_n$  and  $X_n$  in Algorithm 2, we have  $Q_n \geq Q_{n+1} > 0, X_n \geq X_{n+1} > 0$  ( $n = 0, 1, \dots$ ). Moreover, if  $\rho(X_+^{-1}A) < 1$ , then*

$$\limsup_{n \rightarrow \infty} \sqrt[n]{\|X_n - X_+\|} \leq (\rho(X_+^{-1}A))^2, \quad \limsup_{n \rightarrow \infty} \sqrt[n]{\|A_n\|} \leq \rho(X_+^{-1}A).$$

When  $\rho(X_+^{-1}A) = 1$ , it is pointed out in [19] that the sequence  $\{X_n\}$  still converges to  $X_+$  if the sequence  $\{A_n\}$  converges to zero and the sequence  $\{(X_+^{-1}A)^{2^n}\}$  is bounded. It is shown in [9] that  $\{A_n\}$  always converges to zero. However, the sequence  $\{(X_+^{-1}A)^{2^n}\}$  is bounded if and only if all eigenvalues of  $X_+^{-1}A$  on the unit circle are semisimple. The following result is given in [9].

**Theorem 4** *If  $\rho(X_+^{-1}A) = 1$  and all eigenvalues of  $X_+^{-1}A$  on the unit circle are semisimple, then for the sequence  $\{X_n\}$  produced by Algorithm 2*

$$\limsup_{n \rightarrow \infty} \sqrt[n]{\|X_n - X_+\|} \leq \frac{1}{2}.$$

It turns out that  $\{X_n\}$  converges to  $X_+$  even if  $X_+^{-1}A$  has non-semisimple eigenvalues on the unit circle.

**Proposition 5** *If (7) has a positive definite solution and the sequence  $\{X_n\}$  is produced by Algorithm 2, then  $\lim_{n \rightarrow \infty} X_n = X_+$ .*

**PROOF.** We consider the  $m \times m$  block-Toeplitz matrix

$$T_m = \begin{bmatrix} Q & A^T & & & \\ A & Q & \cdots & & \\ & \cdots & \cdots & A^T & \\ & & & A & Q \end{bmatrix}_{m \times m}.$$

By the properties of cyclic reduction (see [3]),  $X_n$  can be viewed as the Schur complement in  $T_{2^n}$  of the lower-right block  $(2^n - 1) \times (2^n - 1)$  matrix. It is well known that this Schur complement can also be obtained by using block Gaussian elimination to eliminate successively the nonzero blocks in the positions  $(2^n - 1, 2^n), (2^n - 2, 2^n - 1), \dots, (1, 2)$ . This means that  $X_n$  can be obtained by  $2^n - 1$  steps of the fixed point iteration (Algorithm 1). That is,  $X_n = Y_{2^n - 1}$ . It follows that  $\{X_n\}$  converges to  $X_+$ .  $\square$

It is pointed out in [19] that  $\lambda_{\min}(Q_n) \geq \lambda_{\min}(T_{2^{n+1}-1})$  and  $\lambda_{\max}(Q_n) \leq \lambda_{\max}(T_{2^{n+1}-1})$ . It then follows from the results in [20] and [22] that

$$\lambda_{\min}(Q_n) \geq s = \inf_{\theta \in [-\pi, \pi]} \lambda_{\min}(f(\theta)), \quad (8)$$

$$\lambda_{\max}(Q_n) \leq t = \sup_{\theta \in [-\pi, \pi]} \lambda_{\max}(f(\theta)), \quad (9)$$

where  $f(\theta) = e^{i\theta}A + Q + e^{-i\theta}A^T$ , and that  $\lambda_{\min}(Q_n) > 0$  when  $s = 0$ . It can be shown (as in [9]) that  $\lim_{n \rightarrow \infty} \lambda_{\min}(Q_n) = 0$  whenever  $\rho(X_+^{-1}A) = 1$ . In view of the remark following Theorem 2,  $\rho(X_+^{-1}A) = 1$  if and only if  $s = 0$ .

#### 4 Application to the quadratic eigenvalue problem

A number of applications of the equation (7) have been mentioned in the literature. Now we can add one more application of the equation (7): the quadratic eigenvalue problem.

**Theorem 6** *The QEP (3) has no eigenvalues on the imaginary axis if and only if  $\psi(\lambda) = \lambda B_0 + B_1 + \lambda^{-1}B_2 > 0$  for all  $\lambda$  on the unit circle, where the matrices  $B_0, B_1, B_2$  are given in (6).*

**PROOF.** We apply the results in the previous section with  $A = B_0$  and  $Q = B_1$ . Recall that  $B_2 = B_0^T$ .

“if” part: Let  $X_+$  be the maximal solution of the matrix equation

$$X + B_0^T X^{-1} B_0 = B_1. \quad (10)$$

Then  $\rho(X_+^{-1} B_0) < 1$  and  $Y = -X_+^{-1} B_0$  is a solution of (5). By the Cayley transform, the matrix equation (2) has a solution with all  $n$  eigenvalues in the open right half plane. These  $n$  eigenvalues are also eigenvalues of the QEP (3). The remaining  $n$  eigenvalues of (3) are obtained by symmetry. Thus, the QEP has no eigenvalues on the imaginary axis.

“only if” part: Let  $\lambda = e^{i\theta}$ ,  $\theta \in [-\pi, \pi]$ . Then  $\psi(\lambda) = f(\theta) = e^{i\theta} B_0 + B_1 + e^{-i\theta} B_2$ . Let the  $n$  real eigenvalues of the Hermitian matrix  $f(\theta)$  be  $\lambda_1(\theta), \lambda_2(\theta), \dots, \lambda_n(\theta)$ , which are  $n$  real-valued continuous functions of  $\theta$ . Since  $f(0) = B_0 + B_1 + B_2 = 4M > 0$ , we have  $\lambda_i(0) > 0$  for  $i = 1, \dots, n$ . Since the QEP (3) has no eigenvalues on the imaginary axis,  $\psi(\lambda)$  is nonsingular for all  $\lambda$  on the unit circle. Therefore,  $\lambda_i(\theta) \neq 0$  for  $i = 1, \dots, n$  and  $\theta \in [-\pi, \pi]$ . By the intermediate value theorem,  $\lambda_i(\theta) > 0$  for  $i = 1, \dots, n$  and  $\theta \in [-\pi, \pi]$ . Thus,  $f(\theta) > 0$  for all  $\theta \in [-\pi, \pi]$ . That is,  $\psi(\lambda) > 0$  for all  $\lambda$  on the unit circle.  $\square$

Therefore, when the QEP (3) has no eigenvalues on the imaginary axis, all its eigenvalues can be found by finding the eigenvalues of  $-X_+^{-1} B_0$ , where  $X_+$  is the maximal solution of (10). We will also consider the QEP (3) for which  $\psi(\lambda) = \lambda B_0 + B_1 + \lambda^{-1} B_2 \geq 0$  for all  $\lambda$  on the unit circle. Since  $\psi(1) = B_0 + B_1 + B_2 = 4M > 0$ ,  $\psi(\lambda)$  is regular and the results in the previous section can still be applied. By replacing  $\psi(\lambda) > 0$  with  $\psi(\lambda) \geq 0$ , we allow some eigenvalues of the QEP to make initial touch on the imaginary axis. Of course, this consideration is enough for us to get some insights when the QEP has no eigenvalues on the imaginary axis, but has some eigenvalues very close to the imaginary axis.

The next result gives a precise relationship between the eigenvalues of the matrix  $-X_+^{-1} B_0$  and those of the matrix polynomial  $\phi(\lambda) = \lambda^2 B_2 + \lambda B_1 + B_0$ . The proof is based on the local Smith form as presented in [7].

**Theorem 7** *Assume that  $\psi(\lambda) \geq 0$  for all  $\lambda$  on the unit circle and let  $X_+$  be the maximal solution of (10). Then the eigenvalues of  $-X_+^{-1} B_0$  are precisely the eigenvalues of  $\phi(\lambda)$  inside or on the unit circle, with the same partial multiplicities for each eigenvalue inside the unit circle and with half of the partial multiplicities for each eigenvalue on the unit circle.*

**PROOF.** The matrix equation (10) is a special discrete algebraic Riccati equation. It is known (see [8] and [9]) that the partial multiplicities of each



eigenvalue of  $-X_+^{-1}B_0$  on the unit circle are half of the partial multiplicities of this eigenvalue for the matrix pencil

$$\lambda \begin{bmatrix} -I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & I & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & -I \\ B_1 & -I & B_2 \\ -B_0 & 0 & 0 \end{bmatrix}.$$

By elementary block operations, this matrix pencil can be reduced to

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & \phi(\lambda) \end{bmatrix}.$$

This proves the assertion for the eigenvalues on the unit circle. Since  $Y = -X_+^{-1}B_0$  is a solution of (5),

$$\phi(\lambda) = (\lambda B_2 + B_2 Y + B_1)(\lambda I - Y).$$

Since the matrix pencil  $\lambda B_2 + B_2 Y + B_1$  does not have eigenvalues inside the unit circle, we know by the local Smith form that the assertion for the eigenvalues inside the unit circle is also true.  $\square$

The above theorem tells us that all eigenvalues of the QEP (3) can be found by finding the eigenvalues of  $-X_+^{-1}B_0$ , under the weaker assumption that  $\psi(\lambda) \geq 0$  for all  $\lambda$  on the unit circle. It is not clear whether  $-X_+^{-1}B_0$  is the only solution of (5) whose eigenvalues are inside or on the unit circle. However, if  $Y$  is any solution of (5) such that  $\rho(Y) \leq 1$  and all eigenvalues of  $Y$  on the unit circle are semisimple, then we necessarily have  $Y = -X_+^{-1}B_0$ . Indeed, if the sequences  $\{A_n\}$  and  $\{X_n\}$  are obtained by Algorithm 2 with  $A = B_0$  and  $Q = B_1$ , then we know from the properties of cyclic reduction (see [3]) that

$$B_0 + X_n Y + A_n^T Y^{2^n+1} = 0, \quad n = 0, 1, \dots \quad (11)$$

Letting  $n \rightarrow \infty$  in (11), we get  $B_0 + X_+ Y = 0$  by Proposition 5 and the fact that  $\lim_{n \rightarrow \infty} A_n = 0$ . Therefore,  $Y = -X_+^{-1}B_0$ . In particular, the matrix  $X$  in (4) must be equal to the real matrix  $(I - X_+^{-1}B_0)(I + X_+^{-1}B_0)^{-1}$ .

It is also easy to establish a precise relationship between the eigenvalues of the matrix polynomial

$$L(\lambda) = \lambda^2 M + \lambda G + K.$$

and the matrix polynomial

$$\phi(\mu) = \mu^2 B_2 + \mu B_1 + B_0.$$

**Proposition 8**  $\mu_0$  is an eigenvalue of  $\phi(\mu)$  if and only if  $\lambda_0 = (1+\mu_0)/(1-\mu_0)$  is an eigenvalue of  $L(\lambda)$ . Moreover, the partial multiplicities of  $\mu_0$  are the same as the partial multiplicities of  $\lambda_0$ .

**PROOF.** We may assume that  $\lambda_0 \neq -1$  (so  $\mu_0 \neq \infty$ ). The result for  $\lambda_0 = -1$  and  $\mu_0 = \infty$  follows readily from the result for  $\lambda_0 = 1$  and  $\mu_0 = 0$ . The proof is again based on the local Smith form. Let the partial multiplicities of  $\lambda_0$  be  $k_1 \leq k_2 \leq \dots \leq k_n$  (including zero partial multiplicities). Then

$$L(\lambda) = E(\lambda) \begin{bmatrix} (\lambda - \lambda_0)^{k_1} & & \\ & \ddots & \\ & & (\lambda - \lambda_0)^{k_n} \end{bmatrix} F(\lambda),$$

where  $E(\lambda)$  and  $F(\lambda)$  are matrix polynomials invertible at  $\lambda_0$ . Now,

$$\phi(\mu) = (1 - \mu)^2 L\left(\frac{1 + \mu}{1 - \mu}\right) = \hat{E}(\mu) \begin{bmatrix} (\mu - \mu_0)^{k_1} & & \\ & \ddots & \\ & & (\mu - \mu_0)^{k_n} \end{bmatrix} \hat{F}(\mu),$$

where

$$\hat{E}(\mu) = (1 - \mu)^2 E\left(\frac{1 + \mu}{1 - \mu}\right)$$

and

$$\hat{F}(\mu) = \begin{bmatrix} \left(\frac{2}{(1-\mu)(1-\mu_0)}\right)^{k_1} & & \\ & \ddots & \\ & & \left(\frac{2}{(1-\mu)(1-\mu_0)}\right)^{k_n} \end{bmatrix} F\left(\frac{1 + \mu}{1 - \mu}\right)$$

are rational matrix functions defined and invertible at  $\mu_0$ . Therefore, the partial multiplicities of  $\mu_0$  are  $k_1 \leq k_2 \leq \dots \leq k_n$ .  $\square$

## 5 Comparison with the linearization approach

The solvent approach for finding all  $2n$  eigenvalues of the QEP (3) can be summarized as follows.

### Algorithm 3

- (1) Let  $B_0 = M + K + G$  and  $B_1 = 2(M - K)$ .
- (2) Apply Algorithm 2 with  $A = B_0$  and  $Q = B_1$  to find the maximal solution  $X_+$  of (10).
- (3) Use the QR algorithm to find eigenvalues  $\mu_i$  ( $i = 1, \dots, n$ ) of  $-X_+^{-1}B_0$ .
- (4) Find eigenvalues  $\lambda_i$  ( $i = 1, \dots, n$ ) of the QEP (3) on the closed right half plane by using

$$\lambda_i = (1 + \mu_i)/(1 - \mu_i), \quad i = 1, \dots, n.$$

- (5) Obtain the remaining  $n$  eigenvalues of the QEP (3) by symmetry.

The algorithm will be useful only when  $B_1 = 2(M - K)$  is not ill-conditioned ( $M$  and  $K$  may still be ill-conditioned). If  $B_1$  is ill-conditioned, we cannot expect much accuracy from Algorithm 3, since Algorithm 3 uses Algorithm 2 in step (2) and  $B_1$  is inverted in the first iteration of Algorithm 2. On the other hand, the matrix  $X_n$  produced by Algorithm 2 is often well-conditioned when  $M - K$  is well-conditioned. This is suggested by  $Q_n > 0$  and

$$Q_n + \sum_{i=0}^{n-1} A_i Q_i^{-1} A_i^T \leq X_n \leq 2(M - K),$$

which is readily seen from Algorithm 2.

In step (1) of Algorithm 3, there is some possibility of cancellation, particularly for the computation of  $B_0$ . This may limit the applicability of the algorithm to some extent.

In step (2) of Algorithm 3, we apply Algorithm 2. In the  $n$ th iteration of Algorithm 2, we need to compute  $A_n^T Q_n^{-1} A_n$ ,  $A_n Q_n^{-1} A_n^T$ , and  $A_n Q_n^{-1} A_n$ . These matrices can be computed as follows. Let  $Q_n = L_n L_n^T$  be the Cholesky factorization of  $Q_n$ ,  $V_n = L_n^{-1} A_n$ , and  $W_n = L_n^{-1} A_n^T$ . Then  $A_n^T Q_n^{-1} A_n = V_n^T V_n$ ,  $A_n Q_n^{-1} A_n^T = W_n^T W_n$ , and  $A_n Q_n^{-1} A_n = W_n^T V_n$ . Therefore, the computational work required for one iteration of Algorithm 2 is about  $\frac{19}{3}n^3$  flops. In the presence of rounding errors, the matrix  $Q_n$  may fail to be positive definite when the QEP has eigenvalues very close to the imaginary axis. In this case, the iteration in Algorithm 2 stops and we take  $X_n$  as an approximation to  $X_+$ . If  $Q_n$  is positive definite and  $\|A_n^T Q_n^{-1} A_n\|_\infty < \epsilon$ , a prescribed tolerance, then we

compute  $X_{n+1}$  as an approximation to  $X_+$  (and skip the computation of  $Q_{n+1}$  and  $A_{n+1}$ ). Normally, 20 iterations of Algorithm 2 is quite enough for us to find the  $2n$  eigenvalues of the QEP by Algorithm 3, with accuracy warranted by the QEP itself.

In step (3) of Algorithm 3, we use the QR algorithm to find eigenvalues of  $-X_+^{-1}B_0$  since the matrix  $X_+$  is typically well-conditioned when  $B_1 = 2(M - K)$  is well-conditioned. To get possibly higher accuracy, we may compute these eigenvalues by applying the QZ algorithm to the pencil  $\lambda X_+ + B_0$ .

In step (4) of Algorithm 3, we may get (computed) eigenvalues with small negative real parts when the QEP has eigenvalues very close to the imaginary axis. In this case, we can simply change these small negative numbers to zero. If the QEP has eigenvalues  $\lambda_i$  with large modulus, then the computed  $\lambda_i$  will have low accuracy since they are obtained from the Cayley transform with  $\mu_i$  close to 1.

We now compare the solvent approach with the linearization approach. Two common linearizations of the QEP (3) are:

$$\lambda \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix} - \begin{bmatrix} 0 & I \\ -K & -G \end{bmatrix} \quad (12)$$

and

$$\lambda \begin{bmatrix} M & G \\ 0 & M \end{bmatrix} - \begin{bmatrix} 0 & -K \\ M & 0 \end{bmatrix}. \quad (13)$$

Each linearization has the same eigenvalues as the QEP. Moreover, the partial multiplicities of each eigenvalue are the same as those for the matrix polynomial

$$L(\lambda) = \lambda^2 M + \lambda G + K.$$

Recall that the eigenvalues of the matrix  $-X_+^{-1}B_0$  are related to those of  $L(\lambda)$  by Theorem 7 and Proposition 8. Thus, qualitatively speaking, computing the eigenvalues of  $-X_+^{-1}B_0$  is no more difficult than computing the eigenvalues of (12) or (13).

For both linearizations, we may use the QZ algorithm to find all  $2n$  eigenvalues. The QZ algorithm has high accuracy. However, it is quite expensive and the computed eigenvalues do not have the right symmetry. The computational work needed for the QZ algorithm is enough for us to carry out Algorithm 3

with 30 iterations for Algorithm 2 in step (2), even if we find the eigenvalues of  $-X_+^{-1}B_0$  in step (3) by applying the QZ algorithm to the pencil  $\lambda X_+ + B_0$ . If we choose to use linearization (13), we may apply an algorithm developed in [2] to compute all  $2n$  eigenvalues. The eigenvalues computed by that algorithm have the right symmetry, with accuracy comparable to that achieved by the QZ algorithm. That algorithm is less expensive than the QZ algorithm in terms of flops, but not necessarily so in terms of timings.

To reduce the computational work involved in the linearization approach, we can reduce the problem of finding the eigenvalues of the pencil (12) to that of finding the eigenvalues of the matrix

$$\begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}G \end{bmatrix}. \quad (14)$$

However, there may be a loss of accuracy in doing this if  $M$  is ill-conditioned. If we use the QR algorithm to find the eigenvalues of (14), the computational work needed is enough for us to carry out Algorithm 3 with about 15 iterations for Algorithm 2 in step (2). The ill-conditioning of the matrix  $M$  may also affect the accuracy of Algorithm 3. Indeed, for the numbers  $s$  and  $t$  defined in (8) and (9), we have  $t/s \geq \text{cond}_2(M)$  since  $f(0) = 4M$ . This means that the matrices  $Q_n$  may be ill-conditioned too. When  $M$  is ill-conditioned but the QEP does not have eigenvalues close to the imaginary axis, the sequence  $\{A_n\}$  in Algorithm 2 often converges to zero quickly (see Theorem 3), so the possible ill-conditioning of  $Q_n$  does not affect the accuracy of  $X_{n+1}$  significantly. The situation for the linearization approach is quite different. If we apply the QR algorithm to the matrix (14) and  $M$  is ill-conditioned, then the input data for the QR algorithm is changed significantly. If the QEP has eigenvalues very close to the imaginary axis, then these eigenvalues cannot be computed with high precision by any method. Indeed, when the QEP has eigenvalues on the imaginary axis, these eigenvalues must have positive even integers as their partial multiplicities (see Theorem 7 and Proposition 8). These eigenvalues are therefore very sensitive to the perturbations in the matrices  $M, G$ , and  $K$  (see [21]).

## 6 Numerical results

In this section we present some numerical results to show that the solvent approach outperforms the linearization approach for the QEP (3). All computations are done in Matlab version 5.3 on a Sun workstation.

**Example 1** Our first example is Example 6.1 of [18]. So we have a QEP (3) with  $n = 100$ . All 200 eigenvalues of the QEP can be found by applying the QZ algorithm to the matrix pencil (13). These eigenvalues are displayed in Figure 6.1 of [18]. The computed eigenvalues appear in quadruplets that approximately have the form  $(\lambda, \bar{\lambda}, -\lambda, -\bar{\lambda})$ . It is reasonable to judge the accuracy of these eigenvalues by the closeness to this form. In this sense, the most accurate quadruplet of eigenvalues computed by the QZ algorithm is

$$\begin{aligned} &0.28165313667904 + 0.87246540561295i, \\ &0.28165313667904 - 0.87246540561294i, \\ &-0.28165313667904 + 0.87246540561295i, \\ &-0.28165313667904 - 0.87246540561295i. \end{aligned}$$

This quadruplet is closest to the imaginary axis but well separated from other eigenvalues of the QEP. The least accurate quadruplet of eigenvalues appears to be

$$\begin{aligned} &1.70067837102530 + 0.03080313392466i, \\ &1.70068051210059 - 0.03080408493533i, \\ &-1.70067813619718 + 0.03080259924921i, \\ &-1.70067813619719 - 0.03080259924919i. \end{aligned}$$

This quadruplet is nearly mixed with other quadruplets. Note that the eigenvalues returned by the QZ algorithm do not appear in conjugate pairs. If we use the QR algorithm to find the eigenvalues of the matrix (14), then the corresponding eigenvalues are

$$\begin{aligned} &0.28165313667905 \pm 0.87246540561294i, \\ &-0.28165313667904 \pm 0.87246540561295i, \end{aligned}$$

and

$$\begin{aligned} &1.70067647383514 \pm 0.03081212542094i, \\ &-1.70067822344813 \pm 0.03080218408502i. \end{aligned}$$

If we use Algorithm 3 with 6 iterations of Algorithm 2 in step (2), then the corresponding eigenvalues are

$$\pm 0.28165313667905 \pm 0.87246540561294i,$$

and

$$\pm 1.70067804782901 \pm 0.03080241119623i.$$

For this example, the solvent approach is least expensive. The accuracy of the eigenvalues computed by Algorithm 3 is no worse than that achieved by the QZ algorithm. The eigenvalues computed by the QR algorithm applied to the matrix (14) have similar accuracy since we have  $\text{cond}_2(M) = 2.84$  for this example.

Example 1 appears to be too easy for Algorithm 3. We now construct an example for which the QEP has eigenvalues close to (or on) the imaginary axis and the matrix  $M$  is ill-conditioned.

**Example 2** We consider the QEP (3) with the matrices  $M, G, K$  defined as follows. Let

$$M_0 = \begin{bmatrix} 10^{-7} & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}, \quad G_0 = \begin{bmatrix} 0 & 1 & & \\ -1 & 0 & & \\ & & 0 & g \\ & & -g & 0 \end{bmatrix}, \quad K_0 = \begin{bmatrix} -1 & & & \\ & -10^{-7} & & \\ & & -4 & \\ & & & -1 \end{bmatrix}.$$

We then generate a random matrix by  $\text{randn}(4)$  and keep 3 digits for its entries. The matrix we get is

$$W = \begin{bmatrix} -0.43 & -1.15 & 0.33 & -0.59 \\ -1.67 & 1.19 & 0.17 & 2.18 \\ 0.13 & 1.19 & -0.19 & -0.14 \\ 0.29 & -0.04 & 0.73 & 0.11 \end{bmatrix}.$$

Now we define

$$M = W^T M_0 W, \quad G = W^T G_0 W, \quad K = W^T K_0 W.$$

If  $g = 3$ , then the eigenvalues of the QEP are

$$\pm 0.70710679886 \pm 0.70710676351i, \quad \pm\sqrt{2}i \text{ (each with multiplicity 2).}$$

If  $g = 2.999999$ , then the eigenvalues of the QEP are

$$\pm 0.70710679886 \pm 0.70710676351i, \quad \pm 0.00122474477 \pm 1.41421303204i.$$

We have the following numerical results for  $g = 3$  ( $g = 2.999999$ ). For the 8 eigenvalues computed by the QZ algorithm, the largest absolute error is  $1.88 \times 10^{-8}$  ( $4.72 \times 10^{-9}$ ). If we use the QR algorithm to find the eigenvalues of the matrix (14), the largest absolute error is  $1.22 \times 10^{-2}$  ( $8.26 \times 10^{-3}$ ). If we use Algorithm 3 with 14 iterations of Algorithm 2 in step (2), the largest absolute error is  $1.53 \times 10^{-9}$  ( $3.96 \times 10^{-9}$ ) for the computed eigenvalues. For this example,  $\text{cond}_2(M) = 1.83 \times 10^8$  and  $\text{cond}_2(M - K) = 27.62$ . The accuracy achieved by applying the QR algorithm to (14) is very low. However, the accuracy achieved by Algorithm 3 is no worse than that achieved by the QZ algorithm.

For this example, the convergence of Algorithm 2 in step (2) of Algorithm 3 is still quite fast, as suggested by Theorems 3 and 4. For Example 2 with  $g = 3$ , each eigenvalue of the QEP on the imaginary axis has partial multiplicity 2. This is the generic case. Of course, it is possible to construct examples for which the eigenvalues of the QEP on the imaginary axis have large even integers as their partial multiplicities. In those situations, we still have convergence for Algorithm 2 in step (2) of Algorithm 3 (see Proposition 5), although the rate of convergence is expected to be much slower. Moreover, the matrix  $Q_n$  in Algorithm 2 will be very ill-conditioned for large  $n$ . This means that it is impossible to get a very good approximation to  $X_+$  in step (2) of Algorithm 3. Nevertheless, Algorithm 3 can still obtain the eigenvalues of the QEP, with accuracy warranted by the QEP itself. Our final example illustrates this point.

**Example 3** We consider the QEP (3) for which the matrices  $M, G, K$  are defined as follows. Let

$$P = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}, \quad S = \begin{bmatrix} 0 & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 \end{bmatrix}, \quad U = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \quad V = -U,$$

$$B_0 = - \begin{bmatrix} P & I \\ & P & I \\ & & P & I \\ & & & P \end{bmatrix}, \quad B_1 = I + B_0^T B_0.$$



We then define the matrices  $M, G, K$  by

$$G = \begin{bmatrix} S & V & & \\ U & S & V & \\ & U & S & V \\ & & U & S \end{bmatrix},$$

$$M = \frac{1}{4}B_1 + \frac{1}{2}(B_0 - G), \quad K = -\frac{1}{4}B_1 + \frac{1}{2}(B_0 - G).$$

It is easily verified that  $M$  is symmetric positive definite,  $K$  is symmetric negative definite, and  $G$  is skew-symmetric. We also have

$$B_0 = M + K + G, \quad B_1 = 2(M - K).$$

The maximal solution of the matrix equation (10) is  $X_+ = I$  (the last statement in Theorem 2 is used here) and the eigenvalues of  $-X_+^{-1}B_0$  are  $\frac{\sqrt{2}}{2}(1 \pm i)$ , each with partial multiplicity 4. Thus, the eigenvalues of the QEP are  $\pm(1 + \sqrt{2})i$ , each with partial multiplicity 8. For the 16 eigenvalues computed by the QZ algorithm, the largest absolute error is  $4.01 \times 10^{-2}$ . If we use the QR algorithm to find the eigenvalues of the matrix (14), the largest absolute error is  $4.77 \times 10^{-2}$ . If we use Algorithm 3 with 10 iterations of Algorithm 2 in step (2), the largest absolute error is  $2.97 \times 10^{-2}$  for the computed eigenvalues. For this example,  $\text{cond}_2(M) = 87.28$ . The accuracy achieved by Algorithm 3 is no worse than that achieved by the other two methods. The lower accuracy in this example is explained by the fact that each of the two exact eigenvalues has partial multiplicity 8 (see [21], for example).

## 7 Conclusions

We have presented an algorithm for finding all eigenvalues of a quadratic eigenvalue problem arising from gyroscopic systems and problems of linear elasticity. The algorithm, which is based on the solvent approach, preserves the Hamiltonian structure of the spectrum of the QEP. It is less expensive than the linearization approach using the QZ algorithm. The computed eigenvalues typically have very good accuracy except those with very large modulus. In most cases, the algorithm is also more accurate and less expensive than the linearization approach using the QR algorithm.

Although the solvent-based algorithm cannot be applied to find the eigenvalues of a strongly stable gyroscopic system, the solvent approach itself still has a

big role to play in verifying the strong stability of a gyroscopic system. Indeed, as shown in [11], the solvent approach can be used to detect a hyperbolic QEP and, in particular, to check a sufficient condition for the strong stability of a gyroscopic system [11,16]. The solvent approach provides a hyperbolicity test that is more efficient than the one given in [14].

## Acknowledgement

The author thanks the referee for his very helpful comments.

## References

- [1] W. N. Anderson, Jr., T. D. Morley, G. E. Trapp, Positive solutions to  $X = A - BX^{-1}B^*$ , *Linear Algebra Appl.* 134 (1990) 53–62.
- [2] P. Benner, R. Byers, V. Mehrmann, H. Xu, Numerical computation of deflating subspaces of skew-Hamiltonian/Hamiltonian pencils, *SIAM J. Matrix Anal. Appl.* 24 (2002) 165–190.
- [3] D. A. Bini, L. Gemignani, B. Meini, Computations with infinite Toeplitz matrices and polynomials, *Linear Algebra Appl.* 343–344 (2002) 21–61.
- [4] G. J. Davis, Numerical solution of a quadratic matrix equation, *SIAM J. Sci. Stat. Comput.* 2 (1981) 164–175.
- [5] J. C. Engwerda, A. C. M. Ran, A. L. Rijkeboer, Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation  $X + A^*X^{-1}A = Q$ , *Linear Algebra Appl.* 186 (1993) 255–275.
- [6] I. Gohberg, P. Lancaster, L. Rodman, *Matrix Polynomials*, Academic Press, New York, 1982.
- [7] I. Gohberg, P. Lancaster, L. Rodman, *Invariant Subspaces of Matrices with Applications*, John Wiley, New York, 1986.
- [8] C.-H. Guo, Newton’s method for discrete algebraic Riccati equations when the closed-loop matrix has eigenvalues on the unit circle, *SIAM J. Matrix Anal. Appl.* 20 (1998) 279–294.
- [9] C.-H. Guo, Convergence rate of an iterative method for a nonlinear matrix equation, *SIAM J. Matrix Anal. Appl.* 23 (2001) 295–302.
- [10] C.-H. Guo, P. Lancaster, Iterative solution of two matrix equations, *Math. Comp.* 68 (1999) 1589–1603.
- [11] C.-H. Guo, P. Lancaster, Algorithms for hyperbolic quadratic eigenvalue problems, submitted for publication.

- [12] N. J. Higham, H.-M. Kim, Numerical analysis of a quadratic matrix equation, *IMA J. Numer. Anal.* 20 (2000) 499–519.
- [13] N. J. Higham, H.-M. Kim, Solving a quadratic matrix equation by Newton’s method with exact line searches, *SIAM J. Matrix Anal. Appl.* 23 (2001) 303–316.
- [14] N. J. Higham, F. Tisseur, P. M. Van Dooren, Detecting a definite Hermitian pair and a hyperbolic or elliptic quadratic eigenvalue problem, and associated nearness problems, *Linear Algebra Appl.* 351–352 (2002) 455–474.
- [15] P. Lancaster, *Lambda-Matrices and Vibrating Systems*, Pergamon Press, Oxford, 1966.
- [16] P. Lancaster, Strongly stable gyroscopic systems, *Electron. J. Linear Algebra* 5 (1999) 53–66.
- [17] H. Langer, Factorization of operator pencils, *Acta Sci. Math. (Szeged)* 38 (1976) 83–96.
- [18] V. Mehrmann, D. Watkins, Structure-preserving methods for computing eigenpairs of large sparse skew-Hamiltonian/Hamiltonian pencils, *SIAM J. Sci. Comput.* 22 (2001) 1905–1925.
- [19] B. Meini, Efficient computation of the extreme solutions of  $X + A^*X^{-1}A = Q$  and  $X - A^*X^{-1}A = Q$ , *Math. Comp.* 71 (2002) 1189–1204.
- [20] M. Miranda, P. Tilli, Asymptotic spectra of Hermitian block Toeplitz matrices and preconditioning results, *SIAM J. Matrix Anal. Appl.* 21 (2000) 867–881.
- [21] J. Moro, J. V. Burke, M. L. Overton, On the Lidskii–Vishik–Lyusternik perturbation theory for eigenvalues of matrices with arbitrary Jordan structure, *SIAM J. Matrix Anal. Appl.* 18 (1997) 793–817.
- [22] S. Serra, Asymptotic results on the spectra of block Toeplitz preconditioned matrices, *SIAM J. Matrix Anal. Appl.* 20 (1998) 31–44.
- [23] J.-G. Sun, S.-F. Xu, Perturbation analysis of the maximal solution of the matrix equation  $X + A^*X^{-1}A = P$ . II, *Linear Algebra Appl.* 362 (2003) 211–228.
- [24] F. Tisseur, K. Meerbergen, The quadratic eigenvalue problem, *SIAM Rev.* 43 (2001) 235–286.
- [25] X. Zhan, J. Xie, On the matrix equation  $X + A^T X^{-1}A = I$ , *Linear Algebra Appl.* 247 (1996) 337–345.