Numer. Math.
DOI 10.1007/s00211-010-0326-3

Numerische Mathematik

Locating a nearest matrix with an eigenvalue of prespecified algebraic multiplicity

Emre Mengi

Received: 27 August 2008 / Revised: 9 June 2010

Abstract The Wilkinson distance of a matrix A is the two-norm of the smallest perturbation E so that $A + E$ has a multiple eigenvalue. Malyshev derived a singular value optimization characterization for the Wilkinson distance. In this work we generalize the definition of the Wilkinson distance as the two-norm of the smallest perturbation so that the perturbed matrix has an eigenvalue of prespecified algebraic multiplicity. We provide a singular value characterization for this generalized Wilkinson distance. Then we outline a numerical technique to solve the derived singular value optimization problems. In particular the numerical technique is applicable to Malyshev's formula to compute the Wilkinson distance as well as to retrieve a nearest matrix with a multiple eigenvalue.

Mathematics Subject Classification (2000) 15A18 · 90C56 · 15A60 · 65K10

1 Introduction

In the book *The Algebraic Eigenvalue Problem* [27, p. 90–93], J.H. Wilkinson suspected that the distance

$$W(A) = \inf\{\|\Delta A\| : (A + \Delta A) \text{ has a multiple eigenvalue}\}$$

The computing resources for this work were supplied through the National Science Foundation Grants DMS-0715146 and DMS-0821816. This work was also supported in part by the TUBITAK (the scientific and technological research council of Turkey) Grant 109T660. Most of this work was completed when the author was holding a S.E.W. assistant professorship in the department of mathematics at the University of California, San Diego.

E. Mengi (✉)
Department of Mathematics, Koc University, Rumelifeneri Yolu,
34450 Sariyer, Istanbul, Turkey
e-mail: emengi@ku.edu.tr

Published online: 26 August 2010

provided an indication of the sensitivity of the eigenvalues. We will refer $W(A)$ as the *Wilkinson distance* of A as it has been done by several other authors in the literature, e.g. [1]. Since $W(A)$ is attained at a matrix $(A + \Delta A)$ that is the limit of a sequence of defective (nondiagonalizable) matrices, $W(A)$ can alternatively be viewed as the distance to the nearest defective matrix. By standard eigenvalue perturbation theory a matrix that is close to a defective matrix has an eigenvalue with large condition number. Conversely Ruhe [24] and Wilkinson [28] have shown that any matrix with an ill-conditioned eigenvalue is close to a defective matrix. It follows that $W(A)$ is a measure of sensitivity of the worst-conditioned eigenvalue of A.

Inspired by the work of Qiu et al. [18], Malyshev [15] deduced the singular value characterization

$$W(A) = \inf_{\lambda \in \mathbb{C}} \sup_{\gamma \in (0,1]} \sigma_{2n-1} \left(\begin{bmatrix} A - \lambda I & \gamma I \\ 0 & A - \lambda I \end{bmatrix} \right). \tag{1.1}$$

Here and elsewhere in this paper $\sigma_j(\cdot)$ denotes the jth largest singular value of its matrix argument. For instance, above σ_{2n-1} is the second smallest singular value of the $2n \times 2n$ matrix argument. In this work, we generalize the result of Malyshev to the distance to the nearest matrix with a multiple eigenvalue of prespecified algebraic multiplicity. Specifically we provide a singular value characterization (summarized in Theorem 2.8) for the problem

$$W_r(A) = \inf\{\|\Delta A\| : (A + \Delta A) \text{ has an eigenvalue of multiplicity } r \text{ or greater}\}$$

in the next section. This distance is discussed in detail in [29], where Wilkinson argued that $W_r(A)$ can be much greater than $W_2(A)$ for $r \geq 3$. In Sect. 3, we outline how one can estimate $W_r(A)$ to low precision exploiting the singular value characterization. The characterizations provided are in the inf–sup form as in (1.1). In (1.1), the inner maximization problem is unimodal. Similarly in the characterizations derived in this paper for $W_r(A)$ any local maximum of the inner maximization is typically a global maximum, so it is easy to locate a global maximum of the inner problem. However, the solution of the outer minimization problem is not straightforward. Section 4 illustrates the effectiveness of the numerical technique suggested in Sect. 3 on some selected examples. Throughout this work $\|\cdot\|$ denotes the matrix or vector two-norm and multiplicity of an eigenvalue always refers to the algebraic multiplicity, i.e. the multiplicity of the eigenvalue as a root of the characteristic polynomial.

2 A singular value characterization for $W_r(A)$

2.1 Rank definition of $W_r(A)$

For the derivation of a singular value characterization the first step is to give an equivalent rank definition of $W_r(A)$. The following result can be deduced from the Jordan factorization of A [11].

Theorem 2.1 *The matrix A has λ as an eigenvalue with algebraic multiplicity r or greater if and only if* $\operatorname{rank}(A-\lambda I)^r \le n-r$.

Let $W_r(A,\lambda)$ denote the distance from A to the nearest matrix with λ as an eigenvalue of multiplicity r or greater. The rank characterization

$$W_r(A,\lambda) = \inf\{\|\Delta A\| : \operatorname{rank}(A+\Delta A-\lambda I)^r \le n-r\} \tag{2.1}$$

is apparent from Theorem 2.1. Furthermore, the definition of $W_r(A)$ can be restated as

$$W_r(A) = \inf_{\lambda\in\mathbb{C}} W_r(A,\lambda).$$

2.2 Linearization of the rank definition

We will make use of the standard result below. See for instance [6, Theorem 2.5.3, p.72] for a proof.

Theorem 2.2 *Let A be an $n\times n$ matrix. Then*

$$\inf\{\|\Delta A\| : \operatorname{rank}(A+\Delta A)\le n-r\} = \sigma_{n-r+1}(A).$$

The difficulty in applying Theorem 2.2 to (2.1) is that the rank definition involves powers of the perturbation ΔA. Therefore, the next step is to modify the rank definition so that its dependence on ΔA is linear.

Theorem 2.3 *Let*

$$\gamma = [\gamma_{1,2}, \gamma_{1,3}, \ldots, \gamma_{1,r}, \ldots, \gamma_{r-1,r}]^T \in \mathbb{C}^{r(r-1)/2}$$

with nonzero components. Then

$$\operatorname{rank}(A-\lambda I)^r \le n-r \quad\Longleftrightarrow\quad \operatorname{rank}\left(\mathcal{A}(A,\lambda,\gamma)\right) \le nr-r$$

where

$$\mathcal{A}(A,\lambda,\gamma) := \begin{bmatrix} A-\lambda I & \gamma_{1,2}I & & & \gamma_{1,r}I \\ 0 & A-\lambda I & & & \gamma_{2,r}I \\ & & \ddots & & \\ & & & A-\lambda I & \gamma_{r-1,r}I \\ 0 & & & 0 & A-\lambda I \end{bmatrix}. \tag{2.2}$$

Proof We deduce the result by showing that the null-spaces of $(A-\lambda I)^r$ and $\mathcal{A}(A,\lambda,\gamma)$ have the same dimension.

Suppose $\{v^k : k = 1,\ldots,m\}$ is a basis for the null space of $(A-\lambda I)^r$ so that $(A-\lambda I)^r v^k = 0,\ k = 1,\ldots,m$. Define $v_1^k := v^k$ and

$$v_{j+1}^k := (A-\lambda I)v_j^k,\quad j = 1,\ldots,r-1.$$

Then $(A-\lambda I)v_r^k = 0$. Now we specify a vector u^k in the null space of $\mathcal{A}(A,\lambda,\gamma)$ corresponding to each v^k in the null space of $(A-\lambda I)^r$. Define

$$u^k := \begin{bmatrix} \left(u_1^k\right)^T & \left(u_2^k\right)^T & \ldots & \left(u_r^k\right)^T \end{bmatrix}^T$$

where $u_r^k := v_r^k$ and u_j^k for $j < r$ is chosen such that

$$(A-\lambda I)u_j^k = -\sum_{l=j+1}^{r} \gamma_{j,l} u_l^k. \tag{2.3}$$

It can be verified by induction that a vector u_j^k for all $j < r$ satisfying (2.3) in the form

$$u_j^k = (-1)^{r-j}\left(\prod_{l=j}^{r-1}\gamma_{l,l+1}\right)v_j^k + \sum_{l=j+1}^{r}\alpha_{j,l}v_l^k. \tag{2.4}$$

exists for some scalars $\alpha_{j,l},\ l = j+1,\ldots,r$. As the base case for $j = r-1$ we have

$$-\gamma_{r-1,r}u_r^k = -\gamma_{r-1,r}v_r^k = (A-\lambda I)(-\gamma_{r-1,r}v_{r-1}^k) = (A-\lambda I)u_{r-1}^k.$$

Now assume that for $j = p,\ldots,r-1$ Eq. (2.3) is satisfied for some $u_j^k,\ldots,u_{r-1}^k$ in the form (2.4). But then

$$\begin{aligned} -\sum_{l=p}^{r}\gamma_{p-1,l}u_l^k &= -\gamma_{p-1,p}u_p^k - \sum_{l=p+1}^{r}\gamma_{p-1,l}u_l^k \\ &= -\gamma_{p-1,p}\,(-1)^{r-p}\left(\prod_{l=p}^{r-1}\gamma_{l,l+1}\right)v_p^k + \tilde{v}_{p+1:r}^k \\ &= (-1)^{r-p-1}\left(\prod_{l=p-1}^{r-1}\gamma_{l,l+1}\right)v_p^k + \tilde{v}_{p+1:r}^k \end{aligned}$$

$$= (A - \lambda I)\left((-1)^{r-p-1}\left(\prod_{l=p-1}^{r-1}\gamma_{l,l+1}\right)v_{p-1}^k + \tilde{v}_{p:r}^k\right)$$
$$= (A - \lambda I)u_{p-1}^k$$

where $\tilde{v}_{j:r}^k \in \text{span}\{v_l^k : l = j, \ldots, r\}$ for $j = p,\ p+1$. Therefore, Eq. (2.3) is satisfied for $j = p-1$ and for some u_{p-1}^k in the form (2.4). It follows from (2.4) that u_j^k is a linear combination of the vectors in the set $\{v_l^k : l = j, \ldots, r\}$ with a nonzero coefficient for v_j^k.

Next we show that the set $\{u^k : k = 1, \ldots, m\}$ is linearly independent. Suppose $\sum_{k=1}^m \beta_k u^k = 0$ for some $\beta_k,\ k = 1, \ldots, m$. First we prove that $\sum_{k=1}^m \beta_k v_j^k = 0$ for all j by induction. As the base case by construction $\sum_{k=1}^m \beta_k v_r^k = \sum_{k=1}^m \beta_k u_r^k = 0$. Now as the inductive hypothesis let us assume $\sum_{k=1}^m \beta_k v_l^k = 0$ for all l greater than or equal to some j. From (2.4) it follows that

$$0 = \sum_{k=1}^m \beta_k u_{j-1}^k = (-1)^{r-j}\left(\prod_{l=j-1}^{r-1}\gamma_{l,l+1}\right)\left(\sum_{k=1}^m \beta_k v_{j-1}^k\right) + \sum_{l=j}^r\left(\alpha_{j-1,l}\left(\sum_{k=1}^m \beta_k v_l^k\right)\right).$$

The last sum above is zero because of the inductive hypothesis and the coefficient of the second to last sum (the product of $\gamma_{l,l+1}$) is nonzero. We deduce that $\sum_{k=1}^m \beta_k v_{j-1}^k = 0$. In particular $\sum_{k=1}^m \beta_k v_1^k = \sum_{k=1}^m \beta_k v^k = 0$. Since the set $\{v^k : k = 1, \ldots, m\}$ is linearly independent, we have $\beta_k = 0,\ k = 1, \ldots, m$. This shows that the set $\{u^k : k = 1, \ldots, m\}$ is linearly independent; therefore, $\dim(\text{Null}(\mathcal{A}(A, \lambda, \gamma))) \geq m = \dim(\text{Null}(A - \lambda I)^r)$.

To complete the proof we assume that $\dim(\text{Null}(\mathcal{A}(A, \lambda, \gamma))) = m$ and need to deduce that $\dim(\text{Null}(A - \lambda I)^r) \geq m$. We start with a basis $\{u^k : k = 1, \ldots, m\}$ for the null space of $\mathcal{A}(A, \lambda, \gamma)$ and partition each vector u^k in the basis into r vectors of equal size

$$u^k = \begin{bmatrix}\left(u_1^k\right)^T & \left(u_2^k\right)^T & \ldots & \left(u_r^k\right)^T\end{bmatrix}^T.$$

It can be verified by induction that the condition $\mathcal{A}(A, \lambda, \gamma)u^k = 0$ implies $(A-\lambda I)^{r-j+1}u_j^k = 0$ for $j = 1, \ldots, r$. Specifically $(A-\lambda I)^r u_1^k = 0$. Define $v^k := u_1^k$. To see that the set $\{v^k : k = 1, \ldots, m\}$ is linearly independent, suppose $\sum_{k=1}^m \beta_k v^k = 0$ for some scalars $\beta_k,\ k = 1, \ldots, m$. Now

$$\mathcal{A}(A, \lambda, \gamma)\left(\sum_{k=1}^m \beta_k u^k\right) = 0,$$

implying

$$\begin{bmatrix} A-\lambda I & \gamma_{1,2}I & & & \gamma_{1,r}I \\ 0 & A-\lambda I & & & \gamma_{2,r}I \\ & & \ddots & & \\ & & & A-\lambda I & \gamma_{r-1,r}I \\ 0 & & & 0 & A-\lambda I \end{bmatrix} \begin{bmatrix} \sum_{k=1}^m \beta_k v^k = 0 \\ \sum_{k=1}^m \beta_k u_2^k \\ \vdots \\ \sum_{k=1}^m \beta_k u_{r-1}^k \\ \sum_{k=1}^m \beta_k u_r^k \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_{1,2}I & & & \gamma_{1,r}I \\ A-\lambda I & & & \gamma_{2,r}I \\ & \ddots & & \\ & & A-\lambda I & \gamma_{r-1,r}I \\ 0 & & 0 & A-\lambda I \end{bmatrix} \begin{bmatrix} \sum_{k=1}^m \beta_k u_2^k \\ \vdots \\ \sum_{k=1}^m \beta_k u_{r-1}^k \\ \sum_{k=1}^m \beta_k u_r^k \end{bmatrix} = 0.$$

But the matrix at the bottom has full column rank. Therefore,

$$\begin{bmatrix} \sum_{k=1}^m \beta_k u_2^k \\ \vdots \\ \sum_{k=1}^m \beta_k u_{r-1}^k \\ \sum_{k=1}^m \beta_k u_r^k \end{bmatrix} = 0 \quad \Longrightarrow \quad \sum_{k=1}^m \beta_k u^k = 0$$

from which we obtain $\beta_k = 0, \ k = 1, \ldots, m$. We conclude that the set $\{v^k : k = 1, \ldots, m\}$ is linearly independent implying $\dim(\text{Null}(A-\lambda I)^r) \geq m = \dim(\mathcal{A}(A, \lambda, \gamma))$. □

Therefore, for any $\gamma \in \mathbb{C}^{r(r-1)/2}$ with nonzero components a linearized rank definition of $W_r(A, \lambda)$ is given by

$$W_r(A, \lambda) = \inf\{\|\Delta A\| : \text{rank}\,(\mathcal{A}(A + \Delta A, \lambda, \gamma)) \leq nr - r\}. \tag{2.5}$$

An application of Theorem 2.2 yields the lower bound

$$W_r(A, \lambda) \geq \sigma_{nr-r+1}(\mathcal{A}(A, \lambda, \gamma)).$$

Theorem 2.2 provides only a lower bound, because the perturbations in the linearized-rank definition are not arbitrary and have block diagonal structure. The inequality above holds for all γ giving us

$$W_r(A, \lambda) \geq \tau_r(A, \lambda) := \sup_\gamma \sigma_{nr-r+1}(\mathcal{A}(A, \lambda, \gamma)), \tag{2.6}$$

a tighter lower bound. In this maximization problem, it is allowable to include vectors γ with zero components because of the continuity of the singular value σ_{nr-r+1}.

2.3 Construction of an optimal perturbation

In this subsection we construct a perturbation ΔA_* with norm $\tau_r(A,\lambda)$ such that $\operatorname{rank}(A+\Delta A_*-\lambda I)^r \le n-r$ under a mild linear independence assumption concerning the block components of the optimal singular vectors and a multiplicity assumption concerning the optimal singular value. Then the rank definition (2.1) implies that $\tau_r(A,\lambda)$ is an upper bound for $W_r(A,\lambda)$ (as well as a lower bound; see (2.6)) yielding the equivalence of $\tau_r(A,\lambda)$ and $W_r(A,\lambda)$.

When constructing the optimal perturbation without loss of generality we assume $\tau_r(A,\lambda)>0$. It is straightforward to verify that $\tau_r(A,\lambda)=0$ if and only if $W_r(A,\lambda)=0$, since

$$\begin{aligned}\tau_r(A,\lambda)=0 &\iff \forall\gamma\ \ \sigma_{nr-r+1}\left(\mathcal{A}(A,\lambda,\gamma)\right)=0\\ &\iff \text{(By Theorem 2.3)}\ \ \operatorname{rank}(A-\lambda I)^r \le n-r\\ &\iff \text{(By definition (2.1))}\ \ W_r(A,\lambda)=0.\end{aligned}$$

Definition 2.4 We say that the linear independence qualification holds at (λ,γ) if there exists a left singular vector $u=[u_1^T\ \ u_2^T\ \ \dots\ \ u_r^T]^T$ associated with the singular value $\sigma_{nr-r+1}\left(\mathcal{A}(A,\lambda,\gamma)\right)$ such that the set $\{u_1,u_2,\dots,u_r\}$ is linearly independent where $u_1,u_2,\dots,u_r\in\mathbb{C}^n$.

Suppose γ_* is a point where $\tau_r(A,\lambda)$ is attained and

$$\begin{aligned}\underline{u}&=[\underline{u}_1^T\ \ \underline{u}_2^T\ \ \dots\ \ \underline{u}_r^T]^T\\ \underline{v}&=[\underline{v}_1^T\ \ \underline{v}_2^T\ \ \dots\ \ \underline{v}_r^T]^T\end{aligned}$$

are left and right singular vectors associated with $\tau_r(A,\lambda)$ so that

$$\mathcal{A}(A,\lambda,\gamma_*)\,\underline{v}=\tau_r(A,\lambda)\underline{u}\quad\text{and}\quad\underline{u}^*\,\mathcal{A}(A,\lambda,\gamma_*)=\tau_r(A,\lambda)\underline{v}^*. \tag{2.7}$$

Let us also introduce the matrices

$$\underline{U}=[\underline{u}_1\ \ \underline{u}_2\ \ \dots\ \ \underline{u}_r]\quad\text{and}\quad\underline{V}=[\underline{v}_1\ \ \underline{v}_2\ \ \dots\ \ \underline{v}_r]. \tag{2.8}$$

Exploiting the fact that $\tau_r(A,\lambda)$ is the optimal singular value $\sigma_{nr-r+1}\left(\mathcal{A}(A,\lambda,\gamma_*)\right)$, in the next subsection we show that $\underline{u}$ and $\underline{v}$ satisfy

$$\underline{U}^*\underline{U}=\underline{V}^*\underline{V} \tag{2.9}$$

under the assumption that $\tau_r(A,\lambda)>0$ and the multiplicity of the optimum singular value $\tau_r(A,\lambda)$ is one. A geometrical interpretation of (2.9) is that the angle between $\underline{u}_i$ and $\underline{u}_j$ is same as the angle between $\underline{v}_i$ and $\underline{v}_j$ so that the set $\{\underline{u}_1,\underline{u}_2,\dots,\underline{u}_r\}$ can be obtained from the set $\{\underline{v}_1,\underline{v}_2,\dots,\underline{v}_r\}$ by a rotation.

In what follows we will prove that the perturbation

$$\Delta A_*=-\tau_r(A,\lambda)\underline{U}\,\underline{V}^+$$

(i) has norm $\tau_r(A, \lambda)$, and
(ii) satisfies $(A + \Delta A_* - \lambda I)^r \underline{V} = 0$ when the linear independence qualification holds at (λ, γ_*)

where $\underline{V}^+$ denotes the Moore–Penrose pseudoinverse of $\underline{V}$. If the linear independence qualification holds at (λ, γ_*) (that is if $\underline{U}$ has full column rank, which is possible if and only if $\underline{V}$ has full column rank by property (2.9)), an implication of (ii) is that $\text{rank}(A + \Delta A_* - \lambda I)^r \leq n - r$.

Recall that by the definition of the Moore–Penrose pseudoinverse $\underline{V}^+ \underline{V}\, \underline{V}^+ = \underline{V}^+$. The matrix $\underline{U}\, \underline{V}^+ = \underline{U}\, \underline{V}^+ \underline{V}\, \underline{V}^+$ when multiplied onto a vector first orthogonally projects the vector onto Range($\underline{V}$), then performs a change of basis into Range($\underline{U}$). Informally the projection cannot increase the length of the vector and the change of basis does not affect the length because of property (2.9). Therefore, $\|\underline{U}\, \underline{V}^+\| = 1$ as claimed in (i) and $\|\Delta A_*\| = \tau_r(A, \lambda)$. A proof of this result is already given in [18, Lemma 2]. We provide an alternative proof below,[1] that gives a better insight into why the transformation $\underline{U}\, \underline{V}^+ x$ cannot stretch the vector x.

Theorem 2.5 *Suppose that* $\underline{U}, \underline{V} \in \mathbb{C}^{n \times r}$ *satisfy* (2.9) *and that* $\underline{U} \neq 0$. *Then* $\|\underline{U}\, \underline{V}^+\| = 1$.

Proof Given any $x \in \mathbb{R}^n$. Notice that

$$\|\underline{U}\underline{V}^+ x\| = \sqrt{x^* \left(\underline{V}^+\right)^* \underline{U}^* \underline{U}\underline{V}^+ x} = \sqrt{x^* \left(\underline{V}^+\right)^* \underline{V}^* \underline{V}\underline{V}^+ x} = \|\underline{V}\underline{V}^+ x\| \leq \|x\|$$

The last inequality follows due to the fact that $\underline{V}\underline{V}^+$ is the orthogonal projector onto Range($\underline{V}$). If x is chosen in Range($\underline{V}$), then $\underline{V}\underline{V}^+ x = x$ proving that $\|\underline{U}\, \underline{V}^+\| = 1$. □

The proof that $(A + \Delta A_* - \lambda I)^r \underline{V} = 0$ is purely algebraic and follows from the definition of the matrix $\mathcal{A}(A, \lambda, \gamma)$ and Eq. (2.7). Here we assume that the linear independence qualification holds at (λ, γ_*), so $\underline{V}$ has full rank meaning $\underline{V}^+ \underline{V} = I$. Specifically the satisfaction of the linear independence qualification implies $\underline{U}\, \underline{V}^+ \underline{v}_j = \underline{U} e_j = \underline{u}_j$, where e_j denotes the jth column of the identity matrix of size r.

Theorem 2.6 *Suppose that the linear independence qualification holds at* (λ, γ_*) *and property* (2.9) *is satisfied. Then for all* $j = 1, \ldots, r$ *the vector* $\underline{v}_j$ *satisfies*

$$(A + \Delta A_* - \lambda I)^{r-j+1} \underline{v}_j = 0.$$

Proof The proof is by induction. As the base case for $j = r$ from the last block row of Eq. (2.7) we obtain

$$(A + \Delta A_* - \lambda I)\underline{v}_r = (A - \lambda I)\underline{v}_r + \Delta A_* \underline{v}_r = \tau_r(A, \lambda)\underline{u}_r - \tau_r(A, \lambda)\underline{u}_r = 0.$$

[1] Thanks to an anonymous referee for pointing out this simpler version of the proof in the initial manuscript.

Now as the inductive hypothesis assume that for $l = j+1, \ldots, r$

$$(A + \Delta A_* - \lambda I)^{r-l+1} \underline{v}_l = 0.$$

From the jth block row of (2.7) we have

$$\begin{aligned}
(A + \Delta A_* - \lambda I)^{r-j+1} \underline{v}_j &= (A + \Delta A_* - \lambda I)^{r-j} (A + \Delta A_* - \lambda I) \underline{v}_j \\
&= (A + \Delta A_* - \lambda I)^{r-j} \left((A - \lambda I) \underline{v_j} + \Delta A_* \underline{v}_j \right) \\
&= (A + \Delta A_* - \lambda I)^{r-j} \\
&\quad \times \left(\tau_r(A, \lambda) \underline{u}_j - \sum_{l=j+1}^{r} \gamma_{j,l} \underline{v}_l - \tau_r(A, \lambda) \underline{u}_j \right) \\
&= - \sum_{l=j+1}^{r} \gamma_{j,l} (A + \Delta A_* - \lambda I)^{l-j-1} \\
&\quad \times \left((A + \Delta A_* - \lambda I)^{r-l+1} \underline{v}_l \right) \\
&= 0
\end{aligned}$$

as desired. Above we used the inductive hypothesis in the second to last equality. □

The claim (ii) is an immediate corollary of Theorem 2.6.

Corollary 2.7 *Suppose that the assumptions of Theorem 2.6 hold so that* $\operatorname{rank}(\underline{V}) = r$. *Then*

$$\operatorname{rank}(A + \Delta A_* - \lambda I)^r \leq n - r.$$

This completes the derivation of the singular value characterization, which is summarized below.

Theorem 2.8 *Recalling the definition (2.2) of* $\mathcal{A}(A, \lambda, \gamma)$, *suppose*

$$\tau_r(A) := \inf_{\lambda \in \mathbb{C}} \sup_{\gamma} \sigma_{nr-r+1} \left(\mathcal{A}(A, \lambda, \gamma) \right) = \inf_{\lambda \in \mathbb{C}} \tau_r(A, \lambda) \tag{2.10}$$

is attained at (λ_*, γ_*) *where the linear independence qualification holds and the multiplicity of* $\sigma_{nr-r+1} \left(\mathcal{A}(A, \lambda_*, \gamma_*) \right)$ *is one. Then*

(i) $W_r(A) = \tau_r(A)$,
(ii) *a nearest matrix to* A *with an eigenvalue of multiplicity* r *is given by*

$$A_* = A - \tau_r(A) \underline{U}\, \underline{V}^+,$$

(iii) *the matrix* A_* *has* λ_* *as an eigenvalue of multiplicity* r,

where $\underline{U}$, $\underline{V}$ *is as defined in* (2.8) *in terms of the singular vectors associated with* $\mathcal{A}(A, \lambda_*, \gamma_*)$.

Clearly the propositions analogous to (i)–(ii) of Theorem 2.8 also hold relating $W_r(A, \lambda)$ and $\tau_r(A, \lambda)$ for a fixed λ rather than $W_r(A)$ and $\tau_r(A)$. This would again require the assumptions that $\tau_r(A, \lambda)$ is attained at a γ_* where the multiplicity of σ_{nr-r+1} $(\mathcal{A}(A, \lambda, \gamma_*))$ is one and the linear independence qualification is satisfied.

2.4 The derivation of $\underline{U}^*\underline{U} = \underline{V}^*\underline{V}$

To deduce (2.9) we will refer to the following standard result due to Rellich [19–23] regarding the eigenvalues of a Hermitian matrix subject to Hermitian, analytic perturbations. (See also Kato's book [10, Sect. 2.6.2].)

Theorem 2.9 (Rellich [19–23]) *Consider an analytic Hermitian matrix-valued function* $A(t) : \mathbb{R} \to \mathbb{C}^{n\times n}$. *The function* $A(t)$ *has the decomposition*

$$A(t) = Q(t)\Lambda(t)Q(t)^*$$

where $Q(t) : \mathbb{R} \to \mathbb{C}^{n\times n}$ *is a unitary analytic matrix-valued function and* $\Lambda(t) : \mathbb{R} \to \mathbb{C}^{n\times n}$ *is a diagonal analytic matrix-valued function.*

In [2, Theorem 1] the theorem above is applied to the matrix

$$\begin{bmatrix} 0 & A(t) \\ A(t)^* & 0 \end{bmatrix}$$

to show the existence of an analytic signed singular value decomposition for real analytic $A(t) = \tilde{U}(t)\tilde{\Sigma}(t)\tilde{V}(t)^T$. Here $\tilde{\Sigma}(t)$ is diagonal, but it can have negative entries along the diagonal. Its extension to complex matrix-valued functions is straightforward.

Theorem 2.10 *Let* $A(t) : \mathbb{R} \to \mathbb{C}^{n\times n}$ *be an analytic matrix-valued function. There exists a decomposition*

$$A(t) = \widetilde{U}(t)\widetilde{\Sigma}(t)\widetilde{V}(t)^*$$

where $\tilde{U}(t) : \mathbb{R} \to \mathbb{C}^{n\times n}$, $\tilde{V}(t) : \mathbb{R} \to \mathbb{C}^{n\times n}$ *are unitary and analytic,* $\widetilde{\Sigma}(t)$ *is diagonal and analytic at all* t.

Suppose $\tilde{u}_\ell(t)$, $\tilde{v}_\ell(t)$ are the ℓth columns of $\widetilde{U}(t)$ and $\widetilde{V}(t)$, respectively and $\tilde{\sigma}_\ell(t)$ is the signed singular value at the ℓth diagonal entry of $\widetilde{\Sigma}(t)$. Using the product rule and $A(t)\tilde{v}_\ell(t) = \tilde{\sigma}_\ell(t)\tilde{u}_\ell(t)$, it is straightforward to deduce

$$\frac{d\tilde{\sigma}_\ell(t)}{dt} = \text{Real}\left(\tilde{u}_\ell(t)^* \frac{dA(t)}{dt}\tilde{v}_\ell(t)\right). \tag{2.11}$$

See the remarks following Lemma 4 in [15] for details.

The quantity $\tau_r(A,\lambda)$ is the maximum value of $g(\gamma) = \sigma_{nr-r+1}(\mathcal{A}(A,\lambda,\gamma))$. In what follows we view $g(\gamma)$ as a mapping from $\mathbb{R}^{r(r-1)}$ to $\mathbb{R}$ by replacing each component $\gamma_{j,k}$ of γ with its real part $\Re\gamma_{j,k}$ and imaginary part $\Im\gamma_{j,k}$. Furthermore we assume that $\sigma_{nr-r+1}(\mathcal{A}(A,\lambda,\gamma_*))$ is strictly greater than zero and has multiplicity one. In this case $\sigma_{nr-r+1}(\mathcal{A}(A,\lambda,\gamma)) = \tilde{\sigma}_\ell(\mathcal{A}(A,\lambda,\gamma))$ for some ℓ and for all γ in a small neighborhood of γ_*. It follows from (2.11) that for $j < k$

$$\begin{aligned}\text{Real}\left(\underline{u}^*\frac{\partial\mathcal{A}(A,\lambda,\gamma_*)}{\partial\,\Re\gamma_{j,k}}\underline{v}\right) &= \text{Real}(\underline{u}_j^*\underline{v}_k) = 0,\\ \text{Real}\left(\underline{u}^*\frac{\partial\mathcal{A}(A,\lambda,\gamma_*)}{\partial\,\Im\gamma_{j,k}}\underline{v}\right) &= \text{Real}(i\underline{u}_j^*\underline{v}_k) = 0.\end{aligned}$$

Therefore, we deduce the following result.

Theorem 2.11 *Suppose $\tau_r(A,\lambda) = \sigma_{nr-r+1}(\mathcal{A}(A,\lambda,\gamma_*)) > 0$ and the multiplicity of $\sigma_{nr-r+1}(\mathcal{A}(A,\lambda,\gamma_*))$ is one. The equality $\underline{u}_j^*\underline{v}_k = 0$ holds for all $j = 1,\ldots,r$ and $j < k \le r$.*

Next we prove the property $\underline{U}^*\underline{U} = \underline{V}^*\underline{V}$ by exploiting Theorem 2.11.

Theorem 2.12 *Under the assumptions of Theorem 2.11 the property $\underline{u}_j^*\underline{u}_k = \underline{v}_j^*\underline{v}_k$ holds for all $j = 1,\ldots,r$ and $k = 1,\ldots,r$.*

Proof Notice that it suffices to show that $\underline{u}_j^*\underline{u}_k = \underline{v}_j^*\underline{v}_k$ only for $j \le k$. From the definition of $\mathcal{A}(A,\lambda,\gamma)$ it follows that

$$(A-\lambda I)\underline{v}_k + \gamma_{k,k+1}\underline{v}_{k+1} + \cdots + \gamma_{k,r}\underline{v}_r = \tau_r(A,\lambda)\underline{u}_k$$

and

$$\underline{u}_j^*(A-\lambda I) + \gamma_{j-1,j}\underline{u}_{j-1}^* + \cdots + \gamma_{1,j}\underline{u}_1^* = \tau_r(A,\lambda)\underline{v}_j^*.$$

By multiplying the first equation by $\underline{u}_j^*$ from the left, then substituting for $\underline{u}_j^*(A-\lambda I)$ in the first equation using the second equation and noting $\underline{u}_l^*\underline{v}_m = 0$ for all $l < m$ (by Theorem 2.11) we obtain

$$\begin{aligned}\underline{u}_j^*(A-\lambda I)\underline{v}_k + \gamma_{k,k+1}\underline{u}_j^*\underline{v}_{k+1} + \cdots + \gamma_{k,r}\underline{u}_j^*\underline{v}_r &= \tau_r(A,\lambda)\underline{u}_j^*\underline{u}_k\\ (\tau_r(A,\lambda)\underline{v}_j^* - \gamma_{j-1,j}\underline{u}_{j-1}^* - \cdots - \gamma_{1,j}\underline{u}_1^*)\underline{v}_k &= \tau_r(A,\lambda)\underline{u}_j^*\underline{u}_k\\ \underline{v}_j^*\underline{v}_k &= \underline{u}_j^*\underline{u}_k\end{aligned}$$

as desired. □

It should be noted that for $r = 2$ in [15] the result above is proved without the assumption that the multiplicity of the optimal singular value is one. In particular the

existence of a pair of left and right singular vectors, $u_\ell(t_*)$, $v_\ell(t_*)$ associated with the unsigned singular value $\sigma_\ell(t_*)$ at a local extremum t_* satisfying

$$\text{Real}\left(u_\ell(t_*)^* \frac{dA(t_*)}{dt} v_\ell(t_*)\right) = 0.$$

is established (even when the multiplicity of $\sigma_\ell(t_*)$ is two or greater). Clearly this result extends in the sense that there exist left and right singular vectors $u_{j,k}(\gamma_*)$, $v_{j,k}(\gamma_*)$ such that

$$\text{Real}\left(u_{j,k}(\gamma_*)^* \frac{\partial \mathcal{A}(A, \lambda, \gamma_*)}{\partial\, \Re\gamma_{j,k}} v_{j,k}(\gamma_*)\right) = 0$$

and similarly with respect to $\Im\gamma_{j,k}$. Unfortunately, the singular vectors $u_{j,k}(\gamma_*)$, $v_{j,k}(\gamma_*)$ in the above equality can differ for different j and k. In the optimal perturbation $\Delta A = -\tau_r(A, \lambda)\underline{U}\underline{V}^+$ it is crucial that $\underline{U}$ and $\underline{V}$ are defined in terms of the block components of a unique consistent pair of singular vectors $\underline{u}$, $\underline{v}$. The proofs of Theorems 2.6 and 2.12 are based on this assumption. For $r = 2$ the multiplicity assumption is not needed, since there is only one ordinary derivative that must be equal to zero at the optimal point and that is necessarily satisfied by a unique consistent pair of left and right singular vectors. The multiplicity assumption concerning the optimal singular values turns out to be essential for $r > 2$. Indeed, in practice for $r > 2$ what we observed is, whenever this multiplicity assumption is violated, $\tau_r(A)$ gives only a lower bound for $W_r(A)$, while $\tau_r(A) = W_r(A)$ as long as the multiplicity assumption holds. The examples in Sect. 4 further illustrate the necessity of the multiplicity assumption. Note that we deduced the property $\underline{U}^*\underline{U} = \underline{V}^*\underline{V}$ only under the multiplicity assumption. This property is satisfied regardless of whether the linear independence qualification holds at (λ, γ_*).

2.5 Comparison with Malyshev's formula

We conclude this section by summarizing the differences between Malyshev's formula (1.1) for $r = 2$ and its generalization provided in this section for any $r \leq n$.

- The inner maximization in Malyshev's formula is over a real variable in the interval $(0, 1]$. In the general formula the inner maximization is over complex variables. The example in Sect. 4.4 illustrates that for $r = 3$ the maximization of the inner problem over real variables can yield values significantly smaller than $W_r(A)$.
- For $r = 2$ the linear independence qualification always holds at a stationary point of the inner maximization problem. In general for $r > 2$ the linear independence qualification usually holds at a stationary point of the inner problem, but not always. An example for which the linear independence qualification is violated is described in Sect. 4.2.
- For $r = 2$ the multiplicity assumption is not needed, *i.e.* even if the multiplicity of the singular value is greater than one at a local maximizer of the inner problem where the singular value takes the value β, the equality $W_2(A, \lambda) = \beta$ holds. In

the general formula the multiplicity assumption is essential. In practice we observe that $\beta < W_r(A, \lambda)$ when the multiplicity assumption is violated at a local maximizer where the singular value takes the value β. See Sect. 4.2 for examples for which the multiplicity assumption is violated.

- In Malyshev's formula the inner maximization is unimodal over the interval (0, 1]. In the general formula any stationary point of the inner problem where the multiplicity assumption and linear independence qualification hold is a global maximizer. However, the inner problem can have stationary points where one or both of these two assumptions are violated. These stationary points are not necessarily global maximizers. Therefore, in general the inner problem is not unimodal. But numerically it is easy to locate a stationary point where both of the assumptions hold. These issues are discussed in Sect. 3.2.

3 Computation

We exploit the singular value characterization (2.10) for the computation of $W_r(A)$. Let us define the function

$$h(\lambda, \gamma) = \sigma_{nr-r+1}\left(\mathcal{A}(A, \lambda, \gamma)\right).$$

Then

$$\tau_r(A, \lambda) = \sup_{\gamma \in \mathbb{C}^{r(r-1)/2}} h(\lambda, \gamma) \quad \text{and} \quad \tau_r(A) = \inf_{\lambda \in \mathbb{C}} \tau_r(A, \lambda).$$

The method we suggest here requires the evaluation of $h(\lambda, \gamma)$ at various λ, γ. The inner maximization, *i.e.* evaluation of $\tau_r(A, \lambda)$ at a given λ, can be performed by a standard smooth optimization technique, as it turns out any local maximizer where the multiplicity and linear independence assumptions are satisfied is a global maximizer. A key observation is that $\tau_r(A, \lambda)$ is Lipschitz with respect to λ. There are derivative-free techniques for the optimization of Lipschitz functions. Most of the current such techniques stem from ideas due to Piyavskii [17] and Shubert [26]. In the subsequent three subsections we address the details of how $h(\lambda, \gamma)$ can be evaluated efficiently, and how the inner maximization and outer minimization problems can be solved. The method is meant only for low precision, up to 5-6 digit accuracy.

3.1 Evaluation of $h(\lambda, \gamma)$

An initial Schur factorization $A = QTQ^*$ reduces the computational work for $h(\lambda, \gamma)$ drastically. Clearly

$$\mathcal{A}(A, \lambda, \gamma) = \mathcal{Q}\,\mathcal{A}(T, \lambda, \gamma)\,\mathcal{Q}^*$$

where $\mathcal{Q} = I \otimes Q$ with $\otimes$ denoting the Kronecker product. Therefore, the singular values of $\mathcal{A}(A, \lambda, \gamma)$ and $\mathcal{A}(T, \lambda, \gamma)$ are the same. We need to compute the rth

smallest singular value of the upper triangular matrix $\mathcal{A}(T, \lambda, \gamma)$ which has size rn. This can be achieved typically at a cost of $O(r^2n^2)$ by means of an iterative method such as shift-invert Lanczos, since given a vector $v \in \mathbb{C}^{2rn}$ the linear system

$$\begin{bmatrix} 0 & \mathcal{A}(T, \lambda, \gamma) \\ \mathcal{A}(T, \lambda, \gamma)^* & 0 \end{bmatrix} u = v$$

can be solved for u by decoupling it into two triangular systems of size rn.

The overall cost of performing p function evaluations is $O(n^3 + pr^2n^2)$. Since the method requires function evaluations at various λ and γ (typically $p \approx 500 - 1{,}000$ for four digit accuracy), the computational burden of the initial Schur factorization is negligible unless n is very large.

3.2 Solution of the inner maximization problem

For a fixed λ suppose γ_* is a stationary point of $h(\lambda, \gamma)$ with respect to γ and the linear independence qualification holds at (λ, γ_*), and suppose also that the multiplicity of the singular value $\sigma_{nr-r+1}(\mathcal{A}(A, \lambda, \gamma_*))$ is one. Then the derivation in the previous section applies to show that $\beta = h(\lambda, \gamma_*) = W_r(A, \lambda)$, so β is indeed the globally maximum value.

In theory, $h(\lambda, \gamma)$ is not differentiable at a γ where the multiplicity of the singular value is greater than one. However, in practice excluding non-generic cases it is very unlikely that a computational technique will iterate over such a point where there are ties in the singular values. Therefore, a standard smooth optimization technique can typically be used to maximize $h(\lambda, \gamma)$ with respect to γ and retrieve $W_r(A, \lambda)$. In practice we observe that the multiplicity assumption and linear independence qualification are usually satisfied at a local maximizer. However, strictly speaking these inner maximization problems are not unimodal, as there can potentially exist stationary points where one or both of these two assumptions are violated. Such stationary points are not necessarily global maximizers. When the smooth optimization technique converges to such a stationary point (which happens very rarely in practice), it can be restarted with a different initial point. For $r = 2$ the maximization is over only one real variable, so the secant method with an appropriate line search can be used. For $r > 2$ the inner maximization is over $(r-1)r/2$ complex variables or (by representing each complex variable with its real and complex parts) over $r(r-1)$ real variables, and the BFGS algorithm is the ideal choice. In our experiments we usually observe fast convergence. In 3-6 iterations BFGS usually computes $W_r(A, \lambda)$ with 5 digit accuracy.

3.3 Solution of the outer minimization problem

A well-known result due to Weyl concerning the eigenvalues of a Hermitian matrix is that they are not very sensitive to Hermitian perturbations. See [8, Theorem 4.3.1, p.181] for a proof of Weyl's theorem below.

Theorem 3.1 (Weyl) *Let A and E be Hermitian $n \times n$ matrices and $\lambda_j(\cdot)$ denote the jth largest eigenvalue of its matrix argument. Then*

$$|\lambda_j(A) - \lambda_j(A+E)| \leq \|E\|.$$

An immediate corollary of Weyl's Theorem is that $h(\lambda, \gamma)$ is Lipschitz with respect to λ with Lipschitz constant one, *i.e.*

$$|h(\lambda + \delta\lambda, \gamma) - h(\lambda, \gamma)| \leq \|(\delta\lambda)I\| \leq |\delta\lambda|. \tag{3.1}$$

Suppose $\tau_r(A, \lambda)$ is attained at γ_*. Then from (3.1) it follows that

$$\begin{aligned}\tau_r(A, \lambda) &= h(\lambda, \gamma_*) \leq h(\lambda + \delta\lambda, \gamma_*) + |\delta\lambda| \leq \sup_{\gamma} h(\lambda + \delta\lambda, \gamma) + |\delta\lambda| \\ &= \tau_r(A, \lambda + \delta\lambda) + |\delta\lambda|.\end{aligned}$$

Therefore, $\tau_r(A, \lambda) - \tau_r(A, \lambda + \delta\lambda) \leq |\delta\lambda|$. Indeed by switching the roles of $\tau_r(A, \lambda)$ and $\tau_r(A, \lambda + \delta\lambda)$ the inequality $\tau_r(A, \lambda + \delta\lambda) - \tau_r(A, \lambda) \leq |\delta\lambda|$ can also be deduced. This yields the following result.

Theorem 3.2 *The function $\tau_r(A, \lambda)$ is Lipschitz with respect to λ with Lipschitz constant one, i.e.*

$$|\tau_r(A, \lambda + \delta\lambda) - \tau_r(A, \lambda)| \leq |\delta\lambda|.$$

At first thought a grid search to minimize $\tau_r(A, \lambda)$ with respect to λ seems applicable. The idea is to evaluate $\tau_r(A, \lambda)$ at all of the nodes of a 2D grid with a discretization length $\epsilon\sqrt{2}$. Then the distance from λ_* (where $\tau_r(A, \lambda)$ takes the minimum value) to one of the grid points is at most ϵ. Consequently, by Theorem 3.2 the minimum value attained by $\tau_r(A, \lambda)$ over the grid points cannot differ from the exact value $\tau_r(A)$ by more than ϵ. The difficulty with this brute-force approach is it usually requires too many function evaluations even for 1-2 digit accuracy.

Two simpler singular value minimization problems of similar Lipschitz nature are

$$\phi(A) = \inf_{\lambda \in \mathbb{R}} \sigma_n(A - \lambda i I) \quad \text{and} \quad \varphi(A, B) = \inf_{\lambda \in \mathbb{C}} \sigma_n([A - \lambda I \;\; B]) \tag{3.2}$$

where $A \in \mathbb{C}^{n \times n}$ and $B \in \mathbb{C}^{n \times m}$ with $m \leq n$. The quantity $\phi(A)$ is known as the distance to instability and is equal to the 2-norm of the smallest perturbation to A so that the perturbed matrix has an eigenvalue on the imaginary axis [14]. On the other hand, the quantity $\varphi(A, B)$ is called the distance to uncontrollability and gives the norm of the smallest perturbation $\|[\Delta A \;\; \Delta B]\|$ so that the dynamical system $\dot{x}(t) = (A + \Delta A)x(t) + (B + \Delta B)u(t)$ is uncontrollable [16]. Efficient grid-based optimization techniques for $\phi(A)$ and $\varphi(A, B)$ are suggested in [3] and [7], respectively. Unfortunately, the techniques of [3] and [7] do not extend to $\tau_r(A)$ in a straightforward fashion.

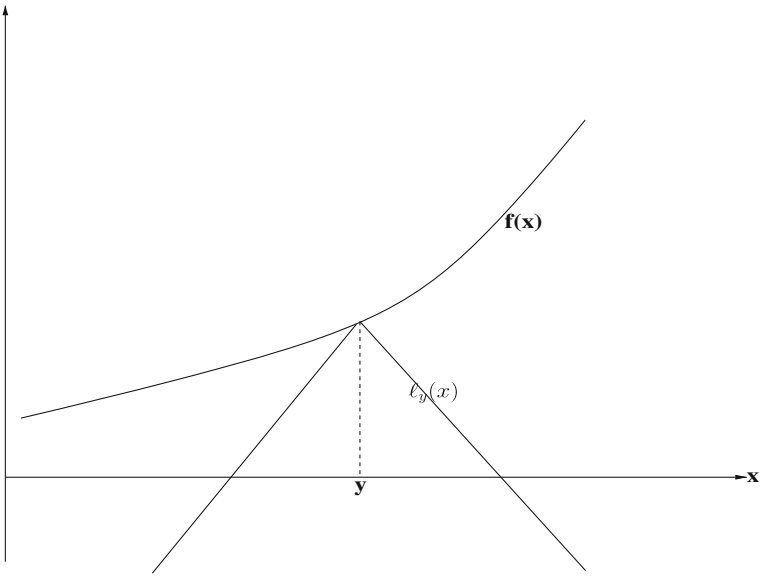

Fig. 1 The piecewise line $\ell_y(x)$ lies below $f(x)$

Piyavskii [17] and Shubert [26] suggested an algorithm for the optimization of Lipschitz functions based on forming a piecewise linear approximation that lies underneath the function. Here we describe the Piyavskii-Shubert algorithm briefly for a univariate function $f(x)$ with Lipschitz constant κ. The algorithm requires function evaluations only and does not require derivatives. The basic observation is that given a point $\bar{x}$, the piecewise linear function

$$\ell_{\bar{x}}(x) = f(\bar{x}) - \kappa|x - \bar{x}|$$

lies below f (see Fig. 1). The algorithm keeps track of the points $\mathcal{S} = \{x_1, x_2, \ldots, x_k\}$ where f is already evaluated. The function $L(x)$

$$L(x) = \max_{x_j \in \mathcal{S}} \ell_{x_j}(x)$$

also lies below $f(x)$. Suppose a global minimizer of f is known to be contained in the interval $[x_1, x_2]$. Lower and upper bounds for the minimum of $f(x)$ are given by

$$l := \min_{x \in [x_1, x_2]} L(x) \leq \inf f(x) \quad \text{and} \quad u := \min_{x_j \in \mathcal{S}} f(x_j) \geq \inf f(x).$$

At each iteration the algorithm

1. finds the minimizer x_* of $L(x)$ in $[x_1, x_2]$,
2. adds this point to the set $\mathcal{S}$,
3. evaluates $f(x)$ at x_*,

4. updates the line $L(x)$,
5. refines the lower and upper bounds l and u.

The algorithm starts with $\mathcal{S} = \{x_1, x_2\}$ and keeps iterating until the lower and upper bounds differ by less than a prespecified tolerance. The first two iterations of the algorithm on a sample function are illustrated in Fig. 2. For a multivariate function the piecewise lines become the hypercones $\ell_{x_j}(x) = f(x_j) - \kappa\|x - x_j\|$. Otherwise the algorithm is similar. (Finding the globally minimum value of a piecewise hypercone turns out to be harder, however there are good lower bounds for the estimation of this minimum value [25].)

We view $\tau_r(A, \lambda)$ as a Lipschitz function over two variables (the real and complex parts of λ) with Lipschitz constant $\kappa = 1$ and apply a modified version of the Piyavskii-Shubert algorithm to minimize $\tau_r(A, \lambda)$ with respect to λ. The reduction in the number of function evaluations (of $\tau_r(A, \lambda)$) is remarkable as compared to a brute-force grid search. While a grid-search over the intervals $\Re\lambda, \Im\lambda \in [-20, 20]$ (We assume that the multiple eigenvalue of a nearest matrix has real and imaginary parts smaller than 20 in absolute value; another choice for these intervals could be $[-c\|A\|, c\|A\|]$ for some positive constant c, that is the interval could be chosen in a relative sense.) would require about 2×10^{10} function evaluations for four digit accuracy, the modified Piyavskii–Shubert algorithm performs typically 100–250 evaluations for the same accuracy.

3.4 Software

The software to estimate $\tau_r(A)$ to four digit precision is implemented in Fortran and a mex file for MacOS running on a PowerPC processor is freely available on the author's webpage.[2]

For the evaluation of the singular value function $h(\lambda, \gamma)$ we use ARPACK [12], which is an implementation of the implicitly restarted Arnoldi method. For the solution of the inner maximization problem the limited memory BFGS code written by Nocedal (specifically an implementation of the algorithm described in [13]) is incorporated. The outer Lipschitz minimization problem is solved by an implementation [5] of the DIRECT algorithm [9]. The DIRECT algorithm is derived from the Piyavskii-Shubert algorithm. It additionally attempts to estimate the Lipschitz constant. Even though in our case the Lipschitz constant is known to be one in advance, locally around the minimizers the slope of the function $\tau_r(A, \lambda)$ is smaller. It is possible that the DIRECT algorithm reduces the number of function evaluations around the minimizer, though we have not tested this systematically.

We should note that in the current implementation the derivatives of $h(\lambda, \gamma)$ with respect to γ (which is required by BFGS to compute $\tau_r(A, \lambda)$) are estimated by finite differences. Analytic formulas for these derivatives in terms of the associated singular vectors (see Sect. 2.4) are available. However, we were not able to obtain these singular vectors reliably. The current implementation would be more efficient if we were able to compute the singular vectors reliably. This would reduce the number of function evaluations when $r > 2$.

[2] http://home.ku.edu.tr/~emengi.

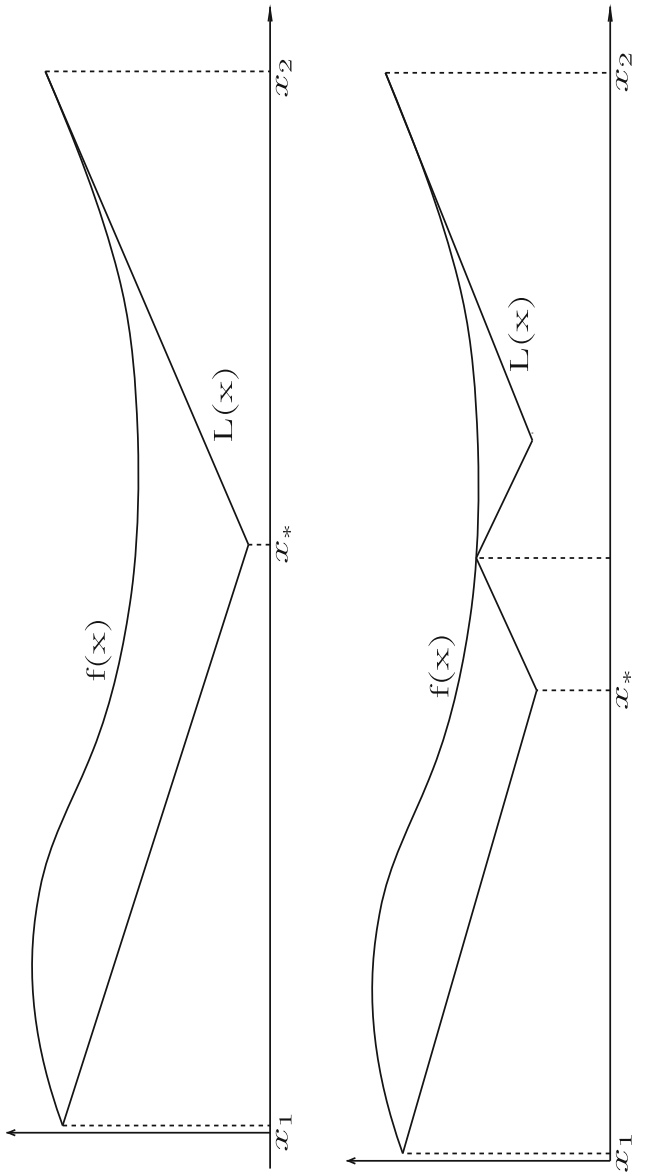

Fig. 2 The first two iterations of the Piyavskii–Shubert algorithm applied to a sample function

4 Numerical experiments

All of the numerical experiments are performed in Matlab 7.4 running on an Apple Power Mac G5 with a 2.5 GHz Quad PowerPC processor (the software is not particularly implemented in a parallel fashion, however Matlab exploits multiple cores to a limited degree), 8 GB memory and Mac OS 10.4.11 operating system. We haven't experimented with the problems with $r > 3$ (even though the software is implemented to work for arbitrary r). This is mainly because the computations would take an excessive amount of time. The increase in the running time is more dramatic with respect to r as compared to n. For $r = 2, 3, 4$, the numbers of real variables of the inner optimization problem are 1, 6, 12, respectively. Note that this should affect only the inner optimization; the outer optimization is over the complex plane regardless of r. One may expect that BFGS would require more iterations to compute $\tau_r(A, \lambda)$ as r increases. In practice we observed a very slight increase in the number of BFGS iterations when r is increased from 2 to 3. A bigger factor affecting the running time is the derivatives estimated by finite differences, which could be avoided if we were able to obtain the singular vectors reliably and use the analytic formulas for the derivatives (see the discussion in Sect. 3.4). But still the slow-down we have experienced is too large to be solely caused by these extra function evaluations for finite differences and the increase in the size of singular value problems. It seems that the most important factor contributing to the slow-down is that retrieving the rth smallest singular value becomes significantly harder as r increases and ARPACK performs considerably more iterations.

4.1 Wilkinson distance and pseudospectra

The ϵ-pseudospectrum of a matrix A is the bounded set consisting of the eigenvalues of all matrices within an ϵ neighborhood of A, i.e.

$$\Lambda_\epsilon(A) = \{\lambda \in \mathbb{C} : \exists E \text{ with } \|E\| < \epsilon \quad \text{s.t. } \det(A + E - \lambda I) = 0\}.$$

If A has n distinct eigenvalues, then for small ϵ the set $\Lambda_\epsilon(A)$ is comprised of n disjoint components each of which contains one of the eigenvalues. It has long been conjectured [4] and recently proved [1] that the smallest ϵ such that two components of $\Lambda_\epsilon(A)$ coalesce is $W_2(A)$. Consider the Hessenberg matrix

$$H = \begin{bmatrix} 3 & -2 & 1 & 4 \\ -1 & -3 & 1 & 1 \\ 0 & -4 & 2 & 1 \\ 0 & 0 & 5 & 1 \end{bmatrix}$$

Our implementation of the technique described in Sect. 3 returns $W_2(H) = 0.5556$. In Fig. 3, the inner curve is the boundary of the ϵ-pseudospectrum for $\epsilon = W_2(H)$. Two components coalesce as expected at $\lambda_* = 1.5181$ (marked with an asterisk in the figure) where the function $\tau_2(H, \lambda)$ is minimized. The point $\lambda_* = 1.5181$ is the double

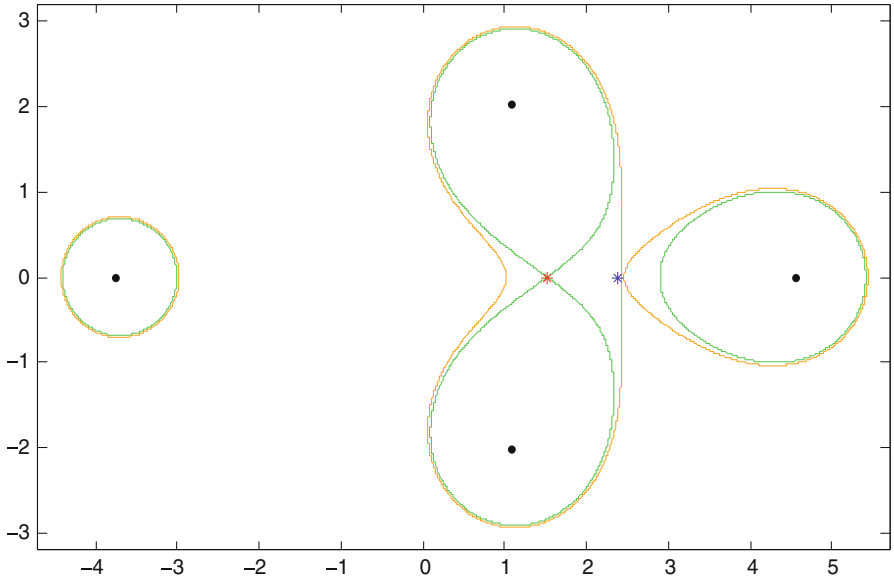

Fig. 3 The *inner* and *outer curves* are the boundaries of the ϵ-pseudospectrum on the complex plane for $\epsilon = W_2(H) = 0.5556$ and $\epsilon = W_3(H) = 0.5731$, respectively. *Black dots* represent the eigenvalues of H

eigenvalue of the nearest matrix (entries of the matrix are rounded to four decimal digits)

$$H_* = \begin{bmatrix} 3.0082 & -2.0306 & 1.0215 & 3.9132 \\ -0.5849 & -3.0152 & 0.9122 & 1.0544 \\ -0.3014 & -3.8648 & 1.9683 & 1.3209 \\ 0.1332 & 0.0947 & 4.8953 & 1.3066 \end{bmatrix}.$$

It seems plausible to conjecture that for $\epsilon = W_3(A)$ a component of the ϵ-pseudospectrum will coalesce with another component which has been formed by two components which already coalesced for smaller ϵ values and contains two eigenvalues. This indeed seems to be the case for the Hessenberg matrix H above. The outer curve in Fig. 3 corresponds to the boundary of the ϵ-pseudospectrum for $\epsilon = W_3(H) = 0.5731$ and the point of coalescence $\lambda_* = 2.3777 - 0.0183i$ (marked with an asterisk) minimizes $\tau_3(H, \lambda)$. (Note that for a real matrix A we have $\tau_r(A, \lambda) = \tau_r(A, \bar{\lambda})$. This indicates that if $\lambda_* = 2.3777 - 0.0183i$ is a global minimizer, so is $\bar{\lambda}_* = 2.3777 + 0.0183i$.) It is likely that the imaginary part of λ_* is due to impreciseness of the estimate and the exact minimizer is purely a real number. The fact that singular values are insensitive to perturbations means that even if $\tau_r(A)$ is accurate to 4 decimal digits, λ_* minimizing $\tau_r(A, \lambda)$ is potentially computed less accurately.

In general the smallest ϵ so that r components of $\Lambda_\epsilon(A)$ coalesce and become one component provides only a lower bound for $W_r(A)$. For instance our computational experience indicates that a component of $\Lambda_\epsilon(A)$ does not always coalesce with another component containing two eigenvalues for $\epsilon = W_3(A)$. Usually this

coalescence occurs for smaller $\epsilon < W_3(A)$. Suppose λ_* is the point of coalescence mentioned and $\lambda_1, \lambda_2, \lambda_3$ are the eigenvalues of A contained in this coalesced component. Let us represent the variation in λ_1, λ_2 and λ_3 with respect to the perturbation E with the functions $\lambda_1(E)$, $\lambda_2(E)$ and $\lambda_3(E)$. The coalescence that occurs for smaller ϵ values than expected is due to the existence of three distinct matrices E_1, E_2 and E_3 each of norm ϵ such that $\lambda_1(E_1) = \lambda_2(E_2) = \lambda_3(E_3) = \lambda_*$. This does not necessarily imply the existence of E of norm ϵ such that $\lambda_1(E) = \lambda_2(E) = \lambda_3(E) = \lambda_*$. (Note that there exists such a perturbation for a point of coalescence λ_* of $\Lambda_\epsilon(A)$ when only two eigenvalues are involved. Specifically, denoting the variation in the eigenvalues λ_1, λ_2 of A that coalesce at λ_* by $\lambda_1(\tilde{E})$, $\lambda_2(\tilde{E})$, it was shown by Alam and Bora [1] that there exists a perturbation E with norm ϵ such that $\lambda_1(E) = \lambda_2(E) = \lambda_*$. This does not seem true when three eigenvalues are involved in the coalescence.) The matrix

$$H_I = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 2 & -2 & -2 \\ 1 & 2 & 3 & -3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

is the inverse of a Hessian matrix generated by the command `gallery('invhess',4)` in Matlab. The computed values are $W_2(H_I) = 0.0328$ and $W_3(H_I) = 1.3972$, respectively. In Fig. 4 on the left, the innermost and outermost curves are the boundaries of the ϵ-pseudospectrum for $\epsilon = W_2(H_I)$ and $\epsilon = W_3(H_I)$, respectively. Again two components of the ϵ-pseudospectrum coalesce for $\epsilon = W_2(H_I)$ at $\lambda_* = 1.97531$ minimizing $\tau_2(H_I, \lambda)$. On the other hand, all of the components of the ϵ-pseudospectrum have already coalesced for $\epsilon < W_3(H_I)$.

To gain insight into what the distance $W_3(A)$ and the complex point λ_* where it is attained mean in terms of pseudospectra, let us generalize the definition of the ϵ-pseudospectrum in the following manner

$$\Lambda_{\epsilon,r}(A) = \{\lambda \in \mathbb{C} : \exists E \text{ with } \|E\| < \epsilon \quad \text{s.t. } \lambda \text{ is an eigenvalue of multiplicity } r \text{ or greater of } (A+E)\}.$$

Specifically for $r = 2$ this definition simplifies to

$$\begin{aligned} \Lambda_{\epsilon,2}(A) &= \{\lambda \in \mathbb{C} : \tau_2(A, \lambda) < \epsilon\} \\ &= \left\{ \lambda \in \mathbb{C} : \sup_{\gamma \in (0,1]} \sigma_{2n-1} \left(\begin{bmatrix} A - \lambda I & \gamma I \\ 0 & A - \lambda I \end{bmatrix} \right) < \epsilon \right\}. \end{aligned}$$

Notice that when $r \geq 2$ the set $\Lambda_{\epsilon,r}(A)$ is empty for small ϵ unless A has a multiple eigenvalue (unlike $\Lambda_\epsilon(A)$ which is non-empty for all $\epsilon > 0$). Furthermore when $r \geq 2$ a component of $\Lambda_{\epsilon,r}(A)$ does not necessarily contain an eigenvalue. (For instance see $\Lambda_{\epsilon,2}(A)$ for the smoke matrix on the right-hand side of Fig. 4. None of the six components whose boundaries are plotted by dotted curves contain an eigenvalue.)

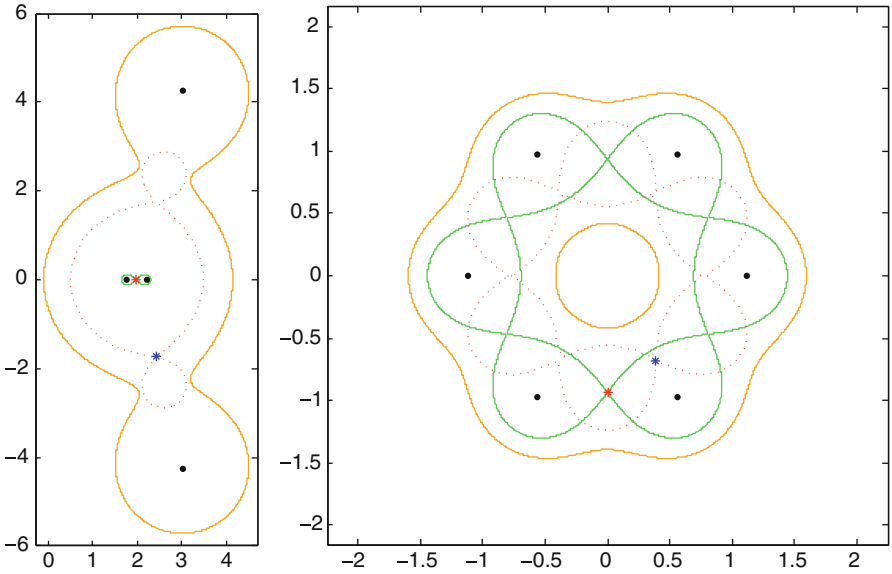

Fig. 4 The boundary of the set $\Lambda_\epsilon(A)$ is depicted for $\epsilon = W_2(A)$ and $\epsilon = W_3(A)$ with solid curves for the inverse Hessian (*left*) and smoke (*right*) examples. The boundary of $\Lambda_{\epsilon,2}(A)$ is plotted for $\epsilon = W_3(A)$ with dotted curves. *Asterisks* mark the computed complex points minimizing $\tau_2(A, \lambda)$ and $\tau_3(A, \lambda)$ with respect to λ. *Black dots* illustrate the eigenvalues of unperturbed matrices

It turns out that a second possibility for the geometric interpretation of $W_3(A)$ is that two components of $\Lambda_{\epsilon,2}(A)$ coalesce for $\epsilon = W_3(A)$. For instance as shown in Fig. 4 on the left two components of $\Lambda_{\epsilon,2}(H_I)$ (plotted as a dotted curve) coalesce for $\epsilon = W_3(H_I)$ at $\lambda_* = 0.3841 \pm 0.6767i$ where $\tau_3(H_I, \lambda)$ is minimized. To illustrate these ideas on a complex matrix consider the 6×6 smoke matrix S that can be generated by the command `gallery('smoke',6)` in Matlab. In Fig. 4 on the right, the boundary of $\Lambda_\epsilon(S)$ is plotted for $\epsilon = W_2(S) = 0.2120$ and $\epsilon = W_3(S) = 0.3270$ as solid curves. The points minimizing $\tau_2(S, \lambda)$ and $\tau_3(S, \lambda)$ are marked with asterisks. Once again two components of $\Lambda_\epsilon(S)$ coalesce at $\lambda_* = -0.9328i$ minimizing $\tau_2(S, \lambda)$ for $\epsilon = W_2(S)$. The set $\Lambda_\epsilon(S)$ is comprised of only one component for $\epsilon = W_3(S)$ and $\lambda_* = 0.3841 - 0.6767i$ minimizing $\tau_3(S, \lambda)$ lies strictly inside of this component. On the other hand two components of $\Lambda_{\epsilon,2}(S)$ (whose boundary is plotted as a dotted curve in Fig. 4 on the right) coalesce at $\lambda_* = 0.3841 - 0.6767i$ for $\epsilon = W_3(S)$. It is also apparent from Fig. 4 that for the smoke matrix each one of $\tau_2(S, \lambda)$ and $\tau_3(S, \lambda)$ have five other global minimizers with respect to λ in addition to λ_* values mentioned above. This means that the closest matrices with an eigenvalue of multiplicity two and three are not unique.

We conjecture that there are two possibilities regarding how the minimizer of $\tau_3(A, \lambda)$ is related to pseudospectra; *two components of* $\Lambda_\epsilon(A)$ *(one containing two of the eigenvalues of A, the other containing only one of the eigenvalues) or* $\Lambda_{\epsilon,2}(A)$ *coalesce at this minimizer for* $\epsilon = W_3(A)$. The latter seems to be the more common case. In general it seems reasonable to expect that two components of one of

the sets $\Lambda_{\epsilon,j}(A)$, $j = 1, \ldots, r-1$ coalesce for $\epsilon = W_r(A)$ at the minimizer of $\tau_r(A, \lambda)$.

4.2 Degenerate cases when $r \geq 3$

We were able to show that $W_r(A) = \tau_r(A)$ under the assumptions that the multiplicity of the optimal singular value is one and the linear independence qualification holds. In practice we observe that usually both of these conditions are satisfied. However, for certain matrices at the computed optimizer (λ_*, γ_*), we observed that one or both of these conditions are not met. For these matrices we can verify that the computed value is only a lower bound for $W_r(A)$ by means of the connection between $W_r(A)$ and the coalesence of the components of pseudospectra. In this subsection we will distinguish the computed value $\tilde{\tau}_r(A)$ from the exact value $\tau_r(A)$. In the examples of this subsection we can only ensure that we retrieve a local maximizer of the inner problem; in theory it is possible that there exist global maximizers of the inner problem where both of the assumptions are satisfied and BFGS converges to an undesired maximizer, so $\tilde{\tau}_r(A) \leq \tau_r(A)$. Furthermore we will denote the computed value of $\tau_r(A, \lambda)$ with $\tilde{\tau}_r(A, \lambda)$. The examples in this section are chosen to demonstrate that a maximizer of the inner problem where both of the assumptions do not hold typically provide only a strict lower bound for $W_r(A)$. In other words both of the assumptions are essential. As mentioned at the opening of this section we experimented with the case $r = 3$ only for computational convenience.

One important class of matrices for which $\tilde{\tau}_r(A)$ seems to provide only a lower bound is the set of Toeplitz matrices. (This assertion is based on numerical experiments on several Toeplitz matrices of various sizes.) Consider for instance

$$T = \begin{bmatrix} 2+i & 1-3i & 2 \\ 1+2i & 2+i & 1-3i \\ 2 & 1+2i & 2+i \end{bmatrix}.$$

In Fig. 5, the boundaries of $\Lambda_\epsilon(T)$ are plotted for $\epsilon = \tau_2(T) = 1.0977$ (inner solid curve) and $\epsilon = \tilde{\tau}_3(T) = 2.7914$ (outer solid curve), while the boundary of $\Lambda_{\epsilon,2}(T)$ is depicted only for $\epsilon = \tilde{\tau}_3(T)$ (dotted curve). The minimizer of $\tilde{\tau}_3(T, \lambda)$ (marked with an asterisk) lies strictly inside $\Lambda_\epsilon(T)$ and outside $\Lambda_{\epsilon,2}(T)$ for $\epsilon = \tilde{\tau}_3(T)$. A point outside the dotted curve can be made an eigenvalue of multiplicity three only with perturbations of norm greater than $\epsilon = \tilde{\tau}_3(T)$ or otherwise the point would be contained inside $\Lambda_{\epsilon,2}(T)$. On the other hand the left-most eigenvalue cannot be moved inside the dotted curve on the right without a perturbation of norm greater than $\epsilon = \tilde{\tau}_3(T)$. Similarly to move the right-most eigenvalue inside the dotted curve on the left a perturbation of norm greater than $\epsilon = \tilde{\tau}_3(T)$ is needed. Therefore there is no perturbation of norm $\tilde{\tau}_3(T)$ or smaller that moves all three eigenvalues onto the same point in the complex plane, that is $\tilde{\tau}_3(T) < W_3(T)$. For this example $\sigma_{3n-1}(\mathcal{A}(T, \lambda_*, \gamma_*)) = \sigma_{3n-2}(\mathcal{A}(T, \lambda_*, \gamma_*)) = 2.7914$ at the computed optimizer (λ_*, γ_*). Therefore, the multiplicity assumption is violated.

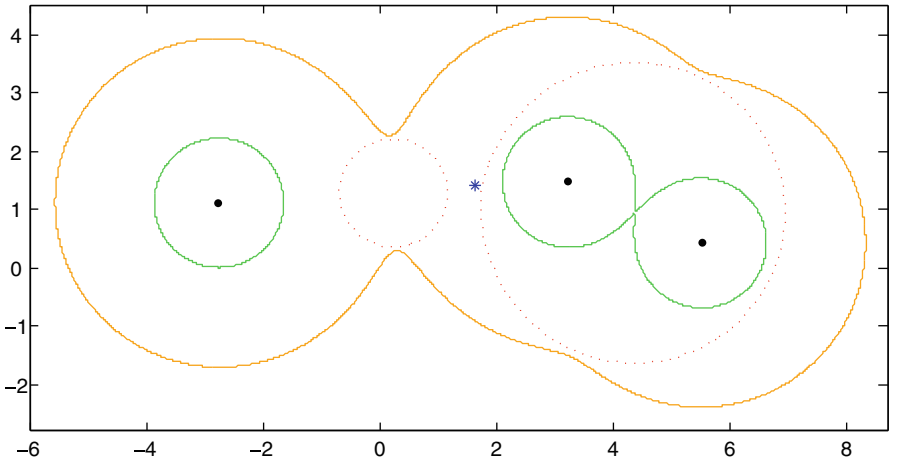

Fig. 5 The sets $\Lambda_\epsilon(T)$ for $\epsilon = \tau_2(T)$ (*inner solid curve*), $\epsilon = \tilde{\tau}_3(T)$ (*outer solid curve*) and $\Lambda_{\epsilon,2}(T)$ for $\epsilon = \tilde{\tau}_3(T)$ (*dotted curve*) are illustrated for the Toeplitz example. The computed minimizer of $\tau_3(T, \lambda)$ with respect to λ is marked with an *asterisk*

The linear independence qualification is violated very rarely in practice. One such example is the 3×3 diagonal matrix $D = \text{diag}(2, 1, 3)$. For this example $\tilde{\tau}_3(D) = 0.3430 < \tau_2(D) = 0.5000 = W_2(D) < W_3(D)$. The computed optimal singular value has multiplicity two, *i.e.* $\sigma_{3n-1}(\mathcal{A}(D, \lambda_*, \gamma_*)) = \sigma_{3n-2}(\mathcal{A}(D, \lambda_*, \gamma_*)) = 0.3430$. The set of right singular vectors associated with these singular values is

$$\text{span}([1 \;\; 1 \;\; 1]^T \otimes [1 \;\; 0 \;\; 0]^T, [1 \;\; 1 \;\; 1]^T \otimes [0 \;\; 1 \;\; 0]^T).$$

Any vector in this set has linearly dependent block components. Neither the multiplicity assumption nor the linear independence qualification holds for D.

4.3 Efficiency of the computational technique

In this subsection we provide running times to compute $\tau_2(A)$. On the simple class of the set of random matrices the Lipschitz based optimization works very efficiently. The cpu-times in seconds for random matrices of various sizes are given in Table 1. The quadratic behavior of the running times with respect to the size of the input matrix becomes apparent for matrices of size greater than 100. Clearly the size of the input matrix does not affect the number of function evaluations listed at the bottom row. Recall that in the table the number of function evaluations refers to the number of times $\tau_2(A, \lambda)$ is computed for various λ. Each $\tau_2(A, \lambda)$ computation involves the maximization of the second smallest singular value with respect to γ.

In general the running times largely depend on the specifics of the input matrix. In Table 2 we list the cpu-times (in seconds) and the number of function evaluations (in parentheses next to each cpu-time) for matrices of various kinds for sizes 5, 10, 20. The matrices in the first four classes are generated by the Matlab command `gallery`.

Table 1 cpu-times and number of function evaluations for random matrices of various size

Size	5	10	20	50	100	200	500	1,000
cpu-time (seconds)	2	4	4	9	12	48	273	1,201
fun evals	127	189	153	115	119	131	113	107

Table 2 cpu-times (in seconds) and number of function evaluations (in parentheses) for various matrices of size 5, 10, 20

	$n = 5$	$n = 10$	$n = 20$
Parter	4.9 (291)	9.1 (199)	33.8 (183)
Riemann	3.4 (173)	4.0 (109)	11.1 (235)
Smoke	4.5 (247)	8.4 (271)	57.3 (903)
Tridiagonal	60.3 (201)	83.8 (125)	115.3 (107)
Wilkinson	46.4 (125)	155.5 (217)	

The Wilkinson matrices are generated by the command `wilkinson`. The remarkable property of each Wilkinson matrix is that its spectrum consists of pairs of eigenvalues that are close to each other. Indeed the Wilkinson matrix of size 20 has an eigenvalue pair that differ from each other by about 8×10^{-14}. The technique described is for low precision, so we did not attempt to compute $\tau_2(A)$ for the Wilkinson matrix of size 20. Considerably more running time is required by the tridiagonal and Wilkinson matrices, which is caused by neither the number of BFGS calls to compute $\tau_2(A, \lambda)$ nor the number of BFGS iterations. It seems the increase in the running time for these matrices is due to ARPACK. (The QR algorithm in LAPACK is likely to run faster for these examples since the sizes of matrices are small.)

4.4 Necessity of optimization over complex variables

In Malyshev's formula the inner maximization is over a real variable (indeed over the unit interval $(0, 1]$). At first thought one may question whether the inner maximization in the more general formula derived in this paper can be restricted to real variables rather than complex variables. Maximizing over real variables in the inner problem often yields values significantly smaller than the ones obtained by maximizing over complex variables. For instance for the matrix

$$G = \begin{bmatrix} 3+i & 2+2i & 5 \\ 4i & 5+2i & -3+4i \\ -2-4i & 1-2i & 3 \end{bmatrix}$$

to compute $W_3(G)$ when we optimize over real variables in the inner problem, we obtain 3.1700. The computed value is $\tau_3(G) = 3.2960$, when the inner problem is optimized over complex variables. In Fig. 6, two components of $\Lambda_{\epsilon,2}(G)$ coalesce at

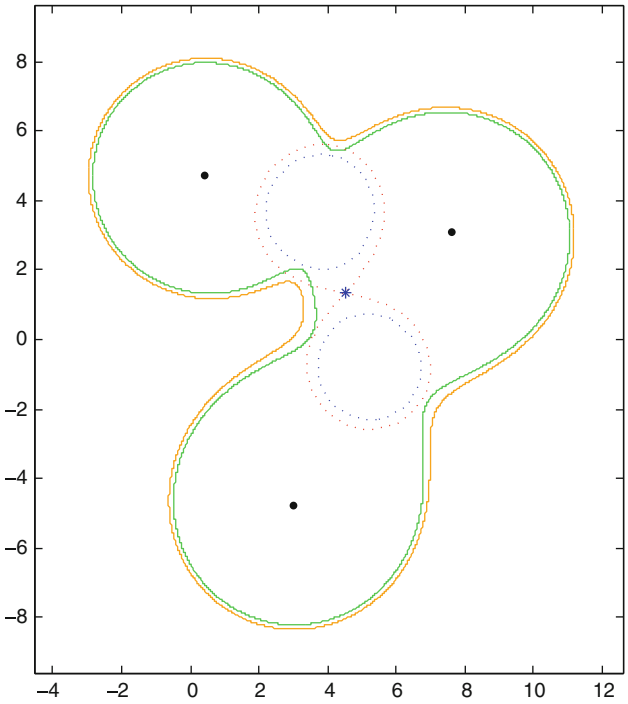

Fig. 6 The boundaries of $\Lambda_{\epsilon,2}(G)$ are depicted with the *dotted curves*. For $\epsilon = \tau_3(G) = 3.2960$ two components of $\Lambda_{\epsilon,2}(G)$ coalesce, while for $\epsilon = 3.1700$ (the optimum value when the inner optimization is performed over real variables) the components of $\Lambda_{\epsilon,2}(G)$ are considerably apart from each other. *Solid curves* correspond to the boundaries of $\Lambda_\epsilon(G)$ for $\epsilon = \tau_3(G)$ and $\epsilon = 3.1700$. *Asterisk* marks the point minimizing $\tau_3(G, \lambda)$

$\lambda_* = 4.5176 + 1.3352i$ (where $\tau_3(G, \lambda)$ attains the smallest value) for $\epsilon = \tau_3(G)$, while these two components are considerably apart from each other for $\epsilon = 3.1700$.

5 Conclusion

In this paper we provided a singular value characterization for the distance in 2-norm to the nearest matrix with an eigenvalue of prespecified algebraic multiplicity. This is a generalization of Malyshev's result for the distance to the nearest matrix with a multiple eigenvalue [15]. The singular value characterizations hold under a mild linear independence and multiplicity assumption.

We also introduced a numerical technique to compute this distance as well as the nearest matrix with an eigenvalue of prespecified algebraic multiplicity. The numerical technique is meant for low precision (up to 5–6 digit accuracy) and exploits the Lipschitz continuity of singular values.

Acknowledgments The author is grateful to Michael L. Overton for providing invaluable comments on the initial version of this manuscript, and two anonymous referees for their suggestions, which greatly improved the readability of this work.

References

1. Alam, R., Bora, S.: On sensitivity of eigenvalues and eigendecompositions of matrices. Linear Algebra Appl. **396**, 273–301 (2005)
2. Bunse-Gerstner, A., Byers, R., Mehrmann, V., Nichols, N.K.: Numerical computation of an analytic singular value decomposition of a matrix valued function. Numer. Math. **60**, 1–39 (1991)
3. Byers, R.: Detecting nearly uncontrollable pairs. In: Kaashoek, M.A., van Schuppen, J.H., Ran, A.C.M. (eds.) Proceedings of the International Symposium MTNS-89, vol. III. North-Holland, Amsterdam (1990)
4. Demmel, J.W.: Computing stable eigendecompositions of matrices. Linear Algebra Appl. **79**, 163–193 (1986)
5. Gablonsky, J.M.: Modifications of the DIRECT algorithm. Ph.D. thesis. North Carolina State University, Raleigh, North Carolina (2001)
6. Golub, G., Van Loan, C.: Matrix Computations, 3rd edn. The John Hopkins University Press, Baltimore (1996)
7. Hinrichsen, D., Motscha, M.: Optimization problems in the robustness analysis of linear state space systems. In: Proceedings of the Seminar on Approximations and Optimization, Havana, LN Mathematics. Springer, Berlin (1987)
8. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, London (1985)
9. Jones, D.R., Perttunen, C.D., Stuckman, B.E.: Lipschitzian optimization without the Lipschitz constant. J. Optim. Theory Appl. **79**(1), 157–181 (1993)
10. Kato, T.: Perturbation Theory for Linear Operators, 2nd edn. Springer, Berlin (1976)
11. Lancaster, P., Tismenetsky, M.: The Theory of Matrices, 2nd edn. Academic Press, London (1985)
12. Lehoucq, R., Sorensen, D., Yang, C.: ARPACK users' guide: solution of large scale eigenvalue problems with implicitly restarted Arnoldi methods (1997)
13. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Math. Program. **45**(3, (Ser. B)), 503–528 (1989)
14. Van Loan, C.F.: How near is a stable matrix to an unstable matrix. Contemp. Math. **47**, 465–477 (1985)
15. Malyshev, A.N.: A formula for the 2-norm distance from a matrix to the set of matrices with multiple eigenvalues. Numer. Math. **83**, 443–454 (1999)
16. Paige, C.C.: Properties of numerical algorithms related to computing controllability. IEEE Trans. Automat. Contr. **26**(1), 130–138 (1981)
17. Piyavskii, S.A.: An algorithm for finding the absolute extremum of a function. USSR Comput. Math. Math. Phys. **12**, 57–67 (1972)
18. Qiu, L., Bernhardsson, B., Rantzer, A., Davison, E.J., Young, P.M., Doyle, J.C.: A formula for computation of the real stability radius. Automatica **31**(6), 879–890 (1995)
19. Rellich, F.: Storungstheorie der spektralzerlegung, I. Math. Ann. **113**, 600–619 (1937)
20. Rellich, F.: Storungstheorie der spektralzerlegung, II. Math. Ann. **113**, 677–685 (1937)
21. Rellich, F.: Storungstheorie der spektralzerlegung, III. Math. Ann. **116**, 555–570 (1939)
22. Rellich, F.: Storungstheorie der spektralzerlegung, IV. Math. Ann. **117**, 356–382 (1940)
23. Rellich, F.: Storungstheorie der spektralzerlegung, V. Math. Ann. **118**, 462–484 (1942)
24. Ruhe, A.: Properties of a matrix with a very ill-conditioned eigenproblem. Numer. Math. **15**, 57–60 (1970)
25. Sergeyev, Y.D., Kvasov, D.E.: Global search based on efficient diagonal partitions and a set of lipschitz constants. SIAM J. Optim. **16**(3), 910–937 (2006)
26. Shubert, B.: A sequential method seeking the global maximum of a function. SIAM J. Numer. Anal. **9**, 379–388 (1972)
27. Wilkinson, J.H.: The Algebraic Eigenvalue Problem. Clarendon Press, Oxford (1965)
28. Wilkinson, J.H.: Note on matrices with a very ill-conditioned eigenproblem. Numer. Math. **19**, 176–178 (1972)
29. Wilkinson, J.H.: On neighbouring matrices with quadratic elementary divisors. Numer. Math. **44**, 1–21 (1984)