BOUNDS FOR THE EXTREMAL EIGENVALUES OF A CLASS OF SYMMETRIC TRIDIAGONAL MATRICES WITH APPLICATIONS[1]

Hannes Buchholzer and Christian Kanzow

Preprint 294 March 2010

University of Würzburg
Institute of Mathematics
Am Hubland, 97074 Würzburg, Germany

e-mail: buchholzer@mathematik.uni-wuerzburg.de
kanzow@mathematik.uni-wuerzburg.de

March 10, 2010

[1]This research was partially supported by the DFG (Deutsche Forschungsgemeinschaft) under grant KA 1296/16-1, 16-2.

Abstract. We consider a class of symmetric tridiagonal matrices which may be viewed as perturbations of Toeplitz matrices. The Toeplitz structure is destroyed since two elements on each off-diagonal are perturbed. Based on a careful analysis of the corresponding characteristic polynomial, we derive sharp bounds for the extremal eigenvalues of this class of matrices in terms of the original data of the given matrix. In this way, we also obtain a lower bound for the smallest singular value of certain matrices. Some numerical results indicate that our bounds are extremely good.

Key Words: Tridiagonal matrices, symmetric matrices, eigenvalues, singular values, extremal eigenvalues.

1 Introduction

Consider a tridiagonal matrix of the form

$$J = \begin{bmatrix} \alpha & \beta & & & & \\ \beta & \alpha & \gamma & & & \\ & \gamma & \alpha & \ddots & & \\ & & \ddots & \ddots & \gamma & \\ & & & \gamma & \alpha & \delta \\ & & & & \delta & \alpha \end{bmatrix} \in \mathbb{R}^{m\times m} \tag{1}$$

with given entries $\alpha, \beta, \gamma, \delta \in \mathbb{R}$. Matrices of this form arise quite frequently in many contexts, and the eigenvalues of such matrices can often be used to compute eigenvalues of more complicated matrices which arise, e.g., from the discretization of partial differential equations.

Our aim is to give sharp bounds for the smallest and largest eigenvalues of such a matrix. To the best of our knowlegde (and somewhat to our own surprise), there is no previous treatment of this problem in the literature. There exist many results for more general matrices like Gershgorin's, Ostrowski's or Brauer's Theorem (see, e.g., [4, 5, 12]) that estimate the area to which the eigenvalues belong to, however, the bounds one obtains from these results for the particular class of matrices considered here are by far too weak. The lower bound on the smallest eigenvalue for the matrix J may also be used to obtain a lower bound for the smallest singular value of a possibly nonsymmetric matrix. This lower bound seems to be much stronger than existing ones, see, e.g., [7, 8, 10].

To obtain our results, we take a closer look at the class of matrices from (1) and exploit heavily the particular structure using a careful analysis of the corresponding characteristic polynomial. To this end, we begin with some preliminary results in Section 2. The main results are contained in Section 3, where, depending on the relative (absolute) sizes of the matrix entries α, β, γ, and δ, we give suitable bounds for the two extremal eigenvalues. We apply our results to the discretization of a partial differential equation in Section 4 where matrices arise that can be decomposed as a Kronecker product of tridiagonal matrices of the above kind. This kind of application was, in fact, the original motivation for our investigations.

Notation: Given an arbitrary matrix $A \in \mathbb{R}^{m\times m}$, we denote by $A^s := \frac{1}{2}\big(A + A^T\big)$ the symmetric part of A. The singular values of A are denoted by $\sigma_i(A)$ $(i = 1, \ldots, m)$ and ordered in such a way that $\sigma_1(A) \le \sigma_2(A) \le \ldots \le \sigma_m(A)$, in particular, $\sigma_1(A)$ and $\sigma_m(A)$ denote the smallest and the largest singular value of A, respectively. Similarly, given a symmetric matrix $A \in \mathbb{R}^{m\times m}$, the corresponding (real) eigenvalues are denoted by $\lambda_i(A)$ $(i = 1, \ldots, m)$ and ordered in such a way that $\lambda_1(A) \le \lambda_2(A) \le \ldots \le \lambda_m(A)$ so that the symbol $\lambda_1(A)$ $(\lambda_m(A))$ always stands for the smallest (largest) eigenvalue of A. We sometimes also write $\lambda_{\min}(A)$ $(\lambda_{\max}(A))$ for the smallest (largest) eigenvalue of A. Given a vector $x \in \mathbb{R}^m$, the symbol $\|x\|$ denotes the Euclidean norm of x. Finally, the identity matrix in $\mathbb{R}^{m\times m}$ is denoted by I_m.

2 Preliminaries

Let us begin by recalling some known facts about symmetric tridiagonal matrices of the form

$$T := \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \beta_2 & \alpha_2 & \beta_3 & & \\ & \beta_3 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_m \\ & & & \beta_m & \alpha_m \end{bmatrix} \in \mathbb{R}^{m\times m}$$

satisfying (without loss of generality) $\beta_k \neq 0$ for all $k = 2, \ldots, m$. Furthermore, let

$$T_k := \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \beta_2 & \alpha_2 & \beta_3 & & \\ & \beta_3 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_k \\ & & & \beta_k & \alpha_k \end{bmatrix} \in \mathbb{R}^{k\times k}$$

be the leading $k \times k$ principal submatrix of T, and let

$$q_k(x) := \det\left(T_k - xI\right) \qquad \forall k = 1, \ldots, m$$

be the corresponding characteristic polynomial. Then the following recursion holds, cf. [4, p. 437]:

$$\begin{aligned} q_0(x) &:= 1, \\ q_1(x) &= \alpha_1 - x, \\ q_{k+1}(x) &= (\alpha_{k+1} - x)q_k(x) - \beta_{k+1}^2 q_{k-1}(x) \quad \forall k = 1, 2, \ldots, m-1. \end{aligned} \tag{2}$$

Furthermore, the next result is also well-known, see [4, Thm. 8.4.1] or [12, Section 5.6].

Theorem 2.1 *(Sturm Sequence Property)*
Assume that $\beta_k \neq 0$ for all $k = 2, \ldots, m$. Then the following statements hold:

(a) The eigenvalues of all principal submatrices T_k are real and simple.

(b) The eigenvalues of T_{k-1} strictly separate the eigenvalues of T_k in the sense that

$$\lambda_1(T_k) < \lambda_1(T_{k-1}) < \lambda_2(T_k) < \ldots < \lambda_{k-1}(T_k) < \lambda_{k-1}(T_{k-1}) < \lambda_k(T_k).$$

(c) Let $w(\lambda)$ denote the number of sign changes in the Sturm sequence $\{q_0(\lambda), q_1(\lambda), \ldots, q_m(\lambda)\}$ *(where we use the convention that vanishing entries $q_k(x) = 0$ are removed from this sequence before counting the sign changes). Then $w(\lambda)$ equals the number of eigenvalues of the matrix T that are strictly less than λ.*

An immediate consequence of the previous result is the following one which can be used to develop the well-known bisection method to compute single eigenvalues of symmetric tridiagonal matrices.

Corollary 2.2 *Let $a, b \in \mathbb{R}$ be given with $a < b$. Then $w(b) - w(a)$ is the number of eigenvalues of the symmetric tridiagonal matrix T lying in the interval $[a, b)$.*

We next want to give a lower bound for the smallest singular value of a given positive (semi-) definite (but asymmetric) matrix A in terms of the smallest eigenvalue of the corresponding symmtric part A^s. We suspect that this result is known, but were not able to find an explicit reference.

Lemma 2.3 *Let $A \in \mathbb{R}^{m\times m}$ be positive semidefinite (not necessarily symmetric). Then $\sigma_1(A) \geq \lambda_1(A^s) \geq 0$.*

Proof. For an arbitrary (not necessarily symmetric or positive definite) matrix A, we have

$$\min_{\|x\|=1} x^T A x = \min_{\|x\|=1} x^T A^s x = \lambda_1(A^s).$$

In particular, the assumed positive semidefiniteness of A implies the inequality $\lambda_1(A^s) \geq 0$.

In order to verify the first inequality, let us define the matrix $B := A - \lambda_1(A^s)I$. This definition implies

$$B^T + B = A^T + A - 2\lambda_1(A^s)I = 2 \cdot \big(A^s - \lambda_1(A^s)I\big).$$

Since $\lambda_1(A^s) \geq 0$ is the smallest eigenvalue of A^s, it follows that 0 is the smallest eigenvalue of B^s. The symmetry of $B^T + B$ therefore gives

$$\min_{\|x\|=1} x^T \big(B^T + B\big) x = 0.$$

Using the fact that the smallest eigenvalue of the symmetric matrix $A^T A$ is given by $(\sigma_1(A))^2$, cf. [3, Thm. 3.3], and taking into account the definition of the matrix B, we therefore obtain

$$\begin{aligned}
(\sigma_1(A))^2 &= \min_{\|x\|=1} x^T A^T A x \\
&= \min_{\|x\|=1} \left[\lambda_1(A^s)^2 x^T x + \lambda_1(A^s) \cdot x^T \left(B^T + B\right) x + x^T B^T B x\right] \\
&= \lambda_1(A^s)^2 + \min_{\|x\|=1} \left[\lambda_1(A^s) \cdot x^T \left(B^T + B\right) x + x^T B^T B x\right] \\
&\geq \lambda_1(A^s)^2 + \lambda_1(A^s) \cdot \min_{\|x\|=1} \left[x^T \left(B^T + B\right) x\right] + \min_{\|x\|=1} \left[x^T B^T B x\right] \\
&= \lambda_1(A^s)^2 + \min_{\|x\|=1} \left[x^T B^T B x\right] \\
&= \lambda_1(A^s)^2 + \left(\sigma_1\left(B\right)\right)^2
\end{aligned}$$

$$\geq \quad \lambda_1(A^s)^2.$$

Taking the square root and using the fact that $\sigma_1(A) \geq 0$ and (as already noted) $\lambda_1(A^s) \geq 0$, we obtain the desired statement. $\square$

Applying Gershgorin's Theorem to $\lambda_1(A^s)$ and using Lemma 2.3 gives the lower bound

$$\sigma_1(A) \geq \min_{i=1,\ldots,n} \left\{ a_{ii} - \frac{1}{2} \sum_{\substack{j=1 \\ j \neq i}}^{m} \left(a_{ij} + a_{ji} \right) \right\}$$

for the smallest singular value of a possibly nonsymmetric matrix A which is precisely the bound given in [7, Theorem 1].

Assume, for the moment, that $A \in \mathbb{R}^{m \times m}$ is symmetric positive definite. Then $\sigma_1(A) = \lambda_1(A) = \lambda_1(A^s)$, so that the inequality from Lemma 2.3 is actually an equality. Now, since both the singular values and the eigenvalues of A and A^s, respectively, depend continuously on the entries of the corresponding matrices, it follows that we still have $\sigma_1(A) \approx \lambda_1(A^s)$ for matrices A which are close to being symmetric, hence the estimates from Lemma 2.3 are likely to provide very sharp bounds in this case. Of course, this is not true for highly asymmetric matrices. However, later, in our applications, we have to deal with matrices which are close to being symmetric.

We next investigate some properties of the one-dimensional mapping

$$f : (0, \infty) \longrightarrow \mathbb{R},\ y \mapsto (\alpha - x) - \frac{\gamma^2}{y} \tag{3}$$

that will play an essential role in Section 3. Here α, γ, and x are given, whereas y is the variable. We are particularly interested in the properties of the corresponding fixed point iteration $y_{k+1} := f(y_k)$ for $k \in \mathbb{N}$. The following result gives all the necessary information.

Lemma 2.4 *Let $z := \alpha - x$. Choose an initial element $y_1 > 0$ and define $y_{k+1} := f(y_k)$ recursively for $k \in \mathbb{N}$. Then the following statements hold:*

Case $z \geq 2|\gamma|$: Here f has a repelling fixed point $f_1 := \frac{z - \sqrt{z^2 - 4\gamma^2}}{2}$ and an attracting fixed point $f_2 := \frac{z + \sqrt{z^2 - 4\gamma^2}}{2}$ which coincide for $z = \pm 2|\gamma|$.

(a) For $y_1 \in (f_1, f_2)$ we have

$$f_1 < y_1 < y_2 < y_3 < \ldots < y_k < y_{k+1} < \cdots < f_2$$

for all $k \in \mathbb{N}$. Furthermore, it holds that $\lim_{k \to \infty} y_k = f_2$.

(b) For $y_1 > f_2$ we have

$$f_2 < \ldots < y_{k+1} < y_k < \ldots < y_3 < y_2 < y_1$$

for all $k \in \mathbb{N}$. Furthermore, it holds that $\lim_{k \to \infty} y_k = f_2$.

(c) *For* $y_1 = f_2$ *we have* $y_k = f_2$ *for all* $k \in \mathbb{N}$.

(d) *For* $y_1 = f_1$ *we have* $y_k = f_1$ *for all* $k \in \mathbb{N}$.

(e) *For* $y_1 \in (0, f_1)$ *we have*

$$f_1 > y_1 > y_2 > y_3 > \dots$$

and there exists a smallest $k_0 \in \mathbb{N}$ *with* $y_{k_0} \leq 0$. *From that on, the sequence is no longer well-defined.*

Case $z < 2\,|\gamma|$: *Here* f *has no fixed points. We have* $y > f(y)$ *for all* $y > 0$, *and for every starting point* $y_1 > 0$, *we obtain*

$$y_1 > y_2 > y_3 > \dots,$$

and there is a smallest $k_0 \in \mathbb{N}$ *with* $y_{k_0} \leq 0$. *From that on, the sequence is no longer well-defined.*

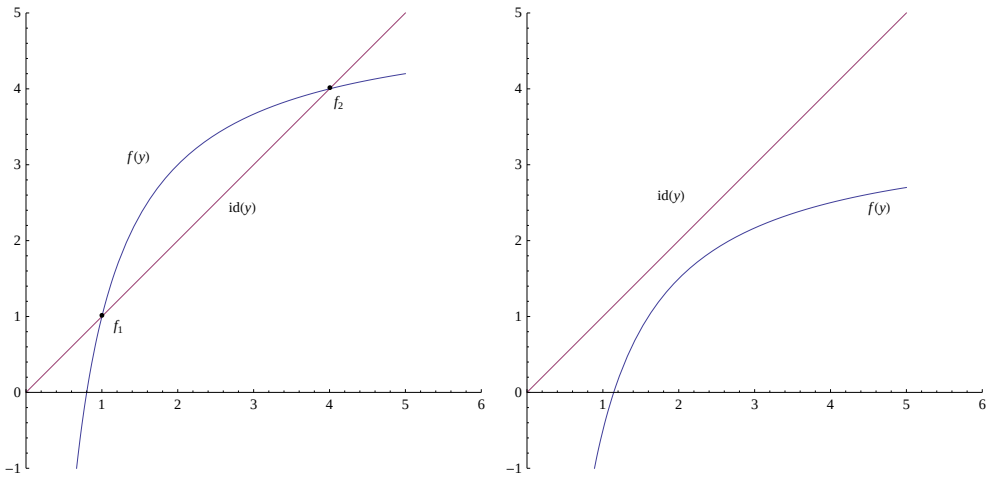

Figure 1: Illustration of Lemma 2.4, left: case 1, right: case 2

Instead of giving the simple proof, we illustrate this result in Figure 1. The left picture shows the first case where we have two (possibly identical) fixed points f_1 and f_2. When $f_1 < f_2$ (so the two fixed points do not coincide), then the derivative f' at the first fixed point is larger than one, hence this fixed point is repelling, whereas the derivative at the second fixed point is smaller than one, hence this fixed point is attracting. The right picture, on the other hand, illustrates the second case where $y > f(y)$ holds for all $y > 0$, so that no fixed points exist.

3 Estimates for the Extremal Eigenvalues

Here we investigate the symmetric tridiagonal matrix J from (1) and assume, without loss of generality, that $\beta \cdot \gamma \cdot \delta \neq 0$ and that $m \geq 4$, since otherwise J is not defined properly. Our aim in this section is to develop accurate estimates for the smallest and largest eigenvalue of J. For the special case where $\beta = \gamma = \delta$, the matrix J becomes a tridiagonal Toeplitz matrix whose eigenvalues are known explicitly and given by

$$\lambda_j = \alpha + 2\,|\gamma| \cos\left(\frac{j}{m+1}\pi\right) \qquad \forall j = 1, \ldots, m, \tag{4}$$

cf. [1, Thm 2.4].

Remark 3.1 Consider, for the moment, once again the special case $\beta = \gamma = \delta$ and let us denote the corresponding Toeplitz matrix by T. Then it follows from (4) that $\lambda_{\min}(T) = \alpha + 2|\gamma| \cos\left(\frac{m}{m+1}\pi\right)$ and $\lambda_{\max}(T) = \alpha + 2|\gamma| \cos\left(\frac{1}{m+1}\pi\right)$. In particular, for increasing dimension $m \to \infty$, we therefore get $\lambda_{\min}(T) \to \alpha - 2|\gamma|$ and $\lambda_{\max}(T) \to \alpha + 2|\gamma|$. For the general case, where β, γ, and δ are not necessarily equal, this still implies that for any bound of the form $\lambda_{\min}(J) \geq \alpha - K$ and $\lambda_{\max}(J) \leq \alpha + K$ for some suitable constant $K > 0$, we must have $K \geq 2|\lambda|$ if this bound should hold for all (sufficiently large) dimensions $m \in \mathbb{N}$. This observation follows from the previous fact by noting that we can reorder the entries of J by a symmetric permutation such the first $m - 2$ principal submatrices of J are Toeplitz matrices T of different dimensions, hence the claim follows from the interlacing property from Theorem 2.1 (b).

For the matrix J, which may be viewed as a (small) perturbation of the Toeplitz case, an analytic representation of the eigenvalues is not known. Our aim is therefore to obtain suitable lower and upper bounds for the extremal eigenvalues of J. Simple estimates can be obtained using Gershgorin's Theorem, see [4, Thm. 7.2.1], which implies that

$$\begin{aligned} \lambda_{\min}(J) &\geq \alpha - \max\big\{2|\gamma|, |\beta| + |\gamma|, |\delta| + |\gamma|\big\} \qquad \text{and} \\ \lambda_{\max}(J) &\leq \alpha + \max\big\{2|\gamma|, |\beta| + |\gamma|, |\delta| + |\gamma|\big\}. \end{aligned}$$

These estimates can be improved using suitable scalings of J, but it seems that the corresponding estimates are still worse than those that we develop in our subsequent theory.

To this end, let J_k be the principal $k \times k$ submatrix of J, and let

$$p_k(x) := \det(J_k - xI) \qquad \forall k = 1, 2, \ldots, m$$

be the corresponding characteristic polynomial. From (2), we obtain that these polynomials satisfy the following recursion:

$$\begin{aligned} p_0(x) &:= 1, \\ p_1(x) &= \alpha - x, \end{aligned}$$

$$\begin{aligned} p_2(x) &= (\alpha - x)\cdot p_1(x) - \beta^2 \cdot p_0(x), \\ p_k(x) &= (\alpha - x)\cdot p_{k-1}(x) - \gamma^2 \cdot p_{k-2}(x) \quad \forall k = 3, \ldots, m-1, \\ p_m(x) &= (\alpha - x)\cdot p_{m-1}(x) - \delta^2 \cdot p_{m-2}(x). \end{aligned} \tag{5}$$

Here, $p_m(x)$ is the characteristic polynomial of J.

In view of Theorem 2.1, the characteristic polynomials $p_k(x)$ have real and single roots. Furthermore, given an arbitrary $\alpha \in \mathbb{R}$, the number $w(\alpha)$, denoting the number of sign changes in the Sturm sequence $p_0(\alpha), p_1(\alpha), \ldots, p_m(\alpha)$, is equal to the number of roots of $p_m(x)$ which are smaller than α.

Based on the above recursion and a simple induction argument, we can easily deduce that the polynomials p_k are symmetric in the following sense:

$$p_k(\alpha - y) = \begin{cases} p_k(\alpha + y), & \text{if } k \text{ is even,} \\ -p_k(\alpha + y), & \text{if } k \text{ is odd.} \end{cases} \tag{6}$$

In particular, for $k = m$, this implies that $\alpha - y$ is an eigenvalue of J if and only if $\alpha + y$ is an eigenvalue of J, hence the eigenvalues of J are distributed symmetrically around the point α. Consequently, $\alpha - K$ is a lower bound for the smallest eigenvalue (for some $K > 0$) if and only if $\alpha + K$ is an upper bound for the largest eigenvalue. Hence we only have to find suitable lower bounds for the smallest eigenvalue of J.

The basic idea to find suitable estimates of K is the following: We will find conditions (on K and, sometimes, also on the dimension m) which guarantee that all the numbers $p_k(\alpha - K)$ have the same sign which is equivalent to saying that all these numbers are positive since $K > 0$ is equivalent to $p_1(\alpha - K) > 0$. Then it follows from our previous considerations that $w(\alpha - K) = 0$, hence all zeros of p_m must be greater or equal to $\alpha - K$. However, since $p_m(\alpha - K) > 0$, all zeros must actually be greater than $\alpha - K$.

Instead of studying the Sturm sequence $\{p_1(x), p_2(x), \ldots, p_m(x)\}$ directly, we consider the quotients

$$r_k(x) := \frac{p_{k+1}(x)}{p_k(x)} \qquad \forall k = 1, 2, \ldots, m-1. \tag{7}$$

Using the recursion of the polynomials $p_k(x)$ in (5), we obtain the corresponding recursion

$$r_1(x) = \frac{(\alpha - x)^2 - \beta^2}{(\alpha - x)}, \tag{8}$$

$$r_{k+1}(x) = (\alpha - x) - \gamma^2 \frac{p_k(x)}{p_{k-1}(x)} = (\alpha - x) - \frac{\gamma^2}{r_k(x)} \quad \text{for } k = 1, 2, \ldots, m-3, \text{ and} \tag{9}$$

$$r_{m-1}(x) = (\alpha - x) - \frac{\delta^2}{r_{m-2}(x)} \qquad \text{for } k = m-2.$$

Based on these quotients, we have the following criterion.

Proposition 3.2 *Let $x < \alpha$. Then every member of the Sturm sequence $p_1(x), \ldots, p_m(x)$ is positive if and only if $r_k(x)$ is positive for all $k = 1, \ldots, m-2$ and $r_{m-2}(x) > h(x)$*

holds, where

$$h\left(x\right) := \frac{\delta^2}{\alpha - x}. \tag{10}$$

Proof. First suppose that all numbers $p_1(x), \ldots, p_m(x)$ are positive. Then (7) immediately implies $r_k(x) > 0$ for all $k = 1, \ldots, m-2$ (and also for $k = m-1$, but this part is not needed for our assertion). Furthermore, since $p_{m-2}(x) > 0$, we have the following equivalences that will also be used in order to verify the converse direction:

$$\begin{aligned} p_m\left(x\right) > 0 \quad &\iff \quad \left(\alpha - x\right) p_{m-1}\left(x\right) - \delta^2 p_{m-2}\left(x\right) > 0 \\ &\iff \quad \left(\alpha - x\right) p_{m-1}\left(x\right) > \delta^2 p_{m-2}\left(x\right) \\ &\overset{p_{m-2}(x)>0}{\iff} \quad \left(\alpha - x\right) \tfrac{p_{m-1}(x)}{p_{m-2}(x)} > \delta^2 \\ &\iff \quad \tfrac{p_{m-1}(x)}{p_{m-2}(x)} > \tfrac{\delta^2}{\alpha - x} \\ &\iff \quad r_{m-2}(x) > h(x). \end{aligned} \tag{11}$$

Since, in the proof of this direction, we have $p_m(x) > 0$, the above chain of equivalences therefore gives $r_{m-2}(x) > h(x)$.

Conversely, assume that $r_1(x), \ldots, r_{m-2}(x)$ are all positive, and that, in addition, we have $r_{m-2}(x) > h(x)$. Since the recursion (7) implies

$$p_{k+1}(x) = r_k(x) p_k(x) \quad \forall k = 1, \ldots, m-1,$$

and since we have $p_1(x) = \alpha - x > 0$ by assumption, we immediately obtain $p_{k+1}(x) > 0$ for all $k = 1, \ldots, m-2$. In particular, we therefore have $p_{m-2}(x) > 0$. The chain of equivalences (11) then shows that we also have $p_m(x) > 0$. □

Note that, in the statement of Proposition 3.2, we could alternatively require the positivity of $r_k(x)$ only for all $k = 1, \ldots, m-3$ since $r_{m-2}(x) > 0$ follows directly from the additional condition $r_{m-2}(x) > h(x)$ due to the fact that $h(x)$ is positive in view of the assumption that $x < \alpha$. We further note that it is indeed enough to consider the positivity of the Sturm sequence $\{p_1(x), \ldots, p_m(x)\}$ instead of $\{p_0(x), p_1(x), \ldots, p_m(x)\}$ since $p_0(x) \equiv 1$ is positive by definition and therefore does not imply additional sign changes.

The interesting part of Proposition 3.2 is the fact that we can characterize the positivity of all members from the Sturm sequence $p_1(x), \ldots, p_m(x)$ in terms of the quotients $r_1(x), \ldots, r_{m-2}(x)$ (together with the function $h(x)$ from (10)). Hence the quotient $r_{m-1}(x)$ is not needed in this characterization which is important since the recursion for r_{m-1} is different from the recursion of all the other quotients $r_k(x)$.

This observation is also useful from the following point of view: We will sometimes consider the dimension m of the given matrix J to be variable, i.e., we consider matrices of the form J with different dimensions. Now, the polynomials p_k and, therefore, also the quotients r_k obviously depend on the dimension of J. However, taking into account the particular structure of J, it follows immediately that, for two diffent dimensions m and $\tilde{m}$ with $m < \tilde{m}$, the quotients $r_k(x)$ $(k = 1, 2, \ldots, m-2)$ are the same for both matrices.

	$r_1(x) < f_1(x)$	$r_1(x) \in \big(f_1(x), f_2(x)\big)$	$r_1(x) > f_2(x)$
$h(x) < f_1(x)$	$\forall m \leq m_0$	$\forall m \in \mathbb{N}$	$\forall m \in \mathbb{N}$
$h(x) \in \big(f_1(x), f_2(x)\big)$	never	$\forall m \geq m_0$	$\forall m \in \mathbb{N}$
$h(x) > f_2(x)$	never	never	$\forall m \leq m_0$

Table 1: Lower bounds x for $\lambda_{\min}(J)$ depending on the sizes of $r_1(x)$ and $h(x)$

We now take a closer look at the recursion (9). The initial element $r_1(x)$ is given by (8), whereas the recursion itself can be written as

$$r_{k+1}(x) = f\big(r_k(x)\big) \qquad \forall k = 1, 2, \ldots, m-3$$

by using the function f from (3). The (fixed point) properties of the mapping f were already discussed in Lemma 2.4. In particular, it follows from this result that, in the only interesting case $x \leq \alpha - 2|\gamma|$, there are two fixed points f_1 and f_2, with f_1 being a repelling fixed point and f_2 being an attracting fixed point. Since these fixed points depend on the given x, we denote them by $f_1(x)$ and $f_2(x)$ from now on. In view of Proposition 3.2 we want the sequence $r_1(x), \ldots, r_{m-3}(x)$ (note that $r_k(x)$ plays the role of y_k in Lemma 2.4) to be positive and, in addition, $r_{m-2}(x) > h(x)$. Obviously, whether these relations hold depend on how the starting point $y_1 = r_1(x)$ and the number $h(x) > 0$ are related to the fixed points of f.

In fact, using Proposition 3.2 and Lemma 2.4, we have the situation from Table 1 whose entries will be explained immediately.

This table assumes (implicitly) that $x \leq \alpha - 2|\gamma|$ and shows in which situation and under which conditions the given x provides a lower bound for the smallest eigenvalue of the matrix J.

More precisely, the table has the following meaning: There are nine cases depending on whether $h(x)$ is smaller than the fixed point $f_1(x)$ or strictly between the two fixed points $f_1(x)$ and $f_2(x)$ or larger than $f_2(x)$, and whether the quotient $r_1(x)$ is smaller than $f_1(x)$, between $f_1(x)$ and $f_2(x)$ or larger than $f_2(x)$. For simplicity of presentation, we do not consider the (often trivial) cases where $h(x)$ or $r_1(x)$ are equal to one of the two fixed points. Then the entry "never" indicates that the given x does not provide a lower bound on the smallest eigenvalue of J regardless of the dimension m of J. The entry "for all $m \in \mathbb{N}$" indicates that the given x is a lower bound of the smallest eigenvalue of J for all dimensions $m \in \mathbb{N}$, whereas the entries "$\forall m \leq m_0$" and "$\forall m \geq m_0$" indicate that the given x provides a lower bound on the smallest eigenvalue of J for all sufficiently small and all sufficiently large m, respectively.

We still have to explain how these entries were obtained. We do not consider all nine cases since the argument is often the same, but let us take a closer view at some of these cases. First, consider the case $h(x) < f_1(x)$ and $r_1(x) < f_1(x)$. Lemma 2.4 then implies that the sequence $r_k(x)$ is monotonically decreasing and eventually becomes negative. Hence, only the first few elements of this sequence are positive, and the additional requirement

$r_{m-2}(x) > h(x)$ can therefore also hold only for sufficiently small (possibly no) dimensions m. In view of Proposition 3.2, it therefore follows that the given x is a lower bound for the smallest eigenvalue of J only for all sufficiently small m, i.e., for all $m \leq m_0$ with some $m_0 \in \mathbb{N}$. This explains the corresponding entry in the upper left corner of Table 1.

Next, consider the case $h(x) < f_1(x)$ and $r_1(x) \in \big(f_1(x), f_2(x)\big)$. Then Lemma 2.4 shows that the sequence $r_k(x)$ is monotonically increasing and converges to the fixed point $f_2(x)$. In particular, $r_k(x)$ is positive for all k, and $r_k(x) > h(x)$ holds for all $k \in \mathbb{N}$, especially, this holds for $k = m - 2$ for any given dimension $m \in \mathbb{N}$. Hence it follows from Proposition 3.2 that the given x provides a lower bound on the smallest eigenvalue of J for all dimensions $m \in \mathbb{N}$ which again explains the corresponding entry in this case.

Now consider the case $h(x) \in \big(f_1(x), f_2(x)\big)$ and $r_1(x) < f_1(x)$. Then Lemma 2.4 shows, in particular, that all quotients $r_k(x)$ stay less than $f_1(x)$, so that the condition $r_{m-2}(x) > h(x)$ never holds in this case regardless of the dimension $m \in \mathbb{N}$. Hence Proposition 3.2 implies that the given x does not provide a lower bound for the smallest eigenvalue for any dimension $m \in \mathbb{N}$.

Finally, consider the case $h(x) \in \big(f_1(x), f_2(x)\big)$ and $r_1(x) \in \big(f_1(x), f_2(x)\big)$. Lemma 2.4 then implies that the sequence $r_1(x), r_2(x), r_3(x), \ldots$ is monotonically increasing and converges to the fixed point $f_2(x)$. Hence, all these quotients are positive, and eventually they are larger than the number $h(x)$. In particular, for sufficiently large dimensions $m \in \mathbb{N}$, we have $r_{m-2}(x) > h(x)$. Hence Proposition 3.2 shows that x is a lower bound for J's smallest eigenvalue for all sufficiently large dimensions m. In addition, the following note holds for this case (which is of particular interest in our further development).

Remark 3.3 Consider once again the case $h(x) \in \big(f_1(x), f_2(x)\big)$ and $r_1(x) \in \big(f_1(x), f_2(x)\big)$. Then it is possible that we already have $r_1(x) > h(x)$. Using a similar reasoning as before, this implies $r_{m-2}(x) > h(x)$ for *all* dimensions $m \in \mathbb{N}$. Consequently, and in addition to the corresponding entry in Table 1, it follows from Proposition 3.2 that x is a lower bound for the smallest eigenvalue of J for *all* dimensions $m \in \mathbb{N}$. — We further note that the condition $r_1(x) > h(x)$ is equivalent to $x < \alpha - \sqrt{\beta^2 + \delta^2}$ (provided that $x < \alpha$).

All the other entries in the Table 1 follow by a similar reasoning. Now, it is clear how to proceed. The previous table gives clear statements on how to get lower bounds for the smallest eigenvalue of J in terms of $r_1(x)$ and $h(x)$ compared to the two fixed points $f_1(x)$ and $f_2(x)$. Our aim is therefore to re-interpret these conditions in terms of the original data of the matrix J. The following technical result investigates these data and shows how $r_1(x)$ is related to the fixed points $f_1(x)$ and $f_2(x)$ depending on the relation between the data β and γ of the matrix J whose dimension m is fixed in this lemma.

Lemma 3.4 *Let $x \leq \alpha - 2\,|\gamma|$, $f_1\,(x)$ and $f_2\,(x)$ be the fixed points of the function f, and let $r_1\,(x)$ be the first quotient from (8). Then the following statements hold:*

1. *Case $|\beta| > \sqrt{2}\,|\gamma|$: Then $r_1\,(x) > f_1\,(x) \iff x < \alpha - \frac{\beta^2}{\sqrt{\beta^2 - \gamma^2}}$, whereas the inequality $r_1\,(x) < f_2\,(x)$ always holds (i.e. this inequality holds for all $x \leq \alpha - 2|\gamma|$).*

2. *Case* $|\beta| = \sqrt{2}\,|\gamma|$*: Then* $f_1(x) < r_1(x) < f_2(x)$ *holds for all* $x < \alpha - 2\,|\gamma|$ *(whereas at the boundary point* $x = \alpha - 2\,|\gamma|$*, we have* $r_1(x) = f_1(x) = f_2(x)$*.)*

3. *Case* $|\gamma| < |\beta| < \sqrt{2}\,|\gamma|$*: Then* $r_1(x) > f_1(x)$ *always holds (i.e. for all* $x \leq \alpha - 2|\gamma|$*), and* $r_1(x) < f_2(x) \iff x < \alpha - \frac{\beta^2}{\sqrt{\beta^2-\gamma^2}}$.

4. *Case* $|\beta| \leq |\gamma|$*: Then we always have* $r_1(x) > f_2(x) \geq f_1(x)$.

Proof. To simplify the notation, we set $z := \alpha - x$, so that $x \leq \alpha - 2\,|\gamma|$ is equivalent to $z \geq 2\,|\gamma|$. We divide the (technical, but completely elementary) proof into three steps: Part (A) contains some facts that will be used in the subsequent parts. In part (B), we study how $r_1(x)$ and $f_1(x)$ relate to each other, and in part (C) we study the relation between $r_1(x)$ and $f_2(x)$.

(A) We start by finding candidates $x \leq \alpha - 2\,|\gamma|$ for which $r_1(x) = f_1(x)$ or $r_1(x) = f_2(x)$ is possible. We do this simultaneously for both fixed points. To this end, note that

$$\begin{aligned} r_1(x) = f_{1/2}(x) &\iff \frac{z^2-\beta^2}{z} = \frac{1}{2}\left(z \mp \sqrt{z^2-4\gamma^2}\right) \\ &\overset{z>0}{\iff} 2z^2 - 2\beta^2 = z^2 \mp z\sqrt{z^2-4\gamma^2} \\ &\iff z^2 - 2\beta^2 = \mp z\sqrt{z^2-4\gamma^2} \\ &\implies z^4 - 4\beta^2 z^2 + 4\beta^4 = z^2\left(z^2-4\gamma^2\right) \\ &\iff \beta^4 = z^2\left(\beta^2-\gamma^2\right). \end{aligned} \tag{12}$$

Hence, for both equations, we have the same necessary condition. It is satisfied for $z_1 = -\frac{\beta^2}{\sqrt{\beta^2-\gamma^2}}$ and $z_2 = +\frac{\beta^2}{\sqrt{\beta^2-\gamma^2}}$. Therefore, the possible candidates are $x_1 = \alpha + \frac{\beta^2}{\sqrt{\beta^2-\gamma^2}}$ and $x_2 = \alpha - \frac{\beta^2}{\sqrt{\beta^2-\gamma^2}}$. We call them the roots of the above equation. Using (12), it is also clear that there are no (real) roots if $|\beta| = |\gamma|$ or $|\beta| < |\gamma|$.

(B) Here we discuss the relation between $r_1(x)$ and the fixed point $f_1(x)$ in terms of the orginal data of the matrix J. To this end, first note that, since $\alpha - x \geq 2|\gamma| > 0$, we have

$$\begin{aligned} r_1(x) > f_1(x) &\iff \frac{(\alpha-x)^2-\beta^2}{(\alpha-x)} > \frac{1}{2}\left((\alpha-x) - \sqrt{(\alpha-x)^2-4\gamma^2}\right) \\ &\iff 2\beta^2 < (\alpha-x)^2 + (\alpha-x)\sqrt{(\alpha-x)^2-4\gamma^2}, \end{aligned}$$

and the last inequality obviously holds for all sufficiently small x. We call this observation (O1).

At this point, we have to discuss several cases:

1. Let $|\beta| > \sqrt{2}\,|\gamma|$. In this case, it turns out, by inserting the two candidate points x_1 and x_2, that x_2 is the only root in the interval $(-\infty, \alpha - 2|\gamma|)$. Together with

observation (O1), it therefore follows that $r_1(x) > f_1(x)$ if $x < x_2$, whereas we have $r_1(x) < f_1(x)$ if $x > x_2$.

2. Let $|\beta| = \sqrt{2}\,|\gamma|$. Here we can write $x_{1/2} = \alpha \pm 2\,|\gamma|$. A simple calculation shows that both candidates are indeed roots. Since $x_2 < x_1$, we obtain from observation (O1) that $r_1(x) > f_1(x)$ for $x < x_2$ (note that $x \leq \alpha - 2\,|\gamma| = x_2$ was the prerequisite of this lemma, so the case $x > x_2$ does not occur), whereas we have $r_1(x) = f_1(x)$ at $x = x_2$ since x_2 is a root of our equation.

3. Let $|\gamma| < |\beta| < \sqrt{2}\,|\gamma|$. Here it is easy to see that x_1 is the only root among the two candidates. Together with observation (O1), we therefore get $r_1(x) > f_1(x)$ for $x < x_1$. However, in this case, we have $x_1 > \alpha - 2\,|\gamma|$. Hence, we obtain $r_1(x) > f_1(x)$ for all $x \leq \alpha - 2\,|\gamma|$.

4. Let $|\beta| \leq |\gamma|$. In this case, there are no roots in view of part (A). It therefore follows from observation (O1) that $r_1(x) > f_1(x)$ holds for all $x \leq \alpha - 2|\gamma|$.

(C) Here we discuss the relation between $r_1(x)$ and the fixed point $f_2(x)$, again in terms of the orginal data of the matrix J. The considerations are similar to those from part (B). To this end, we first note that $r_1(x) > f_2(x)$ is equivalent to $2\beta^2 < z^2\left(1 - \sqrt{1 - 4\gamma^2/z^2}\right) =: g(z)$. Using l'Hospital's rule, we obtain

$$\lim_{z\to\infty} g(z) = \lim_{z\to\infty} \frac{\left(1 - \sqrt{1 - 4\gamma^2/z^2}\right)}{z^{-2}} = \lim_{z\to\infty} \frac{2\gamma^2}{\sqrt{1 - 4\gamma^2/z^2}} = 2\gamma^2\,.$$

Taking into account that $z = \alpha - x$, it follows that $r_1(x) > f_2(x)$ for all x sufficiently small if $|\beta| < |\gamma|$. Similarly, one can show that $r_1(x) < f_2(x)$ for all x sufficiently small if $|\beta| > |\gamma|$ (whereas the case $|\beta| = |\gamma|$ has to be treated separately). We call these statements observation (O2).

Like before, we proceed by considering several cases:

1. Let $|\beta| > \sqrt{2}\,|\gamma|$. Through simple calculation, we get that x_1 is the only root. Observation (O2) therefore implies $r_1(x) < f_2(x)$ for all $x < x_1$. But since $x_1 > \alpha - 2\,|\gamma|$ we even have $r_1(x) < f_2(x)$ for all $x \leq \alpha - 2\,|\gamma|$.

2. Let $|\beta| = \sqrt{2}\,|\gamma|$. Here $x_{1/2} = \alpha \pm 2\,|\gamma|$ are the two roots, but x_1 is greater than $\alpha - 2\,|\gamma|$ and hence irrelevant for our case. Using observation (O2) once again, we get $r_1(x) < f_2(x)$ for all $x < \alpha - 2\,|\gamma|$ as in the previous case, whereas we have $r_1(x) = f_2(x)$ at the boundary point $x = x_2 = \alpha - 2|\gamma|$.

3. Let $|\gamma| < |\beta| < \sqrt{2}\,|\gamma|$. Then x_2 is the only root, and observation (O2) gives $r_1(x) < f_2(x)$ if and only if $x < x_2$.

4. Let $|\beta| = |\gamma|$. We know from part (A) that there are no roots in this case, hence either $r_1(x) < f_2(x)$ or $r_1(x) > f_2(x)$ holds for all $x \leq \alpha - 2|\gamma|$. To decide which of these

two inequalities holds, observation (O2) cannot be applied directly. However, direct calculation shows that $2\beta^2 = 2\gamma^2 < 4\gamma^2 = g(2|\gamma|)$, so that observation (O2) now gives $r_1(\alpha - 2|\gamma|) > f_2(\alpha - 2|\gamma|)$. Hence $r_1(x) > f_2(x)$ holds for all $x \leq \alpha - 2|\gamma|$.

5. Let $|\beta| < |\gamma|$. According to part (A), there are no roots in this case. Together with observation (O2), it follows that $r_1(x) > f_2(x)$ holds for all $x \leq \alpha - 2|\gamma|$.

The statement now follows by summarizing all subcases considered in parts (B) and (C). □

The following result is similar to the previous one (so we skip its proof) and shows how the number $h(x)$ is related to the two fixed points $f_1(x)$ and $f_2(x)$ depending on the original data δ and γ of our matrix J whose dimension m is again assumed to be fixed.

Lemma 3.5 *Let $x \leq \alpha - 2|\gamma|$, $f_1(x)$ and $f_2(x)$ be the fixed points of the function f, and let $h(x)$ be defined by (10). Then the following statements hold:*

1. *Case $|\delta| > \sqrt{2}|\gamma|$: Then $h(x) < f_2(x) \iff x < \alpha - \frac{\delta^2}{\sqrt{\delta^2-\gamma^2}}$, whereas the inequality $h(x) > f_1(x)$ always holds (i.e. for all $x \leq \alpha - 2|\gamma|$).*
2. *Case $|\delta| = \sqrt{2}|\gamma|$: Then $f_1(x) < h(x) < f_2(x)$ for all $x < \alpha - 2|\gamma|$ (and $h(x) = f_1(x) = f_2(x)$ for the boundary point $x = \alpha - 2|\gamma|$).*
3. *Case $|\gamma| < |\delta| < \sqrt{2}|\gamma|$: Then $h(x) > f_1(x) \iff x < \alpha - \frac{\delta^2}{\sqrt{\delta^2-\gamma^2}}$, whereas the inequality $h(x) < f_2(x)$ always holds (i.e. for all $x \leq \alpha - 2|\gamma|$).*
4. *Case $|\delta| \leq |\gamma|$: Then we always have $h(x) < f_1(x) \leq f_2(x)$.*

Now we are going to combine the previous results in order to get estimates for the extremal eigenvalues of the matrix J. We stress, however, that it cannot be avoided that these bounds (in addition to the data of the matrix) sometimes also depend on the dimension m of this matrix, cf. Table 1 and the discussion to derive the entries of this table.

Unfortunately, we have to distinguish several cases in the presentation of our main result. In view of Lemmas 3.4 and 3.5, there are actually 16 different cases to consider, namely those that occur by combining the four possibilities

$$|\beta| > \sqrt{2}|\gamma|, \;\; |\beta| = \sqrt{2}|\gamma|, \;\; |\beta| \in \big(|\gamma|, \sqrt{2}|\gamma|\big), \text{ and } |\beta| \leq |\gamma|$$

from Lemma 3.4 with the corresponding four possibilities

$$|\delta| > \sqrt{2}|\gamma|, \;\; |\delta| = \sqrt{2}|\gamma|, \;\; |\delta| \in \big(|\gamma|, \sqrt{2}|\gamma|\big), \text{ and } |\delta| \leq |\gamma|$$

from Lemma 3.5.

Theorem 3.6 *Define* $\bar{\beta} := \frac{\beta^2}{\sqrt{\beta^2-\gamma^2}}$ *and* $\bar{\delta} := \frac{\delta^2}{\sqrt{\delta^2-\gamma^2}}$. *Then the inequalities*

$$\lambda_{\min}(J) \geq \alpha - K \qquad \text{and} \qquad \lambda_{\max}(J) \leq \alpha + K$$

holds

(a) for all *dimensions* $m \in \mathbb{N}$ *with* K *being the constant from the following table:*

	$\lvert\delta\rvert > \sqrt{2}\lvert\gamma\rvert$	$\lvert\delta\rvert = \sqrt{2}\lvert\gamma\rvert$	$\lvert\delta\rvert \in \big(\lvert\gamma\rvert, \sqrt{2}\lvert\gamma\rvert\big)$	$\lvert\delta\rvert \leq \lvert\gamma\rvert$
$\lvert\beta\rvert > \sqrt{2}\lvert\gamma\rvert$	$\sqrt{\beta^2+\delta^2}$	$\sqrt{\beta^2+\delta^2}$	$\max\left\{\bar{\beta}, \sqrt{\beta^2+\delta^2}\right\}$	$\bar{\beta}$
$\lvert\beta\rvert = \sqrt{2}\lvert\gamma\rvert$	$\sqrt{\beta^2+\delta^2}$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$
$\lvert\beta\rvert \in \big(\lvert\gamma\rvert, \sqrt{2}\lvert\gamma\rvert\big)$	$\max\left\{\bar{\delta}, \sqrt{\beta^2+\delta^2}\right\}$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$
$\lvert\beta\rvert \leq \lvert\gamma\rvert$	$\bar{\delta}$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$

(b) for all sufficiently large *dimensions* $m \in \mathbb{N}$ *with the (usually sharper) constant* K *from the following table:*

	$\lvert\delta\rvert > \sqrt{2}\lvert\gamma\rvert$	$\lvert\delta\rvert = \sqrt{2}\lvert\gamma\rvert$	$\lvert\delta\rvert \in \big(\lvert\gamma\rvert, \sqrt{2}\lvert\gamma\rvert\big)$	$\lvert\delta\rvert \leq \lvert\gamma\rvert$
$\lvert\beta\rvert > \sqrt{2}\lvert\gamma\rvert$	$\max\{\bar{\beta}, \bar{\delta}\}$	$\bar{\beta}$	$\bar{\beta}$	$\bar{\beta}$
$\lvert\beta\rvert = \sqrt{2}\lvert\gamma\rvert$	$\bar{\delta}$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$
$\lvert\beta\rvert \in \big(\lvert\gamma\rvert, \sqrt{2}\lvert\gamma\rvert\big)$	$\bar{\delta}$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$
$\lvert\beta\rvert \leq \lvert\gamma\rvert$	$\bar{\delta}$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$	$2\lvert\gamma\rvert$

Proof. In view of our previous observation, $\alpha - K$ is a lower bound for $\lambda_{\min}(J)$ if and only if $\alpha + K$ is an upper bound for $\lambda_{\max}(J)$ for some $K > 0$. Hence it is enough to verify the lower bounds for the minimum eigenvalue of J. We further note that, in view of Remark 3.1, we (have to) assume throughout this proof that $x \leq \alpha - 2\,|\gamma|$ since there cannot be a lower bound greater than $\alpha - 2\,|\gamma|$ that fits for all (sufficiently large) matrix sizes m.

We begin by stating some elementary inequalities (without proof) that are useful for the subsequent considerations:

(a) If $|\beta| > |\gamma|$, then $\bar{\beta} \geq 2|\gamma|$ and $\bar{\beta} = 2|\gamma|$ holds if and only if $|\beta| = \sqrt{2}|\gamma|$.

(b) If $|\delta| > |\gamma|$, then $\bar{\delta} \geq 2|\gamma|$ and $\bar{\delta} = 2|\gamma|$ holds if and only if $|\delta| = \sqrt{2}|\gamma|$.

(c) If $|\beta| \geq \sqrt{2}|\gamma|$ and $|\delta| \geq \sqrt{2}|\gamma|$, then $\sqrt{\beta^2+\delta^2} \geq \max\{2|\gamma|, \bar{\beta}, \bar{\delta}\}$.

(d) If $|\beta| = \sqrt{2}|\gamma|$ and $|\delta| \in \big(|\gamma|, \sqrt{2}|\gamma|\big)$, then $\bar{\delta} \geq \sqrt{\beta^2+\delta^2}$.

(e) If $|\beta| \in \big(|\gamma|, \sqrt{2}|\gamma|\big)$ and $|\delta| = \sqrt{2}|\gamma|$, then $\bar{\beta} \geq \sqrt{\beta^2+\delta^2}$.

We now verify statements (a) and (b) simultaneously. In principle, we have to consider each of the possible 16 cases separately. However, it will be enough to consider only one of these cases (in fact, one of the more interesting ones), since the remaining cases can be treated in essentially the same way by referring to the corresponding cases from Lemmas 3.4 and 3.5 as well as to the corresponding entries of Table 1.

The case that we consider in more detail is the one where $|\beta| > \sqrt{2}|\gamma|$ and $|\delta| > \sqrt{2}|\gamma|$ holds. Then Lemma 3.4 shows that $r_1(x) < f_2(x)$ holds for all $x \leq \alpha - 2|\gamma|$, whereas $r_1(x) > f_1(x)$ is equivalent to $x < \alpha - \bar{\beta}$. Moreover, Lemma 3.5 shows that $h(x) > f_1(x)$ holds for all $x \leq \alpha - 2|\gamma|$, whereas we have $h(x) < f_2(x)$ if and only if $x < \alpha - \bar{\delta}$. Table 1 therefore shows that, for all $x < \min\{\alpha - \bar{\beta}, \alpha - \bar{\delta}\}$ and all $x \leq \alpha - 2|\gamma|$, this x provides a lower bound for $\lambda_{\min}(J)$ provided that the dimension m is sufficiently large. By continuity, we therefore get the lower bound

$$\lambda_{\min}(J) \geq \min\{\alpha - 2|\gamma|, \alpha - \bar{\beta}, \alpha - \bar{\delta}\} = \alpha - \max\{2|\gamma|, \bar{\beta}, \bar{\delta}\}$$

for all $m \in \mathbb{N}$ sufficiently large. Using observations (a) and (b), this lower bound reduces to

$$\lambda_{\min}(J) \geq \alpha - \max\{\bar{\beta}, \bar{\delta}\}.$$

This is precisely the lower bound given for the case considered here in statement (b).

However, in this particular case, we can also apply Remark 3.3 and obtain a lower bound for $\lambda_{\min}(J)$ for *all* dimensions $m \in \mathbb{N}$ if, in addition, x is chosen in such a way that $r_1(x) > h(x)$. Since this condition is equivalent to $x < \alpha - \sqrt{\beta^2 + \delta^2}$ according to Remark 3.3, it follows, together with our previous considerations, that the lower bound

$$\lambda_{\min}(J) \geq \alpha - \max\left\{2|\gamma|, \bar{\beta}, \bar{\delta}, \sqrt{\beta^2 + \delta^2}\right\}$$

holds for all dimensions $m \in \mathbb{N}$. In view of observation (c), this lower bound boils down to

$$\lambda_{\min}(J) \geq \alpha - \sqrt{\beta^2 + \delta^2}$$

and therefore justifies the corresponding bound given in statement (a). □

We close this section with some remarks about the previous result.

Remark 3.7 (a) Except for the trivial case $|\beta|, |\delta| \leq |\gamma|$, our bounds on the extremal eigenvalues of the matrix J are better than those that come from Gershgorin's Theorem.

(b) The case $|\beta| = |\delta| = \sqrt{2}|\gamma|$ gives $\alpha - 2|\gamma|$ and $\alpha + 2|\gamma|$ as lower and upper bounds for $\lambda_{\min}(J)$ and $\lambda_{\max}(J)$, respectively. However, in this case these bounds are exact, i.e. $\lambda_{\min}(J) = \alpha - 2|\gamma|$ and $\lambda_{\max}(J) = \alpha - 2|\gamma|$. This follows from the recursion (5) which, in this case, gives for $p_0(x) = 1, p_1(x) = 2|\gamma|, p_2(x) = 2|\gamma|^2, p_k(x) = 2|\gamma|^k$ for all $k = 3, \ldots, m-1$ and $p_m(x) = 0$ for $x := \alpha - 2|\gamma|$.

(c) Consider a matrix of the form

$$A = \begin{bmatrix} \alpha & \bar{\beta} & & & & \\ \hat{\beta} & \alpha & \bar{\gamma}_3 & & & \\ & \hat{\gamma}_3 & \alpha & \ddots & & \\ & & \ddots & \ddots & \bar{\gamma}_{m-1} & \\ & & & \hat{\gamma}_{m-1} & \alpha & \bar{\delta} \\ & & & & \hat{\delta} & \alpha \end{bmatrix}.$$

Define $\beta, \gamma \in \mathbb{R}$ in such a way that $\hat{\beta} \times \bar{\beta} = \beta^2$ and $\hat{\delta} \times \bar{\delta} = \delta^2$. Suppose that there is an element $\gamma \in \mathbb{R}$ with $\hat{\gamma}_i \times \bar{\gamma}_i = \gamma^2$ for all $i = 3, \ldots, m-1$. Then the characteristic polynomials of all principal submatrices of A coincide with the polynomials $p_k(x)$ from (5). Consequently, all the previous considerations for the matrix J also hold for the nonsymmetric matrix A. In particular, the same bounds for the extremal eigenvalues are valid for A.

(d) Statement (b) of Theorem 3.6 holds only for all sufficiently large dimensions m, say, for all $m \geq m_0$. Here, the smallest dimension m_0 can be computed in the following way: We are in the situation where $r_1(x) > f_1(x)$ and $h(x) < f_2(x)$ for $x = \alpha - K$, K the bound given in the tables of Theorem 3.6. Then the sequence $r_1(x), r_2(x), r_3(x), \ldots$ is monotonically increasing and converges to $f_2(x)$. So there exists a smallest integer s such that $r_s(x) > h(x)$. Then $m_0 = s + 2$ is the required dimension since $r_s(x)$ determines the behaviour of $p_{s+2}(x)$. Hence we need to compute $r_1(x)$ and $h(x)$ as well as (if still necessary) the other quotients $r_k(x)$ for $k \geq 2$ via the corresponding recursion (9) until, for the first time, $r_s(x)$ is greater than $h(x)$.

(e) Theorem 3.6 (b) shows that $\lambda_{\min}(J) \geq \alpha - K$ holds for all dimensions $m \geq m_0$ with the constant K given in the corresponding table and a sufficiently large dimension $m_0 \in \mathbb{N}$ that can be computed via the previous remark. However, in some cases it might be enough to satisfy a weaker bound of the form $\lambda_{\min}(J) \geq \alpha - \tilde{K}$ for some $\tilde{K} \geq K$. This bound is certainly satisfied for all dimensions $m \geq m_0$, but it might already be satisfied for smaller dimensions, say, for all $m \geq \tilde{m}_0$ with some $\tilde{m}_0 \leq m_0$. The practical computation of $\tilde{m}_0$ can be done as in (d) with $x = \alpha - K$ replaced by $x = \alpha - \tilde{K}$.

4 Application

In this section we want to discuss an example which demonstrates the usefulness of the previous results. To this end, let us consider the linear homogeneous partial differential

equation (PDE) of second order

$$\theta \cdot \frac{\partial c\,(t,x,y)}{\partial t} - \beta_l \cdot q \cdot \frac{\partial^2 c\,(t,x,y)}{\partial x^2} - \beta_t \cdot q \cdot \frac{\partial^2 c\,(t,x,y)}{\partial y^2} + q \cdot \frac{\partial c\,(t,x,y)}{\partial x} = 0$$

on $[0,T]\times\Omega$, where $[0,T]$ for some $T > 0$ denotes the time interval and $\Omega = [0,\omega_x]\times[0,\omega_y] \subseteq \mathbb{R}^2$ for some constants ω_x, ω_y denotes the spatial domain. In addition, we assume that we have boundary conditions described by a Dirichlet condition on the left border and by Neumann conditions on the other boundaries of the domain. This PDE describes, for example, the convection and diffusion of a chemical species in the ground water, where c is the concentration of this species. The scalar constants $\theta, q, \beta_l, \beta_t > 0$ are used to specify some further properties of the given problem. For some background material regarding this particular application, we refer the interested reader to [9, 2].

Since we have a rectangular domain, the simplest discretization is by finite differences. To this end, we denote by h the step size in the spatial directions x and y, and by τ the step size for the discretization in time. Then we have $n = \frac{\omega_x}{h}$ unknown points in each grid row (for $x = 0$ the values are known by the Dirichlet boundary condition) and $m + 1$ unknown points in each grid column, with $m := \frac{\omega_y}{h}$. With $c_{i,j} := c\,(t_l, i \cdot h, j \cdot h)$ and $c_{i,j}^{old} := c\,(t_{l-1}, i \cdot h, j \cdot h)$ we denote the concentrations of the species at the discretized point (ih, jh) in the current time step $t_l = l \cdot \tau$ and the previous time step, respectively.

To obtain a suitable finite difference approximation of the original PDE, we use forward differences for the first term $\frac{\partial c(t,x,y)}{\partial t}$ (which is the only part that includes a derivate with respect to time), resulting in the first-order Euler approximation

$$\frac{c_{i,j} - c_{i,j}^{old}}{\tau}$$

in every grid point $(x_i, y_i) := (ih,\ jh)$. On the other hand, for the second-order derivative $-\beta_l q \frac{\partial^2 c(t,x,y)}{\partial x^2}$ we use the standard central difference approximation. We also apply a second-order central difference approximation to the first-order derivative $q\frac{\partial c(t,x,y)}{\partial x}$. The resulting approximation in each grid row $j = 0, \ldots, m$ for the inner grid points $i = 2, \ldots, n-1$ is

$$\left(-\frac{\beta_l q}{h^2} - \frac{q}{2h}\right) c_{i-1,j} + \frac{2\beta_l q}{h^2} c_{i,j} + \left(-\frac{\beta_l q}{h^2} + \frac{q}{2h}\right) c_{i+1,j},$$

while for $i = 1$ the value of $c_{0,j}$ is known from the Dirichlet boundary condition, so we obtain the approximation

$$\frac{2\beta_l q}{h^2} c_{1,j} + \left(-\frac{\beta_l q}{h^2} + \frac{q}{2h}\right) c_{2,j},$$

whereas for $i = n$ we get, taking into account the Neumann boundary condition on the right side of the domain, the discretization

$$\left(-2\frac{\beta_l q}{h^2}\right) c_{n-1,j} + \frac{2\beta_l q}{h^2} c_{n,j} = 0\,.$$

To write these expressions in matrix notation, we define the vectors

$$c^j := (c_{1,j}, c_{2,j}, \ldots, c_{n,j})^T \qquad \forall j = 1, \ldots, m-1$$

and the $n \times n$ matrix

$$L_x := a_x \cdot M_x \quad \text{with} \quad M_x := \begin{bmatrix} 2 & -1+b & & & & & \\ -1-b & 2 & -1+b & & & & \\ & -1-b & \ddots & \ddots & & & \\ & & \ddots & 2 & -1+b & & \\ & & & -1-b & 2 & -1+b & \\ & & & & -2 & 2 & \end{bmatrix}$$

and

$$a_x := a_x(h) := \frac{\beta_l \cdot q}{h^2} \qquad \text{and} \qquad b := b(h) := \frac{h}{2\beta_l}.$$

Now the resulting equations for each grid row $j = 0, \ldots, m$ read

$$L_x \cdot c^j \, .$$

Similarly, applying the standard central difference approximation to the second-order derivative $-\beta_t q \frac{\partial^2 c(t,x,y)}{\partial y^2}$, we obtain in each grid column $i = 1, \ldots, n$ for the inner points $j = 1, \ldots, m-1$ (each boundary point in y direction has a Neumann condition) the discretized equation

$$-\frac{\beta_t q}{h^2} c_{i,j-1} + \frac{2\beta_t q}{h^2} c_{i,j} - \frac{\beta_t q}{h^2} c_{i,j+1} \, ,$$

whereas on the lower bound $j = 0$ and the upper bound $j = m$, we have

$$\frac{2\beta_t q}{h^2} c_{i,0} - \frac{2\beta_t q}{h^2} c_{i,1} \quad \text{and} \quad -\frac{2\beta_t q}{h^2} c_{i,m-1} + \frac{2\beta_t q}{h^2} c_{i,m},$$

respectively. Like before we define for every grid column $i = 1, \ldots, n$ the vectors

$$c_i := (c_{i,0}, c_{i,1}, \ldots, c_{i,m})^T$$

and the $(m+1) \times (m+1)$ matrix

$$L_y := a_y \cdot M_y \quad \text{with} \quad M_y := \begin{bmatrix} 2 & -2 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -2 & 2 \end{bmatrix}, \quad a_y := a_y(h) := \frac{\beta_t \cdot q}{h^2}.$$

Using this notation, we can write these expressions in every grid column $i = 1, \ldots, n$ in matrix-vector form as

$$L_y \cdot c_i \,.$$

We now want to write all these $n \cdot (m+1)$ equations in one linear system. To this end, we order our unknowns $c_{i,j}$ in time step l lexicographically into one big vector by stacking together the grid rows c^j into

$$c_h = (c_{1,0}, c_{2,0}, \ldots, c_{n,0}, c_{1,1}, \ldots, c_{n,1}, \ldots c_{1,m-1}, c_{2,m-1}, \ldots, c_{n,m-1})^T \,.$$

We do the same for the previous time step and call the resulting vector c_h^{old}, whose entries are no longer unknowns in the current time step t_l. We now formulate the matrix of the full linear system with the help of L_x and L_y. Writing $l^y_{i,j}$ for the elements of L_y and defining the $n(m+1) \times n(m+1)$ matrix

$$L_h = \begin{bmatrix} L_x + l^y_{1,1} I_n & l^y_{1,2} I_n & & & & \\ l^y_{2,1} I_n & L_x + l^y_{2,2} I_n & l^y_{2,3} I_n & & & \\ & l^y_{3,2} I_n & L_x + l^y_{3,3} I_n & \ddots & & \\ & & \ddots & \ddots & l^y_{m,m+1} I_n & \\ & & & l^y_{m+1,m} I_n & L_x + l^y_{(m+1),(m+1)} I_n \end{bmatrix},$$

the resulting linear system in one time step becomes

$$\left(\theta I_{n(m+1)} + \tau L_h\right) \cdot c_h = \theta \cdot c_h^{old} \,. \tag{13}$$

In order to solve this sparse linear system efficiently by an iterative solver, we study the properties of the system matrix $\left(\theta I_{n(m+1)} + \tau L_h\right)$. In particular, we are interested in the smallest singular value to know whether the matrix is nonsingular and to compute the condition number. Moreover, we would like to show that this matrix is positive definite (although nonsymmetric), because this guarantees that, e.g., the restarted GMRES solver used in our numerical test is known to converge in this case, cf. [11].

To this end we shortly review the notions of the Kronecker product and the Kronecker sum together with some basic properties. The following results can be found in [6, Section 4.2 and 4.4]. Let $A = (a_{i,j})$ be a real $k \times k$ matrix and let $B = (b_{i,j})$ be a real $l \times l$ matrix. Then the *Kronecker product* is defined as the $kl \times kl$ matrix

$$A \otimes B = \begin{bmatrix} a_{1,1}B & a_{1,2}B & \cdots & a_{1,k}B \\ a_{2,1}B & a_{2,2}B & \cdots & a_{2,k}B \\ \vdots & & & \vdots \\ a_{k,1}B & a_{k,2}B & \cdots & a_{k,k}B \end{bmatrix}.$$

Given another $l \times l$ matrix C, it holds that

$$A \otimes B + A \otimes C = A \otimes (B + C) \,.$$

Similarly, we also have

$$(A + B) \otimes C = A \otimes C + B \times C$$

for all matrices A, B, C of appropriate dimension. For any real scalar r, we obviously have

$$r \cdot (A \otimes B) = (rA) \otimes B = A \otimes (rB) \ .$$

In addition, it is known that

$$\big(A \otimes B\big)^T = A^T \otimes B^T$$

holds for all suitable matrices A, B. Finally, the *Kronecker sum* of $A \in \mathbb{R}^{k \times k}$ and $B \in \mathbb{R}^{l \times l}$ is defined as the $kl \times kl$ matrix

$$I_l \otimes A + B \otimes I_k \, .$$

The eigenvalues $\mu_{i,j}$ of the Kronecker sum are given by $\lambda_i(A) + \lambda_j(B)$ for all $i = 1, \ldots, k$ and all $j = 1, \ldots, l$.

Now let us go back to our example. Using the notion of the Kronecker sum, the matrix L_h can be written as $L_h = I_{m+1} \otimes L_x + L_y \otimes I_n$, so that the matrix of the linear system (13) becomes

$$L\,(\tau, h) := \theta I_{(m+1)n} + \tau\,(I_{m+1} \otimes L_x + L_y \otimes I_n)\,.$$

We want to compute a lower bound for the smallest singular value of this matrix. To this end, we first give a lower bound for the smallest eigenvalue of the corresponding symmetric part which is given by

$$L^s\,(\tau, h) = \theta I_{(m+1)n} + \tau\,\big(I_{m+1} \otimes L_x^s + L_y^s \otimes I_n\big) \ .$$

The previous considerations show that the smallest eigenvalue of this symmetric part is given by

$$\lambda_1\,(L^s\,(\tau, h)) = \theta + \tau\lambda_1\,(L_x^s) + \tau\lambda_1\,\big(L_y^s\big)\,. \tag{14}$$

Hence we obtain a lower bound for the smallest eigenvalue $\lambda_1\,(L^s\,(\tau, h))$ by calculating lower bounds for $\lambda_1\,(L_x^s)$ and $\lambda_1\,\big(L_y^s\big)$. Since both matrices L_x^s and L_y^s have the structure of the matrix J from (1), we can apply the theory from the previous section. Note, however, that these bounds depend on our step size h. We will show that, for suitable choices of these step sizes, the matrix $L^s(\tau, h)$ has only positive eigenvalues. This implies that the (nonsymmetric) system matrix $L(\tau, h)$ itself is positive definite (recall that a nonsymmetric matrix A is positive definite if and only if its symmetric part A^s is positive definite). Furthermore, Lemma 2.3 then also gives a lower bound for the smallest singular value of $L(\tau, h)$.

Before we proceed, we note that it would alternatively be possible to consider the nonsymmetric matrix $L(\tau, h)$ directly since Remark 3.7 (c) can be applied in our particular application. The subsequent analysis, however, deals with the symmetric part $L^s(\tau, h)$ and calculates a lower bound for the smallest eigenvalue using the representation from (14).

Let us first consider the matrix $L_x^s = a_x(h) \cdot M_x^s$. We now give a lower bound for the smallest eigenvalue of the $n \times n$ matrix

$$M_x^s = \begin{bmatrix} 2 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & \ddots & \ddots & & & \\ & & \ddots & 2 & -1 & & \\ & & & -1 & 2 & -1.5 + 0.5 \cdot b & \\ & & & & -1.5 + 0.5 \cdot b & 2 & \end{bmatrix}.$$

We adapt the results from the previous section and get the following corollary.

Corollary 4.1 *If $b \in \left[3 - 2\sqrt{2},\, 3 + 2\sqrt{2}\right]$ then $\lambda_1(M_x^s) \geq 0$ holds for all $n \geq 4$.*
If $b < 3 - 2\sqrt{2}$ or $b > 3 + 2\sqrt{2}$ then $\lambda_1(M_x^s) \geq 2 - \frac{d^2}{\sqrt{d^2-1}}$ holds for all $n \geq 4$, where $d = -1.5 + 0.5b$ is the perturbed entry of M_x^s.

Proof. We first consider the case $b \in \left[3 - 2\sqrt{2},\, 3 + 2\sqrt{2}\right]$ which is equivalent with $|-1.5 + 0.5b| \leq \sqrt{2}$. With Theorem 3.6 (a) applied in the case "$|\beta| \leq |\gamma|$ and $|\delta| \leq \sqrt{2}|\gamma|$", we get the first estimate. The case $b < 3 - 2\sqrt{2}$ or $b > 3 + 2\sqrt{2}$ is equivalent to $|-1.5 + 0.5b| > \sqrt{2}$. Using Theorem 3.6 (a) once again, but applied in the case "$|\beta| \leq |\gamma|$ and $|\delta| > \sqrt{2}|\gamma|$", we obtain $2 - \frac{d^2}{\sqrt{d^2-1}}$ as a lower bound. The restriction regarding the dimension is simply due to the fact that all considerations in the previous section implicitly assumed that the matrices are at least 4×4-dimensional. □

Note that $2 - \frac{d^2}{\sqrt{d^2-1}}$ in the previous Corollary is always negative in the case where it is applied. Hence the corresponding matrix M_x^s is not necessarily positive definite in this case.

Similarly, we now study the $(m+1) \times (m+1)$ matrix $L_y^s = a_y(h) \cdot M_y^s$. We therefore give a lower bound for the smallest eigenvalue of

$$M_y^s = \begin{bmatrix} 2 & -1.5 & & & \\ -1.5 & 2 & \ddots & & \\ & -1 & \ddots & -1 & \\ & & \ddots & 2 & -1.5 \\ & & & -1.5 & 2 \end{bmatrix}.$$

To achieve the most accurate bounds, we distinguish different matrix sizes.

Corollary 4.2 *If $m \geq 65$ then $\lambda_1\left(M_y^s\right) \geq -0.0125$.*
If $m \geq 26$ then $\lambda_1\left(M_y^s\right) \geq -0.015$.
If $m \geq 16$ then $\lambda_1\left(M_y^s\right) \geq -0.02$.
If $m \geq 12$ then $\lambda_1\left(M_y^s\right) \geq -0.025$.

Proof. Theorem 3.6 (b) applied in the case "$|\beta| > \sqrt{2}|\gamma|$ and $|\delta| > \sqrt{2}|\gamma|$" shows that $\lambda_1\left(M_y^s\right) \geq 2 - \frac{2.25}{\sqrt{1.25}} \approx -0.01246$ holds for all sufficiently large m. Replacing this lower bound by the less restrictive numbers $-0.0125, -0.015, -0.02$ and -0.025, respectively, we obtain the desired statements in a way described in Remark 3.7 (e). □

Using (14), we therefore obtain

$$\begin{aligned}\lambda_{min}\left(L^s\left(\tau,h\right)\right) &= \theta + \tau a_x\left(h\right)\lambda_{min}\left(M_x^s\right) + \tau a_y\left(h\right)\lambda_{min}\left(M_y^s\right) \\ &= \theta + \tau\frac{\beta_l \cdot q}{h^2}\lambda_{min}\left(M_x^s\right) + \tau\frac{\beta_t \cdot q}{h^2}\lambda_{min}\left(M_y^s\right) .\end{aligned}$$

From Lemma 2.3 we know that $\sigma_{min}\left(L\left(\tau,h\right)\right) \geq \lambda_{min}\left(L^s\left(\tau,h\right)\right)$ if $L^s\left(\tau,h\right)$ is positive definite, which is equivalent to $\lambda_{min}\left(L^s\left(\tau,h\right)\right) > 0$. Recall that $b = b\left(h\right) = \frac{h}{2\beta_l}$ and therefore $b\left(h\right) > 0$ for all $h > 0$. Taking into account the two different cases considered in Corollary 4.1, we obtain the lower bound

$$\lambda_{min}\left(L^s\left(\tau,h\right)\right) \geq \theta + \tau\frac{\beta_t \cdot q}{h^2}\lambda_{min}\left(M_y^s\right) \quad \text{for } h \in \left[\left(1.5-\sqrt{2}\right)4\beta_l, \left(1.5+\sqrt{2}\right)4\beta_l\right],$$

whereas we have

$$\lambda_{min}\left(L^s\left(\tau,h\right)\right) \geq \theta + \tau\frac{\beta_t \cdot q}{h^2}\lambda_{min}\left(M_y^s\right) + \tau\frac{\beta_l \cdot q}{h^2} \cdot \left(2 - \frac{d^2}{\sqrt{d^2-1}}\right)$$

for $h \notin \left[\left(1.5-\sqrt{2}\right)4\beta_l, \left(1.5+\sqrt{2}\right)4\beta_l\right]$, where $d = -1.5 + 0.5b$. The possibly negative eigenvalues $\lambda_{min}\left(M_x^s\right)$ and $\lambda_{min}\left(M_y^s\right)$ get amplified by the numbers $\frac{\beta_l \cdot q}{h^2} > 0$ and $\frac{\beta_t \cdot q}{h^2} > 0$, respectively. These factors increase for $h \to 0$. Suppose a time step size $\tau > 0$ is given. Then we need to calculate a minimal step size h_0 such that

$$\theta + \tau\frac{\beta_l \cdot q}{h_0^2}\lambda_{min}\left(M_x^s\right) + \tau\frac{\beta_t \cdot q}{h_0^2} \cdot \lambda_{min}\left(M_y^s\right) > 0$$

holds and therefore our matrix $L\left(\tau,h\right)$ is positive definite and nonsingular. Then we can solve our linear system with all step sizes $h \geq h_0$. Here it is important to have an accurate lower bound for $\lambda_{min}\left(M_x^s\right)$ and $\lambda_{min}\left(M_y^s\right)$ so that we can use step sizes h as small as possible.

Example 4.3 We set $\omega_x = 10$ and $\omega_y = 6$ and therefore use the domain $\Omega = [0,10] \times [0,6]$. We further use the scalars $\tau = 0.1$, $\beta_l = 0.3$, $\beta_t = 0.03$, $q = 0.18$ and $\theta = 0.3$. Depending on the choice of h, we now get different matrix sizes and eigenvalues. In the following table we compare the lower bound of $\lambda_{min}\left(L^s(\tau,h)\right)$ according to our theory (column 'λ_{min} lower bound') with the exact eigenvalue calculated from the corresponding system matrix with the MATLAB function eigs (column 'λ_{min} exact').

h	n	m	size	λ_{min} exact	λ_{min} lower bound
0.5	20	12	260	0.300412963667855	0.2999550000
0.2	50	30	1550	0.300213215599023	0.2997975000
0.1	100	60	6100	0.299416896200867	0.2991835345
0.05	200	120	24200	0.289617314238473	0.2896089457
0.02	500	300	150500	0.170625390123707	0.1705729827
0.01	1000	600	600600	−0.324076096559832	−0.3242857259

We see that the lower bounds obtained from our theory are very sharp. In fact, a rounding process after the first three digits gives identical values for all different matrix sizes.

From Lemma 2.3 we know that our estimate for $\lambda_{min}\left(L^s(\tau,h)\right)$ is also a lower bound for $\sigma_{min}\left(L(\tau,h)\right)$ as long as $L^s(\tau,h)$ is positive semidefinite, i.e., for all step sizes except $h = 0.01$. However, it is clear from Lemma 2.3 that this lower bound will be much less accurate, especially when the matrix $L(\tau,h)$ is far away from being symmetric (this will be the case for smaller values of h). Nevertheless, we will give a comparison of our lower bound for σ_{min} with prior results in this area. To this end, let us define the values $r_k(A) := \sum_{j=1,j\neq k}^{n} |a_{kj}|$ and $c_l(A) := \sum_{i=1,i\neq l}^{n} |a_{il}|$ for an arbitrary matrix $A = [a_{ij}] \in \mathbb{R}^{n\times n}$. Then, Johnson [7, Theorem 3] showed that

$$\sigma_{min}(A) \geq \min_{i=1,\ldots,n} \left\{ |a_{ii}| - \frac{r_i(A) + c_i(A)}{2} \right\}. \tag{15}$$

whereas Johnson and Szulc [8, Theorem 2] proved the lower bound

$$\sigma_{min}(A) \geq \min_{i=1,\ldots,n} \left\{ \sqrt{|a_{ii}|^2 + \left(\frac{r_k(A) - c_k(A)}{2}\right)^2} - \frac{r_k(A) + c_k(A)}{2} \right\}. \tag{16}$$

Reference [8] actually gives some further formulas which, however, all coincide for our particular matrix $L(\tau,h)$. Another interesting lower bound was given by Qi [10, Theorem 3]:

$$\sigma_{min}(A) \geq \max\left\{0, \min\{l_1,\ldots,l_n\}\right\} \qquad \text{with}$$
$$l_i := \min\left\{ \sqrt{a_{ii}^2 - a_{ii}r_i(A) + \frac{c_i(A)^2}{4}} - \frac{c_i(A)}{2}, \sqrt{a_{ii}^2 - a_{ii}c_i(A) + \frac{r_i(A)^2}{4}} - \frac{r_i(A)}{2} \right\}. \tag{17}$$

In the following table, we compare these estimates with our estimate for $L(\tau,h)$ for different step sizes h.

h	σ_1 exact	σ_1 l.b.	(15)	(16)	(17)
0.5	0.30046	0.29996	0.29712	0.29713	0.29689
0.2	0.30031	0.29980	0.24825	0.25049	0.22481
0.1	0.30005	0.29918	0.048	0.06919	0.00000
0.05	0.29978	0.28960	−0.798	−0.68005	0.08769
0.02	0.30006	0.17057	−6.9	−6.04810	0.20315

The column entitled "σ_1 l.b." is our lower bound. We see that the estimates from (15)–(17) all become zero or negative at a certain stage (and, hence, are useless as a lower bound for the smallest singular value). Furthermore, our lower bound is (much) better in almost all situations.

References

[1] A. Böttcher and S. Grudsky: *Spectral Properties of Banded Toeplitz Matrices.* SIAM, Philadelphia, PA, 2005.

[2] H. Buchholzer, C. Kanzow, P. Knabner, and S. Kräutle: *The semismooth Newton method for the solution of reactive transport problems including mineral precipitation-dissolution reactions.* Technical Report, University of Würzburg, Würzburg, Germany, January 2010, submitted for publication.

[3] J.W. Demmel: *Applied Numerical Linear Algebra.* SIAM, Philadelphia, PA, 1997.

[4] G.H. Golub and Ch.F. van Loan: *Matrix Computations.* The Johns Hopkins University Press, Baltimore, MD, second edition 1989.

[5] R. Horn and C. Johnson: *Matrix Analysis.* Cambridge University Press, 1991.

[6] R. Horn and C. Johnson: *Topics in Matrix Analysis.* Cambridge University Press, 1991.

[7] C. Johnson: *A Gersgorin-type lower bound for the smallest singular value.* Linear Algebra and its Applications 112, 1989, pp. 1–7.

[8] C. Johnson and T. Szulc: *Further lower bounds for the smallest singular value.* Linear Algebra and its Applications 272, 1998, pp. 169–179.

[9] S. Kräutle: *General multi-species reactive transport problems in porous media: Efficient numerical approaches and existence of global solutions.* Habilitation Thesis, University of Erlangen, Germany, 2008.

[10] L. Qi: *Some simple estimates for singular values of a matrix.* Linear Algebra and its Applications 56, 1984, pp. 105–119.

[11] Y. Saad: *Iterative Methods for Sparse Linear Systems.* SIAM, Philadelphia, PA, second edition 2003.

[12] J. Stoer and R. Bulirsch: *Introduction to Numerical Analysis.* Springer, New York, NY, third edition 2002.