ON THE SPECTRA OF CERTAIN MATRICES GENERATED BY INVOLUTORY AUTOMORPHISMS

ARIEH ISERLES * AND ANTONELLA ZANNA †

Abstract. Let $A = P + K$ be an $n \times n$ complex matrix with $P = \frac{1}{2}(A - HAH)$ and $K = \frac{1}{2}(A + HAH)$, H being a unitary involution. Having characterised all unitary involutions, we investigate the spectral structure of P and K and, in particular, characterise the eigenvalues of K as zeros of a rational function and prove that, for normal A, $\sigma(K)$ resides in the convex hull of $\sigma(A)$. We also demonstrate that this need not be true when A is not normal.

Key words. Inner automorphism, involution, Lie algebra, Lie triple system

AMS subject classifications. 15A18

1. Introduction. The goal of this paper is to explore a number of issues in matrix analysis which have arisen in research at the interface of geometric numerical integration and computational linear algebra. Although the results of this paper stand alone and they do not require any elaboration of these issues, the latter are important in motivating and setting the backdrop for our work, hence we commence by reviewing them briefly.

Let $\mathfrak{g}$ be a matrix Lie algebra. The approximation of $\exp A$, where $A \in \mathfrak{g}$, is a central step in most numerical methods for the solution of differential equations evolving in Lie groups [5]. The purpose of such "Lie-group solvers" is to propagate the solution within the Lie group $\mathcal{G}$, say, whose Lie algebra is $\mathfrak{g}$. Therefore, it is of critical importance that the approximate exponential resides in $\mathcal{G}$ whenever the argument lives in $\mathfrak{g}$. Unfortunately, many standards technique to approximate the exponential fail to respect the structure for some Lie groups. In particular, all such methods fail to map an arbitrary element of $\mathfrak{sl}(n)$ to $\mathrm{SL}(n)$ for $n \geq 3$ [7]. This has motivated a new breed of methods, designed to respect Lie-group structure: [1, 2] and, in particular, [8] and [6]. The latter two publications are based on a factorization of $\exp A$ for $A \in \mathfrak{g}$ using *Generalized Polar Decomposition (GPD).* Thus, let κ be a Lie-algebra automorphism, hence a linear operator such that $\kappa([A, B]) = [\kappa(A), \kappa(B)]$, $A, B \in \mathfrak{g}$. In addition, we assume that it is an involution, i.e. $\kappa(\kappa(A)) = A$ for all $A \in \mathfrak{g}$. It has been proposed in [8] to represent the Lie algebra $\mathfrak{g}$ as a direct sum of two linear spaces,

$$\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k},$$

where

$$\mathfrak{p} = \{X \in \mathfrak{g} \,:\, \kappa(X) = -X\}$$

is a Lie triple system (i.e. $[X, [Y, Z]] \in \mathfrak{p}$ for all $X, Y, Z \in \mathfrak{p}$), while

$$\mathfrak{k} = \{X \in \mathfrak{g} \,:\, \kappa(X) = X\}$$

*Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, England, email A.Iserles@damtp.cam.ac.uk.

†Department of Informatics, University of Bergen, Høyteknologisenteret, N-5020 Bergen, Norway, email Antonella.Zanna@ii.uib.no.

is a subalgebra of $\mathfrak{g}$. Specifically, we let $P = \frac{1}{2}[A - \kappa(A)] \in \mathfrak{p}$, $K = \frac{1}{2}[A + \kappa(A)] \in \mathfrak{k}$. It is possible to prove that there exist functions $X(t)$ and $Y(t)$, $t \geq 0$, evolving in $\mathfrak{p}$ and $\mathfrak{k}$ respectively, such that

$$\exp(tA) = \exp(X(t)) \exp(Y(t)).$$

The functions X and Y can be evaluated as a linear combination of commutators in the free algebra generated by $\{P, K\}$,

$$\begin{aligned} X(t) &= tP - \tfrac{1}{2}t^2[P,K] - \tfrac{1}{6}t^3[K,[P,K]] + \tfrac{1}{24}t^4([P,[P,[P,K]]] \\ &\qquad - [K,[K,[P,K]]]) + \mathcal{O}(t^5), \\ Y(t) &= tK - \tfrac{1}{12}t^3[P,[P,K]] + \mathcal{O}(t^5). \end{aligned}$$

In a practical algorithm, the series above are truncated and the calculation is accompanied by a number of linear-algebraic techniques which, while being of no direct relevance to the theme of the present paper, are critically important in reducing computational complexity [6].

Suppose that, in addition, the dimension of $\mathfrak{p}$ is small, so that it is easy to compute $\exp(X(t))$ exactly, while $\dim \mathfrak{k} < \dim \mathfrak{g}$. The idea is to iterate this procedure, representing $\mathfrak{k}$ as a direct sum of a Lie triple system and a sub-algebra using another involutory automorphism and so on. The ultimate outcome is a representation

$$\exp(tA) = \exp(X_1(t)) \exp(X_2(t)) \cdots \exp(X_m(t)), \tag{1.1}$$

where each $\exp(X_k(t))$ can be evaluated with relative ease.

Suppose that A is a 'nice' matrix: the real parts of its eigenvalues are relatively small. Even cursory examination of the numerical stability of the representation (1.1) (or its approximation) demonstrates that it is important for the matrix

$$K = K(A) = \tfrac{1}{2}[A + \kappa(A)] \tag{1.2}$$

to share this 'niceness', otherwise we might be generating very large matrices in the course of computation, thereby losing accuracy in finite arithmetic. This brings us to the central theme of this paper, the connection between the spectra of A and those of P and K.

In Section 2 we characterise all involutory inner automorphisms in $\mathrm{U}(n)$. Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r \in \mathbb{C}^n$, where $r \leq n$, be vectors of unit length, orthogonal to each other. Then

$$\kappa(A) = HAH, \qquad H = I - 2\sum_{k=1}^{r} \boldsymbol{u}_k \boldsymbol{u}_k^*, \tag{1.3}$$

is a unitary involutory automorphism. Moreover, every involutory, unitary inner automorphism of $\mathrm{M}_n[\mathbb{C}]$ can be represented in the form (1.3) for some orthogonal vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r$. Thus, such automorphisms are a natural generalisation of a similarity transformation by the familiar Householder reflection. Furthermore, we show in Section 2 an interesting correspondence between the free group generated by the involutory matrices $I - 2\boldsymbol{u}_k \boldsymbol{u}_k^*$, $k = 1, 2, \ldots, n$, and the additive group $\mathbb{Z}_n^2$ of binary n-tuples, with addition defined *modulo* 2.

Section 3 is devoted to our main result. We prove that, for a *normal* matrix A, the eigenvalues of K reside in the convex hull of the eigenvalues of A. This confirms that if A is 'nice', so is K: in particular, if A is a stable matrix, this feature is shared by K.

Finally, in Section 4 we demonstrate that normalcy of A is crucial. Once we consider general matrices $A \in \mathrm{M}_n[\mathbb{C}]$, the result of Section 3 is no longer valid and the eigenvalues of K might well reside outside the convex hull of $\sigma(A)$.

2. Involutory automorphisms in $\mathrm{U}(n)$**.** An *automorphism* in the general linear algebra $\mathrm{M}_n[\mathbb{C}]$ is a linear function $\kappa : \mathrm{M}_n[\mathbb{C}] \to \mathrm{M}_n[\mathbb{C}]$ such that $\kappa([A,B]) = [\kappa(A), \kappa(B)]$ for all $A, B \in \mathrm{M}_n[\mathbb{C}]$. κ is an *inner* automorphism if it is of the form $\kappa(A) = HAH^{-1}$ for some nonsingular matrix $H \in \mathrm{M}_n[\mathbb{C}]$ and an *involution* if $\kappa(\kappa(A)) = A$ for all $A \in \mathrm{M}_n[\mathbb{C}]$. Thus, an involutory inner automorphism is $\kappa(A) = HAH$, where H is itself an involutory matrix, $H^2 = I$. Motivated by numerical stability, we are concerned solely with unitary matrices H. Therefore, the main focus of this section is on the unitary involutory inner automorphism

$$\kappa(A) = HAH, \quad A \in \mathrm{M}_n[\mathbb{C}], \quad \text{where} \quad H \in \mathrm{U}(n), \ \ H^2 = I. \tag{2.1}$$

THEOREM 2.1. *Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_s\}$ be an orthonormal basis of a subspace of $\mathbb{C}^n$, where $r \in \{0, 1, \ldots, n\}$. Then*

$$H = I - 2\sum_{k=1}^{s} \boldsymbol{u}_k \boldsymbol{u}_k^* \tag{2.2}$$

is an involution in $\mathrm{U}(n)$*. Moreover, every involution in* $\mathrm{U}(n)$ *can be represented in the form (2.2).*

Proof. It is straightforward to prove that any H defined by (2.2) is unitary and Hermitian, hence a unitary involution. This proves the first, trivial statement of the theorem.

Next, let us assume that $H \in \mathrm{U}(n)$ is an involution, therefore H is Hermitian. This implies that also the matrix $I - H$ is Hermitian. Suppose that the eigenvalues and normalised eigenvectors of $I - H$ are $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$ and $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n \in \mathbb{C}^n$, respectively. It follows that

$$I - H = \sum_{k=1}^{n} \alpha_k \boldsymbol{u}_k \boldsymbol{u}_k^*.$$

Exploiting orthogonality of eigenvectors, we deduce that

$$H^2 = I - 2\sum_{k=1}^{n} \alpha_k \boldsymbol{u}_k \boldsymbol{u}_k^* + \sum_{k=1}^{n}\sum_{l=1}^{n} \alpha_k \alpha_l (\boldsymbol{u}_k^* \boldsymbol{u}_l) \boldsymbol{u}_k \boldsymbol{u}_l^* = I - \sum_{k=1}^{n} (\alpha_k - 2)\alpha_k \boldsymbol{u}_k \boldsymbol{u}_k^*.$$

Since $H^2 = I$ and the matrices $\boldsymbol{u}_k \boldsymbol{u}_k^*$ are linearly independent, it follows that $\alpha_k \in \{0, 2\}$. Without loss of generality, we may assume that $\alpha_1 = \cdots = \alpha_s = 2$, $\alpha_{s+1} = \cdots = \alpha_n = 0$, and this results in the representation (2.2) and completes the proof of the theorem. ◻

Note that, once we have characterised all unitary involutions in $\mathrm{U}(n)$, we immediately obtain from (2.1) a characterisation of all unitary involutory inner automorphisms of $\mathrm{M}_n[\mathbb{C}]$.

Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ be a unitary basis of $\mathbb{C}^n$, $\|\boldsymbol{u}_k\| = 1$. Given a vector $\boldsymbol{\theta} \in \mathbb{Z}_2^n$, we set

$$G_{\boldsymbol{\theta}} = I - 2\sum_{k=1}^{n} \theta_k \boldsymbol{u}_k \boldsymbol{u}_k^*.$$

Note that, by the last theorem, $G_{\boldsymbol{\theta}}$ is a unitary involution. It is trivial to verify that

$$G_{\boldsymbol{\theta}} G_{\boldsymbol{\varphi}} = G_{\boldsymbol{\theta}+\boldsymbol{\varphi} \bmod 2}.$$

Therefore,

$$\boldsymbol{G} = \{G_{\boldsymbol{\theta}} \,:\, \boldsymbol{\theta} \in \mathbb{Z}_2^n\}$$

is an Abelian multiplicative group, isomorphic to $\mathbb{Z}_2^n$. Another easy observation is that the space of all unitary involutions (2.2), for all $r = 0, 1, \ldots, n$, is a free group generated by $G_{\boldsymbol{e}_k}$, $k = 1, 2, \ldots, n$. Needless to say, $r = 0$ yields the identity matrix and it is easy to prove that $G_{1,1,\ldots,1} = -I$. The latter is a consequence of the well-known identity,

$$\sum_{k=1}^{n} \boldsymbol{u}_k \boldsymbol{u}_k^* = I,$$

known in quantum chemistry as *resolution of identity.*

3. The spectrum of a normal matrix K**.** The purpose of this section is to study the eigenvalues $\mu_1, \mu_2, \ldots, \mu_n$ and eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ of the matrix

$$K = \tfrac{1}{2}[A + \kappa(A)], \tag{3.1}$$

where $A \in \mathrm{M}_n[\mathbb{C}]$ and κ is a unitary involutory inner automorphism, $\kappa(A) = HAH$.

Since $H \in \mathrm{U}(n)$, it has a full set of unitary eigenvectors, therefore $H = QDQ^*$, where $Q \in \mathrm{U}(n)$ and D is diagonal. Moreover, since $H^2 = I$, necessarily, without loss of generality,

$$D = \begin{bmatrix} I_{r\times r} & O_{r\times s} \\ O_{s\times r} & -I_{s\times s} \end{bmatrix},$$

where $r + s = n$. Letting $\tilde{A} = Q^*AQ$, $\tilde{K} = Q^*KQ$, we deduce at once from (3.1) that

$$\tilde{K} = \tfrac{1}{2}(\tilde{A} + D\tilde{A}D) = \begin{bmatrix} \tilde{A}_{1,1} & O \\ O & \tilde{A}_{2,2} \end{bmatrix}, \qquad \text{where} \qquad A = \begin{bmatrix} \tilde{A}_{1,1} & \tilde{A}_{1,2} \\ \tilde{A}_{2,1} & \tilde{A}_{2,2} \end{bmatrix}$$

and $\tilde{A}_{1,1} \in \mathrm{M}_r[\mathbb{C}]$, $\tilde{A}_{2,2} \in \mathrm{M}_s[\mathbb{C}]$. We thus conclude that

$$\sigma(K) = \sigma(\tilde{K}) = \sigma(\tilde{A}_{1,1}) \cup \sigma(\tilde{A}_{2,2}). \tag{3.2}$$

Moreover, if $\tilde{K}\tilde{\boldsymbol{v}} = \mu\tilde{\boldsymbol{v}}$ then

$$\tilde{\boldsymbol{v}} = \begin{bmatrix} \tilde{\boldsymbol{v}}_1 \\ \tilde{\boldsymbol{v}}_2 \end{bmatrix}, \qquad \tilde{\boldsymbol{v}}_1 \in \mathbb{C}^r, \;\; \tilde{\boldsymbol{v}}_2 \in \mathbb{C}^s$$

and either

$$\tilde{A}_{1,1}\tilde{\boldsymbol{v}}_1 = \mu\tilde{\boldsymbol{v}}_1, \qquad \tilde{\boldsymbol{v}}_2 = \boldsymbol{0}$$

or

$$\tilde{\boldsymbol{v}}_1 = \boldsymbol{0}, \qquad \tilde{A}_{2,2}\tilde{\boldsymbol{v}}_2 = \mu\tilde{\boldsymbol{v}}_2.$$

PROPOSITION 3.1. *Let $\boldsymbol{v}$ be an eigenvector of K. Then it is also an eigenvector of H.*

Proof. Let $\tilde{\boldsymbol{v}} = Q^*\boldsymbol{v}$, $\boldsymbol{w} = H\boldsymbol{v}$ and $\tilde{\boldsymbol{w}} = Q^*\boldsymbol{w}$. Note that

$$K\boldsymbol{v} = \mu\boldsymbol{v} \qquad \Rightarrow \qquad (A + HAH)\boldsymbol{v} = 2\mu\boldsymbol{v} \qquad \Rightarrow \qquad (HA + AH)\boldsymbol{v} = 2\mu\boldsymbol{w}$$

therefore $K\boldsymbol{w} = \mu\boldsymbol{w}$. Thus, $\boldsymbol{w} \neq \boldsymbol{0}$ is also an eigenvector of K corresponding to the same eigenvalue μ.

We have

$$\tilde{\boldsymbol{w}} = Q^*(QDQ^*)Q\tilde{\boldsymbol{v}} = D\tilde{\boldsymbol{v}} = \begin{bmatrix} \tilde{\boldsymbol{v}}_1 \\ -\tilde{\boldsymbol{v}}_2 \end{bmatrix},$$

hence

$$\boldsymbol{w} = Q \begin{bmatrix} \tilde{\boldsymbol{v}}_1 \\ -\tilde{\boldsymbol{v}}_2 \end{bmatrix}.$$

Recall that either $\tilde{\boldsymbol{v}}_1$ or $\tilde{\boldsymbol{v}}_2$ is a zero vector. If $\tilde{\boldsymbol{v}}_2 = \boldsymbol{0}$ then

$$H\boldsymbol{v} = \boldsymbol{w} = Q \begin{bmatrix} \tilde{\boldsymbol{v}}_1 \\ \tilde{\boldsymbol{v}}_2 \end{bmatrix} = Q\tilde{\boldsymbol{v}} = \boldsymbol{v},$$

while, by the same token, if $\tilde{\boldsymbol{v}}_1 = \boldsymbol{0}$ then

$$H\boldsymbol{v} = -\boldsymbol{v}.$$

This concludes the proof. □

An alternative formulation of Proposition 3.1 is that the eigenvectors of K reside in one of the linear spaces

$$\mathcal{K}_n = \{\boldsymbol{x} \in \mathbb{C}^n \,:\, H\boldsymbol{x} = \boldsymbol{x}\} \qquad \text{or} \qquad \mathcal{P}_n = \{\boldsymbol{x} \in \mathbb{C}^n \,:\, H\boldsymbol{x} = -\boldsymbol{x}\}.$$

Without loss of generality we assume that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r \in \mathcal{K}_n$ and $\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, \ldots, \boldsymbol{v}_n \in \mathcal{P}_n$. Of course, $\mathcal{K}_n \oplus \mathcal{P}_n = \mathbb{C}^n$.

Recalling the characterisation (2.2) of unitary involutions, we denote

$$\mathcal{U} = \operatorname{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_s\}$$

(the double use of the integer variable s is, as will be apparent soon, not a variable overload). Since

$$H\boldsymbol{v} = \boldsymbol{v} - 2\sum_{k=1}^{s}(\boldsymbol{u}_k^*\boldsymbol{v})\boldsymbol{u}_k,$$

we deduce that

$$\begin{aligned} \boldsymbol{v} \in \mathcal{P}_n \qquad &\Rightarrow \qquad \boldsymbol{v} = \sum_{k=1}^{s}(\boldsymbol{u}_k^*\boldsymbol{v})\boldsymbol{u}_k \qquad &&\Rightarrow \qquad \mathcal{P}_n = \mathcal{U}, \\ \boldsymbol{v} \in \mathcal{K} \qquad &\Rightarrow \qquad \boldsymbol{u}_k^*\boldsymbol{v} = 0, \quad k = 1, 2, \ldots, s \qquad &&\Rightarrow \qquad \mathcal{K}_n = \mathcal{U}^{\perp}. \end{aligned}$$

So far we have allowed an arbitrary $A \in \mathrm{M}_n[\mathbb{C}]$. In the remainder of this section we stipulate that A is *normal,* hence $A^*A = AA^*$ and its normalised eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ form a unitary basis of $\mathbb{C}^n$. We denote the eigenvalues of A by $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively.

THEOREM 3.2. *Let $A \in \mathrm{M}_n[\mathbb{C}]$ be a normal matrix, $H \in \mathrm{U}(n)$ an involution and $K = \frac{1}{2}(A + HAH)$. Then*

$$\sigma(K) \subset \operatorname{conv} \sigma(A), \tag{3.3}$$

where $\sigma(B)$ is the spectrum of the matrix B, while $\operatorname{conv} \Omega$ *is the (closed) convex hull of $\Omega \in \mathbb{C}$.*

Proof. Let $\sigma(K) = \{\mu_1, \mu_2, \ldots, \mu_n\}$ and let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be the corresponding normalised eigenvectors. For clarity, whenever possible we dispense with a subscript and let $K\boldsymbol{v} = \mu\boldsymbol{v}$. Recall that either $\boldsymbol{v} \in \mathcal{K}_n$ or $\boldsymbol{v} = \mathcal{P}_n$.

If $\boldsymbol{v} \in \mathcal{K}_n$ then $H\boldsymbol{v} = \boldsymbol{v}$, hence

$$\mu\boldsymbol{v} = K\boldsymbol{v} = \tfrac{1}{2}(A\boldsymbol{v} + HA\boldsymbol{v}),$$

therefore

$$\mu = \tfrac{1}{2}\boldsymbol{v}^*(A\boldsymbol{v} + HA\boldsymbol{v}) = \boldsymbol{v}^* A\boldsymbol{v}.$$

The conclusion is true also for $\boldsymbol{v} \in \mathcal{P}_n$, since the sign of the second term changes twice. Therefore

$$\boldsymbol{v}^* A\boldsymbol{v} = \mu. \tag{3.4}$$

Since A is normal, it is true that

$$A = \sum_{k=1}^{n} \lambda_k \boldsymbol{w}_k \boldsymbol{w}_k^*.$$

Therefore

$$\mu = \boldsymbol{v}^* A\boldsymbol{v} = \sum_{k=1}^{n} \lambda_k |\boldsymbol{v}^* \boldsymbol{w}_k|^2$$

and we deduce that μ is a convex linear combination of $\lambda_1, \lambda_2, \ldots, \lambda_n$. Consequently, $\mu \in \operatorname{conv} \sigma(A)$ and (3.3) is valid. □

The implications of Theorem 3.2 are clear: if A is a stable matrix then so is K and the ℓ_2 logarithmic norm of K, an appropriate measure of its stability, cannot exceed that of A. Fig. 3.1 displays the eigenvalues (and their convex hull) of a 100×100 normal matrix A and of the corresponding matrix K, with $s = 5$. Both A and H have been randomly generated, using matrices from Gaussian unitary ensemble. Evidently, the eigenvalues of K are consistent with Theorem 3.2. Moreover, they appear to be a moderate perturbation of the eigenvalues of A. We believe that this is in general the case when either s or $n - s$ is significantly smaller than n. Thus, Fig. 3.1 depicts the eigenvalues of the same 100×100 matrix A but different K, generated randomly using $s = 50$. Another structural detail is apparent, it recurs in other computational experiments and is further illustrated in Fig. 3.3: the eigenvalues of K 'shrink' toward the centre of the convex hull. It is unclear by this stage whether this behaviour is a stochastic artifact or whether $\sigma(K)$ can be always confined to significantly smaller geometric structure inside $\operatorname{conv}\sigma(A)$.

A measure of support for our observation that, for small s, $\sigma(K)$ is in general a moderate perturbation of $\sigma(A)$ is provided by the following result.

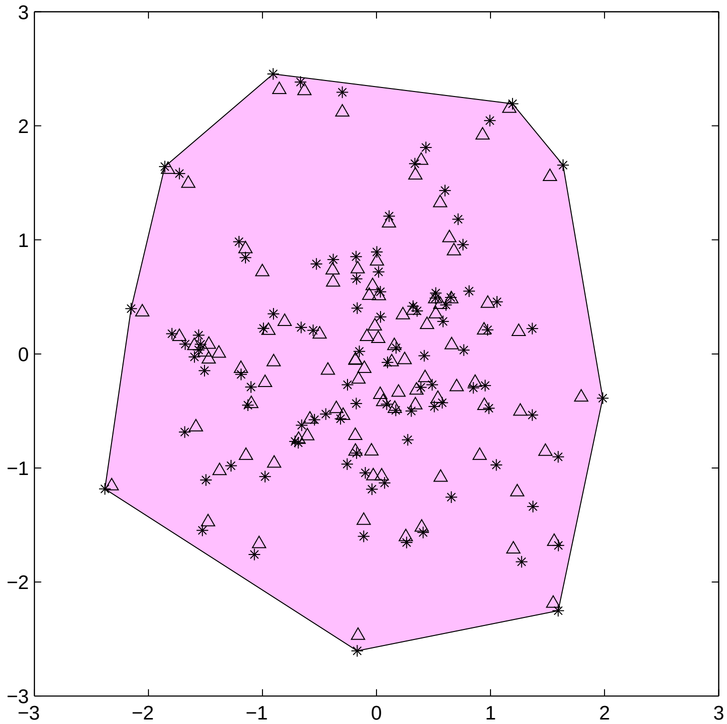

FIGURE 3.1. *The eigenvalues of A (denoted by asterisks) and of K (denoted by triangles) for $n = 100$ and $s = 5$.*

LEMMA 3.3. *Assume that A is a normal matrix. Then,*

$$\|A - K\|_{\mathsf{F}} \le \frac{\sqrt{2s}}{2} \operatorname{diam} \operatorname{conv} \sigma(A), \tag{3.5}$$

where $\operatorname{diam} \Omega$ *is the diameter of the set* $\Omega \subset \mathbb{C}$.

Proof. For simplicity sake, let us denote

$$H = I - 2UU^*,$$

where $U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_s]$.

We commence by noting that $A - K = P$, where $P = \frac{1}{2}(A - HAH)$ has been defined in Section 1. Moreover,

$$P = UU^*A + AUU^* - 2UU^*AUU^* = U[A^*U - UU^*A^*U]^* + [AU - UU^*AU]U^*,$$

therefore P is a rank-$2s$ matrix with (at most) $2s$ nonzero eigenvalues, $\kappa_1, \kappa_2, \ldots, \kappa_{2s}$. We deduce that

$$\|A - K\|_{\mathsf{F}}^2 = \sum_{i=1}^{2s} |\kappa_i|^2.$$

It is easy to verify that, if κ_i is an eigenvalue of P corresponding to the eigenvector $\boldsymbol{y}_i$, then

$$P(H\boldsymbol{y}_i) = -HP\boldsymbol{y}_i = -H\kappa_i \boldsymbol{y}_i = -\kappa_i(H\boldsymbol{y}_i),$$

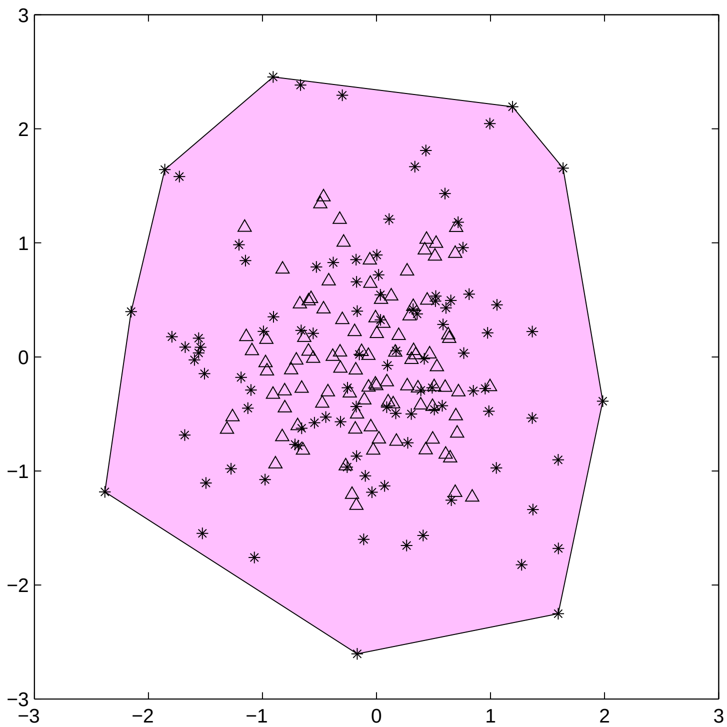

FIGURE 3.2. *The eigenvalues of A (denoted by asterisks) and of K (denoted by triangles) for $n = 100$ and $s = 50$.*

that is, $-\kappa_i$ is also an eigenvalue of P corresponding to the eigenvector $H\boldsymbol{y}_i$. Hence, assuming that the eigenvalues of P are labelled so that $\kappa_{i+s} = -\kappa_i$, we obtain

$$\|A - K\|_{\mathsf{F}}^2 = 2\sum_{i=1}^{s} |\kappa_i|^2 \le 2s \max_{i=1,\ldots,s} |\kappa_i|^2 = 2s\,[\rho(P)]^2,$$

where $\rho(\,\cdot\,)$ denotes the spectral radius. Since the spectrum of P is symmetric with respect to the origin, we deduce that $\rho(P) = \max_{i=1,\ldots,s} |\kappa_i| = \frac{1}{2}$ diam conv$\sigma(P)$.

Recall that, if $B, C \in \mathrm{M}_n[\mathbb{C}]$ are normal, with $\sigma(B) = \{\beta_i : i = 1,\ldots,n\}$ and $\sigma(C) = \{\gamma_i : i = 1,\ldots,n\}$, then

$$\sigma(B + C) \subseteq \mathrm{conv}\{(\beta_i + \gamma_j) : i, j = 1,\ldots,n\}$$

[4]. Since A and HAH are both normal and, HAH being similar to A, share the same eigenvalues λ_i, $i = 1,\ldots,n$, we have

$$\sigma(P) = \sigma[\tfrac{1}{2}(A - HAH)] \subseteq \tfrac{1}{2}\mathrm{conv}\{(\lambda_i - \lambda_j) : i, j = 1,\ldots,n\}.$$

We observe that the set $\mathrm{conv}\{(\lambda_i - \lambda_j) : i, j = 1,\ldots,n\}$ has a point symmetry at the origin: if $(\lambda_i - \lambda_j)$ is a vertex of the convex hull, so is $-(\lambda_i - \lambda_j)$. Therefore,

$$\mathrm{diam\ conv}\{(\lambda_i - \lambda_j) : i, j = 1,\ldots,n\} = 2 \max_{i,j=1,\ldots,n} |\lambda_i - \lambda_j| = 2\mathrm{diam\ conv}\sigma(A).$$

Thus,

$$\rho(P) = \tfrac{1}{2}\mathrm{diam\ conv}\sigma(P) \le \tfrac{1}{2}\mathrm{diam\ conv}\sigma(A)$$

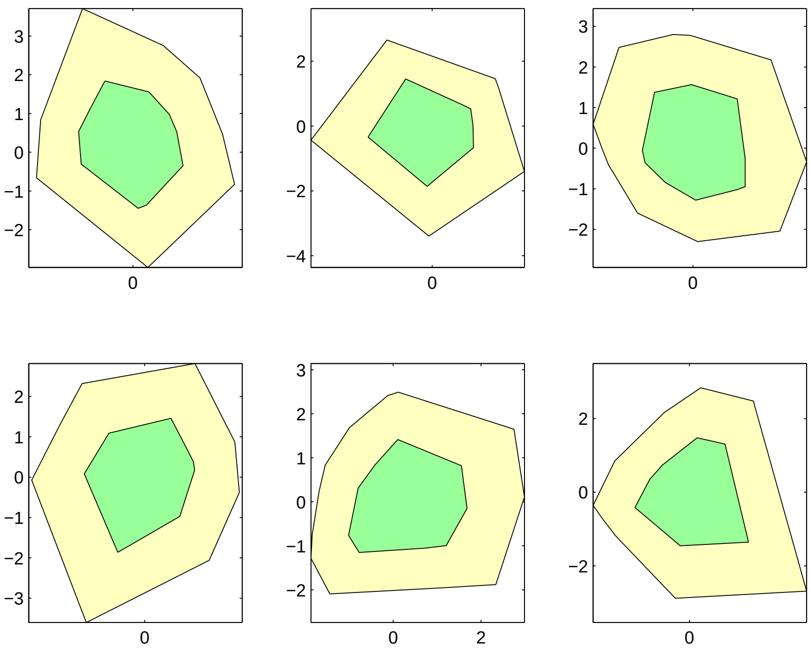

FIGURE 3.3. *The convex hulls of $\sigma(A)$ and $\sigma(K)$ for six random matrices (A being normal) with $n = 200$ and $s = 100$.*

and we conclude that

$$\|A - K\|_{\mathsf{F}}^2 \leq \frac{s}{2} \operatorname{diam} \operatorname{conv} \sigma(A),$$

from which (3.5) follows by taking square roots. □

It is worthwhile to mention that (3.5) is sharp in the case $s = 1$. Letting $H = I - 2\boldsymbol{u}\boldsymbol{u}^*$, a simple calculation reveals that P has rank two and that its eigenvalues are

$$\kappa_{1,2} = \pm\sqrt{\boldsymbol{u}^* A^2 \boldsymbol{u} - (\boldsymbol{u}^* A \boldsymbol{u})^2}.$$

Hence,

$$\|A - K\|_{\mathsf{F}}^2 = \|P\|_{\mathsf{F}}^2 = |\kappa_1|^2 + |\kappa_2|^2 = 2\rho(P)^2.$$

Assume that the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A and the corresponding normalised eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are ordered so that

$$|\lambda_1 - \lambda_n| = \max_{k,l=1,\ldots,n} |\lambda_k - \lambda_l|$$

and take $\boldsymbol{u} = \frac{\sqrt{2}}{2}(\boldsymbol{x}_1 + \boldsymbol{x}_n)$. It is readily observed that $\boldsymbol{u}^* A^k \boldsymbol{u} = \frac{1}{2}(\lambda_1^k + \lambda_n^k)$, $k \geq 0$, hence

$$\rho(P) = \max\{|\kappa_1|, |\kappa_2|\} = \frac{1}{2}|\lambda_1 - \lambda_n| = \frac{1}{2} \operatorname{diam} \operatorname{conv} \sigma(A),$$

from which (3.5) follows as an equality.

The implication of (3.3) for $s=1$ is that the eigenvalues of A and K can be ordered so that, *on average,*

$$|\lambda_k-\mu_k|=\mathcal{O}\big(n^{-1}\big)\,,\qquad k=1,2,\ldots. \tag{3.6}$$

This can be extended to $s\geq 2$, as long as s is small in comparison with n, since P is always of rank $2s$.

Although (3.6) is only a probabilistic statement, not an absolute estimate, it goes some way toward explaining the phenomenon that we have observed in Fig. 3.1.

Another interesting connection between $\sigma(A)$ and $\sigma(K)$ is highlighted in our next result.

LEMMA 3.4. *Let $\boldsymbol{v}\in\mathcal{K}_n$ be an eigenvector of K with eigenvalue μ. Then either $\mu\in\sigma(A)$ or it is a zero of the rational function*

$$\psi(x)=\sum_{k=1}^{n}\frac{|\zeta_k|^2}{\lambda_k-x} \tag{3.7}$$

where

$$\boldsymbol{z}=\sum_{k=1}^{s}\beta_k\boldsymbol{u}_k=\sum_{l=1}^{n}\zeta_l\boldsymbol{x}_l$$

$\beta_k=\boldsymbol{u}_k^*A\boldsymbol{v}$, $k=1,2,\ldots,s$, and $\boldsymbol{x}_l$, $l=1,2,\ldots,n$ are the eigenvectors of A.

Proof. Suppose first that $A\boldsymbol{v}\in\mathcal{K}_n$. Then $H\boldsymbol{v}=\boldsymbol{v}$, $HA\boldsymbol{v}=A\boldsymbol{v}$, $K\boldsymbol{v}=\mu\boldsymbol{v}$ and (3.1) imply that $A\boldsymbol{v}=\mu\boldsymbol{v}$, hence $\mu\in\sigma(A)$. Let us turn our attention to the other case, namely $A\boldsymbol{v}\cap\mathcal{P}_n\neq\{\boldsymbol{0}\}$. Since $\mathcal{K}_n=\mathcal{U}^\perp$, we choose an arbitrary unitary basis of $\mathcal{U}^\perp$, namely $\{\boldsymbol{u}_{s+1},\boldsymbol{u}_{s+2},\ldots,\boldsymbol{u}_n\}$. Thus,

$$\boldsymbol{v}=\sum_{k=s+1}^{n}\alpha_k\boldsymbol{u}_k,\qquad \alpha_k=\boldsymbol{u}_k^*\boldsymbol{v},\;\; k=s+1,s+2,\ldots,n,\quad \|\boldsymbol{\alpha}\|=1. \tag{3.8}$$

It follows at once from the definition of K that

$$\mu\boldsymbol{v}=K\boldsymbol{v}=\tfrac{1}{2}(A+HAH)\boldsymbol{v}=\tfrac{1}{2}(A\boldsymbol{v}+HA\boldsymbol{v})=A\boldsymbol{v}-\sum_{l=1}^{s}(\boldsymbol{u}_l^*A\boldsymbol{v})\boldsymbol{u}_l.$$

Therefore

$$\boldsymbol{v}=(A-\mu I)^{-1}\sum_{l=1}^{s}(\boldsymbol{u}_l^*A\boldsymbol{v})\boldsymbol{u}_l.$$

Comparison with (3.8) yields

$$\sum_{l=s+1}^{n}\alpha_l\boldsymbol{u}_l=(A-\mu I)^{-1}\sum_{l=1}^{s}\beta_l\boldsymbol{u}_l,$$

where $\beta_l=\boldsymbol{u}_l^*A\boldsymbol{v}$. Multiplying with $\boldsymbol{u}_k^*$, $k=1,2,\ldots,s$, on the left, $\boldsymbol{v}\in\mathcal{U}^\perp$ implies that

$$0=\boldsymbol{u}_k^*\boldsymbol{v}=\sum_{l=1}^{s}\beta_l\boldsymbol{u}_k^*(A-\mu I)^{-1}\boldsymbol{u}_l,\qquad k=1,2,\ldots,s.$$

Letting $\Phi_{k,l} = \boldsymbol{u}_k^*(A-\mu I)^{-1}\boldsymbol{u}_l$, $k,l = 1,2,\ldots,s$, we thus deduce $\Phi\boldsymbol{\beta} = \boldsymbol{0}$, consequently $\boldsymbol{\beta}^*\Phi\boldsymbol{\beta} = 0$. Written in longhand, this is equivalent to

$$\sum_{k=1}^{s}\sum_{l=1}^{s}\bar{\beta}_k\beta_l\boldsymbol{u}_k^*(A-\mu I)^{-1}\boldsymbol{u}_l = 0.$$

Therefore

$$\boldsymbol{z}^*(A-\mu I)^{-1}\boldsymbol{z} = 0, \qquad \text{where} \qquad \boldsymbol{z} = \sum_{k=1}^{s}\beta_k\boldsymbol{u}_k.$$

If $\boldsymbol{z} = \boldsymbol{0}$ then $A\boldsymbol{v} \in \mathcal{K}_n$, a possibility that we have already ruled out. Therefore $\boldsymbol{z} \neq \boldsymbol{0}$.

We expand $\boldsymbol{z}$ in the eigenvectors of A,

$$\boldsymbol{z} = \sum_{k=1}^{n}\zeta_k\boldsymbol{x}_k.$$

Therefore

$$0 = \boldsymbol{z}^*(A-\mu I)^{-1}\boldsymbol{z} = \sum_{k=1}^{n}\frac{|\zeta_k|^2}{\lambda_k-\mu},$$

and this proves (3.7). □

Lemma 3.4 comes into its own when A is Hermitian, when also, trivially, K is Hermitian, in which case we recover some known results in linear algebra, namely the *Weyl theorem* on eigenvalues of a sum of Hermitian matrices and the *interlacing eigenvalues theorem for bordered matrices* [3].

Let us assume that $s = 1$, thus, that H is a Householder reflection. Ordering the eigenvalues of A as $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ and those of K as $\mu_1 \le \mu_2 \le \cdots \le \mu_n$, we already know from Theorem 3.2 that $\mu_1,\mu_2,\ldots,\mu_n \in [\lambda_1,\lambda_n]$. Assume further that the spectrum of A is distinct, $\sigma(A)\cap\sigma(K) = \emptyset$ and that all the ζ_ks in Lemma 3.4 are nonzero – in other words, that $\boldsymbol{u}_1$ is not orthogonal to an eigenvector of A.

The function ψ from (3.7), being a rational function of type $(n-1)/n$, has exactly $n-1$ real zeros for distinct λ_ks. Since $\dim\mathcal{K}_n = n-1$, it follows that *all* its zeros are also eigenvalues of K. It is a trivial observation, though, that ψ changes sign in every interval of the form $(\lambda_k,\lambda_{k+1})$, $k = 1,2,\ldots,n-1$. Therefore, there must be at least one μ_l in each interval of this form and a trivial counting argument demonstrates that there must be a single μ_l is each interval, except for one interval that encloses two μ_ls. In other words, there exists $p \in \{1,2,\ldots,n-1\}$ such that

$$\begin{aligned} &\mu_k \in (\lambda_k,\lambda_{k+1}), \qquad k = 1,2,\ldots,p, \\ &\mu_k \in (\lambda_{k-1},\lambda_k), \qquad k = p+1,p+2,\ldots,n. \end{aligned} \tag{3.9}$$

This ‘almost interlace’ property is illustrated in Fig. 3.4.

The restrictive conditions (distinct eigenvalues, $\sigma(A)\cap\sigma(K) = \emptyset$ and $\boldsymbol{u}_1^\top\boldsymbol{x}_k \neq 0$) can be all removed by a limiting argument, except that open intervals in (3.9) need be replaced by closed intervals.

The situation is more complicated for $s \ge 2$. Thus, for example, for $s = 2$ similar argument implies that there must be a μ_l, corresponding to a $\boldsymbol{v} \in \mathcal{K}_n$, in exactly $n-2$ intervals of the form $(\lambda_k,\lambda_{k+1})$. Moreover, eigenvectors in $\mathcal{P}_n$ contribute two extra μ_ks, which can reside anywhere in $[\lambda_1,\lambda_n]$.

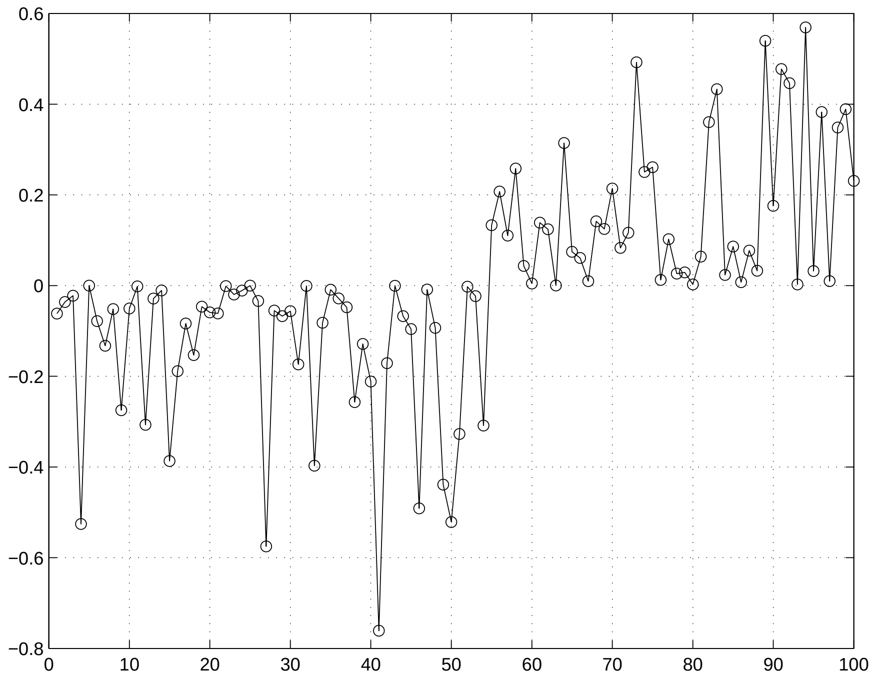

FIGURE 3.4. *The sequence* $\boldsymbol{\lambda} - \boldsymbol{\mu}$ *for a random* 100×100 *Hermitian matrix* A.

4. The non-normal case. Since $K\boldsymbol{v} = \mu\boldsymbol{v}$, $\|\boldsymbol{v}\| = 1$, implies in the proof of Theorem 3.2 that $\mu = \boldsymbol{v}^* A\boldsymbol{v}$, we deduce that $\mu \in \mathrm{W}(A)$, where $\mathrm{W}(B)$ is the *numerical range* or *field of values* of a matrix $B \in \mathrm{M}_n[\mathbb{C}]$ [4], and that $\sigma(K)$ lies in a closed ball of radius $\|A\|$. This, however, falls short of establishing a connection between $\sigma(K)$ and $\sigma(A)$ for a general (rather than normal) $A \in \mathrm{M}_n[\mathbb{C}]$. In this section we demonstrate that normalcy, although used just once in the proof of Theorem 3.2, is not an artifact of the method of proof: it is possible to find non-normal matrices A for which $\sigma(K) \not\subset \mathrm{conv}\,\sigma(A)$.

Specifically, we let

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & 0 & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} = \sum_{k=1}^{n-1} \boldsymbol{e}_k \boldsymbol{e}_{k+1}^\top \in \mathrm{M}_n[\mathbb{R}]. \tag{4.1}$$

Our first observation is that $\|A\| = 1$, therefore $|\mu_k| \le 1$, $k = 1, 2, \ldots, n$. (We retain the notation from Section 3.)

Let $H = I - 2\boldsymbol{u}\boldsymbol{u}^*$, $\|\boldsymbol{u}\| = 1$, therefore $s = 1$. We choose $\boldsymbol{v} \in \mathcal{K}_n$ and suppose first that $A\boldsymbol{v} \in \mathcal{K}_n$. As apparent from the proof of Theorem 3.2, this implies $A\boldsymbol{v} = \mu\boldsymbol{v}$, therefore $\mu = 0$ and $\boldsymbol{v} = \pm\boldsymbol{e}_1$. Since $H\boldsymbol{e}_1 = \boldsymbol{e}_1 - 2\bar{u}_1\boldsymbol{u}$, this takes place if and only if $u_1 = 0$. As soon as we rule this out, μ is necessarily a zero of the function

$$\psi(x) = \boldsymbol{u}^*(A - xI)^{-1}\boldsymbol{u},$$

which we have defined in the proof of Lemma 3.4: this is true regardless of A being normal. Nilpotency of A implies that

$$\psi(x) = -\sum_{k=0}^{n-1} (\boldsymbol{u}^* A^k \boldsymbol{u}) x^{-k-1}. \tag{4.2}$$

Thus, we seek zeros away from the origin of the function

$$\varphi(x) = -x^n \psi(x) = \sum_{k=0}^{n-1} (\boldsymbol{u}^* A^k \boldsymbol{u}) x^{n-k-1}.$$

To concentrate of a specific example, let us choose $\boldsymbol{u} = n^{-1/2}\mathbf{1}$. Therefore $\boldsymbol{u}^\top A^k \boldsymbol{u} = (n-k)/n$ and

$$\varphi(x) = \frac{1}{n} \sum_{k=0}^{n-1} (k+1) x^k$$

yields

$$n(1-x)^2 \varphi(x) = n x^{n+1} - (n+1) x^n + 1. \tag{4.3}$$

Let $r\mathrm{e}^{\mathrm{i}\theta}$ be a zero of φ, $r \in (0,1]$. Then (4.3) implies that

$$r^{2n} = \frac{1}{|n+1-nr\mathrm{e}^{\mathrm{i}\theta}|^2} = \frac{1}{(n+1)^2 - 2n(n+1)r\cos\theta + n^2 r^2} \geq \frac{1}{[(1+r)n+1]^2}.$$

Therefore

$$r \geq \frac{1}{[(1+r)n+1]^{1/n}} \geq \frac{1}{(2n+1)^{1/n}}.$$

Since the one eigenvalue not covered by this analysis is $\boldsymbol{u}^* A \boldsymbol{u} = 1 - 1/n$, it follows that

$$\sigma(K) \subset \{z \in \mathbb{C} \,:\, (2n+1)^{-1/n} \leq |z| \leq 1 - 1/n\}.$$

As a matter of fact, it is possible to prove, with extra effort, that $\rho(K) = 1 - 1/n$. Yet, this is not necessary to the observation that $\sigma(K)$ extends well outside $\mathrm{conv}\sigma(A) = \{0\}$.

The case $\boldsymbol{u} = n^{-1/2}\mathbf{1}$ is generic, in the following sense. Whenever $s = 1$, necessarily $\boldsymbol{u}^* A \boldsymbol{u} \in \sigma(K)$, since $\boldsymbol{u}$ spans the one-dimensional linear space $\mathcal{P}_n$. Therefore, unless $\boldsymbol{u}^* A \boldsymbol{u} = \sum_{k=1}^{n-1} \bar{u}_k u_{k+1} = 0$, it is true that $\sigma(K)$ contains points outside the origin and the inclusion (3.3) does not hold.

Acknowledgments. A number of colleagues was generous in their advice on the subject matter of this paper and we are happy to acknowledge discussions with Brad Baxter (Birkbeck College, University of London), John Butcher (University of Auckland) and Ernst Hairer (Université de Genevé).

REFERENCES

[1] E. Celledoni and A. Iserles, *Approximating the exponential from a Lie algebra to a Lie group*, Math. Comp., 69 (2000), pp. 1457–1480.

[2] ———, *Methods for the approximation of the matrix exponential in a Lie-algebraic setting*, IMA J. Num. Anal., 21 (2001), pp. 463–488.

[3] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[4] ———, *Topics in Matrix Analysis*, Cambridge University Press, 1991.

[5] A. Iserles, H. Munthe-Kaas, S. P. Nørsett, and A. Zanna, *Lie-group methods*, Acta Numerica, 9 (2000), pp. 215–365.

[6] A. Iserles and A. Zanna, *Efficient computation of the matrix exponential by generalized polar decompositions*, Tech. Rep. NA2002/09, University of Cambridge, 2002.

[7] F. Kang and Z.-J. Shang, *Volume-preserving algorithms or source-free dynamical systems*, Numer. Math., 71 (1995), pp. 451–463.

[8] H. Munthe-Kaas, R. G. W. Quispel, and A. Zanna, *Generalized polar decompositions on Lie groups with involutive automorphisms*, Found. Comp. Math., 1 (2001), pp. 297–324.