SIAM J. Matrix Anal. Appl. (SIMAX), 18(1):83–103, 1997

BOUNDS FOR THE COMPONENTWISE DISTANCE TO THE NEAREST SINGULAR MATRIX

S. M. RUMP †

Abstract. The normwise distance of a matrix A to the nearest singular matrix is well known to be equal to $\|A\|/\mathrm{cond}(A)$ for norms being subordinate to a vector norm. However, there is no hope to find a similar formula or even a simple algorithm for computing the componentwise distance to the nearest singular matrix for general matrices. This is because Rohn and Poljak [7] showed that this is an NP-hard problem.

Denote the minimum Bauer-Skeel condition number achievable by column scaling by κ. Demmel [3] showed that κ^{-1} is a *lower* bound for the componentwise distance to the nearest singular matrix. In this paper we prove that $2.4 \cdot n^{1.7} \cdot \kappa^{-1}$ is an *upper* bound. This extends and proves a conjecture by N. J. Higham and J. Demmel. We give an explicit set of examples showing that an upper bound cannot be better than $n \cdot \kappa^{-1}$. Asymptotically, we show that $n^{1+\ln 2+\varepsilon} \cdot \kappa^{-1}$ is a valid upper bound.

Key words. componentwise distance, NP-hardness, optimal Bauer-Skeel condition number

AMS subject classifications. 65F35, 15A60

0. Introduction. Let A be an n by n matrix and denote its smallest singular value by $\sigma_n(A)$. It is well known that the distance to the nearest singular matrix in the 2-norm or Frobenius norm is equal to $\sigma_n(A)$. More general, for any consistent matrix norm $\|\cdot\|$ being subordinate to a vector norm we have

$$\min\{\,\|\delta A\| \mid A+\delta A \text{ singular}\,\} = \frac{1}{\|A^{-1}\|} = \frac{\|A\|}{\mathrm{cond}(A)}. \tag{1}$$

An appropriate δA of rank 1 can be explicitly calculated (cf. [**?**], [13]). Such a perturbation does, in general, alter each component of A. In many practical applications, one may be interested in leaving specific components such as system zeros unaltered, for example, if the matrix arises from some discretisation scheme. More general, this leads to the question of the componentwise distance to the nearest singular matrix. The componentwise distance may be weighted by some nonnegative matrix E. More precisely, we define

$$\sigma(A,E) := \min\{\,\alpha \in \mathbb{R} \mid A+\widetilde{E} \quad \text{singular where} \quad |\widetilde{E}_{ij}| \le \alpha \cdot E_{ij} \quad \text{for all } i,j\,\}. \tag{2}$$

If no such α exists, we set $\sigma(A,E) := \infty$. For singular matrices, $\sigma(A,E) = 0$ for every weight matrix E. Specific values of E are $E = |A|$ for relative perturbations, or $E = (\mathbf{1})_{nn}$ for absolute perturbations. Among others, the componentwise distance to the nearest singular matrix was discussed in [8], [11], [10], and in [3]. In [8] we also find a first approach towards an estimation of the nearness to singularity in a norm not being subordinate to a vector norm, namely $\|A\| := \max_{i,j} |A_{ij}|$.

We cannot expect to find a formula or even a simple algorithm forq calculating $\sigma(A,E)$. This is because Rohn and Poljak [7] proved that computation of $\sigma(A,E)$ is NP-complete. For an outline of their proof see also [3]. Nevertheless, we may find bounds for $\sigma(A,E)$, and for classes of matrices even explicit formulas.

Another view of $\sigma(A,E)$ is the maximum value such that the interval matrix $[A-\alpha E,\ A+\alpha E]$ is nonsingular for $\alpha < \sigma(A,E)$. The interval matrix is defined as being the set of all matrices $\widetilde{A}$ with $A_{ij} - \alpha E_{ij} \le \widetilde{A}_{ij} \le A_{ij} + \alpha E_{ij}$ for all i,j, or in short notation $A - \alpha E \le \widetilde{A} \le A + \alpha E$. The interval matrix is called nonsingular if every matrix $\widetilde{A} \in [A-\alpha E,\ A+\alpha E]$ is nonsingular. In a very interesting paper [9], Rohn gave 13 necessary and sufficient criteria for $[A-E,\ A+E]$ being nonsingular.

†Technische Informatik III, TU Hamburg-Harburg, EißendorferStraße 38, 21071 Hamburg, Germany

A thorough discussion of $\sigma(A,E)$ can be found in the very interesting paper [3]. Demmel [3] proved that $\sigma(A,E)$ is equal to the inverse of $\min \kappa(AD,\, ED)$, the minimum taken over all diagonal matrices D, where $\kappa(A,E) := \| \, |A^{-1}| \cdot E\|$ denotes the Bauer-Skeel condition number. For any p-norm, he proves

$$\min_D \ \kappa(AD,\, ED) = \rho(\, |A^{-1}| \cdot E),$$

extending a result by Bauer [1]. In other words, the minimum Bauer-Skeel condition number achievable by column scaling is equal to the inverse of $\rho(\, |A^{-1}| \cdot E)$. Demmel and N. J. Higham conjecture that $\rho(\, |A^{-1}| \cdot E)$ and $\sigma(A,E)$ are not too far apart. They conjecture for relative perturbations existence of some constant $\gamma \in \mathbb{R}$, possibly depending on the dimension, with

$$\sigma(A, |A|\,) \leq \frac{\gamma}{\rho(\, |A^{-1}| \cdot |A|\,)}. \tag{3}$$

In this paper, our main goal is to show existence of such constants $\gamma(n)$ and to derive lower and upper bounds for $\gamma(n)$. First, we show that $\sigma(A,E) \geq \sigma_n(A)$ for $\|E\|_2 = 1$. A corresponding result for other norms is given in §2. However, this bound can be arbitrarily weak. Following we give some new bounds for $\sigma(A,E)$.

In §4 a perturbation formula for determinants is stated which is the key to prove an upper bound of $\gamma(n)$.

In §5 we will prove $\gamma \geq n$. In §6, for arbitrary weight matrices E we prove

$$\frac{1}{\rho(\, |A^{-1}| \cdot E)} \leq \sigma(A,E) \leq \frac{\gamma(n)}{\rho(\, |A^{-1}| \cdot E)} \qquad \text{with} \qquad \gamma(n) = c \cdot n^{\alpha} \tag{4}$$

for $c = 2.4$ and $\alpha = 1.7$. Moreover, for $n \to \infty$ we show that for every $\varepsilon > 0$, α can be replaced by $1 + \ln 2 + \varepsilon$. In view of $\gamma \geq n$, we conjecture $\gamma = n$.

In [3], Demmel gave reasons to be interested in the componentwise distance to the nearest singular matrix. In §2, we add a lower and upper componentwise error bound for the solution of a linear system $Ax = b$ subject to componentwise perturbations of the matrix and the right hand side. Such upper bounds are known in the literature and are valid for nonsingular A and $|\widetilde{A} - A| \leq E$ with $\rho(|\, A^{-1}| \cdot E) < 1$. We derive a componentwise bound for the *minimum* perturbation of the solution subject to finite perturbations of A and b. (4) shows that those estimates cover perturbation matrices $\widetilde{A}$ not too far from the next singular matrix.

The paper is organized as follows. In §1 we introduce the used notation. In §2 follows a componentwise lower and upper perturbation bound for finite componentwise perturbations of a linear system. In §3, lower bounds on $\sigma(A,E)$ are given. For orthogonal matrices we show that γ (see (4)) is at least of the order of $\sqrt{n}$.

In §4, a Sherman-Morrison-Woodbury like perturbation theorem for determinants is given. In fact, this is an *equality* for finite perturbations of a matrix. In §5 we derive upper bounds on $\sigma(A,E)$. For E being of rank 1, such as for absolute perturbations, we show $\gamma(n) \leq n$, and for relative perturbations we give a set of matrices $A \in M_n(R)$ with $\gamma(n) = n$. For a class of matrices including M-matrices wie prove $\gamma(n) = 1$, i.e. $\sigma(A,E) = \rho(\, |A^{-1}| \cdot E)^{-1}$.

In §6 the results are extended to obtain an explicit upper bound on $\gamma(n)$ for general A and E, and in §7 those bounds are quantified into (4). We close with the conjecture that (4) is valid for $\gamma(n) = n$ for all A, E. If this is true, the set of matrices given in §5 would imply that inequality (4) with $\gamma(n) = n$ is sharp.

1. Notation. In the following we list some notation from matrix theory, cf. for example [6], [5]. $V_n(\mathbb{R})$ denotes the set of vectors with n real components, $M_{m,n}(\mathbb{R})$ the set of real m by n matrices, and $M_n(\mathbb{R}) = M_{n,n}(\mathbb{R})$. The components of a matrix $A \in M_n(\mathbb{R})$ are referred by A_{ij} or $A_{i,j}$. For short notation, components of A^{-1} are referred by A^{-1}_{ij}. $(\mathbf{1})$ denotes a vector with all components equal to 1, $(\mathbf{1})_{nn} \in M_n(\mathbb{R})$ the matrix with all columns equal to $(\mathbf{1})$.

Q_{kn} denotes the set of strictly increasing sequences of k integers chosen from $\{1, \ldots, n\}$. For $\omega \in Q_{kn}$, we denote $\omega = (\omega_1, \ldots, \omega_k)$. For $C \in M_n(\mathbb{R})$, $\omega \in Q_{kn}$, $C[\omega] \in M_k(\mathbb{R})$ denotes the k by k submatrix of C lying in rows and columns ω. A sequence $\zeta = (i_1, \ldots, i_k)$, $k \geq 1$ of mutually different integers $i_\nu \in \{1, \ldots, n\}$ is

called a *cycle*. We identify the cycles $(i_1, \ldots, i_k)$ and $(i_p, \ldots, i_k, i_1, \ldots, i_{p-1})$, where $1 \le p \le k$. It is $|\zeta| := k$. A *full cycle* ζ on $\{1, \ldots, n\}$ is a cycle ζ with $|\zeta| = n$.

For $C \in M_n(\mathbb{R})$ and a cycle $\zeta = (i_1, \ldots, i_k)$ on $\{1, \ldots, n\}$, we put

$$\Pi_\zeta(C) := C_{i_1 i_2} \cdot \ldots \cdot C_{i_{k-1} i_k} \cdot C_{i_k i_1},$$

the *cycle product* for ζ. Note the last factor in the product. Therefore, $|\Pi_\zeta(C)|^{1/|\zeta|}$ is the geometric mean of the elements of the cycle product. Each diagonal element C_{ii} is a cycle product, namely of the cycle (i). (Here our definition differs from Engel/Schneider [4]).

With one exception, throughout the paper, *absolute value* and *comparison* is used *componentwise*. For example, for $A, B \in M_n(\mathbb{R})$,

$$|A| \le B \quad \text{means} \quad |A_{ij}| \le B_{ij} \quad \text{for} \quad 1 \le i, j \le n.$$

The exception are cycles $\zeta = (i_1, \ldots, i_k)$, where $|\zeta| = k$. The *singular values* of a matrix $A \in M_n(\mathbb{R})$ are denoted in decreasing order with increasing indices, i.e. $\sigma_1(A) \ge \ldots \ge \sigma_n(A) \ge 0$.

For $A, E \in M_n(\mathbb{R})$, $E \ge 0$, $\sigma(A, E)$ denotes the *componentwise distance*, weighted by E, *to the nearest singular matrix* (cf. (0.2)).

For finite $\sigma(A, E)$, the set of all matrices $\widetilde{A} \in M_n(\mathbb{R})$ with $|\widetilde{A} - A| \le \sigma(A, E) \cdot E$ is compact. For every nonsingular $\widetilde{A}$ there is a neighbourhood of $\widetilde{A}$ consisting only of nonsingular matrices. Therefore

$$\sigma(A, E) < \infty \;\Rightarrow\; \exists\, \delta A \in M_n(\mathbb{R}) : \;\; |\delta A| = \sigma(A, E) \cdot E \;\text{ and }\; A + \delta A \;\text{ singular},$$

showing that we are allowed to use a minimum in the definition (0.2) of $\sigma(A, E)$. ρ denotes the spectral radius, whereas ρ_0 denotes the *real spectral radius*:

$$B \in M_n(\mathbb{R}) : \;\; \rho_0(B) := \max \left\{ |\lambda| \;\middle|\; \lambda \in \mathbb{R} \;\text{ is an eigenvalue of }\; B \right\}.$$

If B has no real eigenvalues, we set $\rho_0(B) := 0$. I denotes the identity matrix of proper dimension, especially $I_k \in M_k(\mathbb{R})$ denotes the k by k identity matrix. A *signature matrix* S is a diagonal matrix with diagonal entries $+1$ or -1, i.e. $|S| = I$.

We frequently use standard results from matrix and Perron-Frobenius theory such as

$$\begin{aligned} &A \in M_{nk}(\mathbb{R}), B \in M_{kn}(\mathbb{R}) \;\Rightarrow \\ &\qquad\qquad \text{The set of nonzero eigenvalues of } AB \text{ and } BA \text{ are identical,} \end{aligned} \tag{5}$$

cf. Theorem 1.3.20 in [5], and

$$\begin{aligned} &A \in M_n(\mathbb{R}) \text{ and } A \ge 0,\; x \in V_n(\mathbb{R}) \text{ with } x > 0 \;\Rightarrow \\ &\qquad \min_i \frac{(Ax)_i}{x_i} \le \rho(A) \le \max_i \frac{(Ax)_i}{x_i}. \end{aligned} \tag{6}$$

The latter can be found in [2].

2. Finite perturbations for a linear system. Calculating bounds on $\sigma(A, E)$ can be motivated, for example, by looking at linear systems with finite perturbations of the input data. For a linear system $Ax = b$ consider the perturbed system $\widetilde{A}\widetilde{x} = \widetilde{b}$ with $\delta A := \widetilde{A} - A$, $\delta b := \widetilde{b} - b$, $\delta x := \widetilde{x} - x$. Then for nonsingular A,

$$A \cdot (I + A^{-1} \cdot \delta A) \cdot (\widetilde{x} - x) = \widetilde{A} \cdot (\widetilde{x} - x) = \widetilde{b} - \widetilde{A}x = \delta b - \delta A \cdot x. \tag{7}$$

If $\rho(A^{-1} \cdot \delta A) < 1$, then $I + A^{-1} \cdot \delta A$ and $\widetilde{A} = A \cdot (I + A^{-1} \cdot \delta A)$ are nonsingular, and (7) implies

$$\delta x = (I + A^{-1} \cdot \delta A)^{-1} \cdot A^{-1} \cdot (\delta b - \delta A \cdot x). \tag{8}$$

If $\rho\big(|A^{-1}| \cdot \Delta A\big) < 1$ then $I - |A^{-1}| \cdot \Delta A$ is an M-matrix. If the perturbations δA, δb are componentwise bounded by $|\delta A| \le \Delta A$, $|\delta b| \le \Delta b$ then (2.2) implies

$$|\delta x| \le (I - |A^{-1}| \cdot \Delta A)^{-1} \cdot |A^{-1}| \cdot (\Delta b + \Delta A \cdot |x|\,). \tag{9}$$

For given weight matrix ΔA, consider the set of matrices with componentwise distance from A weighted by ΔA not greater than σ:

$$\widetilde{A} \in U_\sigma(A, \Delta A) \;\Leftrightarrow\; |\widetilde{A} - A| \le \sigma \cdot \Delta A.$$

For $\sigma \le \rho(\,|A^{-1}| \cdot \Delta A)$, Perron-Frobenius-Theory yields

$$\rho(I - A^{-1} \cdot \widetilde{A}) = \rho\big(A^{-1} \cdot (A - \widetilde{A})\big) \le \rho(\,|A^{-1}| \cdot \Delta A) < 1,$$

and therefore regularity of all $\widetilde{A} \in U_\sigma(A, \Delta A)$. The bound (2.3) requires $|A^{-1}| \cdot \Delta A$ to be convergent, whereas (8) is valid for $\rho(A^{-1} \cdot \delta A) < 1$. Therefore we may ask, how far a matrix $\widetilde{A}$ with $\rho(\,|A^{-1}| \cdot |\widetilde{A} - A|\,) \ge 1$ can be from the nearest singular matrix. An answer to this question shows how strong the assumption $\rho(\,|A^{-1}| \cdot \Delta A) < 1$ is.

3. Lower bounds on $\sigma(A, E)$. A simple and well-known lower bound on $\sigma(A, E)$ is

$$\frac{1}{\rho(\,|A^{-1}| \cdot E)} \le \sigma(A, E) \quad \text{for all nonsingular } A \in M_n(\mathbb{R}),\ 0 \le E \in M_n(\mathbb{R}). \tag{10}$$

This can be seen using Perron-Frobenius Theory and

$$\begin{aligned} \rho(\,|A^{-1}| \cdot E) < 1 \quad &\Rightarrow \quad \rho(A^{-1} \cdot \delta A) < 1 \quad \text{for all} \quad |\delta A| \le E \\ &\Rightarrow \quad A + \delta A = A \cdot (I + A^{-1} \cdot \delta A) \quad \text{is nonsingular.} \end{aligned}$$

Another lower bound is (cf. [12], Theorem 1.8, p. 75)

$$\frac{\sigma_n(A)}{\sigma_1(E)} \le \sigma(A, E). \tag{11}$$

This can be generalized in the following way.

THEOREM 3.1. *Let $\|\cdot\|$ be a matrix norm subordinate to an absolute vector norm $\|\cdot\|$. Then for nonsingular $A \in M_n(\mathbb{R})$ and $0 \le E \in M_n(\mathbb{R})$,*

$$\frac{1}{\|A^{-1}\| \cdot \|E\|} \le \sigma(A, E). \tag{12}$$

(12) is especially valid for all p-norms. For absolute norms such as 1-norm and ∞-norm,

$$\frac{1}{\|A^{-1}\| \cdot \|E\|} \le \frac{1}{\rho(\,|A^{-1}| \cdot E)} \le \sigma(A, E), \tag{13}$$

whereas for the 2-norm

$$\frac{1}{\|A^{-1}\|_2 \cdot \|E\|_2} \le \frac{\sqrt{n}}{\rho(\,|A^{-1}| \cdot E)}. \tag{14}$$

Proof. To prove (12), let $\delta A \in M_n(\mathbb{R})$ with $|\delta A| \le \sigma \cdot E$ for $\sigma < (\,\|A^{-1}\| \cdot \|E\|\,)^{-1}$. The vector norm is absolute implying $\|x\| = \|\,|x|\,\|$ and $|x| \le |y| \Rightarrow \|x\| \le \|y\|$ for $x, y \in V_n(\mathbb{R})$ (cf. [13], Theorem II.1.2). Let $x \in V_n(\mathbb{R})$ with $\|x\| = 1$ and $\|\delta A\| = \|\delta A \cdot x\|$. Then

$$\begin{aligned} \|\delta A\| = \|\delta A \cdot x\| = \|\,|\delta A \cdot x|\,\| &\le \|\sigma \cdot E \cdot |x|\,\| \le \|\sigma \cdot E\| \cdot \|\,|x|\,\| \\ &= \sigma \cdot \|E\| < \|A^{-1}\|^{-1}. \end{aligned} \tag{15}$$

For every $0 \ne y \in V_n(\mathbb{R})$ holds $\|y\| \le \|A^{-1}\| \cdot \|Ay\|$, and (15) yields

$$\|\delta A \cdot y\| \le \|\delta A\| \cdot \|y\| < \|A^{-1}\|^{-1} \cdot \|y\| \le \|Ay\|, \quad \text{and therefore} \quad (A + \delta A) \cdot y \ne 0.$$

Therefore, $A + \delta A$ is nonsingular for $|\delta A| \leq \sigma \cdot E$, and $\sigma < (\,\|A^{-1}\| \cdot \|E\|\,)^{-1}$, proving (12). For absolute matrix norms,

$$\rho(\,|A^{-1}| \cdot E) \leq \|\,|A^{-1}| \cdot E\| \leq \|A^{-1}\| \cdot \|E\|$$

proving (13). For the 2-norm holds

$$\rho(\,|A^{-1}| \cdot E) \leq \|\,|A^{-1}|\,\|_2 \cdot \|E\|_2 \leq \|\,|A^{-1}|\,\|_F \cdot \|E\|_2 = \|A^{-1}\|_F \cdot \|E\|_2 \leq \sqrt{n} \cdot \|A^{-1}\|_2 \cdot \|E\|_2,$$

proving (14) and the theorem. □

(13) shows that for absolute matrix norms such as the 1-norm or ∞-norm, the bound (12) cannot be better than (10). The 2-norm is not absolute, and (14) shows that the lower bound (11) for $\sigma(A, E)$ may be better up to a factor $\sqrt{n}$ than (10). In fact, we can identify a class of matrices for which this improvement is approximately achieved.

Let $Q \in M_n(\mathbb{R})$ be orthogonal, and consider absolute perturbations $E = (\mathbf{1})_{nn}$. Then (11) yields

$$\frac{\sigma_n(Q)}{\sigma_1\big((\mathbf{1})_{nn}\big)} = \frac{1}{n} \leq \sigma(Q, E).$$

On the other hand, $E = (\mathbf{1})_{nn} \in M_n(\mathbb{R})$ and $x = (\mathbf{1}) \in V_n(\mathbb{R})$ imply

$$\big\{\,|Q^{-1}| \cdot E\big\}^T \cdot x = E \cdot |Q| \cdot x = \Big(\sum_{i,j} |Q_{ij}|\Big) \cdot x,$$

and (6) yields $\rho(\,|Q^{-1}| \cdot E) = \sum\limits_{i,j} |Q_{ij}|$. If Q is an orthogonalized random matrix with components uniformly distributed in $[-1, 1]$, then $|Q_{ij}| \approx n^{-1/2}$. Thus, for the ratio between the two lower bounds (11) and (10) we obtain

$$\big\{\,\sigma_n(Q)/\sigma_1(E)\,\big\} \,/\, \big\{\,1/\rho(\,|Q^{-1}| \cdot E)\,\big\} \approx n^{-1} \cdot n^2 \cdot n^{-1/2} = \sqrt{n}.$$

The same heuristic holds for $E = |Q|$, cf. [12]. For every Hadamard matrix ($H \in M_n(\mathbb{R})$ with $H^T H = n \cdot I$) the ratio is equal to $\sqrt{n}$. This sheds a first light on a possible quantity $\gamma(n)$ such that (4) holds. In §5 we will prove $\gamma(n) \geq n$.

Example 3.2. *The lower bound (11) may be arbitrarily weak. Consider*

$$A = \begin{pmatrix} 2\varepsilon & -\varepsilon \\ -\varepsilon & 1 \end{pmatrix} \quad \text{and} \quad E = |A| \quad \text{for some } \varepsilon > 0.$$

A is a diagonally dominant M-matrix. As we will see in (5.5), A being M-matrix implies equality in (10), i.e. $\sigma(A, |A|\,) = \rho(\,|A^{-1}| \cdot |A|\,) = 1 + 0(\sqrt{\varepsilon})$. On the other hand, $\sigma_2(A)/\sigma_1(\,|A|\,) = 2\varepsilon + 0(\varepsilon^2)$ underestimates $\sigma(A, |A|\,)$ arbitrarily. This corresponds to $\sigma_2(A) = 2\varepsilon + 0(\varepsilon^2)$. That means, the normwise distance in the 2-norm or Frobenius norm to the nearest singular matrix can be arbitrarily small compared to a componentwise distance.

4. A perturbation theorem for determinants. A lower bound on $\sigma(A, E)$ is obtained by proving *regularity* of a set of matrices. This was done in §3 by using spectral properties. To obtain an *upper* bound on $\sigma(A, E)$, we may construct a specific perturbation δA with $|\delta A| \leq \sigma_0 \cdot E$, $\sigma_0 \in \mathbb{R}$ such that $A + \delta A$ is singular. This proves $\sigma(A, E) \leq \sigma_0$. Another possibility to obtain an upper bound on $\sigma(A, E)$ is the following. If $|\delta A| \leq \sigma_0 \cdot E$ and $\det(A) \cdot \det(A + \delta A) \leq 0$, then a continuity argument yields $\sigma(A, E) \leq \sigma_0$. Therefore we state the following explicit formula for the relative change of the determinant of a matrix subject to a rank-k perturbation. It is a Sherman-Morrison-Woodbury like perturbation formula for determinants.

Lemma 4.1. *Let $A \in M_n(\mathbb{R})$ and $U, V \in M_{n,k}(\mathbb{R})$ be given. Then for nonsingular A,*

$$\det(A + UV^T) = \det(A) \cdot \det(I_k + V^T A^{-1} U), \tag{16}$$

where I_k denotes the k by k identity matrix.

Proof. It is

$$\det(A + UV^T) = \det(A) \cdot \det(I_n + A^{-1}UV^T).$$

Denoting the eigenvalues of $X \in M_n(\mathbb{R})$ by $\lambda_i(X)$ implies

$$\det(I_n + A^{-1}UV^T) = \prod_{i=1}^{n} \lambda_i(I_n + A^{-1}UV^T) = \prod_{i=1}^{n} \{1 + \lambda_i(A^{-1}UV^T)\}.$$

The set of nonzero eigenvalues of $A^{-1}UV^T$ and $V^TA^{-1}U$ are identical (see (5)), thus proving the lemma. □

This lemma has a nice and for itself interesting corollary.

COROLLARY 4.2. *Let $A \in M_n(\mathbb{R})$ and $u, v \in V_n(\mathbb{R})$. Then for nonsingular A,*

$$\det(A + uv^T) = \det(A) \cdot (1 + v^TA^{-1}u). \tag{17}$$

For arbitrary $A \in M_n(\mathbb{R})$ holds ($\mathrm{adj}(A)$ denotes the adjoint of A),

$$\det(A + uv^T) = \det(A) + v^T \cdot \mathrm{adj}(A) \cdot u. \tag{18}$$

The corollary shows that the relative change of the determinant is linear for rank-1 perturbations of the matrix. The second well-known formula follows, for example, by a continuity argument using $A \cdot \mathrm{adj}(A) = \det(A) \cdot I$.

5. Upper bounds on $\sigma(A, E)$. The perturbation lemma for determinants given in §4 allows for other lower bounds on $\sigma(A, E)$. The first result can be found in [8], Corollary 5.1, (iii).

THEOREM 5.1. *Let $A \in M_n(\mathbb{R})$ be nonsingular and $E \in M_n(\mathbb{R})$ with $E \geq 0$. Then*

$$\sigma(A, E) \leq \frac{1}{\max\limits_i \{\, |A^{-1}| \cdot E \,\}_{ii}}\ , \tag{19}$$

where 0^{-1} is interpreted as ∞.

Proof. Set $\alpha := \max\limits_i \{\, |A^{-1}| \cdot E \,\}_{ii} \neq 0$ and let i be an index, for which this maximum is achieved. Denote the $i\nu$-th component of A^{-1} by $A^{-1}_{i\nu}$ and define $u \in V_n(\mathbb{R})$ by $u_\nu := -\alpha^{-1} \cdot \mathrm{sign}(A^{-1}_{i\nu}) \cdot E_{\nu i}$. Then

$$e_i^T \cdot A^{-1} \cdot u = -\alpha^{-1} \cdot \sum_{\nu=1}^{n} |A^{-1}_{i\nu}| \cdot E_{\nu i} = -1, \tag{20}$$

and Corollary 4.2 implies $\det(A + u \cdot e_i^T) = 0$. Now $|ue_i^T| \leq \alpha^{-1} \cdot E$ yields $\sigma(A, E) \leq \alpha^{-1}$. □

EXAMPLE 5.2. *The upper bound (19) can be arbitrarily weak. Consider*

$$A = \begin{pmatrix} \varepsilon & 0 & 1 & 1 \\ 0 & \varepsilon & 1 & 1 \\ 1 & 1 & \varepsilon & 0 \\ 1 & 1 & 0 & \varepsilon \end{pmatrix}, E = |A| \quad \text{with} \quad |A^{-1}| \cdot |A| \approx \begin{pmatrix} 1 & 1 & 1/\varepsilon & 1/\varepsilon \\ 1 & 1 & 1/\varepsilon & 1/\varepsilon \\ 1/\varepsilon & 1/\varepsilon & 1 & 1 \\ 1/\varepsilon & 1/\varepsilon & 1 & 1 \end{pmatrix}, \tag{21}$$

where the components of $|A^{-1}| \cdot |A|$ are accurate up to a relative error ε. Then (19) gives $\sigma(A, |A|\,) \leq 1 + 0(\varepsilon)$. On the other hand,

$$\det(A + \varepsilon \cdot \delta A) = 0 \qquad \text{for } \delta A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

showing $\sigma(A, |A|) \leq \varepsilon$.

In Theorem 5.1, a rank-1 perturbation is used to prove (19). In a normwise sense, the minimum distance to the nearest singular matrix is achieved by a rank-1 perturbation. This is no longer true for componentwise distances, as will be shown by the following example.

EXAMPLE 5.3. *According to Corollary 4.2, the smallest σ such $A + \sigma e$ is singular with $|e| \leq \sigma \cdot |A|$ and rank(e) = 1 is given by $\sigma = |\widehat{\varphi}|^{-1}$, where $\widehat{\varphi}$ is an optimal value of the constraint optimization problem*

$$\varphi(u,v) := v^T A^{-1} u = \text{Min!} \quad \text{subject to} \quad |uv^T| \leq |A|.$$

In Example 5.2, partition the vectors $u, v \in V_4(\mathbb{IR})$ into two vectors $U_i, V_i \in V_2(\mathbb{IR})$, $i \in \{1,2\}$, either having 2 components. That means $u = (U_1, U_2)^T$, $v = (V_1, V_2)^T$. Let

$$|uv^T| \leq |A|, \quad \textit{i.e.} \quad U_i V_i^T \leq \varepsilon \cdot I \quad \text{and} \quad U_i V_j^T \leq (\mathbf{1})_{22} \quad \text{for } 1 \leq i,\ j \leq 2,\ i \neq j.$$

The large elements of A^{-1} are in the upper left and lower right 2 by 2 block:

$$A^{-1} \approx \begin{pmatrix} X & Y \\ Y & X \end{pmatrix} \quad \text{with} \quad X = \frac{1}{2\varepsilon} \cdot \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad Y = \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

up to a relative error of the order ε. Therefore

$$\begin{aligned} |v^T A^{-1} u| &\leq |V_1^T X U_1| + |V_2^T X U_2| + |V_1^T Y U_2| + |V_2^T Y U_1| \\ &\leq 2\varepsilon \cdot \sum |X_{ij}| + 2 \cdot \sum |Y_{ij}| \leq 6 \end{aligned}$$

Therefore, Corollary 4.2 implies that the minimum distance to the nearest singular matrix subject to rank-1 perturbations weighted by $|A|$ is at least $1/6$ compared to $\sigma(A, |A|) \leq \varepsilon$. This observation sheds light on the difficulties to calculate $\sigma(A, E)$ or to find upper bounds for it.

One may define the rank-k componentwise distance to the nearest singular matrix as follows

$$\sigma_k(A,E) := \min\{\, \alpha \in \mathbb{IR} \mid A + \widetilde{E} \text{ singular for } |\widetilde{E}| \leq \alpha \cdot E \text{ and } \operatorname{rank}(\widetilde{E}) \leq k \,\}.$$

We use $\operatorname{rank}(\widetilde{E}) \leq k$ *because E may be rank-deficient. We have just seen in Example 5.3 that $\sigma_2(A,E)/\sigma_1(A,E)$ may be arbitrarily small.* ∎

Given the lower bound (10), one may ask whether there exist finite constants $\gamma(n) \in \mathbb{IR}$ only depending on n such that

$$\frac{1}{\rho(\,|A^{-1}| \cdot E)} \leq \sigma(A,E) \leq \frac{\gamma(n)}{\rho(\,|A^{-1}| \cdot E)} \tag{22}$$

for all nonsingular $A \in M_n(\mathbb{IR})$ and $0 \leq E \in M_n(\mathbb{IR})$. This question has been raised in [3] and answered for some classes of matrices. The main purpose of this paper is to derive bounds for $\gamma(n)$. This will be done by using Lemma 4.1. For this purpose we need the following result by Rohn (for notation see §1).

THEOREM 5.4. **(Rohn)** *For nonsingular $A \in M_n(\mathbb{IR})$ and $0 \leq E \in M_n(\mathbb{IR})$ holds*

$$\frac{1}{\max\limits_{S_1,S_2} \rho_0(S_1 A^{-1} S_2 E)} = \sigma(A,E),$$

where ρ_0 denotes the real spectral radius and the maximum is taken over all signature matrices. 1/0 is interpreted as ∞.

Proof. cf. [9]. □

We start with a theorem bounding $\gamma(n)$ for general weight matrices E, and identify a class of matrices with $\gamma(n) = 1$.

THEOREM 5.5. *For nonsingular $A \in M_n(\mathbb{IR})$ and $0 \leq E \in M_n(\mathbb{IR})$, the following is true.*

i) Assume a matrix $S \in M_n(\mathbb{IR})$ of rank 1 exists with

$$S_{ij} = \begin{cases} +1 & \text{if} \quad A_{ij}^{-1} > 0 \\ -1 & \text{if} \quad A_{ij}^{-1} < 0 \\ +1 \text{ or } -1 & \text{if} \quad A_{ij}^{-1} = 0 \end{cases} .$$

Then (22) holds with $\gamma(n) = 1$.

ii) If $0 < \eta \leq |E_{ij}| \leq \zeta$ *for all* $1 \leq i, j \leq n$, *then (22) holds with* $\gamma(n) = n \cdot \zeta / \eta$.

Proof. Let $S = uv^T$ with $u, v \in V_n(\mathbb{R})$, $|u| = |v| = (\mathbf{1})$. Defining $S_1 = \mathrm{diag}(u)$, $S_2 = \mathrm{diag}(v)$, we have $S_1 A^{-1} S_2 = |A^{-1}|$ and Rohn's characterization in Theorem 5.4 proves the first part. W.l.o.g. assume $\sigma(A, E) < \infty$. It is $\eta \cdot \|A^{-1}\|_\infty \leq \max_i (\,|A^{-1}| \cdot E)_{ii}$ and $\rho(\,|A^{-1}| \cdot E) \leq \|A^{-1}\|_\infty \cdot \|E\|_\infty \leq n \cdot \zeta \cdot \|A^{-1}\|_\infty$. Thus, Theorem 5.1 proves the second part and therefore the theorem. □

For important classes of matrices such as nonnegative invertible matrices, among them all M-matrices, we already have a precise formula for $\sigma(A, E)$:

$$(23) \qquad A \in M_n\mathbb{R} \text{ nonnegative invertible, } 0 \leq E \in M_n(\mathbb{R}) \Rightarrow \ \sigma(A, E) = \frac{1}{\rho(\,|A^{-1}| \cdot E)}.$$

Example 5.6. *If constants* $\gamma(n)$ *with (22) exist at all, we can give a lower bound on* $\gamma(n)$ *by means of the following. Define* $A \in M_n(\mathbb{R})$ *by*

$$(24) \qquad A := \begin{pmatrix} 1 & & & & & & s \\ 1 & 1 & & & 0 & & \\ & 1 & 1 & & & & \\ & & 1 & 1 & & & \\ & & & \ddots & \ddots & & \\ & 0 & & & & 1 & \\ & & & & & 1 & 1 \end{pmatrix} \quad \text{with } s := (-1)^{n+1}.$$

The determinant of A calculates to

$$\det(A) = \prod_{i=1}^{n} A_{ii} + (-1)^{n+1} \cdot \Pi_\zeta(A) = 2, \text{ where } \zeta = (1, \ldots, n)$$

and $\prod_\zeta(A) = A_{12} \cdot A_{23} \cdot \ldots \cdot A_{n-1,n} \cdot A_{n1}$. *If the elements of A are afflicted with relative perturbations, i.e.* $E = |A|$, *then only the 1's and s change. Therefore, any* $\widetilde{A}$ *with* $|\widetilde{A} - A| \leq \sigma \cdot |A|$ *with* $\sigma < 1$ *is nonsingular, and therefore* $\sigma(A, |A|\,) = 1$. *On the other hand,* $|A^{-1}| \cdot |A| = (\mathbf{1})_{nn}$ *and* $\rho(\,|A^{-1}| \cdot |A|\,) = n$. *This proves the following lemma.*

Lemma 5.7. *If constants* $\gamma(n) \in \mathbb{R}$ *with (22) for every nonsingular* $A \in M_n(\mathbb{R})$ *and* $0 \leq E \in M_n(\mathbb{R})$ *exist at all, then* $\gamma(n) \geq n$.

Next we show that $\gamma(n) \leq n$ for E being of rank 1. For the proof we use Corollary 4.2, which is a consequence of Lemma 4.1 for $k = 1$. In the remaining part of the paper, we will extend this proof to $k > 1$ to obtain upper bounds for $\gamma(n)$ and for general A, E.

Theorem 5.8. *Let nonsingular* $A \in M_n(\mathbb{R})$ *and* $0 \leq E \in M_n(\mathbb{R})$ *with* $E = uv^T$ *for some* $u, v \in V_n(\mathbb{R})$, $u, v \geq 0$. *Then*

$$\frac{1}{\rho(\,|A^{-1}| \cdot E)} \leq \sigma(A, E) \leq \frac{n}{\rho(\,|A^{-1}| \cdot E)} \ .$$

Proof. According to Theorem 5.4 and using (1.1),

$$(25) \qquad \sigma(A, E)^{-1} = \max_{S_1, S_2} \rho_0(S_1 A^{-1} S_2 u v^T) = \max_{S_1, S_2} v^T S_1 A^{-1} S_2 u,$$

where the maximum is taken over all signature matrices S_1, S_2. For any i, $1 \leq i \leq n$, we can choose appropriate signature matrices S_1, S_2 such that $v^T S_1 A^{-1} S_2 u \geq v_i \cdot (\,|A^{-1}| \cdot u)_i$. Using (25) this yields

$$\sigma(A, E)^{-1} \geq \max_i \ v_i \cdot (\,|A^{-1}| \cdot u)_i.$$

On the other hand, using (1.1),

$$\rho(\,|A^{-1}|\cdot E) = \rho(\,|A^{-1}|\cdot uv^T) = v^T\cdot|A^{-1}|\cdot u \le n\cdot\max_i\ v_i\cdot(\,|A^{-1}|\cdot u)_i. \qquad \square$$

COROLLARY 5.9. *For nonsingular $A\in M_n(\mathbb{R})$ and absolute perturbations, i.e. $E=(\mathbf{1})_{nn}$, estimation (22) holds with $\gamma(n)=n$.*

6. Estimation of $\gamma(n)$. To make further progress in the estimation of $\gamma(n)$ we show that for nonsingular A, $\sigma(A,E)$ depends continuously on A and E. Using this we can restrict the class of matrices A and E to matrices with only nonzero components. For the proof we can hardly use a simple continuity argument on $\rho_0(S_1A^{-1}S_2E)$ in connection with Theorem 5.4. This is because the search domain is restricted by E, and the (in absolute value) largest *real* eigenvalue may be multiple and become complex under arbitrarily small perturbations.

LEMMA 6.1. *For nonsingular $A\in M_n(\mathbb{R})$, $\sigma(A,E)$ depends continuously on A and E.*

Proof. For $\sigma(A,E)=\infty$ we show that $\sigma(\widetilde{A},\widetilde{E})$ becomes unbounded for $\widetilde{A}\to A$, $\widetilde{E}\to E$. A compactness and continuity argument shows that for every finite $0<c\in\mathbb{R}$:

$$\forall\ |e|\le c\cdot E: \qquad |\det(A+e)|\ge\delta>0.$$

For every $\widetilde{A}$, $\widetilde{E}$ close enough to A, E, this implies $|\det(\widetilde{A}+\widetilde{e})|\ge\delta/2>0$ for every $|\widetilde{e}|\le c\cdot\widetilde{E}$, and hence $\sigma(\widetilde{A},\widetilde{E})>c$.

Assume $\sigma:=\sigma(A,E)<\infty$. We will show that for small enough $\varepsilon>0$, there exists some $\delta>0$ such that both of the following statements are true:

$$\forall\ e\in M_n(\mathbb{R}):\ \ |e|\le(\sigma-\varepsilon)\cdot E \quad\Rightarrow\quad \det(A)\cdot\det(A+e)>\delta, \tag{26}$$

$$\exists\ e\in M_n(\mathbb{R}):\ \ |e|\le(\sigma+\varepsilon)\cdot E \ \text{ and }\ \det(A)\cdot\det(A+e)<-\delta. \tag{27}$$

(26) is seen as follows. For $\varepsilon>0$, the set of matrices $A+e$ with $|e|\le(\sigma-\varepsilon)\cdot E$ is nonempty and compact. Hence, $\det(A)\cdot\det(A+e)$ achieves a minimum on this set. By definition of σ, this minimum is positive. To see (27), observe that $\det(A)\cdot\det(A+e)\ge0$ for all $|e|\le\sigma\cdot E$. For any index pair i,j, the determinant $\det(A+\varepsilon\cdot e_ie_j^T)$ depends linearly on ε. Now proceed as follows. There is some e such that $A+e$ is singular and $|e|=E$. If for an index pair i,j, the determinant $\det(A+e)$ is independent on e_{ij}, then replace e_{ij} by 0. At each step of this process, $\det(A+e)=0$ and $|e|\le E$. The definition of $\sigma(A,E)<\infty$ implies that during this process we must arrive at some e and an index pair k, l, such that $\det(A+e)$ is not constant when changing e_{kl}. Then defining $e'\in M_n(\mathbb{R})$ by $e'_{ij}:=e_{ij}$ for $(i,j)\ne(k,l)$ and $e'_{kl}:=e_{kl}\cdot(1+\varepsilon')$ for small $\varepsilon'>0$ proves (27).

Now the continuity of the determinant implies for $\widetilde{A}$, $\widetilde{E}$ close enough to A,E,

$$\forall\ |\widetilde{e}|\le(\sigma-\varepsilon)\cdot\widetilde{E}: \qquad \det(\widetilde{A})\cdot\det(\widetilde{A}+\widetilde{e})>\delta/2 \ \text{ and}$$

$$\exists\ |\widetilde{e}|\le(\sigma+\varepsilon)\cdot\widetilde{E}: \qquad \det(\widetilde{A})\cdot\det(\widetilde{A}+\widetilde{e})<-\delta/2,$$

and therefore $\sigma(A,E)-\varepsilon<\sigma(\widetilde{A},\widetilde{E})<\sigma(A,E)+\varepsilon$. $\square$

COROLLARY 6.2. *If (22) holds for each $E>0$, then it holds for each $E\ge0$.*

Our goal for this chapter is to prove the following upper bound for $\sigma(A,E)$. The quantities φ_t occuring in this estimation will be quantified and estimated in §7.

PROPOSITION 6.3. *Let $A,E\in M_n(\mathbb{R})$ with A nonsingular and $E\ge0$ be given. Define recursively $\varphi_1:=1$, $\varphi_2:=1$ and $\varphi_t\in\mathbb{R}$, $2<t\in\mathbb{N}$ to be the (unique) positive root of*

$$P_t(x)\in\mathbb{R}[x] \quad\text{with}\quad P_t(x):=x^{t-1}-x^{t-2}-\sum_{\nu=1}^{t-1}\varphi_\nu^\nu\cdot x^{t-1-\nu}. \tag{28}$$

Then

$$\sigma(A,E) \le \frac{n\cdot\varphi_n}{\rho(\,|A^{-1}|\cdot E)}. \tag{29}$$

Therefore, the quantities $\gamma(n)$ defined in (5.4) satisfy

$$\gamma(1)=1,\quad \gamma(2)=2 \quad \text{and} \tag{30}$$
$$n\le\gamma(n)\le n\cdot\varphi_n.$$

The proof divides into several parts and needs some preparatory lemmata. First, we will construct a specific rank-k perturbation in order to be able to apply Lemma 4.1 to bound $\gamma(n)$ for general A, E. We use the same principle as in the proof of Theorem 5.1 adapted to rank-k perturbations.

LEMMA 6.4. *Let nonsingular $A\in M_n(\mathbb{R})$ and $0\le E\in M_n(\mathbb{R})$ be given, and set $C:=|A^{-1}|\cdot E$. For $1\le k\le n$ define*

$$i' := \begin{cases} i+1 & \text{for } 1\le i<k \\ 1 & \text{for } i=k \end{cases} \tag{31}$$

and $U,V\in M_{n,k}(\mathbb{R})$ by

$$U_{\nu i'} := \operatorname{sign}(A^{-1}_{i\nu})\cdot E_{\nu i'} \quad\text{and}\quad V_{\mu i}:=\delta_{\mu i}$$

for $1\le\mu,\nu\le n$, $1\le i\le k$ and the Kronecker symbol δ. Set $\widetilde{C}:=V^TA^{-1}U$. Then

i) $|\widetilde{C}|\le C[\omega]$ for $\omega=(1,\dots,k)$.
ii) $\widetilde{C}_{ii'}=C_{ii'}$ for $1\le i\le k$.
iii) $|UV^T|\le E$.
iv) $\sigma(A,E)\le\{\rho_0(\widetilde{C})\}^{-1}$, where 0^{-1} is interpreted as ∞.

Proof. For $1\le i,j\le k$ follows

$$|\,(V^TA^{-1}U)_{ij}| = |\sum_{\nu=1}^n\sum_{\mu=1}^n V_{\mu i}A^{-1}_{\mu\nu}U_{\nu j}| \le \sum_{\nu=1}^n |A^{-1}_{i\nu}|\cdot E_{\nu j} = C_{ij},$$

and therefore $|\widetilde{C}|\le C[\omega]$ and i). For $1\le i\le k$ holds

$$\widetilde{C}_{ii'} = (V^TA^{-1}U)_{ii'} = \sum_{\nu=1}^n\sum_{\mu=1}^n V_{\mu i}A^{-1}_{\mu\nu}U_{\nu i'} = \sum_{\nu=1}^n A^{-1}_{i\nu}\cdot\operatorname{sign}(A^{-1}_{i\nu})\cdot E_{\nu i'} = C_{ii'}.$$

and therefore ii). For $1\le\mu,\nu\le n$ holds

$$|\,(UV^T)_{\nu\mu}| = |\sum_{i=1}^k U_{\nu i}V_{\mu i}|,$$

such that $|(UV^T)_{\nu\mu}| = E_{\nu\mu}$ for $1\le\mu\le k$, and $|(UV^T)_{\nu\mu}|=0$ for $k+1\le\mu\le n$. This proves iii). For $\lambda:=\rho_0(\widetilde{C})>0$, it is $\det(\lambda\cdot I - s\cdot\widetilde{C})=0$ for $s=-1$ or $s=1$. Lemma 4.1 implies

$$\det(A - s\cdot\lambda^{-1}\cdot UV^T) = \det(A)\cdot\det(I_k - s\cdot\lambda^{-1}\cdot V^TA^{-1}U) = 0.$$

Together with iii) and the definition (0.2) of $\sigma(A,E)$, this proves iv) and the theorem. □

Our aim is to construct a rank-k perturbation of A with large real spectral radius. Then Lemma 4.1 allows to give an upper bound on $\sigma(A,E)$. A first step is the following, first generalization of Theorem 5.1. It will later yield the precise value for $\gamma(2)$.

THEOREM 6.5. *Let $A\in M_n(\mathbb{R})$ be nonsingular and $E\in M_n(\mathbb{R})$ with $E\ge 0$. For $C:=|A^{-1}|\cdot E$ holds*

$$\sigma(A,E)\le\frac{1}{\max\limits_{i,j}\sqrt{C_{ij}\cdot C_{ji}}}. \tag{32}$$

Proof. For $i = j$, (32) has been proved in Theorem 5.1. Reordering of indices puts the cycle (i, j), for which the maximum in (6.7) is achieved, into the cycle (1,2), and Lemma 6.4 proves for $i \neq j$ existence of a 2 by 2 matrix $\widetilde{C} = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ with $0 \leq \beta = C_{ij}$, $0 \leq \gamma = C_{ji}$, $|\alpha| \leq C_{ii}$, $|\delta| \leq C_{jj}$, and $\sigma(A, E) \leq \rho_0(\widetilde{C})^{-1}$. If $|\alpha\delta| \geq \beta\gamma$, then $\sqrt{C_{ii} \cdot C_{jj}} \geq \sqrt{C_{ij} \cdot C_{ji}}$ and Theorem 5.1 yields (32). Otherwise, $\det(\widetilde{C}) < 0$. The characteristic polynomial of $\widetilde{C}$ is $\lambda^2 - \mathrm{trace}(\widetilde{C}) \cdot \lambda + \det(\widetilde{C})$, so that the eigenvalues of $\widetilde{C}$ are $\frac{1}{2} \cdot \Big\{ \mathrm{trace}(\widetilde{C}) \pm \sqrt{\mathrm{trace}(\widetilde{C})^2 - 4 \cdot \det(\widetilde{C})} \Big\} = \frac{1}{2} \cdot \Big\{ \alpha + \delta \pm \sqrt{(\alpha - \delta)^2 + 4\beta\gamma} \Big\}$ are both real. The absolute value of one of them is not less than $\sqrt{\beta\gamma}$, i.e. $\sigma(A, E) \leq \rho_0(\widetilde{C})^{-1} \leq (\beta\gamma)^{-1/2}$. □

The idea of the proof of Theorem 6.5 is the following: for a given cycle of C of length 2, a suitable rank-2 perturbation of A is constructed which allows to prove an upper bound of $\sigma(A, E)$ by using Lemma 6.4. In the following we will carry this idea to cycles of C of length k, $1 \leq k \leq n$.

First, we will identify a class of matrices for which we can give explicit *lower* bounds for their real spectral radius. The class of matrices is constructed in such a way that the matrices given in Lemma 6.4 can be used to bound $\sigma(A, E)$ from above.

LEMMA 6.6. *Let nonnegative $C \in M_k(\mathbb{R})$ and some $0 < a \in \mathbb{R}$ be given. Define $\varphi_1 := 1$, $\varphi_2 := 1$, and for $t > 2$ define recursively $\varphi_t \in \mathbb{R}$ to be the positive zero of*

$$P_t(x) \in \mathbb{R}[x] \quad \text{with} \quad P_t(x) := x^{t-1} - x^{t-2} - \sum_{\nu=1}^{t-1} \varphi_\nu^\nu \cdot x^{t-1-\nu}. \tag{33}$$

Suppose

$$\forall\ 1 \leq \mu < k \quad \forall\ \overline{\omega} \in \Gamma_{\mu k} : \qquad (\Pi_{\overline{\omega}}(C))^{1/\mu} \leq \varphi_\mu \cdot a, \tag{34}$$

and for $\omega = (1, \ldots, k)$,

$$|\Pi_\omega(C)|^{1/k} \geq \varphi_k \cdot a. \tag{35}$$

Then, for i' defined as in (6.6) and every $\widetilde{C} \in M_k(\mathbb{R})$ with

$$|\widetilde{C}| \leq C \quad \text{and} \quad \widetilde{C}_{ii'} = C_{ii'} \quad \text{for} \quad 1 \leq i \leq k, \tag{36}$$

holds

$$\rho_0(\widetilde{C}) \geq a.$$

Proof. The **proof** divides in the following parts. First, we transform C into a standard form such that all $C_{ii'}$ in the cycle $(1, \ldots, k)$ in (35) are equal. Second, we bound C by a circulant, show regularity of that matrix and $\det(\widetilde{C} - \lambda I) \neq 0$ for all $0 \leq \lambda < a$. Finally, the sign of the determinant of any $\widetilde{C}$ with (36) is determined, from which the lemma follows.

The case $k = 1$ is trivial; for $k = 2$ the proof of $\rho_0(\widetilde{C}) \geq a$ is included in the proof of Theorem 6.5.

Assume $k > 2$, and set $b := |\Pi_\omega(C)|^{1/k}$. Direct computation shows that any similarity transformation of C by a diagonal matrix D leaves all cycle products invariant.

Thus (34) and (35) remain valid for any diagonal D with positive diagonal entries. Define diagonal $D \in M_k(\mathbb{R})$ by

$$f_i := b^{-1} \cdot C_{ii'} \quad \text{and} \quad D_{ii} := \prod_{\nu=i}^{k} f_\nu \quad \text{for } 1 \leq i \leq k.$$

We show that w.l.o.g. C can be replaced by $D^{-1}CD$. We have $f_i > 0$, and (6.10) implies $D_{11} = 1$. It is

$$(D^{-1}CD)_{ii'} = \left(\prod_{\nu=i}^{k} f_\nu^{-1} \right) \cdot C_{ii'} \cdot \left(\prod_{\nu=i'}^{k} f_\nu \right) = C_{ii'} \cdot f_i^{-1} = b. \tag{37}$$

If $\widetilde{C} \in M_k(\mathbb{R})$ is any matrix satisfying (36), then $|D^{-1}\widetilde{C}D| \leq D^{-1}CD$, and (37) yields $(D^{-1}\widetilde{C}D)_{ii'} = (D^{-1}CD)_{ii'} = b$. Since the set of eigenvalues of $\widetilde{C}$ and $D^{-1}\widetilde{C}D$ are identical, we can restrict our attention to matrices $C \in M_n(\mathbb{R})$, $C \geq 0$ and

$$C_{ii'} = b \quad \text{for} \quad 1 \leq i \leq k. \tag{38}$$

Set

$$C = \begin{pmatrix} c_{1,1} & b & c_{1,k-1} & \dots & & c_{1,2} \\ c_{2,2} & c_{2,1} & b & c_{2,k-1} & \dots & c_{2,3} \\ & \dots & c_{3,1} & b & \dots & \\ c_{k-1,k-1} & c_{k-1,k-2} & & \dots & & b \\ b & c_{0,k-1} & & \dots & & c_{0,1} \end{pmatrix}. \tag{39}$$

Let $\mu \in \mathbb{N}$, $1 \leq \mu < k$ be given and define $\overline{\omega} \in \Gamma_{\mu k}$ by $\overline{\omega} = (1, \dots, \mu)$. Then setting $q := a/b$, (34) implies

$$c_{\mu\mu} \cdot \prod_{i=1}^{\mu-1} b \;\leq\; (\varphi_\mu \cdot a)^\mu \quad \text{and therefore} \quad c_{\mu\mu} \leq b \cdot \varphi_\mu^\mu \cdot q^\mu.$$

Applying the same argument successively for $\overline{\omega} = \big(i, (i+1) \bmod \mu, \dots, (i+\mu) \bmod \mu\big)$ yields

$$c_{i,\mu} \leq b \cdot \varphi_\mu^\mu \cdot q^\mu \quad \text{for all} \quad 0 \leq i < k,\ 1 \leq \mu < k. \tag{40}$$

Therefore,

$$C \leq b \cdot \begin{pmatrix} c_1 & 1 & c_{k-1} & & & c_2 \\ c_2 & c_1 & 1 & c_{k-1} & \dots & c_3 \\ & & c_1 & 1 & & \\ & \dots & & \dots & \dots & \\ c_{k-1} & c_{k-2} & & \dots & & 1 \\ 1 & c_{k-1} & & & & c_1 \end{pmatrix} =: b \cdot \overline{C} \tag{41}$$

with $c_\mu := \varphi_\mu^\mu \cdot q^\mu$ for $1 \leq \mu < n$.

Let $\widetilde{C} \in M_k(\mathbb{R})$ with (36) be given, and let $\lambda \in \mathbb{R}$ with $0 \leq \lambda < a$. Next we show that all matrices $\widetilde{C} - \lambda I$ are nonsingular. By assumption (6.11) and using (6.16),

$$|\widetilde{C} - \lambda I| \leq C + \lambda \cdot I \leq b \cdot \overline{C} + \lambda I \quad \text{and} \quad (\widetilde{C} - \lambda I)_{ii'} = \widetilde{C}_{ii'} = C_{ii'} = b. \tag{42}$$

By (41) and (33), using $q := a/b \leq \varphi_k^{-1}$ from (35) and $\varphi_2 = 1$, we have for $k \geq 3$,

$$\begin{aligned} \lambda + b \cdot \sum_{\nu=1}^{k-1} c_\nu &< b \cdot \left\{ q + \sum_{\nu=1}^{k-1} \varphi_\nu^\nu \cdot q^\nu \right\} \leq b \cdot \left\{ \varphi_k^{-1} + \sum_{\nu=1}^{k-1} \varphi_\nu^\nu \cdot \varphi_k^{-\nu} \right\} = \\ &= b \cdot \varphi_k^{-k+1} \cdot \left\{ \varphi_k^{k-2} + \sum_{\nu=1}^{k-1} \varphi_\nu^\nu \cdot \varphi_k^{k-\nu-1} \right\} = b \cdot \varphi_k^{-k+1} \cdot \varphi_k^{k-1} = b. \end{aligned} \tag{43}$$

This shows that the element $b = \widetilde{C}_{ii'} = C_{ii'}$ strictly dominates the sum of the absolute values of the other components in each row of $C + \lambda I$ and of $\widetilde{C} - \lambda I$. That means, multiplication by a suitable permutation matrix produces a strictly diagonally dominant matrix and proves regularity of every $\widetilde{C} - \lambda I$ with $\widetilde{C}$ satisfying (36) and $0 \leq \lambda < a$.

We proved that for every $\widetilde{C} \in M_k(\mathbb{R})$ with (36), the determinant of $\widetilde{C} - \lambda I$ is nonzero for $0 \leq \lambda < a$. Therefore, the value of the characteristic polynomial $p(\lambda) = \det(\lambda I - \widetilde{C})$ of $\widetilde{C}$ has the same sign for $0 \leq \lambda < a$. Now $p(\lambda) \to +\infty$ for $\lambda \to +\infty$. Therefore, the lemma is proved if we can show $p(0) < 0$, because in this case the characteristic polynomial must intersect with the real axis for some $\lambda^* \geq a$, thus proving $\rho_0(\widetilde{C}) \geq \lambda^* \geq a$.

We already proved that *every* matrix $\widetilde{C}$ satisfying (36) is nonsingular. Therefore

$\operatorname{sign}\big(p(0)\big) = \operatorname{sign}\big(\det(-B)\big)$ for every matrix B with $|B| \le C$ and $B_{ii'} = C_{ii'} = b$. Define

$$B_{ij} := \begin{cases} C_{ii'} & \text{for } j = i' \\ 0 & \textit{otherwise} \end{cases} .$$

Then $\operatorname{sign}\big(\det(B)\big) = (-1)^{k+1}$ and therefore $\operatorname{sign}\big(p(0)\big) = (-1)^{2k+1} = -1$. The theorem is proved. ∎

EXAMPLE 6.7. *One can show that, at least for odd n, the bounds in Lemma 6.6 are sharp in the sense that there are examples with equality in (34) and (35) such that $\widetilde{C}$ with (36) exists with $\rho_0(\widetilde{C}) = a$. Consider*

$$C := \begin{pmatrix} a & b & c \\ c & a & b \\ b & c & a \end{pmatrix} \quad \text{with} \quad b := \varphi_3 \cdot a \quad \text{and} \quad c := a/\varphi_3 \quad \text{and} \quad \widetilde{C} := \begin{pmatrix} -a & b & -c \\ -c & -a & b \\ b & -c & -a \end{pmatrix} .$$

Then $C_{11} = \varphi_1 \cdot a = a$, $\sqrt{C_{12}C_{21}} = \varphi_2 \cdot a = a$ and $(C_{12}C_{23}C_{31})^{1/3} = \varphi_3 \cdot a$. $\widetilde{C}$ is a circulant, and its eigenvalues compute to $P(\varepsilon^k)$, $k = 0, 1, 2$ where $\varepsilon = e^{2\pi i/3}$ and $P(x) = bx^2 - cx - a$ (cf.[6]). It is $b - c - a = b(1 - \varphi_2^2 q^2 - \varphi_1 q) = b \cdot \varphi_3^{-1} = a$ with $q := a/b$. The other two eigenvalues are complex, thus $\rho_0(\widetilde{C}) = a$. The example extends to odd $n \in \mathbb{N}$.

The combination of Lemma 6.4, Theorem 6.5, and Lemma 6.6 gives the key to construct a rank-k perturbation of A to achieve an upper bound for $\sigma(A, E)$. The following theorem is the generalization of Theorems 5.1 and 6.5 for cycles of length k, $1 \le k \le n$.

THEOREM 6.8. *Let $A, E \in M_n(\mathbb{R})$ with nonsingular A and $E \ge 0$ be given and define $C := |A^{-1}| \cdot E$. For $1 \le k \le n$ and any $\omega \in \Gamma_{kn}$ set*

$$0 \ne \tau := \big(\Pi_\omega(C)\big)^{1/k} . \tag{44}$$

Then for φ_k defined as in Lemma 6.6,

$$\sigma(A, E) \le \varphi_k/\tau.$$

In other words, φ_k divided by the geometric mean of the elements of any cycle of C bounds $\sigma(A, E)$ from above.

Proof. Let some $\omega \in \Gamma_{kn}$ and τ from (44) be given and set $a := \tau/\varphi_k$. If for $k = 1$ or $k = 2$ there exists some $\overline{\omega} \in \Gamma_{kn}$ with $\{\Pi_{\overline{\omega}}(C)\}^{1/k} \ge \varphi_k \cdot a$, then $\varphi_1 = \varphi_2 = 1$ and Theorem 5.1 and Theorem 6.5 imply $\sigma(A, E) \le a^{-1} = \varphi_k/\tau$. Therefore, we may assume $\{\Pi_{\overline{\omega}}(C)\}^{1/k} < \varphi_k \cdot a$ for all $\overline{\omega} \in \Gamma_{kn}$ and $k \in \{1, 2\}$. Hence, there is some $m \in \mathbb{N}$, $2 \le m \le k$ such that

$$\forall\ 1 \le \mu < m \quad \forall\ \overline{\omega} \in \Gamma_{\mu n} : (\Pi_{\overline{\omega}}(C))^{1/\mu} \le \varphi_\mu \cdot a,$$

and

$$\exists\ \widetilde{\omega} \in \Gamma_{mn} : \quad (\Pi_{\overline{\omega}}(C))^{1/m} \ge \varphi_m \cdot a.$$

After suitable rearrangement of indices we may assume $\widetilde{\omega} = (1, \ldots, m)$, and Lemma 6.4 yields a matrix $\widetilde{C} \in M_m(\mathbb{R})$ with properties i) and ii) of Lemma 6.4 and $\sigma(A, E) \le \rho_0(\widetilde{C})^{-1}$. But Lemma 6.6 shows for all such matrices $\rho_0(\widetilde{C}) \ge a = \tau/\varphi_m$. Regarding $m \le k$, the theorem is proved if we can show

$$t \in \mathbb{N} \quad \Rightarrow \quad \varphi_t \le \varphi_{t+1}. \tag{45}$$

We know $\varphi_1 = \varphi_2 = 1$, from the definition (6.8) we see $\varphi_3 = 1 + \sqrt{2}$, and for $t \ge 3$

$$P_{t+1}(x) = x \cdot P_t(x) - \varphi_t^t.$$

Hence $P_{t+1}(\varphi_t) < 0$ and $\varphi_{t+1} > \varphi_t$. The theorem is proved. ∎

Theorem 6.8 reduces the problem of finding upper bounds of $\sigma(A, E)$ to finding proper cycles of some length k of $|A^{-1}| \cdot E$ with large geometric mean corresponding to a suitable rank-k perturbation. This is done in the following proof of Proposition 6.3.

Proof. of Proposition 6.3. Corollary 6.2 allows us to assume $E > 0$. Therefore, $|A^{-1}| \cdot E$ is positive, and Perron-Frobenius Theory yields existence of a positive eigenvector $x \in V_n(\mathbb{R})$ with $|A^{-1}| \cdot E \cdot x = \rho(\,|A^{-1}| \cdot E) \cdot x$, $\rho(\,|A^{-1}| \cdot E) > 0$. Define the diagonal matrix $D_x \in M_n(\mathbb{R})$ by $(D_x)_{ii} := x_i$. We may replace A by $A \cdot D_x$ and E by $E \cdot D_x$, because for any nonsingular diagonal matrices D_1, D_2, $\sigma(A, E) = \sigma(D_1 A D_2, D_1 E D_2)$. This is because $|\delta A| \leq \sigma \cdot E$ *iff* $|D_1 \cdot \delta A \cdot D_2| \leq \sigma \cdot |D_1 E D_2|$ and $A + \delta A$ is singular *iff* $D_1 A D_2 + D_1 \cdot \delta A \cdot D_2$ is singular (cf. [3]). Then

$$C := |\,(A \cdot D_x)^{-1}| \cdot E \cdot D_x = D_x^{-1} \cdot |A^{-1}| \cdot E \cdot D_x \quad \text{and} \quad C \cdot (\mathbf{1}) = \rho(\,|A^{-1}| \cdot E) \cdot (\mathbf{1}).$$

That means C is a multiple of a row stochastic matrix. Set $\rho := \rho(\,|A^{-1}| \cdot E)$.

Denote an index of the maximal component of C in row i by m_i. Then either $\{\, m_i \mid 1 \leq i \leq n \,\} = \{1, \ldots, n\}$ or, there is a cycle $m_j, m_{j+1}, \ldots, m_{j+k-1}, m_{j+k} = m_j$ of length k. That means, with a suitable renumbering, there is some $k \in \mathbb{N}$, $1 \leq k \leq n$ such that for the upper left k by k principal submatrix of C holds

$$C_{ii'} \geq \rho/n \quad \text{for} \quad 1 \leq i \leq k, \tag{46}$$

where i' is defined as in (6.6).

Then Theorem 6.8, (6.20) and (6.21) imply for $\omega = (1, \ldots, k)$,

$$\sigma(A, E) \leq \varphi_k \cdot \{\Pi_\omega(C)\}^{-1/k} \leq n \cdot \varphi_k/\rho \leq n \cdot \varphi_n/\rho. \qquad \square$$

In the remaining §7, we will replace the bound (30) by giving explicit bounds for $\gamma(n)$ only depending on n. An asymptotic bound will be given as well.

7. Explicit bounds for $\gamma(n)$. The main result in §6 is the upper bound (30) in Proposition 6.3. This bound is given in terms of φ_k, the positive zeros of the polynomial P_t defind in (28). In the remaining part of the paper we will give bounds on $\gamma(n)$ showing the dependence on n by a simple function. Moreover, the asymptotic behaviour of $\gamma(n)$ for $n \to \infty$ is given.

The polynomials $P_t(x) \in \mathbb{R}[x]$ defined in (28) satisfy

$$P_t(x) = x^{t-1} - x^{t-2} - \sum_{\nu=1}^{t-1} \varphi_\nu^\nu \cdot x^{t-1-\nu} \quad \text{and} \quad P_t(\varphi_t) = 0 \ \text{ for } \ t > 2. \tag{47}$$

Therefore, for $n \geq 3$,

$$\varphi_n^{-1} + \sum_{i=1}^{n-1} \varphi_i^i \cdot \varphi_n^{-i} = 1. \tag{48}$$

By (6.20), $x + \sum\limits_{i=1}^{n-1} \varphi_i^i \cdot x^i$ is strictly increasing for $x > 0$. Hence, for $x > 0$,

$$x + \sum_{i=1}^{n-1} \varphi_i^i \cdot x^i \leq 1 \quad \text{implies} \quad x \leq \varphi_n^{-1}, \quad \text{that is} \quad \varphi_n \leq x^{-1}. \tag{49}$$

We are aiming on a bound of the form

$$\varphi_k \leq c \cdot k^\alpha \tag{50}$$

for some constants c and α. To determine c and α, we notice that if (7.4) is satisfied for $1 \leq k < n$, then

$$\sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^{\alpha_i} \leq 1 - c^{-1} \cdot n^{-\alpha} \quad \text{implies} \quad \varphi_n \leq c \cdot n^\alpha. \tag{51}$$

This is because the left hand side of (7.5) yields

$$1 \geq c^{-1} \cdot n^{-\alpha} + \sum_{i=1}^{n-1} i^{\alpha i} \cdot n^{-\alpha i} \geq (c \cdot n^\alpha)^{-1} + \sum_{i=1}^{n-1} \varphi_i^i \cdot (c \cdot n^\alpha)^{-i},$$

and (7.3) implies $(c \cdot n^\alpha)^{-1} \leq \varphi_n^{-1}$.

Therefore, our first step is to derive upper bounds for

$$\sum_{i=1}^{n-1} \sigma_i \quad \text{with} \quad \sigma_i := \left(\frac{i}{n}\right)^{i\alpha}. \tag{52}$$

σ_i depends on n and α. We use the abbreviation σ_i for fixed n and α and omit extra parameters for better readability. In order to estimate the sum (52), we will split it into 3 parts, which will be bounded individually. For $i \geq 1$ holds

$$\frac{\sigma_{i+1}}{\sigma_i} = \left(\frac{i+1}{n}\right)^{(i+1)\alpha} \cdot \left(\frac{n}{i}\right)^{(i+1)\alpha} \cdot \left(\frac{i}{n}\right)^{\alpha} = \left(\frac{i}{n}\right)^{\alpha} \cdot \left(1+\frac{1}{i}\right)^{(i+1)\alpha} > \left(\frac{i}{n}\right)^{\alpha} \cdot e^{\alpha},$$

and therefore

$$\sigma_i < \left(\frac{n}{i \cdot e}\right)^{\alpha} \cdot \sigma_{i+1} \quad \text{for} \quad i \geq 1. \tag{53}$$

For all $\beta \in \mathbb{R}$ with $1 < \beta < e$ and $k := \lceil \frac{n \cdot \beta}{e} \rceil$ holds $k - 1 < \frac{n \cdot \beta}{e} \leq k$. Then (53) gives

$$\sum_{i=k}^{n-1} \sigma_i < \sigma_{n-1} + \sum_{i=k}^{n-2} \left(\frac{n}{i \cdot e}\right)^{\alpha} \cdot \sigma_{i+1} = \sigma_{n-1} + \sum_{i=k+1}^{n-1} \left(\frac{n}{(i-1) \cdot e}\right)^{\alpha} \cdot \sigma_i,$$

and therefore

$$\sigma_{n-1} - \sigma_k > \sum_{i=k+1}^{n-1} \left\{1 - \left(\frac{n}{(i-1) \cdot e}\right)^{\alpha}\right\} \cdot \sigma_i \geq \sum_{i=k+1}^{n-1} \left\{1 - \left(\frac{n}{k \cdot e}\right)^{\alpha}\right\} \cdot \sigma_i,$$

and $\frac{n}{k \cdot e} \leq \beta^{-1}$ yields

$$\sigma_{n-1} - \sigma_k > (1 - \beta^{-\alpha}) \cdot \sum_{i=k+1}^{n-1} \sigma_i.$$

By choice, $\beta > 1$, and $\alpha \geq 0$ implies $(1 - \beta^{-\alpha})^{-1} > 1$. Therefore,

$$\sum_{i=k}^{n-1} \sigma_i < (1 - \beta^{-\alpha})^{-1} \cdot \sigma_{n-1} = (1 - \beta^{-\alpha})^{-1} \cdot \left(\frac{n-1}{n}\right)^{(n-1)\alpha} =: \mu_n \tag{54}$$

holds for every $\alpha \geq 0$, $1 < \beta < e$ and $k := \lceil \frac{n\beta}{e} \rceil$. This is the first part of the sum (7.6) for a suitable k to be determined. Define

$$f(x) := \left(\frac{x}{n}\right)^{x\alpha} \quad \text{with} \quad f'(x) = \left(\frac{x}{n}\right)^{x\alpha} \cdot \left\{\alpha \cdot \ln \frac{x}{n} + \alpha\right\}.$$

For $x > 0$, $f(x)$ has exactly one minimum at $x = \frac{n}{e}$. Then $f(i) = \sigma_i$ shows

$$\sigma_k \geq \sigma_l \quad \text{for} \quad 1 \leq k \leq l \leq \frac{n}{e}, \qquad \text{and} \qquad \sigma_k \leq \sigma_l \quad \text{for} \quad \frac{n}{e} \leq k \leq l \leq n-1. \tag{55}$$

Set $M := \lceil n/e \rceil$. Then $k = \lceil \frac{n\beta}{e} \rceil$ satisfies $M \leq k \leq n$, and (55) implies for $n \geq 3$,

$$\sum_{i=M}^{k-1} \sigma_i \leq (k - M) \cdot \sigma_{k-1} < \left(\frac{n\beta}{e} + 1 - \frac{n}{e}\right) \cdot f\left(\frac{n\beta}{e}\right) < \frac{n\beta}{e} \cdot f\left(\frac{n\beta}{e}\right)$$

(56)

$$\leq n \cdot \left(\frac{\beta}{e}\right)^{n\frac{\beta\alpha}{e}+1} =: \nu_n.$$

This is the second part of the sum (7.6). Finally, (7.9) implies

$$\sum_{i=1}^{M-1} \sigma_i < \left(\frac{1}{n}\right)^{\alpha} + \left(\frac{2}{n}\right)^{2\alpha} + \left(\frac{3}{n}\right)^{3\alpha} + \left(\frac{4}{n}\right)^{4\alpha} \cdot \left(\frac{n}{e}\right) =: \xi_n, \tag{57}$$

which is the third part of the sum (7.6). The inequalities (54), (56) and (57) together yield

$$n^{-\alpha} + \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^{i\alpha} \leq n^{-\alpha} + \mu_n + \nu_n + \xi_n \quad \text{for} \quad n \geq 3. \tag{58}$$

Next, we show that all three sequences μ_n, ν_n, ξ_n are decreasing for large enough n. $\left(1+\frac{1}{n}\right)^n$ is monotonically increasing for $n \geq 1$, therefore for $n \geq 2$,

$$\left(\frac{n+1}{n}\right)^{n\alpha} \geq \left(\frac{n}{n-1}\right)^{(n-1)\alpha} \;\Rightarrow\; \mu_{n+1} \leq \mu_n.$$

Suppose

$$n_0 \geq \left\{ \left(\frac{e}{\beta}\right)^{\frac{\beta\alpha}{e}} - 1 \right\}^{-1}. \tag{59}$$

Then for $n \geq n_0$,

$$1+\frac{1}{n} \leq \left(\frac{e}{\beta}\right)^{\frac{\beta\alpha}{e}} \;\Rightarrow\; (n+1)\cdot\left(\frac{\beta}{e}\right)^{\frac{\beta\alpha}{e}} \leq n \;\Rightarrow (n+1)\cdot\left(\frac{\beta}{e}\right)^{(n+1)\frac{\beta\alpha}{e}+1} \leq n\cdot\left(\frac{\beta}{\alpha}\right)^{n\frac{\beta\alpha}{e}+1},$$

and therefore $\nu_{n+1} \leq \nu_n$ for $n \geq n_0$ with n_0 satisfying (59). Finally, for $n \geq 1$ and $\alpha > 0.25$, $1-4\alpha < 0$ and therefore

$$(n+1)^{1-4\alpha} \leq n^{1-4\alpha} \;\Rightarrow\; \left(\frac{4}{n+1}\right)^{4\alpha}\cdot\left(\frac{n+1}{e}\right) \leq \left(\frac{4}{n}\right)^{4\alpha}\cdot\left(\frac{n}{e}\right) \;\Rightarrow\; \xi_{n+1} \leq \xi_n.$$

Summarizing, this proves the following lemma.

LEMMA 7.1. *Define $\varphi_1 := 1$, $\varphi_2 := 1$ and recursively φ_n to be the positive zero of $P_n(x)$ given in (47). Let constants $c, \alpha \in \mathbb{R}$, $\alpha \geq \ln 2$ and $3 \leq n_0 \in \mathbb{N}$ be given with $\varphi_n \leq c\cdot n^\alpha$ for $n < n_0$. If a constant $\beta \in \mathbb{R}$, $1 < \beta < e$ exists such that (59) is satisfied and μ_n, ν_n, ξ_n defined in (54), (56), (57) satisfy*

$$n^{-\alpha} + \mu_n + \nu_n + \xi_n \leq 1 \quad \text{for} \quad n = n_0, \tag{60}$$

then

$$\varphi_n \leq c\cdot n^\alpha \quad \text{for all} \quad n \in \mathbb{N}.$$

Proof. (50) is satisfied for $1 \leq k < n$, and (58) and (7.14) prove the left hand side of (51) for $n = n_0$, and therefore (7.4) for $k = n$. The quantities $n^{-\alpha}$, μ_n, ν_n and ξ_n are decreasing for increasing n. Thus, (7.14) and therefore (7.4) is valid for all $n \geq n_0$. By assumption, $\varphi_n \leq c\cdot n^\alpha$ for $n < n_0$ as well. ∎

For example, for $\beta := 2.697$, $\alpha := 0.7$ and $n_0 := 3000$, one checks by explicit calculation $\varphi_n \leq 2.321\cdot n^\alpha$ for $1 \leq n \leq n_0$. The lower bound (59) for n_0 is less than 183, $\mu_n < 0.992$, $\nu_n < 0.0003$, $\xi_n < 0.0038$, and $n^{-\alpha} < 0.0038$ for $n = n_0$. This proves the following result.

COROLLARY 7.2. *For all $n \geq 1$, $\varphi_n \leq 2.321\cdot n^{0.7}$. The difference $2.321\cdot n^{0.7} - \varphi_n$ is less than 2.8 for $1 \leq n < 20$, and less than 2.0 for $20 \leq n \leq 2000$.*

Summarizing, Corollary 7.2, Proposition 6.3, and Lemma 5.7 prove the following result.

PROPOSITION 7.3. *Let $A, E \in M_n(\mathbb{R})$ with nonsingular A and $E \geq 0$ be given. Then for all $n \geq 1$*

$$\frac{1}{\rho(\,|A^{-1}|\cdot E)} \leq \sigma(A,E) \leq \frac{\gamma(n)}{\rho(\,|A^{-1}|\cdot E)},$$

with

$$n \leq \gamma(n) \leq 2.321\cdot n^{1.7}.$$

The lower bound for $\gamma(n)$ is sharp.

Finally, we will show the asymptotic behaviour of upper bounds for $\gamma(n)$. Let $\alpha := \ln(2+2\eta)$, $\eta > 0$. For any $1 < \beta < e$ and $n \to \infty$,

$$n^{-\alpha} \to 0, \;\; \mu_n \to (1-\beta^{-\alpha})^{-1}\cdot e^{-\alpha}, \quad \nu_n \to 0 \quad \text{and} \quad \xi_n \to 0.$$

For $\ln\beta := (2+\eta)/(2+2\eta)$, a short computation yields

$$(1-\beta^{-\alpha})^{-1}\cdot e^{-\alpha} = \frac{2+\eta}{2+4\eta+2\eta^2} < 1.$$

Hence, for this β and large enough n_0, (59) holds and

$$n^{-\alpha} + \mu_n + \nu_n + \xi_n < 1 \quad \text{for all} \quad n \geq n_0.$$

Therefore, for large enough c with $\varphi_n \leq c\cdot n^\alpha$ for $n < n_0$, Lemma 7.1 implies that $\varphi_n \leq c\cdot n^\alpha$ for all $n \in \mathbb{N}$. Using $\alpha > \ln 2$ proves the following.

PROPOSITION 7.4. *Let $\gamma(n)$ be defined as follows:*

$$\gamma(n) := \inf\{\, \sigma(A,E)\cdot\rho(|A^{-1}|\cdot E) \mid A \in M_n(\mathbb{R}) \quad \text{nonsingular and} \quad 0 \leq E \in M_n(\mathbb{R})\,\}.$$

Then $\gamma(n)$ is finite for all $n \in \mathbb{N}$. Moreover, for any $\varepsilon > 0$ there exists some $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ holds

$$n \leq \gamma(n) \leq n^{1+\ln 2+\varepsilon}. \tag{61}$$

The lower bound in (61) is sharp. †)

In his paper [3], Demmel showed that for the Bauer-Skeel condition number $\kappa(A,E) := \| |A^{-1}|\cdot E\|$ with any p-norm, $1 \leq p \leq \infty$, there holds

$$\frac{1}{\rho(|A^{-1}|\cdot E)} = \frac{1}{\min\limits_D \kappa(AD,\, ED)}\,,$$

where the minimum is taken over all diagonal D. Thus, Proposition 7.3 and Proposition 7.4 prove that the componentwise relative distance to the nearest singular matrix for any weight matrix $E \geq 0$ is not too far from the reciprocal of the smallest condition number achievable by column scaling. The evidence presented in this paper leads us to the following conjecture.

CONJECTURE 7.5. *For all nonsingular $A \in M_n(\mathbb{R})$ and $0 \leq E \in M_n(\mathbb{R})$ holds*

$$\frac{1}{\rho(|A^{-1}|\cdot E)} \leq \sigma(A,E) \leq \frac{n}{\rho(|A^{-1}|\cdot E)}\,. \tag{62}$$

If the conjecture is true, Lemma 5.7 shows that it is sharp.

Acknowledgments. The author wishes to thank Jiri Rohn for many helpful comments.

REFERENCES

[1] F. BAUER, *Optimally scaled matrices*, Numerische Mathematik 5, (1963), pp. 73–87.
[2] L. COLLATZ, *Einschließungssatz für die charakteristischen Zahlen von Matrizen*, Math. Z., 48 (1942), pp. 221–226.
[3] J. DEMMEL, *The Componentwise Distance to the Nearest Singular Matrix*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 10–19.
[4] G. ENGEL AND H. SCHNEIDER, *The Hadamard-Fischer Inequality for a Class of Matrices Defined by Eigenvalue Monotonicity*, Linear and Multilinear Algebra 4, (1976), pp. 155 – 176.
[5] R. HORN AND C. JOHNSON, *Matrix Analysis*, Cambridge University Press, 1985.
[6] M. MARCUS AND H. MINC, *A survey of matrix theory and matrix inequalities*, Dover publications, New York, 1992.
[7] S. POLJAK AND J. ROHN, *Checking Robust Nonsingularity Is NP-Hard*, Math. of Control, Signals, and Systems 6, (1993), pp. 1–9.
[8] J. ROHN, *Nearness of Matrices to Singularity*, KAM Series on Discrete Mathematics and Combinatories, (1988).
[9] ———, *Systems of Linear Interval Equations*, Linear Algebra Appl. 126, (1989), pp. 39–78.
[10] ———, *Interval Matrices: Singularity and Real Eigenvalues*, SIAM J. Matrix Anal. Appl. 14, (1993), pp. 82–91.
[11] S. RUMP, *Estimation of the Sensitivity of Linear and Nonlinear Algebraic Problems*, Linear Algebra and its Applications 153, (1991), pp. 1–34.
[12] ———, *Verification Methods for Dense and Sparse Systems of Equations*, in Topics in Validated Computations — Studies in Computational Mathematics, J. Herzberger, ed., Elsevier, Amsterdam, 1994, pp. 63–136.

†) Note added in proof: In the meantime it has been shown by the author that $n \leq \gamma(n) \leq (3+2\sqrt{2})\cdot n$.

[13] G. Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, 1990.