A structure preserving Sort-Jacobi algorithm for computing eigenvalues or singular values is presented. It applies to an arbitrary semisimple Lie algebra on its (-1)-eigenspace of the Cartan involution. Local quadratic convergence for arbitrary cyclic schemes is shown for the regular case. The proposed method is independent of the representation of the underlying Lie algebra and generalizes well-known normal form problems such as e.g. the symmetric, Hermitian, skew-symmetric, symmetric and skew-symmetric $\mathbb{R}$-Hamiltonian eigenvalue problem and the singular value decomposition.

A Sort-Jacobi Algorithm for Semisimple Lie Algebras

M. Kleinsteuber

1. Introduction

Since its introduction in 1846, [24], variants of the Jacobi algorithm have been applied to many structured eigenvalue problems (EVP), including e.g. the real skew-symmetric eigenvalue decomposition (EVD), [?, ?, ?], computations of the singular value decomposition (SVD) [28], non-symmetric EVPs [4,8,34,37], complex symmetric EVPs [9], and the computation of eigenvalues of normal matrices [13]. For extensions to different types of generalized EVPs, we refer to [5,38]. For applications of Jacobi–type methods to problems in systems theory, see [18,?, ?]. We also refer to [3] for an extensive list of structured eigenvalue problems and relevant literature.

In contrast to earlier work, we extend the classical concept of a Jacobi-algorithm towards a unified Lie algebraic approach to structured EVDs, where structure of a matrix is defined by being an element of a Lie algebra (or of a suitably defined sub-structure).

To the best of the author's knowledge, Wildberger [39] has been the first who proposed a generalization of the non-cyclic classical Jacobi algorithm on arbitrary compact Lie algebras and succeeded in proving global convergence of the algorithm. The well–known classification of compact Lie algebras shows that this approach essentially includes structure preserving Jacobi-type methods for the real skew-symmetric, the complex skew-Hermitian, the real skew-symmetric Hamiltonian, the complex skew-Hermitian Hamiltonian eigenvalue problem, and some exceptional cases.

Wildberger's work has been subsequently extended by Kleinsteuber et al. [25], where local quadratic convergence for general cyclic Jacobi schemes is shown to hold for arbitrary regular elements in a compact Lie algebra.

Explicitly, our setting here is the following. Let G be a semisimple Lie group and $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be the Cartan decomposition of its Lie algebra. We propose a Jacobi–type method that "diagonalizes" an element $S \in \mathfrak{p}$ by conjugation with some $k \in K$, where $K \subset G$ is the Lie subgroup corresponding to $\mathfrak{k}$. To see that the analysis of Jacobi-type methods on compact Lie algebras $\mathfrak{k}$ appears as a special case of our result, note that $\mathfrak{k}^{\mathbb{C}} := \mathfrak{k} \oplus i\mathfrak{k}$ is the Cartan decomposition of $\mathfrak{k}^{\mathbb{C}}$, the complexification of $\mathfrak{k}$.

A characteristic feature of all known Jacobi-type methods is that they act to minimize the distance to diagonality while preserving the eigenvalues. Conventional Jacobi algorithms, like cyclic Jacobi algorithms for symmetric matrices, Kogbetliantz's method for the SVD, cf. [28], methods for the skew-symmetric EVD, cf. [15], and for the Hamiltonian EVD, cf. [29], as well as recent methods for the perplectic EVD, cf. [30], are all based on reducing the sum of squares of off-diagonal entries (the so-called off-norm). It is well known to the numerical linear algebra community, cf. [23], [31], that sorting the diagonal elements after each step within the algorithm speeds up the convergence, yet the off-norm does not take this sorting into account. Moreover, the off-norm function that is to be minimized has a complicated critical point structure and several global minima, what makes the analysis of such *non-sorting* Jacobi methods considerably more complicated if clustered eigenvalues occur.

Both difficulties might be remedied by a better choice of cost function that measures the distance to diagonality. In fact, Brockett's trace function turns out to be a more appropriate distance measure than the off-norm function. In [2], R.W. Brockett showed that the gradient flow of the trace function can be used to diagonalize a symmetric matrix and simultaneously sort the eigenvalues. This trace function has also been considered by e.g. M.T. Chu, [6] associated with gradient methods for matrix factorizations, and subsequently by many others. For a systematic critical point analysis of the trace function in a Lie group setting, we refer to [7]. See also [35] for more recent results on this topic in the framework of reductive Lie groups.

In this paper we propose and analyze the Sort-Jacobi algorithm for a large class of structured matrices that, besides the compact Lie algebra cases, essentially includes the normal form problems quoted in Table 1, cf. [27] for the defining representations of the corresponding Lie algebras. These cases arise from the well-known classification of simple Lie algebras. Cyclic Jacobi-type methods for some of these cases have been discussed earlier, e.g. for the symmetric/Hermitian EVD see [11,23], for the skew-symmetric EVD see [15,23,33], for the real and complex SVD see [23,28], for the real symmetric and skew-symmetric EVD see [10], for the Hermitian $\mathbb{R}$-Hamiltonian EVD see [4], and for the perplectic EVD see [30]. Note that the methods proposed in this paper exclusively use one-parameter transformations and therefore slightly differ from the algorithms in [4,10,15,23,33], where block Jacobi methods are used, i.e. multiparameter transformations that annihilate more than one pair of off-diagonal elements at the same time.

The local convergence behavior of the Sort-Jacobi algorithm for the above mentioned classes is examined and local quadratic convergence is proved for the regular case, independent of any cyclic scheme. A local convergence analysis for the irregular case, i.e. where eigenvalues/singular values occur in clusters is more subtle and will be the matter of a subsequent publication.

This paper is organized as follows. In Section 2 we discuss a Lie algebraic version of the aforementioned trace function and propose the Sort-Jacobi algorithm in full generality. Local quadratic convergence is proven in Section 3 for the regular case. The Sort-Jacobi algorithm is exemplified for the case of the exceptional Lie algebra of derivations of the complex octonians in Section 4.

2. The Linear Trace Function and the Cyclic Sort-Jacobi Algorithm

Throughout this paper, $\mathfrak{g}$ denotes a real semisimple Lie algebra of matrices. In the case of complex semisimple Lie algebras, we consider their realification. Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be the Cartan decomposition of $\mathfrak{g}$ into $(+1)$- and (-1)-eigenspace of the corresponding Cartan involution θ. Denote by $GL(\mathfrak{g})$ the general linear group of $\mathfrak{g}$, let $\mathrm{End}(\mathfrak{g})$ denote the set of endomorphisms of $\mathfrak{g}$ and let $\mathrm{ad}(\mathfrak{g}) \subset \mathrm{End}(\mathfrak{g})$ be the adjoint representation of $\mathfrak{g}$. Let $\mathrm{Int}_{\mathfrak{g}}(\mathfrak{k})$ be the analytic subgroup of $GL(\mathfrak{g})$ with Lie algebra $\mathrm{ad}_{\mathfrak{g}}(\mathfrak{k}) \subset \mathrm{ad}(\mathfrak{g})$.

$\mathfrak{g}$	$\mathfrak{k}$	EVD/SVD
$\mathfrak{sl}(n,\mathbb{R})$	$\mathfrak{so}(n,\mathbb{R})$	symmetric EVD
$\mathfrak{sl}(n,\mathbb{C})$	$\mathfrak{su}(n)$	Hermitian EVD
$\mathfrak{so}(n,\mathbb{C})$	$\mathfrak{so}(n,\mathbb{R})$	skew-symmetric EVD
$\mathfrak{su}^*(2n)$	$\mathfrak{sp}(n)$	Hermitian Quaternion EVD, i.e. of $\begin{bmatrix} S & \Psi^* \\ \Psi & \overline{S} \end{bmatrix}, S,\Psi \in \mathbb{C}^{n\times n}, \mathrm{tr}S = 0, S = S^*, \Psi = -\Psi^\top$
$\mathfrak{so}(p,q)$	$\mathfrak{so}(p,\mathbb{R}) \oplus \mathfrak{so}(q,\mathbb{R})$	real SVD, skew-sym. persymmetric EVD
$\mathfrak{su}(p,q)$	$\mathfrak{s}(\mathfrak{u}(p) \oplus \mathfrak{u}(q))$	complex SVD, Hermitian $\mathbb{R}$-Hamiltonian EVD
$\mathfrak{so}^*(2n)$	$\mathfrak{u}(n)$	skew-Takagi factorization, i.e. SVD of $B = -B^\top \in \mathbb{C}^{n\times n}$
$\mathfrak{sp}(n,\mathbb{R})$	$\mathfrak{u}(n)$	symmetric Hamiltonian EVD, Takagi factorization
$\mathfrak{sp}(n,\mathbb{C})$	$\mathfrak{sp}(n)$	Hermitian $\mathbb{C}$-Hamiltonian EVD
$\mathfrak{sp}(p,q)$, $p \geq q$	$\mathfrak{sp}(p) \oplus \mathfrak{sp}(q)$	(p,q)-Hamiltonian SVD, i.e. SVD of $\begin{bmatrix} B & -\overline{F} \\ F & \overline{B} \end{bmatrix}$, $B, F \in \mathbb{C}^{p\times q}$
$\mathfrak{g}_2$	$\mathfrak{su}(2) \oplus \mathfrak{su}(2)$	cf. Section 4

Table 1
Cartan-decompositions of simple Lie algebras and corresponding matrix factorizations.

Following the idea in [22] and [25], the Jacobi algorithm is formulated as an optimization task on the $\mathrm{Int}_{\mathfrak{g}}(\mathfrak{k})$-*adjoint orbit* of an element $S \in \mathfrak{p}$, i.e.

$$\mathcal{O}(S) = \{\varphi(S) \mid \varphi \in \mathrm{Int}_{\mathfrak{g}}(\mathfrak{k})\}. \tag{2.1}$$

Note, that $\mathcal{O}(S) \subset \mathfrak{p}$. In a next step, the cost function is concretized to a Lie algebraic version of the linear trace function considered by Brockett [2]. The resulting algorithm is the Sort-Jacobi Algorithm.

2.1. Givens Rotations

The aim of this subsection is to find a suitable Lie algebraic generalization of Givens rotations, cf. [14], 5.1.8. This requires the concept of the restricted root space decomposition of $\mathfrak{g}$.

Let $\mathfrak{a} \subset \mathfrak{p}$ be a maximal Abelian subalgebra. We denote by $\mathfrak{a}^*$ the dual space of $\mathfrak{a}$. For $\lambda \in \mathfrak{a}^*$, let

$$\mathfrak{g}_\lambda := \{X \in \mathfrak{g} \mid \mathrm{ad}_H X = \lambda(H)X \text{ for all } H \in \mathfrak{a}\}. \tag{2.2}$$

If $\lambda \neq 0$ and $\mathfrak{g}_\lambda \neq 0$, the vector space $\mathfrak{g}_\lambda$ is called *restricted-root space* and λ is called *restricted root*. The set of restricted roots is denoted by Σ. A vector $X \in \mathfrak{g}_\lambda$ is called a *restricted-root vector*. The following result will proof to be useful. Note, that

$$B_\theta \colon \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}, \quad (X,Y) \mapsto -\kappa(X,\theta Y), \tag{2.3}$$

where κ is the Killing form on $\mathfrak{g}$, admits a scalar product. Moreover, $B_\theta\big|_{\mathfrak{k}\times\mathfrak{k}} = -\kappa\big|_{\mathfrak{k}\times\mathfrak{k}}$ and $B_\theta\big|_{\mathfrak{p}\times\mathfrak{p}} = \kappa\big|_{\mathfrak{p}\times\mathfrak{p}}$.

Theorem 2.1 (Restricted-root space decomposition)**.** *The real semisimple Lie algebra* $\mathfrak{g}$ *decomposes orthogonally with respect to* B_θ *into*

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \sum_{\lambda \in \Sigma} \mathfrak{g}_\lambda \tag{2.4}$$

and $\mathfrak{g}_0 = \mathfrak{z}_{\mathfrak{k}}(\mathfrak{a}) \oplus \mathfrak{a}$, *where* $\mathfrak{z}_{\mathfrak{k}}(\mathfrak{a}) := \{X \in \mathfrak{k} \mid [X, H] = 0 \text{ for all } H \in \mathfrak{a}\}$ *denotes the centralizer of* $\mathfrak{a}$ *in* $\mathfrak{k}$*. Furthermore, for* $\lambda, \mu \in \Sigma \cup \{0\}$*, we have*

(a) $[\mathfrak{g}_\lambda, \mathfrak{g}_\mu] \subseteq \begin{cases} \mathfrak{g}_{\lambda+\mu} & \text{if } \lambda + \mu \in \Sigma \cup \{0\} \\ 0 & \text{else.} \end{cases}$

(b) $\theta(\mathfrak{g}_\lambda) = \mathfrak{g}_{-\lambda}$*, and hence* $\lambda \in \Sigma \iff -\lambda \in \Sigma$*.*

Proof. Cf. [27], Ch. VI, Prop. 6.40. Similar to the above, it is possible to decompose complex semisimple Lie algebras into a maximal Abelian subalgebra and the so called *root spaces* (in contrast to *restricted*-root spaces), cf. [27], Sec. 1, Ch. II. In this context, the term *Cartan subalgebra* arises. Note, that although a Cartan subalgebra is related to the maximal Abelian subspace $\mathfrak{a} \subset \mathfrak{p}$, they do not coincide. A further investigation is not relevant for our purposes and we refer to the literature. The word *restricted* roots is due to the fact, that they are the nonzero restrictions to $\mathfrak{a}$ of the (ordinary) roots of the complexification $\mathfrak{g}^{\mathbb{C}}$. Cf. [27], Ch. VI, Prop. 6.47 and the subsequent remark. Note that the restricted-root space decomposition can be equivalently computed via the eigenspaces of a single operator ad_H for a generic element $H \in \mathfrak{a}$ with pairwise distinct roots. Such elements are dense in $\mathfrak{a}$ since they are obtained by omitting from $\mathfrak{a}$ the finitely many hyperplanes $\{H \in \mathfrak{a} \mid \lambda_i(H) - \lambda_j(H) = 0\}$, $\lambda_i \neq \lambda_j$.

Let λ be a restricted root and denote by $H_\lambda \in \mathfrak{a}$ its dual, i.e.

$$\lambda(H) = \kappa(H_\lambda, H) \quad \text{for all } H \in \mathfrak{a}. \tag{2.5}$$

By [27], Ch. VI, Prop. 6.52, every nonzero $E_\lambda \in \mathfrak{g}_\lambda$ can be normalized such that

$$T_\lambda := [E_\lambda, \theta(E_\lambda)] = -\frac{2}{|\lambda|^2} H_\lambda. \tag{2.6}$$

Furthermore, let

$$\Omega_\lambda := E_\lambda + \theta(E_\lambda) \quad \text{and} \quad \overline{\Omega}_\lambda := E_\lambda - \theta(E_\lambda). \tag{2.7}$$

Note that $\Omega_\lambda \in \mathfrak{k}$, $\overline{\Omega}_\lambda \in \mathfrak{p}$, $\|\overline{\Omega}_\lambda\|^2 = \|\Omega_\lambda\|^2 = \frac{2}{|\lambda|^2}$ and $\lambda(T_\lambda) = -2$. We call $\exp(t\Omega_\lambda)$ a *Givens-rotation*. The following example justifies this definition that generalizes (5.1.7) in [14].

Example 2.2. Consider the case where $\mathfrak{g} := \mathfrak{sl}(n, \mathbb{R})$ and the Cartan involution yields the decomposition into skew symmetric and symmetric matrices with the diagonal matrices as the maximal Abelian subalgebra. Denote by X_{ij} the (i, j)-entry of the matrix X. Then the roots are given by

$$\lambda_{ij}(H) = H_{ii} - H_{jj}, \quad i \neq j.$$

Recall that the Killing form κ is given by $\kappa(X,Y) = 2n\mathrm{tr}(XY)$. Therefore,

$$H_{\lambda_{ij}} = \frac{1}{2n}(e_i e_i^\top - e_j e_j^\top)$$

where e_i denotes the i-th standard basis vector and

$$\lambda_{ij}(H_{\lambda_{ij}}) = |\lambda_{ij}|^2 = \frac{1}{n}.$$

Hence $T_{\lambda_{ij}} = -e_i e_i^\top + e_j e_j^\top$ and $E_{\lambda_{ij}} = \pm e_i e_j^\top$. Depending on the choice of $E_{\lambda_{ij}}$, either $\Omega_{\lambda_{ij}} = e_i e_j^\top - e_j e_i^\top$ or $\Omega_{\lambda_{ij}} = e_j e_i^\top - e_i e_j^\top$.

2.2. Cyclic Jacobi on $\mathcal{O}(S)$

We introduce a notion of positivity on $\mathfrak{a}^*\setminus\{0\}$. A subset $\mathcal{P}$ of $\mathfrak{a}^*\setminus\{0\}$ consists of positive elements if for any $l \in \mathfrak{a}^*\setminus\{0\}$ exactly one of l and $-l$ is in $\mathcal{P}$ and the sum and any positive multiple of elements in $\mathcal{P}$ is again in $\mathcal{P}$. We denote the set of positive restricted roots by Σ^+. Theorem 2.1 assures that $\lambda \in \Sigma$ if and only if $-\lambda \in \Sigma$ and, moreover, that Σ is finite. Thus a set of positive roots is obtained by a hyperplane through the origin in $\mathfrak{a}^*$ that does not contain any root and defining all roots on one side to be positive. Hence partitioning Σ into $\Sigma^+ \cup \Sigma^-$, where $\Sigma^- := \Sigma \setminus \Sigma^+$ is the set of *negative roots*, is not unique. Now let

$$\mathfrak{k}_\lambda := \{X + \theta(X) \mid X \in \mathfrak{g}_\lambda\} \tag{2.8}$$

denote the orthogonal projection of $\mathfrak{g}_\lambda$ onto $\mathfrak{k}$. If $\{E^{(1)}, \dots, E^{(l)}\}$ is a basis of $\mathfrak{g}_\lambda$ with all $E^{(i)}$ normalized as in (2.6), then

$$\mathcal{B}_\lambda := \{E^{(i)} + \theta(E^{(i)}) \mid i = 1, \dots, l\}$$

is an orthogonal basis of $\mathfrak{k}_\lambda$, normalized in terms of Eq. (2.7). The union of these basis over all $\lambda \in \Sigma^+$ yields an orthogonal basis of

$$\sum_{\lambda\in\Sigma^+} \mathfrak{k}_\lambda = \mathfrak{g}_0^\perp \cap \mathfrak{k} \tag{2.9}$$

and will further be denoted by

$$\mathcal{B} := \{\Omega_1, \dots, \Omega_m\} := \bigcup_{\lambda\in\Sigma^+} \mathcal{B}_\lambda. \tag{2.10}$$

For $Y \in \mathfrak{g}$, we denote the *adjoint action* of $\exp(Y)$ by $\mathrm{Ad}_{\exp Y}$. Note, that

$$\mathrm{Ad}_{\exp Y}(X) = \exp(\mathrm{ad}_Y)(X) = \exp(Y)X\exp(-Y) \quad \text{for all } X \in \mathfrak{g}. \tag{2.11}$$

We are now ready to explain a Jacobi Sweep for maximizing a function f on the $\mathrm{Int}_{\mathfrak{g}}(\mathfrak{k})$ orbit of $S \in \mathfrak{p}$. Note, that a minimization task is analogously defined.

Cyclic Jacobi Sweep. Let f be some real valued function on $\mathcal{O}(S)$. Define for $\Omega \in \mathcal{B}$ the search directions

$$r_\Omega\colon\ \mathbb{R}\times\mathcal{O}(S) \longrightarrow \mathcal{O}(S), \quad (t, X) \longmapsto \mathrm{Ad}_{\exp t\Omega} X, \tag{2.12}$$

and let the *step-size* $t_*^{(i)}(X)$ be defined as the local maximum of $f \circ r_{\Omega_i}(X,t)$ with smallest absolute value. To achieve uniqueness, we choose $t_*^{(i)}(X)$ to be nonnegative if $t_*^{(i)}(X)$ as well as $-t_*^{(i)}(X)$ fulfill this condition. Note, that $t_*^{(i)}(X)$ is well defined since $f \circ r_\Omega(X,t)$ is periodic. This follows directly from the following Lemma.

Lemma 2.3. *The one–parameter subgroups* $\varphi_\lambda : \mathbb{R} \to \mathrm{Int}_{\mathfrak{g}}(\mathfrak{k}),\ t \mapsto \mathrm{Ad}_{\exp t\Omega_\lambda}$ *are isomorphic to the circle* $S^1 := \{\mathrm{e}^{\mathrm{i}t} \mid t \in \mathbb{R}\}$.

Proof. Let $\mathfrak{u} = \mathfrak{k} \oplus \mathrm{i}\mathfrak{p}$ be the compact real form of the complexification of $\mathfrak{g}$ and denote $U := \mathrm{Int}(\mathfrak{u})$ the inner automorphims of $\mathfrak{u}$. Correspondingly,

$$\mathfrak{s} := \langle \Omega_\lambda, \mathrm{i}\overline{\Omega}_\lambda, \mathrm{i}T_\lambda \rangle_{LA} \subset \mathfrak{u}$$

is the compact real form of the complexification of the Lie algebra $\langle \Omega_\lambda, \overline{\Omega}_\lambda, T_\lambda \rangle_{LA}$. Consider now the closure of $\varphi_\lambda(\mathbb{R})$ in $\mathrm{Int}(\mathfrak{g})$. Since $\mathrm{Int}_{\mathfrak{g}}(\mathfrak{k})$ is compact, $\overline{\varphi_\lambda(\mathbb{R})} \subset \mathrm{Int}_{\mathfrak{g}}(\mathfrak{k})$. Moreover, $\mathrm{Int}_{\mathfrak{g}}(\mathfrak{k})$ is a closed subset of U and hence $\varphi_\lambda(\mathbb{R})$ is closed in $\mathrm{Int}_{\mathfrak{g}}(\mathfrak{k})$ if and only if it is closed in U. It is easily seen, that the Lie algebra $\mathfrak{s}$ is isomorphic to $\mathfrak{su}(2)$ and hence so is $\mathrm{ad}(\mathfrak{s})$. The analytic subgroup $S \subset U$ with Lie algebra $\mathrm{ad}(\mathfrak{s})$ is closed in U, because U is compact and $\mathrm{ad}(\mathfrak{s})$ is semisimple, cf. [32], Corollary 2. Therefore the closure $\overline{\varphi_\lambda(\mathbb{R})}$ is contained in S. Since every compact Abelian analytic Lie group is a torus, cf. [27], Ch. I.12, Corollary 1.103,

$$\overline{\varphi_\lambda(\mathbb{R})} = S^1 \times \cdots \times S^1.$$

On the other hand, for dimensional reasons, the only torus contained in S is S^1, so $\overline{\varphi_\lambda(\mathbb{R})} = S^1$. Therefore

$$\overline{\varphi_\lambda(\mathbb{R})} = \varphi_\lambda(\mathbb{R}),$$

since both Lie groups are connected and have the same Lie algebra and are therefore identical, cf. [17], Ch. II, Thm. 2.1. Thus $\varphi_\lambda(\mathbb{R}) = S^1$. □

A *sweep* on $\mathcal{O}(S)$ is the map

$$s\colon \mathcal{O}(S) \longrightarrow \mathcal{O}(S), \tag{2.13}$$

explicitly given as follows. Set $X_k^{(0)} := X \in \mathcal{O}(S)$.

$$\begin{aligned}
X_k^{(1)} &:= r_{\Omega_1}\left(t_*^{(1)}\left(X_k^{(0)}\right), X_k^{(0)}\right)\\
X_k^{(2)} &:= r_{\Omega_2}\left(t_*^{(2)}\left(X_k^{(1)}\right), X_k^{(1)}\right)\\
X_k^{(3)} &:= r_{\Omega_3}\left(t_*^{(3)}\left(X_k^{(2)}\right), X_k^{(2)}\right)\\
&\ \ \vdots\\
X_k^{(m)} &:= r_{\Omega_m}\left(t_*^{(m)}\left(X_k^{(m-1)}\right), X_k^{(m-1)}\right),
\end{aligned}$$

and set $s(X) := X_k^{(m)}$. The **Jacobi algorithm** consists of iterating sweeps:

1. Assume that we already have $X_0, X_1, \ldots, X_k \in \mathcal{O}(S)$ for some $k \in \mathbb{N}$.
2. Put $X_{k+1} := s(X_k)$ and continue with the next sweep.

Note, that by construction, a Jacobi sweep does not work in directions $\Omega \in \mathfrak{z}_{\mathfrak{k}}(\mathfrak{a})$.

Although the cost function has not been specified yet, some remarks for the comparison with the Jacobi algorithm for diagonalizing symmetric matrices are in order. If the above algorithm is intended for minimizing the off-norm function

$$\mathrm{off}(X) = \|X - X_0\|^2, \tag{2.14}$$

where X_0 denotes the orthogonal projection of X onto $\mathfrak{a}$, then, following example 2.2, it generalizes the well known cyclic Jacobi algorithm for symmetric EVP. In that case, $t_*^{(i)}(X)$ is the smallest angle such that the corresponding Givens rotation diagonalizes the corresponding (2×2)-subproblem.

If the above algorithm is intended for maximizing the trace function, which is defined below, it generalizes the cyclic Sort-Jacobi method for the symmetric EVP, cf. [23] and [31].

2.3. The trace function

Let $N \in \mathfrak{a}$ with $\lambda(N) < 0$ for all $\lambda \in \Sigma^+$. Our goal now is to minimize the distance function

$$\mathcal{O}(S) \longrightarrow \mathbb{R}, \quad X \longrightarrow B_\theta(X - N, X - N). \tag{2.15}$$

This simplifies to

$$\begin{aligned} B_\theta(X - N, X - N) &= \kappa(X - N, X - N) = \kappa(X, X) + \kappa(N, N) - 2\kappa(X, N) \\ &= \kappa(S, S) + \kappa(N, N) - 2\kappa(X, N) \end{aligned}$$

because of $B_\theta|_\mathfrak{p} = \kappa|_\mathfrak{p}$ and the Ad-invariance of κ. Minimizing the function defined in Eq. (2.15) is therefore equivalent to maximizing the following function.

Definition 2.4. *Let $S \in \mathfrak{p}$ and let κ denote the Killing form. The* trace function *is given by*

$$f\colon\ \mathcal{O}(S) \longrightarrow \mathbb{R}, \quad X \longmapsto \kappa(X, N). \tag{2.16}$$

Proposition 2.5. *(a) X is a critical point of the trace function* (2.16) *if and only if $X \in \mathfrak{a}$. In particular, there are only finitely many critical points.*

(b) The trace function (2.16) *has exactly one maximum, say Z, and one minimum, say $\widetilde{Z}$, and $\lambda(Z) \leq 0$, $\lambda(\widetilde{Z}) \geq 0$ for all $\lambda \in \Sigma^+$.*

Proof. (a) To compute the critical points, let $\Omega \in \mathfrak{k}$ and denote by $\xi = \mathrm{ad}_\Omega X$ an arbitrary tangent vector in $T_X\mathcal{O}(S)$. The ad-invariance of κ yields

$$Df(X)\xi = \kappa(\xi, N) = \kappa(\mathrm{ad}_\Omega X, N) = \kappa(\Omega, \mathrm{ad}_X N).$$

The Killing form is negative definite on $\mathfrak{k}$ and hence

$$Df(X) = 0 \quad \Longleftrightarrow \quad [X, N] = 0.$$

Since N is a regular element, it follows $X \in \mathfrak{a}$, cf. [27], Sec. VI., Lemma 6.50. Now $|\mathcal{O}(S) \cap \mathfrak{a}|$ is finite, cf. [17], Ch. VII, Thm. 2.12, and hence f has only finitely many critical points.

(b) We compute the Hessian H_f at the critical points X. Again, let $\xi = \mathrm{ad}_\Omega X$ be tangent to $\mathcal{O}(S)$ at X. Decompose $\Omega \in \mathfrak{k}$ according to Eq. (2.10) into

$$\Omega = \Omega_0 + \sum_{i=1}^{m} d_i \Omega_i,$$

where $\Omega_i \in \mathcal{B}$, $\Omega_0 \in \mathfrak{z}_{\mathfrak{k}}(\mathfrak{a})$ and denote by λ_i the positive restricted root with $\Omega_i \in \mathfrak{k}_{\lambda_i}$. Then

$$\begin{aligned} \mathsf{H}_f(X)(\xi,\xi) &= \tfrac{\mathrm{d}^2}{\mathrm{d}t^2}\big|_{t=0} \kappa(\mathrm{Ad}_{\exp(t\Omega)} X, N) \\ &= \kappa(\mathrm{ad}_\Omega^2 X, N) = -\kappa(\mathrm{ad}_\Omega X, \mathrm{ad}_\Omega N) \\ &= -\sum_{i=1}^{m} \lambda_i(X)\, \lambda_i(N)\, \kappa(d_i \overline{\Omega}_i, d_i \overline{\Omega}_i) \qquad (2.17) \\ &= -\sum_{i=1}^{m} \lambda_i(X)\, \lambda_i(N)\, \frac{2d_i^2}{|\lambda_i|^2}. \end{aligned}$$

By assumption, $\lambda(N) < 0$ for all $\lambda \in \Sigma^+$, so a necessary condition for a local maximum Z is that $\lambda(Z) \leq 0$ for all $\lambda \in \Sigma^+$. The orbit $\mathcal{O}(S)$ intersects the closure of the Weyl chamber

$$C^- := \{H \in \mathfrak{a} \mid \lambda(H) < 0 \text{ for all } \lambda \in \Sigma^+\}$$

exactly once, cf. [17], Ch. VII, Thm. 2.12 and [26]. Hence Z is the only local maximum of the function and by compactness of $\mathcal{O}(S)$ it is the unique global maximum. A similar argument proves the existence of a unique minimum, having all positive roots greater or equal to zero. □

2.4. Explicit step-size selection

We restrict the trace function (2.16) to the orbits of one-parameter subgroups in order to explicitly compute the step size t_*. Let $\overline{\Omega}_\lambda$, Ω_λ and T_λ as in Eq. (2.7), $X \in \mathfrak{p}$. Let

$$\mathrm{p}\colon \mathfrak{g} \longrightarrow \mathfrak{g}_0, \quad \mathrm{p}(X) = X_0 \qquad (2.18)$$

denote the orthogonal projection with respect to B_θ. Similar to Eq. (2.8), define

$$\mathfrak{p}_\lambda := \{X - \theta(X) \mid X \in \mathfrak{g}_\lambda\} \qquad (2.19)$$

as the orthogonal projection of $\mathfrak{g}_\lambda$ onto $\mathfrak{p}$.

Theorem 2.6. *Let*

$$c_\lambda := \frac{\kappa(X, \overline{\Omega}_\lambda)}{\kappa(\overline{\Omega}_\lambda, \overline{\Omega}_\lambda)} \qquad (2.20)$$

be the $\overline{\Omega}_\lambda$*-coefficient of* X*. Then*

$$\mathrm{p}\left(\mathrm{Ad}_{\exp t\Omega_\lambda} X\right) = X_0 + g(t)T_\lambda,$$

where $X_0 := \mathrm{p}(X)$ *and* g *is given by*

$$g\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad t \longmapsto \tfrac{1}{2}\lambda(X_0)\Big(1-\cos(2t)\Big) - c_\lambda \sin(2t). \tag{2.21}$$

The proof of Theorem 2.6 is based on the following two lemmas.

Lemma 2.7. *The following identities hold for* $n \in \mathbb{N}_0$*.*

(a) $\mathrm{ad}_{\Omega_\lambda}^{2n+1}\overline{\Omega}_\lambda = (-1)^n 2^{2n+1}(-T_\lambda)$,

(b) $\mathrm{ad}_{\Omega_\lambda}^{2n} T_\lambda = (-1)^n 2^{2n} T_\lambda$.

Proof. (a) The first formula is shown by induction. For $n = 0$, we compute

$$[\Omega_\lambda, \overline{\Omega}_\lambda] = [E_\lambda, \theta(E_\lambda)] - [\theta(E_\lambda), E_\lambda] = 2[E_\lambda, \theta(E_\lambda)] = 2T_\lambda.$$

Assume now that it is true for $n \geq 0$. Then

$$\mathrm{ad}_{\Omega_\lambda}^{2n+3}\overline{\Omega}_\lambda = -2\mathrm{ad}_{\Omega_\lambda}^{2n+2} T_\lambda = -4\mathrm{ad}_{\Omega_\lambda}^{2n+1}\overline{\Omega}_\lambda = (-1)^{n+1}2^{2n+3}(-T_\lambda),$$

and the formula is shown for $n+1$.

(b) The second identity follows from (a) by a straightforward calculation. It is clearly true for $n = 0$. Now let $n \geq 1$. Then

$$\mathrm{ad}_{\Omega_\lambda}^{2n} T_\lambda = 2\mathrm{ad}_{\Omega_\lambda}^{2n-1}\overline{\Omega}_\lambda = 2(-1)^{n-1}2^{2n-1}(-T_\lambda) = (-1)^n 2^{2n} T_\lambda. \quad \square$$

Lemma 2.8. *Let* λ, μ *be positive restricted roots with* $\lambda \neq \mu$*. Then* $\mathrm{p}(\mathrm{ad}_{\Omega_\mu}^k \overline{\Omega}_\lambda) = 0$ *for all* $k \in \mathbb{N}$*.*

Proof. The proof is done by induction, separately for the even and the odd case. The assumption is clearly true for $n = 0$ and $n = 1$ by Theorem 2.1. Now let $H \in \mathfrak{a}$ be arbitrary. Then, by the induction hypothesis,

$$\begin{aligned}\kappa(\mathrm{ad}_{\Omega_\mu}^k \overline{\Omega}_\lambda, H) &= \mu(H)\kappa(\mathrm{ad}_{\Omega_\mu}^{k-1}\overline{\Omega}_\lambda, \overline{\Omega}_\mu) = \mu(H)\kappa(\mathrm{ad}_{\Omega_\mu}^{k-2}\overline{\Omega}_\lambda, [\overline{\Omega}_\mu, \Omega_\mu]) \\ &= \mu(H)\kappa(\mathrm{ad}_{\Omega_\mu}^{k-2}\overline{\Omega}_\lambda, 2T_\mu) = 0,\end{aligned}$$

since $T_\mu \in \mathfrak{a}$. This completes the proof. $\square$

Proof of Theorem 2.6. For all $t \in \mathbb{R}$ we have the identity

$$\mathrm{Ad}_{\exp t\Omega_\lambda} X = \exp(\mathrm{ad}_{t\Omega_\lambda})X = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathrm{ad}_{\Omega_\lambda}^k X. \tag{2.22}$$

It is shown that, if we decompose $X \in \mathfrak{p}$ into its $\mathfrak{p}_\lambda$-components then the only summands in Eq. (2.22) that affect the projection onto $\mathfrak{a}$ are X_0 and $c_\lambda\overline{\Omega}_\lambda$. First, assume that $\overline{\Omega}'_\lambda, \overline{\Omega}_\lambda \in \mathfrak{p}_\lambda$ and $\kappa(\overline{\Omega}_\lambda, \overline{\Omega}'_\lambda) = 0$. Then we have for all $H \in \mathfrak{a}$ that

$$0 = \lambda(H)\kappa(\overline{\Omega}_\lambda, \overline{\Omega}'_\lambda) = \kappa(\overline{\Omega}_\lambda, [H, \Omega'_\lambda]) = \kappa([\Omega'_\lambda, \overline{\Omega}_\lambda], H)$$

and hence $[\Omega'_\lambda, \overline{\Omega}_\lambda] \in \mathfrak{a}^\perp$. Therefore, Theorem 2.1 implies that $[\Omega'_\lambda, \overline{\Omega}_\lambda] \in \mathfrak{p}_{2\lambda}$ if $2\lambda \in \Sigma$ and is zero otherwise. We can apply Lemmas 2.7 and 2.8 and compute

$$\begin{aligned}
\mathrm{p}\left(\mathrm{Ad}_{\exp t\Omega_\lambda} X\right) &= \mathrm{p}\left(\sum_{k=0}^{\infty} \frac{t^k}{k!} \mathrm{ad}^k_{\Omega_\lambda} X\right) = \mathrm{p}\left(\sum_{k=0}^{\infty} \frac{t^k}{k!} \mathrm{ad}^k_{\Omega_\lambda} (X_0 + c_\lambda \overline{\Omega}_\lambda)\right) = \\
&= X_0 + \mathrm{p}\left(\sum_{k=1}^{\infty} \frac{t^k}{k!} \mathrm{ad}^{k-1}_{\Omega_\lambda} [\Omega_\lambda, X_0]\right) + c_\lambda \mathrm{p}\left(\sum_{k=1}^{\infty} \frac{t^k}{k!} \mathrm{ad}^k_{\Omega_\lambda} \overline{\Omega}_\lambda\right) = \\
&= X_0 - \lambda(X_0)\mathrm{p}\left(\sum_{k=1}^{\infty} \frac{t^k}{k!} \mathrm{ad}^{k-1}_{\Omega_\lambda} \overline{\Omega}_\lambda\right) + c_\lambda \mathrm{p}\left(\sum_{k=1}^{\infty} \frac{t^k}{k!} \mathrm{ad}^k_{\Omega_\lambda} \overline{\Omega}_\lambda\right) = \\
&= X_0 - \lambda(X_0) \sum_{k=0}^{\infty} \frac{t^{2k+2}}{(2k+2)!} \mathrm{ad}^{2k+1}_{\Omega_\lambda} \overline{\Omega}_\lambda + c_\lambda \sum_{k=0}^{\infty} \frac{t^{2k+1}}{(2k+1)!} \mathrm{ad}^{2k+1}_{\Omega_\lambda} \overline{\Omega}_\lambda.
\end{aligned}$$

Again by Lemma 2.7, the last sum simplifies to

$$\begin{aligned}
\sum_{k=0}^{\infty} \frac{t^{2k+1}}{(2k+1)!} \mathrm{ad}^{2k+1}_{\Omega_\lambda} \overline{\Omega}_\lambda &= -\sum_{k=0}^{\infty} \frac{t^{2k+1}}{(2k+1)!} (-1)^k 2^{2k+1} T_\lambda \\
&= -\sum_{k=0}^{\infty} (-1)^k \frac{(2t)^{2k+1}}{(2k+1)!} T_\lambda = -\sin(2t) T_\lambda.
\end{aligned}$$

Furthermore,

$$\sum_{k=0}^{\infty} \frac{t^{2k+2}}{(2k+2)!} \mathrm{ad}^{2k+1}_{\Omega_\lambda} \overline{\Omega}_\lambda = -\sum_{k=0}^{\infty} \frac{t^{2k+2}}{(2k+2)!} (-1)^k 2^{2k+1} T_\lambda.$$

Now we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_{k=0}^{\infty} \frac{t^{2k+2}}{(2k+2)!} (-1)^k 2^{2k+1} = \sum_{k=0}^{\infty} (-1)^k \frac{(2t)^{2k+1}}{(2k+1)!} = \sin(2t),$$

and

$$\sum_{k=0}^{\infty} \frac{t^{2k+2}}{(2k+2)!} (-1)^k 2^{2k+1} = \tfrac{1}{2}(1 - \cos(2t)),$$

and therefore

$$\mathrm{p}\left(\mathrm{Ad}_{\exp t\Omega_\lambda} X\right) = X_0 + \tfrac{1}{2}\lambda(X_0)(1 - \cos(2t)) T_\lambda - c_\lambda \sin(2t) T_\lambda. \quad \square$$

In the next lemma we analyze the variation of the $\mathfrak{a}$-component of $\mathrm{Ad}_{\exp t\Omega_\lambda}$ in more precise terms by discussing the function (2.21).

Lemma 2.9. *The function* $g(t) = \frac{1}{2}\lambda(X_0)(1-\cos(2t)) - c_\lambda \sin(2t)$ *is* π*-periodic and is either constant or possesses on* $(-\frac{\pi}{2}, \frac{\pi}{2}]$ *exactly one minimum* $t_{\min}$ *and one maximum* $t_{\max}$*. In this case*

$$\begin{aligned} \sin 2t_{\min} &= \frac{2c_\lambda}{\sqrt{4c_\lambda^2 + \lambda(X_0)^2}}, & \cos 2t_{\min} &= \frac{\lambda(X_0)}{\sqrt{4c_\lambda^2 + \lambda(X_0)^2}}, \\ \sin 2t_{\max} &= -\sin 2t_{\min}, & \cos 2t_{\max} &= -\cos 2t_{\min}, \end{aligned} \tag{2.23}$$

and $g(t_{\min}) = \frac{1}{2}\lambda(X_0) - \frac{1}{2}\sqrt{4c_\lambda^2 + \lambda(X_0)^2}$ *and* $g(t_{\max}) = \frac{1}{2}\lambda(X_0) + \frac{1}{2}\sqrt{4c_\lambda^2 + \lambda(X_0)^2}$.

Proof. The first assertion is trivial and we only need to prove Eqs. (2.23). Substituting $v := \sin 2t$ and $u := \cos 2t$ into the function $g(t) = \frac{1}{2}\lambda(X_0)\Big(1 - \cos(2t)\Big) - c_\lambda \sin(2t)$, leads to the following optimization task.

$$\begin{aligned} \text{Minimize/Maximize} \quad & \tfrac{1}{2}\lambda(X_0)(1-u) - c_\lambda v \\ \text{subject to} \quad & u^2 + v^2 = 1. \end{aligned} \tag{2.24}$$

We use the Lagrangian multiplier method to find the solutions. Let

$$L_m(u,v) := \tfrac{1}{2}\lambda(X_0)(1-u) - c_\lambda v + m(u^2+v^2-1)$$

be the Lagrangian function with multiplier m. By assumption, $g(t)$ is not constant and therefore the system of equations

$$\begin{aligned} D_u L_m(u,v) = \quad & -\tfrac{1}{2}\lambda(X_0) + 2mu &= 0 \\ D_v L_m(u,v) = \quad & -c_\lambda + 2mv &= 0 \\ & u^2 + v^2 &= 1 \end{aligned}$$

has exactly the two solutions

$$\begin{aligned} (u_1, v_1, m_1) &= \left(\frac{\lambda(X_0)}{\sqrt{4c_\lambda^2 + \lambda(X_0)^2}}, \frac{2c_\lambda}{\sqrt{4c_\lambda^2 + \lambda(X_0)^2}}, \tfrac{1}{2}\sqrt{4c_\lambda^2 + \lambda(X_0)^2} \right) \quad \text{and} \\ (u_2, v_2, m_2) &= \left(-\frac{\lambda(X_0)}{\sqrt{4c_\lambda^2 + \lambda(X_0)^2}}, -\frac{2c_\lambda}{\sqrt{4c_\lambda^2 + \lambda(X_0)^2}}, -\tfrac{1}{2}\sqrt{4c_\lambda^2 + \lambda(X_0)^2} \right). \end{aligned} \tag{2.25}$$

An inspection of the Hessian of $L_{m_i}(u_i, v_i)$ for $i = 1, 2$ and noting that $(u_1, v_1) = -(u_2, v_2)$ completes the proof of the first assumption. The last assertion is proven by a straightforward computation. □

The next theorem provides explicit formulas for the sine and cosine of the step-size selections. It allows the direct implementation of the sorting Givens-rotations and generalizes Prop. 6.1.1 in [23]. Note, that in contrast to the restrictions on the rotation angles used for conventional (non-sorting) Jacobi-methods, cf. [11] for details, the whole interval $(-\frac{\pi}{2}, \frac{\pi}{2}]$ is considered for rotation angles and the resulting rotation sorts the entries on the diagonal.

Theorem 2.10. *Let f be the trace function* (2.16) *and let $g(t)$ as in Eq.* (2.21)*. Then the following holds.*

(a) $f\left(\mathrm{Ad}_{\exp t\Omega_\lambda} X\right) = \kappa(X_0, N) - \frac{2\lambda(N)}{|\lambda|^2} g(t)$.

(b) Let $t_* \in (-\frac{\pi}{2}, \frac{\pi}{2}]$ *be the (local) maximum of* $f\left(\mathrm{Ad}_{\exp t\Omega_\lambda} X\right)$*. Then*

$$\cos 2t_* = -\frac{\lambda(X_0)}{\sqrt{4c_\lambda^2 + \lambda(X_0)^2}}, \qquad \sin 2t_* = -\frac{2c_\lambda}{\sqrt{4c_\lambda^2 + \lambda(X_0)^2}}$$

and hence

$$\cos t_* = \sqrt{\frac{1+\cos 2t_*}{2}}, \qquad \sin t_* = \begin{cases} \sqrt{\frac{1-\cos 2t_*}{2}} & \text{if } \sin 2t_* \geq 0 \\ -\sqrt{\frac{1-\cos 2t_*}{2}} & \text{if } \sin 2t_* < 0. \end{cases}$$

(c) $\lambda\left(\mathrm{p}\left(\mathrm{Ad}_{\exp t_*\Omega_\lambda} X\right)\right) = -\frac{1}{2}\sqrt{4c^2 + \lambda(X_0)^2} \leq 0$.

Proof. (a) The orthogonality of p yields

$$f(X) = \kappa(X, N) = \kappa(\mathrm{p}(X), N).$$

Let $T_\lambda = [E_\lambda, \theta(E_\lambda)]$ be defined as in Eq. (2.6). By Theorem 2.6,

$$\begin{aligned} f\left(\mathrm{Ad}_{\exp t\Omega_\lambda} X\right) &= \kappa(X_0 + g(t)T_\lambda, N) \\ &= \kappa(X_0, N) - \frac{2\lambda(N)}{|\lambda|^2} g(t), \end{aligned} \tag{2.26}$$

where $g(t) = \frac{1}{2}\lambda(X_0)(1-\cos(2t)) - c_\lambda \sin(2t)$ and the last identity holds since by definition $T_\lambda = -\frac{2}{|\lambda|^2} H_\lambda$.

(b) Since by assumption $\lambda(N) < 0$, the second statement now follows immediately by Lemma 2.9 and a standard trigonometric argument.

(c) is an immediate consequence of Lemma 2.9. □

We present a Matlab-like pseudo code of the algorithm. Let $\mathcal{B} = \{\Omega_1, \ldots, \Omega_m\}$ be as in Eq. (2.10) and let λ_i denote the restricted root with $\Omega_i \in \mathfrak{k}_{\lambda_i}$.

Partial Step of Sort-Jacobi Sweep. For a given $X \in \mathfrak{p}$, the following algorithm computes $(\sin t_*, \cos t_*, \sin 2t_*, \cos 2t_*)$, such that

$$\widetilde{X} := \mathrm{Ad}_{\exp(t_*\Omega_i)} X$$

has no $\overline{\Omega}_i$-component and such that $\lambda_i(\widetilde{X}_0) \leq 0$.

function: $(\cos t_*, \sin t_*, \cos 2t_*, \sin 2t_*) = \mathrm{givens}(X, \Omega_i)$
    Set $c_\lambda := \overline{\Omega}_i$-coefficient of X.

Set $X_0 := \mathrm{p}(X)$.
Set $dis := \sqrt{\lambda_i(X_0)^2 + 4c_\lambda^2}$.
if $dis \neq 0$
    Set $(\cos 2t_*, \sin 2t_*) := -\dfrac{1}{dis}\,(\lambda_i(X_0), 2c_\lambda)$.
else
    Set $(\cos 2t_*, \sin 2t_*) := (1, 0)$.
endif
Set $\cos t_* := \sqrt{\frac{1+\cos 2t_*}{2}}$.
if $\sin 2t_* \geq 0$
    Set $\sin t_* = \sqrt{\frac{1-\cos 2t_*}{2}}$.
else
    Set $\sin t_* = -\sqrt{\frac{1-\cos 2t_*}{2}}$.
endif
end givens

The algorithm is designed to compute $(\cos t_*, \sin t_*, \cos 2t_*, \sin 2t_*)$ of the step size t_* since this is natural by the chosen normalization of the sweep directions Ω_i, cf. Eq. (2.7). Nevertheless, depending on the underlying matrix representation, we can not exclude that the Givens rotations $\exp t\Omega_i$ have entries of the type $(\cos rt, \sin rt)$ with $r \neq 1, 2$. In this case it is advisable to use standard trigonometric arguments to compute $\cos rt_*$, $\sin rt_*$ respectively, by means of $\cos t_*, \sin t_*, \cos 2t_*, \sin 2t_*$.

Sort-Jacobi Algorithm. Denote by $d(X) := ||X - X_0||^2$ the squared distance from X to the maximal Abelian subalgebra $\mathfrak{a}$. Let $\mathcal{B} = \{\Omega_1, \dots, \Omega_m\}$ be as in Eq. (2.10). Given a Lie algebra element $S \in \mathfrak{p}$ and a tolerance $tol > 0$, the following algorithm overwrites S by $\varphi(S)$ where $\varphi \in \mathrm{Int}_{\mathfrak{g}}(\mathfrak{k})$ and $d(\varphi(S)) \leq tol$.

Set $\varphi :=$ identity.
while $d(S) > tol$
    for $i = 1 : m$
        $(\cos t_*, \sin t_*, \cos 2t_*, \sin 2t_*) := \mathrm{givens}(S, \Omega_i)$.
        $\varphi := \mathrm{Ad}_{\exp t_*\Omega_i} \circ \varphi$.
        $S := \mathrm{Ad}_{\exp t_*\Omega_i} S$.
    endfor
endwhile

3. Local Quadratic Convergence for the Regular Case

As already mentioned in the introduction, the local convergence analysis of the Jacobi algorithm that includes the irregular case proves to be tricky and will be treated separately in a forthcoming paper. For the symmetric EVP, the local convergence proof for special cyclic sweeps in the irregular case of van Kempen, [36], has been supplemented in [16].

The local convergence analysis for the regular case that is discussed here is based on the investigation of a more general setting for Jacobi-type methods on manifolds, cf. [22,26]. We particularize these results to the situation at hand and prove in a first step that, for regular elements, the step-size selections $t_*^{(i)}(X)$ are smooth in a neighborhood of a maximum of the trace function.

Lemma 3.1. *Let $Z \in \mathcal{O}(S) \cap \mathfrak{a}$ be a critical point of the trace function f with $\lambda(Z) < 0$. Denote by $\mathsf{H}_f(Z)$ the Hessian of f at Z. Then the step-size selection $t_*^{(i)}(X)$ is smooth in a neighborhood of Z if $\Omega_i \in \mathfrak{k}_\lambda$. In this case, the derivative is given by*

$$Dt_*^{(i)}(Z)(\xi) = -\frac{\mathsf{H}_f(Z)(\xi_i, \xi)}{\mathsf{H}_f(Z)(\xi_i, \xi_i)} = -d_i,$$

where $\xi = [Z, \Omega]$ and $d_i = \frac{\kappa(\Omega, \Omega_i)}{\kappa(\Omega_i, \Omega_i)}$ is the Ω_i-coefficient of Ω.

Proof. The main argument is the Implicit Function Theorem, cf. [1], Theorem 2.5.7. Recall r_{Ω_i} (2.12) and define the C^∞-function

$$\psi\colon \mathbb{R} \times \mathcal{O}(S) \longrightarrow \mathbb{R}, \quad \psi(t, X) := \frac{\mathrm{d}}{\mathrm{d}t}\left(f \circ r_{\Omega_i}(t, X)\right).$$

By the chain rule we have

$$\psi(t, X) = Df\left(r_{\Omega_i}(t, X)\right) r'_{\Omega_i}(t, X). \tag{3.1}$$

Since Z is a local maximum of $f\left(r_{\Omega_i}(t, X)\right)$ it follows that

$$\psi(0, Z) = 0.$$

Differentiating ψ with respect to the first variable yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi(t, X)\Big|_{(0,Z)} = \frac{\mathrm{d}^2}{\mathrm{d}t^2} f \circ r_{\Omega_i}(t, X)\Big|_{(0,Z)} = \mathsf{H}_f(Z)(\xi_i, \xi_i), \tag{3.2}$$

where $\xi_i := [Z, \Omega_i] = r'_{\Omega_i}(0, Z) \in T_Z\mathcal{O}(S)$. By Eq. (2.17),

$$\mathsf{H}_f(Z)(\xi_i, \xi_i) = -\lambda(Z)\,\lambda(N)\,\frac{2}{|\lambda|^2} < 0.$$

Now the Implicit Function Theorem yields that there exists a neighborhood U' of Z and a unique smooth function $l\colon U' \longrightarrow \mathbb{R}$ such that $\psi(l(X), X) = 0$ for all $X \in U'$. Since $\psi(t_*^{(i)}(Z), Z) = 0$, it follows from the uniqueness of l that there exists a suitable neighborhood $U \subset U'$ of X such that $l(X) = t_*^{(i)}(X)$ for all $X \in U$. Differentiating ψ with respect to the second variable yields together with Eq. (3.1)

$$\begin{aligned} D_X\psi(t, X)\Big|_{t=0,X=Z}\xi &= D_X\Big(\frac{\mathrm{d}}{\mathrm{d}t} f \circ r_{\Omega_i}(t, X)\Big)\Big|_{t=0,X=Z}\xi \\ &= D_X\left(Df(X) r'_{\Omega_i}(0, X)\right)\Big|_{X=Z}\xi \\ &= \mathsf{H}_f(Z)(\xi_i, \xi). \end{aligned}$$

By symmetrizing Eq. (2.17) it is easy to check that

$$\mathsf{H}_f(Z)(\xi_i, \xi) = -\lambda(Z)\lambda(N)\frac{2d_i}{|\lambda|^2}.$$

Since $\psi(t_*^{(i)}(X), X) = 0$ for all $X \in U$,

$$0 = D\psi(t_*^{(i)}(X), X)\Big|_{X=Z}\xi = \frac{\mathrm{d}}{\mathrm{d}t}\psi(t_*^{(i)}(Z), Z) \cdot Dt_*^{(i)}(Z)\xi + D_X\psi(t_*^{(i)}(Z), Z)\xi$$

and the assertion follows. □

The Sort-Jacobi algorithm converges locally quadratically fast to the maximum Z in the regular case. It is hence a generalization of the well-known result of Henrici, cf. [21], who proved local quadratic convergence for one particular type of cyclic sweep for the Hermitian EVP.

We call $S \in \mathfrak{p}$ *regular*, if there is no restricted root that annihilates the maximum of the trace function (2.16).

Theorem 3.2. *Denote by $f\colon \mathcal{O}(S) \longrightarrow \mathbb{R}$ the trace function and let Z be a maximum of f. If S is a regular element, then the Sort-Jacobi Algorithm is locally quadratic convergent to Z.*

Proof. The proof consists essentially of two parts. In the first part we show that the Hessian of f in Z is nondegenerate and that for the sweep directions Ω_i, the set $\{[Z, \Omega_i]\}$ forms a basis of $T_Z\mathcal{O}(S)$ that is orthogonal with respect to the Hessian of f. In the second part it is shown that this orthogonality is sufficient for local quadratic convergence. We only sketch how a Taylor series argument applies and refer to [26] for details.

The Hessian is given by

$$\mathsf{H}_f(Z)(\xi, \xi) = \kappa(\mathrm{ad}^2_\Omega Z, N) = -\kappa(\mathrm{ad}_Z\Omega, \mathrm{ad}_N\Omega),$$

cf. Eq. (2.17). We have $[Z, \Omega_i] = \lambda(Z)\overline{\Omega}_i$ and since by assumption Z is regular, $[Z, \Omega_i] \neq 0$ for all Ω_i. Orthogonality with respect to the Hessian is shown straightforwardly, since for $\Omega_i = E_\lambda + \theta(E_\lambda)$ and $\Omega_j = E_\mu + \theta(E_\mu)$ the orthogonality of the Ω_i implies for $i \neq j$

$$\mathsf{H}_f(Z)(\mathrm{ad}_Z\Omega_i, \mathrm{ad}_Z\Omega_j) = \tfrac{1}{2}\big(\lambda(Z)\ \mu(N) + \lambda(N)\mu(Z)\big)\ \kappa(\overline{\Omega}_i, \overline{\Omega}_j) = 0.$$

Now let $\xi \in T_Z\mathcal{O}(S)$ denote an arbitrary tangent space element. The derivative of one Givens-rotation $r_{\Omega_i}(t_*^{(i)}(X), X)$ in Z is given by

$$\begin{aligned} D\left(r_{\Omega_i}(t_*^{(i)}(X), X)\Big|_{X=Z}\right)\xi &= Dr_{\Omega_i}(t, X)|_{(t,X)=(t_*^{(i)}(Z),Z)} \circ D(t_*^{(i)}(X), \mathrm{id})|_{X=Z}\xi \\ &= Dt_*^{(i)}(Z)(\xi)\xi_i + \xi \end{aligned}$$

since $t_*^{(i)}(Z) = 0$. Therefore, by Lemma 3.1,

$$Dr_{\Omega_i}(Z)\xi = \xi - \frac{\mathsf{H}_f(Z)(\xi_i, \xi)}{\mathsf{H}_f(Z)(\xi_i, \xi_i)}\xi_i.$$

Thus $Dr_{\Omega_i}(Z)$ is a projection operator that – by orthogonality of the ξ_i's with respect to H_f – maps ξ into $(\mathbb{R}\xi_i)^\perp$. The composition of these projection operators is the zero map. Since Z is a fixed point, i.e. $r_i(Z) = Z$ for all $i = 1, \ldots, N$, we conclude

$$Ds(Z) = D(r_m \circ \cdots \circ r_1)(Z) = 0.$$

Consequently, a sweep defines a smooth map on a neighborhood of Z with vanishing derivative. Now reformulating everything in local coordinates, Taylor's Theorem yields

$$s(X) = s(Z) + Ds(Z)(X - Z) + \tfrac{1}{2}D^2 f(\xi)(X - Z, X - Z),$$

where $\xi \in \overline{U}$, a suitable compact neighborhood of Z. Using that $s(Z) = Z$, it follows

$$||s(X) - Z|| \leq \sup_{\xi \in \overline{U}} ||D^2 s(\xi)|| \cdot ||X - Z||^2.$$

Thus the algorithm induced by s converges locally quadratically fast to Z. □

Although the order in which the different elementary rotations Ω_i are worked off is irrelevant for the proof as long as regular elements are considered, and although regular elements form a dense subset in $\mathfrak{p}$, it is worth to point out that in practice, the ordering *does* matter for convergence speed. In fact, the relevance of the ordering is the bigger, the more the eigenvalues/singular values of S are clustered.

4. Example - The Exceptional Case $\mathfrak{g}_2$

To illustrate the previous results, consider the Lie algebra of derivations of the complex octonians, cf. [12]. We deduce a Sort-Jacobi algorithm arising from one of its real forms. Note, that this example is not isomorphic to any of the other cases listed in Table 1. Consider the 14-dimensional real Lie algebra

$$\mathfrak{g}_2 := \left\{ \begin{bmatrix} 0 & \sqrt{2}b^\top & \sqrt{2}c^\top \\ -\sqrt{2}c & M & B \\ -\sqrt{2}b & C & -M^\top \end{bmatrix} \,\middle|\, b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, B = \begin{bmatrix} 0 & -b_3 & b_2 \\ b_3 & 0 & -b_1 \\ -b_2 & b_1 & 0 \end{bmatrix}, \right.$$
$$\left. c = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}, C = \begin{bmatrix} 0 & -c_3 & c_2 \\ c_3 & 0 & -c_1 \\ -c_2 & c_1 & 0 \end{bmatrix}, b_i, c_i \in \mathbb{R}, M \in \mathbb{R}^{3\times 3}, \operatorname{tr} M = 0 \right\}. \tag{4.1}$$

We work with the following basis of $\mathfrak{g}_2$. Let E_{ij} denote the (7×7)-matrix with (i, j)-entry 1 and 0 elsewhere.

$$\begin{aligned} &X_1 := \sqrt{2}(E_{16} - E_{31}) + E_{54} - E_{72}; && X_2 := E_{23} - E_{65}; \\ &X_3 := \sqrt{2}(E_{15} - E_{21}) + E_{73} - E_{64}; && X_4 := \sqrt{2}(E_{14} - E_{71}) + E_{35} - E_{26}; \\ &X_5 := E_{34} - E_{76}; && X_6 := E_{24} - E_{75}; \\ &Y_i := -X_i^\top, \ i = 1, \ldots, 6; && \\ &H_1 := E_{22} - E_{44} - E_{55} + E_{77}; && H_2 := E_{33} - E_{44} - E_{66} + E_{77}. \end{aligned} \tag{4.2}$$

By help of the Killing form we compute the Cartan involution and the corresponding Cartan decomposition of $\mathfrak{g}_2$.

Proposition 4.1. *The Killing form on* $\mathfrak{g}_2$ *is given by*

$$\kappa_{\mathfrak{g}_2}(X,Y) = 4\mathrm{tr}(XY). \tag{4.3}$$

Proof. Using the commutator relations of the basis (4.2) one can easily construct matrix representations of the adjoint operators $\mathrm{ad}_{X_i}, \mathrm{ad}_{Y_i}, \mathrm{ad}_{H_j} \in \mathbb{R}^{14\times 14}$. It is straightforward to check that for all $Z, \widetilde{Z} \in \{X_1,\dots,X_6,Y_1,\dots,Y_6,H_1,H_2\}$ the relation

$$\kappa(Z,\widetilde{Z}) = \mathrm{tr}(\mathrm{ad}_Z \circ \mathrm{ad}_{\widetilde{Z}}) = 4\mathrm{tr}(Z\widetilde{Z})$$

holds. Hence for arbitrary $X \in \mathfrak{g}_2$ we have $\kappa(X,X) = 4\mathrm{tr}(X^2)$. The claim now follows by symmetrizing. □

Corollary 4.2. *A Cartan involution on* $\mathfrak{g}_2$ *is given by* $\theta(X) = -X^\top$. *Correspondingly, the Cartan decomposition is* $\mathfrak{g}_2 = \mathfrak{k} \oplus \mathfrak{p}$ *with*

$$\mathfrak{k} = \mathfrak{g}_2 \cap \mathfrak{so}(7,\mathbb{R}), \quad \mathfrak{p} = \{X \in \mathfrak{g}_2 \mid X^\top = X\}. \tag{4.4}$$

Proof. For $\theta(X) = -(X)^\top$, the bilinear form

$$B_\theta(X,Y) = -\kappa(X,\theta(Y)) = 4\mathrm{tr}(XY^\top)$$

is an inner product of $\mathfrak{g}_2$. Therefore θ is a Cartan involution. Obviously, for $\Omega \in \mathfrak{k}$ and $\overline{\Omega} \in \mathfrak{p}$, one has $\theta(\Omega) = \Omega$ and $\theta(\overline{\Omega}) = -\overline{\Omega}$. □

With respect to the maximal Abelian subspace

$$\mathfrak{a} := \{a_1H_1 + a_2H_2 \mid a_i \in \mathbb{R}\} \subset \mathfrak{p},$$

we can choose the set of positive restricted roots by

$$\boxed{\begin{array}{lll} \lambda_1 := a_2, & \lambda_2 := a_1 - a_2, & \lambda_3 := a_1, \\ \lambda_4 := a_1 + a_2, & \lambda_5 := a_1 + 2a_2, & \lambda_6 := 2a_1 + a_2. \end{array}} \tag{4.5}$$

The corresponding restricted-root spaces are given by

$$\mathfrak{g}_{\lambda_i} = \mathbb{R}X_i, \quad \mathfrak{g}_{-\lambda_i} = \mathbb{R}Y_i, \quad i = 1,\dots,6. \tag{4.6}$$

We now present a Sort-Jacobi algorithm that diagonalizes an element $S \in \mathfrak{p}$, preserving the special structure of $\mathfrak{p}$. Note, that for $i = 1,\dots,6$ we have $\theta(X_i) = Y_i$ and the $X_i \in \mathfrak{g}_{\lambda_i}$ are normalized such that $\lambda_i([X_i,\theta(X_i)]) = \lambda_i([X_i,Y_i]) = -2$ for all $i = 1,\dots,6$. Let

$\Omega_i := X_i + \theta(X_i) = X_i + Y_i \in \mathfrak{k}$. The Givens rotations are

$$\exp(t\Omega_1) = \begin{bmatrix} \cos 2t & 0 & \frac{\sin 2t}{\sqrt{2}} & 0 & 0 & \frac{\sin 2t}{\sqrt{2}} & 0 \\ 0 & \cos t & 0 & 0 & 0 & 0 & \sin t \\ -\frac{\sin 2t}{\sqrt{2}} & 0 & \frac{1}{2}+\frac{1}{2}\cos 2t & 0 & 0 & -\frac{1}{2}+\frac{1}{2}\cos 2t & 0 \\ 0 & 0 & 0 & \cos t & -\sin t & 0 & 0 \\ 0 & 0 & 0 & \sin t & \cos t & 0 & 0 \\ -\frac{\sin 2t}{\sqrt{2}} & 0 & -\frac{1}{2}+\frac{1}{2}\cos 2t & 0 & 0 & \frac{1}{2}+\frac{1}{2}\cos 2t & 0 \\ 0 & -\sin t & 0 & 0 & 0 & 0 & \cos t \end{bmatrix},$$

$$\exp(t\Omega_2) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos t & \sin t & 0 & 0 & 0 & 0 \\ 0 & -\sin t & \cos t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos t & \sin t & 0 \\ 0 & 0 & 0 & 0 & -\sin t & \cos t & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\exp(t\Omega_3) = \begin{bmatrix} \cos 2t & \frac{\sin 2t}{\sqrt{2}} & 0 & 0 & \frac{\sin 2t}{\sqrt{2}} & 0 & 0 \\ -\frac{\sin 2t}{\sqrt{2}} & \frac{1}{2}+\frac{1}{2}\cos 2t & 0 & 0 & -\frac{1}{2}+\frac{1}{2}\cos 2t & 0 & 0 \\ 0 & 0 & \cos t & 0 & 0 & 0 & -\sin t \\ 0 & 0 & 0 & \cos t & 0 & \sin t & 0 \\ -\frac{\sin 2t}{\sqrt{2}} & -\frac{1}{2}+\frac{1}{2}\cos 2t & 0 & 0 & \frac{1}{2}+\frac{1}{2}\cos 2t & 0 & 0 \\ 0 & 0 & 0 & -\sin t & 0 & \cos t & 0 \\ 0 & 0 & \sin t & 0 & 0 & 0 & \cos t \end{bmatrix},$$

$$\exp(t\Omega_4) = \begin{bmatrix} \cos 2t & 0 & 0 & \frac{\sin 2t}{\sqrt{2}} & 0 & 0 & \frac{\sin 2t}{\sqrt{2}} \\ 0 & \cos t & 0 & 0 & 0 & -\sin t & 0 \\ 0 & 0 & \cos t & 0 & \sin t & 0 & 0 \\ -\frac{\sin 2t}{\sqrt{2}} & 0 & 0 & \frac{1}{2}+\frac{1}{2}\cos 2t & 0 & 0 & -\frac{1}{2}+\frac{1}{2}\cos 2t \\ 0 & 0 & -\sin t & 0 & \cos t & 0 & 0 \\ 0 & \sin t & 0 & 0 & 0 & \cos t & 0 \\ -\frac{\sin 2t}{\sqrt{2}} & 0 & 0 & -\frac{1}{2}+\frac{1}{2}\cos 2t & 0 & 0 & \frac{1}{2}+\frac{1}{2}\cos 2t \end{bmatrix},$$

$$\exp(t\Omega_5) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos t & \sin t & 0 & 0 & 0 \\ 0 & 0 & -\sin t & \cos t & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cos t & \sin t \\ 0 & 0 & 0 & 0 & 0 & -\sin t & \cos t \end{bmatrix},$$

$$\exp(t\Omega_6) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos t & 0 & \sin t & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -\sin t & 0 & \cos t & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos t & 0 & \sin t \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -\sin t & 0 & \cos t \end{bmatrix}.$$

The implementation of the Sort-Jacobi algorithm is now straightforward. As an example, the regular element

$$S_{reg} = \begin{bmatrix} 0 & -3.17415 & -3.90421 & -4.63169 & 3.17415 & 3.90421 & 4.63169 \\ -3.17415 & 0.993208 & 3.14172 & 2.55770 & 0 & 3.27510 & -2.76069 \\ -3.90421 & 3.14172 & 3.224433 & -1.97516 & -3.27510 & 0 & 2.24446 \\ -4.63169 & 2.55770 & -1.97516 & -4.23754 & 2.76069 & -2.24446 & 0 \\ 3.17415 & 0 & -3.27510 & 2.76069 & -0.993208 & -3.14172 & -2.5577 \\ 3.90421 & 3.27510 & 0 & -2.24446 & -3.14172 & -3.24433 & 1.97516 \\ 4.63169 & -2.76069 & 2.24446 & 0 & -2.5577 & 1.97516 & 4.23754 \end{bmatrix}$$

is almost diagonalized after 3 sweeps (off-norm $< 10^{-10}$). It converges to the diagonal matrix

$$Z_{reg} = \mathrm{diag}\Big[\, 0, -9.12818, -1.97129, 11.0995, 9.12818, 1.97129, -11.0995 \,\Big].$$

Irregular elements show the same convergence behavior. In all simulations, at most 3 sweeps were required for quasi diagonalization (off-norm $< 10^{-10}$).

5. Conclusions and further remarks

Lie theory provides us both with a unified treatment and a coordinate free approach to Jacobi-type methods. In particular, it allows a formulation of Jacobi methods that is independent of the underlying matrix representation. Thus we can bring generality to a subject which has been dominated by case-by-case studies. Given a representation of a semisimple Lie algebra and a regular element S in the (-1)-eigenspace of the Cartan involution, the previous results allow to straightforwardly formulate a locally quadratic convergent Sort-Jacobi algorithm that diagonalizes S. Now although isomorphic representations yield isomorphic algorithms in an algebraic sense, this does not mean that these algorithms are equivalent from a numerical point of view. In particular, backward stability and relative accuracy might not be preserved under (algebraically) isomorphic algorithms.

Acknowledgments

This work presents partial results from my PhD thesis, [26]. I am very grateful to my supervisor U. Helmke, Würzburg, for his support and guidance. Many thanks also to G. Dirr for a lot of fruitful discussions.

REFERENCES

1. R. Abraham, J.E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications.* 2^{nd} *Ed.* Applied Mathematical Sciences 75. Springer, New York, 1988.
2. R.W. Brockett. Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems. *Lin. Algebra & Appl.*, 146:79–91, 1991.
3. A. Bunse-Gerstner, R. Byers, and V. Mehrmann. A chart of numerical methods for structured eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 13(2):419–453, 1992.
4. R. Byers. A Hamiltonian-Jacobi algorithm. *IEEE Transactions on Automatic Control*, 35(5):566-570, 1990.

5. J.-P. Charlier and P. Van Dooren. A Jacobi-like algorithm for computing the generalized Schur form of a regular pencil. *J. Comput. Appl. Math.*, 27(1/2):17–36, 1989.
6. M.T. Chu and K. Driessel. The projected gradient method for least squares matrix approximations with spectral constraints. *SIAM J. Numer. Anal.*, 27:1050–1060, 1990.
7. J.J. Duistermaat, J.A.C. Kolk, and V.S. Varadarajan. Functions, Flows and Oscillatory Integrals on Flag Manifolds and Conjugacy Classes in Real Semisimple Lie Groups. *Compositio Mathematica*, 49:309–398, 1983.
8. P.J. Eberlein. A Jacobi-like method for the automatic computation of eigenvalues and eigenvectors of an arbitrary matrix. *J. Soc. Indust. Appl. Math.*, 10(1):74–88, 1962.
9. P.J. Eberlein. On the diagonalization of complex symmetric matrices. *J. Inst. Maths. Appl.*, 7:377–383, 1971.
10. H. Faßbender, D.S. Mackey, and N. Mackey. Hamilton and Jacobi come full circle: Jacobi algorithms for structured Hamiltonian eigenproblems. *Lin. Algebra & Appl.*, 332-334:37–80, 2001.
11. G.E. Forsythe and P. Henrici. The cyclic Jacobi method for computing the principal values of a complex matrix. *Transactions of the American Math. Soc.*, 94:1–23, 1960.
12. W. Fulton and J. Harris. *Representation Theory* Springer, New York 1991.
13. H.H. Goldstine and L.P. Horwitz. A procedure for the diagonalization of normal matrices. *J. of the ACM*, 6:176–195, 1959.
14. G. Golub and C. F. van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, 1996.
15. D. Hacon. Jacobi's method for skew-symmetric matrices. *SIAM J. Matrix Anal. Appl.*, 14(3):619–628, 1993.
16. V. Hari. On sharp quadratic convergence bounds for the serial Jacobi methods. *Numer. Math.*, 60:375–406, 1991.
17. S. Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces.* Academic Press, San Diego, 1978.
18. U. Helmke and K. Hüper. The Jacobi Method: A Tool for Computation and Control. In: Systems and Control in the Twenty-First Century. Eds: C.I. Byrnes, and B.N. Datta, and D.S. Gilliam, and C.F. Martin. pp. 205–228, Birkhäuser Boston, 1997.
19. U. Helmke and K. Hüper. A Jacobi-type method for computing balanced realizations. *Systems & Control Letters*, 39:19–30, 2000.
20. U. Helmke, K. Hüper, and J.B. Moore. Computation of Signature Symmetric Balanced Realizations. *J. Global Optimization*, 27:135–148, 2003.
21. P. Henrici. On the speed of convergence of cyclic and quasicyclic Jacobi methods for computing eigenvalues of Hermitian matrices. *J. Soc. Indust. Appl. Math*, 6(2):144–162, 1958.
22. K. Hüper. A Dynamical System Approach to Matrix Eigenvalue Algorithms. In: Mathematical Systems Theory in Biology, Communications, Computation, and Finance. Eds: J. Rosenthal and D. S. Gilliam. IMA Vol 134, 257–274, Springer 2003.
23. K. Hüper. *Structure and convergence of Jacobi-type methods for matrix computations.* Ph.D. thesis, Technical University of Munich, June 1996.
24. C.G.J. Jacobi. Über ein leichtes Verfahren, die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen. *Crelle's J. für die reine und ange-*

wandte Mathematik, 30:51–94, 1846.
25. M. Kleinsteuber, U. Helmke, and K. Hüper. Jacobi's Algorithm on Compact Lie Algebras. *SIAM J. Matrix Anal. Appl.* 26(1):42–69, 2004.
26. M. Kleinsteuber *Jacobi-Type Methods on Semisimple Lie Algebras - A Lie Algebraic Approach to the Symmetric Eigenvalue Problem*. PhD thesis, Dep. of Mathematics, University of Würzburg, January 2006. `http://www.mathematik.uni-wuerzburg.de/~kleinsteuber/`
27. A.W. Knapp. *Lie Groups Beyond an Introduction,* 2^{nd} *Ed.* Birkhäuser, Boston, 2002.
28. E.G. Kogbetliantz. Solution of linear equations by diagonalization of coefficient matrix. *Quart. Appl. Math.*, 13:1231–132, 1955.
29. N. Mackey. Hamilton and Jacobi meet again: Quaternions and the eigenvalue problem. *SIAM J. Scientific and Statistical Computing*, 5:470–475, 1984.
30. D.S. Mackey, N. Mackey, and D.M. Dunlavy. Structure preserving algorithms for perplectic Eigenproblems. *Electronic J. Linear Algebra*, 13:10–39, 2005.
31. W.F. Mascarenhas. On the convergence of the Jacobi method for arbitrary orderings. *SIAM J. Matrix Anal. Appl.*, 16:1197–1209, 1995.
32. G.D. Mostov. The extensibility of local Lie groups of transformations and groups on surfaces. *Ann. of Math.*, 52:606–636, 1950.
33. M.H.C. Paardekooper. An eigenvalue algorithm for skew-symmetric matrices. *Num. Math.*, 17:189–202, 1971.
34. G.W. Stewart. A Jacobi-like algorithm for computing the Schur decomposition of a nonhermitian matrix. *SIAM J. Sci. Stat. Comput.*, 6(4):853–864, 1985.
35. T.Y. Tam. Gradient flows and double bracket equations. *Differ. Geom. Appl.*, 20(2):209–224, 2004.
36. H.P.M. van Kempen. On the quadratic convergence of the special cyclic Jacobi method. *Num. Math.*, 9:19–22, 1966.
37. K. Veselić. On a class of Jacobi-like procedures for diagonalizing arbitrary real matrices. *Num. Math.*, 33:157–172, 1979.
38. K. Veselić. A Jacobi eigenreduction algorithm for definite matrix pairs. *Num. Math.*, 64:241–269, 1993.
39. N.J. Wildberger. Diagonalization in compact Lie algebras and a new proof of a theorem of Kostant. *Proc. of the AMS*, 119(2):649–655, 1993.