JACOBI ANGLES FOR SIMULTANEOUS DIAGONALIZATION.

JEAN-FRANÇOIS CARDOSO* AND ANTOINE SOULOUMIAC†

Abstract. Simultaneous diagonalization of several matrices can be implemented by a Jacobi-like technique. This note gives the required Jacobi angles in close form.

Key words. Simultaneous diagonalization, Jacobi iterations, eigenvalues, eigenvectors, structured eigenvalue problem.

AMS subject classifications. 65F15, 65-04.

Introduction. Simultaneous diagonalization of several commuting matrices has been recently considered in [1], mainly motivated by stability and convergence concerns. Exact or approximate simultaneous diagonalization was also independently introduced as a solution to a statistical identification problem [2] (see [3] for a later paper in English). The simultaneous diagonalization algorithm described in these papers is an extension of the Jacobi technique: a joint diagonality criterion is iteratively optimized under plane rotations. The purpose of this note is to complement [1] by giving a close form expression for the optimal Jacobi angles.

1. Jacobi angles in close form. Consider a set $\mathcal{A} = \{A_k | k = 1, K\}$ of K complex $N \times N$ matrices. When the matrices in $\mathcal{A}$ are normal commuting matrices, their off-diagonal terms can be set to zero by a unitary transform, thus simultaneously diagonalizing the set $\mathcal{A}$. Define, as in [1],

$$\textbf{off}(A) \stackrel{\text{def}}{=} \sum_{1 \leq i \neq j \leq N} |a_{ij}|^2 \tag{1}$$

where a_{ij} denotes the (i,j)-th entry of matrix A. Simultaneous diagonalization may be obtained by minimizing the composite objective $\sum_{k=1,K} \textbf{off}(U A_k U^H)$ by a unitary matrix U. The extended Jacobi technique for simultaneous diagonalization constructs U as a product of plane rotations globally applied to all the matrices in $\mathcal{A}$.

Denote $R(i,j,c,s)$ the complex rotation matrix equal to the identity matrix but for the following entries

$$\begin{pmatrix} r_{ii} & r_{ij} \\ r_{ji} & r_{jj} \end{pmatrix} = \begin{pmatrix} c & \bar{s} \\ -s & \bar{c} \end{pmatrix} \quad \text{with} \quad c, s \in \mathbf{C} \quad \text{and} \quad |c|^2 + |s|^2 = 1. \tag{2}$$

It is desired, for each choice of $i \neq j$, to find the complex angles c and s which minimize the following objective function

$$O(c,s) \stackrel{\text{def}}{=} \sum_{k=1,K} \textbf{off}\left(R(i,j,c,s) A_k R^H(i,j,c,s)\right) \tag{3}$$

For a given pair (i,j) of indices, a 3×3 real symmetric matrix G is defined as

$$G \stackrel{\text{def}}{=} \text{Real}\left(\sum_{k=1,K} h^H(A_k) h(A_k)\right) \tag{4}$$

* Télécom Paris / CNRS URA 820 / GdR TdSI, 46 rue Barrault, 75634 Paris Cedex 13, France (cardoso@sig.enst.fr).

† Schlumberger Measurement & Systems / ETL, B.P. 620-05, 50 Avenue Jean Jaurès, 92542 Montrouge Cedex, France (souloum@montrouge.smr.slb.com). The author was supported by Thomson-CSF/RGS/STS while this study was conducted.

$$h(A) \stackrel{\text{def}}{=} [a_{ii} - a_{jj} \ , \ a_{ij} + a_{ji} \ , \ \mathrm{i}(a_{ji} - a_{ij})]. \tag{5}$$

For any set $\mathcal{A}$ of $N \times N$ matrices, commuting or not, real or not, and regardless of symmetry properties such as hermitianity, unitarity, normality, etc., the following theorem allows the Jacobi angles to be computed in close form.

THEOREM 1. *Under constraint $|c|^2 + |s|^2 = 1$, the objective function $O(c,s)$ is minimized at*

$$c = \sqrt{\frac{x+r}{2r}} \qquad s = \frac{y - \mathrm{i}z}{\sqrt{2r(x+r)}} \qquad r = \sqrt{x^2+y^2+z^2}, \tag{6}$$

where $[x,y,z]^T$ is any eigenvector associated to the largest eigenvalue of G.

Proof. Let a_{ij} and a'_{ij} respectively denote the (i,j)-th entry of matrices A and $A' = R(i,j,c,s)AR^H(i,j,c,s)$. Since $a'_{kk} = a_{kk}$ for $k \neq i$ and $k \neq j$ for plane rotations on the pair (i,j), the following invariance holds:

$$\mathbf{off}(A') + |a'_{ii}|^2 + |a'_{jj}|^2 = \mathbf{off}(A) + |a_{ii}|^2 + |a_{jj}|^2 \tag{7}$$

because unitary transforms preserve the norm $\sum_{kl} |a_{kl}|^2$. Hence, minimization of $\mathbf{off}(RAR^H)$ is seen, by (7), to be equivalent to *maximization* of $|a'_{ii}|^2 + |a'_{jj}|^2$. The latter, in turn, is equivalent to the maximization of $|a'_{ii} - a'_{jj}|^2$, as seen by the identity $2(|a'_{ii}|^2 + |a'_{jj}|^2) = |a'_{ii} + a'_{jj}|^2 + |a'_{ii} - a'_{jj}|^2$ and by $a'_{ii} + a'_{jj} = a_{ii} + a_{jj}$ (invariance of the trace under unitary transforms). One finds

$$a'_{ii} - a'_{jj} = (|c|^2 - |s|^2)(a_{ii} - a_{jj}) + 2csa_{ij} + 2\bar{s}\bar{c}a_{ji} \tag{8}$$

which is better rewritten as the inner product $a'_{ii} - a'_{jj} = h(A)v(c,s)$, between the complex 1×3 vector $h(A)$ defined in (5) and the 3×1 *real* vector $v(c,s)$ defined as

$$v(c,s)^T \stackrel{\text{def}}{=} [\ |c|^2 - |s|^2 \ , \ cs + \bar{c}\bar{s} \ , \ \mathrm{i}(cs - \bar{c}\bar{s})\ ]. \tag{9}$$

Hence, the Jacobi angles minimizing $O(c,s)$ are those maximizing

$$\sum_{k=1,K} |h(A_k)v(c,s)|^2 = v(c,s)^T \left(\sum_{k=1,K} h^H(A_k)h(A_k) \right) v(c,s). \tag{10}$$

Note that the 3×3 matrix on the right hand side of (10) is hermitian: its imaginary part[1] is skew-symmetric and consequently contributes nothing to a quadratic form in the *real* vector $v(c,s)$. Therefore the right hand side of (10) also is $v(c,s)^T G v(c,s)$. Next,we recognize that

$$\{v(c,s) | c,s \in \mathbf{C}, |c|^2 + |s|^2 = 1\} = \{[x,y,z]^T | x,y,z \in \mathbf{R}, x^2+y^2+z^2 = 1\}. \tag{11}$$

Thus minimization of $O(c,s)$ under the constraint $|c|^2 + |s|^2 = 1$ is equivalent to the maximization of a real 3×3 quadratic form under unit norm constraint. The solution is known to be given by any unit norm eigenvector of G associated to the (possibly degenerate) maximum eigenvalue. Now, if $[x,y,z]^T$ is a non zero eigenvector of G, not necessarily normed to unity, associated to the largest eigenvalue, its normalization and the inversion of relation (9) yields expression (6) by choosing c real positive. This choice is always possible since for any real angle ϕ, one has $v(c,s) = v(ce^{\mathrm{i}\phi}, se^{-\mathrm{i}\phi})$.□

[1] This imaginary part is zero in the special case where $\mathcal{A}$ contains only hermitian matrices.

2. Remarks on implementation and approximate simultaneous diagonalization. Regarding implementation, the following remarks are in order. **(i)** When $\mathcal{A}$ is a set of *real* symmetric matrices, the rotation parameters c and s are real: the last component of each vector $h(A_k)$ then is zero and G can be reduced to a 2×2 matrix by deleting its last row and last column: Theorem 1 then is similar to Theorem 6.1 of [1]. **(ii)** For the sake of numerical stability, the Jacobi technique should be restricted to 'inner rotations' [1]. In our setting, it corresponds to choosing an eigenvector with $x \geq 0$. **(iii)** Since matrix G is only 3×3, its dominant eigenvector may be computed explicitly. However, lacking a close form expression with proven stability, a standard numerical eigenvalue method should be preferred for the sake of numerical stability. **(iv)** It seems sensible to initialize the Jacobi algorithm for simultaneous diagonalization of a set $\mathcal{A}$ with the unitary matrix obtained as the (plain) diagonalizer of some matrix in $\mathcal{A}$. This initialization turns the spurious stationary point of the Jacobi algorithm given in eqs. (9-10) of [1] into a well behaved set.

We conclude with a few words about the relevance of *approximate* simultaneous diagonalization. There is a current trend in signal and data processing of extracting information from the eigenstructure of matrices which are functions of the available data. In some cases of interest, there is a set $\mathcal{A}_T$ of matrix-valued statistics computed from a number of T available samples with the property that, almost surely, the limit set $\mathcal{A}_\infty$ contains commuting matrices; the common eigenstructure could then be computed from any member of the set $\mathcal{A}_\infty$ or from some linear combinations of matrices in $\mathcal{A}_\infty$. In practice though, only a finite number of samples is available and the matrices in $\mathcal{A}_T$ do not exactly share the same eigenstructure. Determining the eigenstructure of interest from only one matrix in $\mathcal{A}_T$ is not satisfactory because, besides relying on an arbitrary choice, it amounts to discarding the information contained in the other matrices of $\mathcal{A}_T$. Also, it may happen that each matrix of $\mathcal{A}_\infty$ has some degenerate eigenvalues but that the whole set $\mathcal{A}_\infty$ has well determined common eigenvectors. Hence, from a statistical point of view, it is very desirable, for the sake of accuracy and robustness, to rather define the 'average eigenstructure' of $\mathcal{A}_T$. Optimizing a joint diagonality criterion, possibly appropriately weighted, offers a quantitative definition of such an average eigenstructure. In this stochastic context, we note that the criterion can only be minimized but cannot generally be driven to zero: the 'average eigenstructure' is well defined but corresponds only to an approximate simultaneous diagonalization.

Note: A MATLAB implementation of the extended Jacobi technique for simultaneous diagonalization is freely available upon request from `cardoso@sig.enst.fr`.

Acknowledgments: Comments by an anonymous referee improved the first version of this note.

REFERENCES

[1] A. Bunse-Gerstner, R. Byers, and V. Mehrmann, *Numerical methods for simultaneous diagonalization*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 927–949.

[2] A. Souloumiac and J.-F. Cardoso, *Comparaison de méthodes de séparation de sources*, in Proc. GRETSI, Juan les Pins, France, 1991, pp. 661–664.

[3] J.-F. Cardoso and A. Souloumiac, *Blind beamforming for non Gaussian signals*, IEE Proceedings-F, 140 (1993), pp. 362–370.