Available online at www.sciencedirect.com
ScienceDirect
Digital Signal Processing 16 (2006) 468–478
Digital Signal Processing
www.elsevier.com/locate/dsp

A centroid (Karcher mean) approach to the joint approximate diagonalisation problem: The real symmetric case

Jonathan H. Manton

Department of Electrical and Electronic Engineering, The University of Melbourne, Victoria 3010, Australia

Available online 8 August 2005

Abstract

A real symmetric matrix is diagonalisable by a suitable orthonormal change of basis. The joint approximate diagonalisation problem is to find an orthonormal change of basis which simultaneously diagonalises, or approximately diagonalises, two or more real symmetric matrices. A number of important signal processing problems require the computation of a joint approximate diagonaliser. However, no algorithm to date is guaranteed to find the optimal diagonaliser. This paper reformulates the diagonalisation problem as a convex optimisation problem on a Riemannian manifold and is thus able to guarantee global convergence to the optimal diagonaliser.

Keywords: Eigendecomposition; Joint eigenvectors; Optimisation on manifolds; Blind source separation

1. Introduction

Given symmetric matrices $C_1, \ldots, C_p \in \mathbb{R}^{n\times n}$, the joint diagonalisation problem is to find an orthonormal matrix $X \in \mathbb{R}^{n\times n}$ such that the matrices $X^T C_i X$ for $i = 1, \ldots, p$, are diagonal. Such an X exists if and only if the C_i share the same eigenvectors, in which case X can be found relatively easily from the eigenvectors of the C_i. In practice though [1–5], the C_i are observed in noise, and the joint diagonaliser X of the C_i given only the matri-

E-mail address: j.manton@ieee.org.

1051-2004/$ – see front matter
doi:10.1016/j.dsp.2005.06.003

ces $A_i = C_i + E_i$, where the E_i represent random error matrices, cannot be found by a straightforward eigendecomposition.

If the error distributions are known, the maximum likelihood estimate of X is obtained by solving the constrained optimisation problem: Given $A_1, \ldots, A_p$, find the $X, C_1, \ldots, C_p$, which maximise the joint likelihood of $(E_1, \ldots, E_p)$, where $E_i = A_i - C_i$, subject to the equality constraints that the $X^T C_i X$ are diagonal matrices and $X^T X = I$. This is called joint approximate diagonalisation, emphasising that X only approximately diagonalises the observed matrices A_i. Different distributions for the E_i, and estimators other than the maximum-likelihood estimator, lead to different definitions of an optimal joint approximate diagonaliser. All algorithms for computing these various diagonalisers iteratively minimise a cost function and may converge to a local minimum, hence their performance cannot be guaranteed [3,6,7].

This paper proposes a new definition of joint approximate diagonalisation which is geometrically meaningful, irrespective of any particular signal processing application. This geometrical insight sheds light on the strengths and limitations of joint approximate diagonalisation techniques in general. For instance, just like antipodal points on a sphere do not have a unique centroid, in certain cases the joint approximate diagonaliser cannot be defined uniquely either. Moreover, the proposed definition is chosen to ensure the optimal diagonaliser with respect to this definition can be computed reliably in practice.

Although the proposed algorithm computes a diagonaliser which is optimal according to the geometric definition, the diagonaliser may not be optimal for a given signal processing application. Nevertheless, in such situations, the intention is for the proposed algorithm either to be used directly as a sub-optimal but reliable solution—reliable albeit sub-optimal algorithms are often preferable—or as a means of computing a good initial guess for a subsequent local optimisation routine, such as if the maximum-likelihood estimate is sought.

It is emphasised the ideas in this paper are in their infancy and, as noted throughout the paper, there are numerous ways the results can be either generalised or tailored to specific applications. The first of such notes is that the complex-valued case will be studied elsewhere; while the geometric definition of joint approximate diagonalisation easily generalises, different techniques are required for a mathematical analysis.

2. Geometric joint approximate diagonalisation

The joint approximate diagonalisation problem is to find an orthonormal matrix $X \in \mathbb{R}^{n\times n}$ such that the $X^T A_i X$ are approximately diagonal, where the A_i are symmetric matrices. There are numerous ways of defining what approximately diagonal means. This section introduces a geometric definition, the motivation being that under this definition, the optimal diagonaliser can be found reliably in practice by minimising a convex function on a manifold.

Intuitively, previous definitions of optimal joint diagonalisation led to difficult optimisation problems because the joint diagonalisation problem is a generalisation of the matrix eigenvalue problem, and to date, attempting to solve the matrix eigenvalue problem by minimising a cost function (without any form of implicit or explicit deflation taking place) is not competitive with current best algorithms. This suggests the inherent matrix eigen-

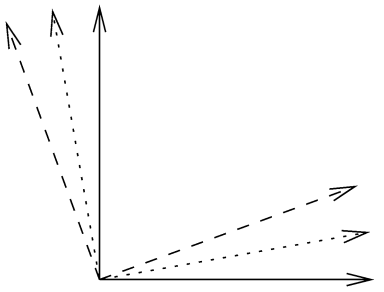

Fig. 1. If the solid lines represent the eigenvectors of one 2×2 matrix and the dashed lines the eigenvectors of another, then the dotted lines represent the proposed estimate of the joint eigenvectors.

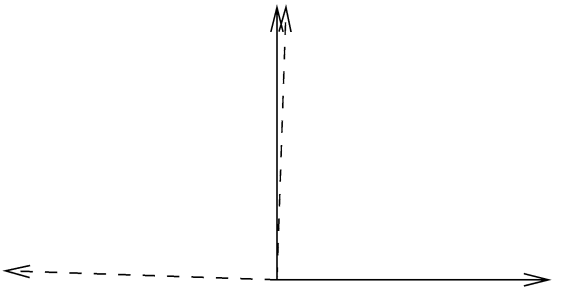

Fig. 2. As drawn, the angle between the solid and the dashed eigenvectors is close to 90°. However, replacing the dashed eigenvector on the left by its negative results in a very small angle between the two pairs of eigenvectors.

value problem should be factored out of the joint diagonalisation problem. This can be done either at the start, whereby the joint diagonaliser is expressed as some function of the eigendecompositions of the A_i, or at the end, whereby the joint diagonaliser X is found by computing the eigendecomposition of a matrix A, where A is some function of the A_i.

The former approach is taken. Given the symmetric matrices $A_1, \ldots, A_p \in \mathbb{R}^{n\times n}$, let $X_i \in \mathbb{R}^{n\times n}$ be orthonormal matrices such that the $X_i^T A_i X_i$ are diagonal. The columns of X_i are the eigenvectors of A_i, so X_i is unique up to permutation and sign change of its columns unless A_i has a multiple eigenvalue; this latter situation is temporarily ignored. If the A_i can be jointly diagonalised exactly then suitable permutations and sign changes of the columns of the X_i exist such that $X_1 = \cdots = X_p$.

If a small amount of noise is present, after suitable permutations and sign changes, the X_i can be expected to be close to each other in some metric, say $d(\cdot\,,\cdot)$. It is proposed to define the joint approximate diagonaliser X as the centroid of the X_i, namely

$$X = \arg\min_{X\in\mathcal{SO}_n} f(X), \quad f(X) = \sum_{i=1}^{p} d^2(X, X_i),\ X_i \in \mathcal{SO}_n, \tag{1}$$

where $\mathcal{SO}_n = \{X \in \mathbb{R}^{n\times n} \mid X^T X = I,\ \det X = 1\}$ denotes the set of orthonormal matrices with unit determinant, called the special orthogonal group. This determinant condition is explained presently. Figure 1 illustrates the idea when $n = p = 2$. Figure 2 shows the need for permutations and sign changes.

It remains to define $d(\cdot\,,\cdot)$. Let the columns of the orthonormal matrices X and Y represent the eigenvectors of two matrices. As in Fig. 1, the distance $d(X, Y)$ should represent how much X needs to be rotated for it to equal Y. Since rotations have unit determinant, it is only possible to rotate X to equal Y if $\det X = \det Y$. An orthonormal matrix has determinant ± 1 and the determinant can be changed by a sign change of one column. This explains why, without loss of generality, (1) assumes $X_i \in \mathcal{SO}_n$. (Mathematically, the orthogonal group, when considered as a topological space, consists of two connected components, the special orthogonal group being the component containing the identity matrix [8].) Now, since $\mathcal{SO}_n$ is a Lie group, there is a natural distance function $d_{\mathcal{SO}_n}(\cdot\,,\cdot)$

on it, induced from the unique bi-invariant metric [9,10], which does indeed measure the amount of rotation:

$$d_{\mathcal{SO}_n}(X,Y) = \frac{1}{\sqrt{2}}\left\|\log\left(XY^T\right)\right\|, \quad X, Y \in \mathcal{SO}_n, \tag{2}$$

where log is the principal matrix logarithm and $\|\cdot\|$ the Frobenius norm.

Let $\mathcal{G}_n$ be the sub-group of $\mathcal{SO}_n$ such that $X \in \mathcal{G}_n$ if and only if there exists a permutation matrix $P \in \mathcal{SO}_n$ and a diagonal matrix $D \in \mathcal{SO}_n$ with $X = PD$. To take into account the correct permutation and sign change, define

$$d(X,Y) = \min_{P,Q\in\mathcal{G}_n} d_{\mathcal{SO}_n}(XP, YQ) = \min_{Q\in\mathcal{G}_n} d_{\mathcal{SO}_n}(X, YQ). \tag{3}$$

Definition 1. Given symmetric $A_i \in \mathbb{R}^{n\times n}$ for $i = 1,\ldots,p$, let $X_i \in \mathcal{SO}_n$ be such that $X_i^T A_i X_i$ is diagonal. A geometric joint approximate diagonaliser of the A_i is any X minimising (1), with d defined in (3).

Section 3 uses the geometry of $\mathcal{SO}_n$ to establish conditions for the geometric joint approximate diagonaliser to be unique.

2.1. Ordered joint approximate diagonalisation

If an A_i has a multiple eigenvalue then Definition 1 is apparently unsatisfactory because the non-uniqueness of X_i extends beyond permutations and sign changes of its columns. The mathematical solution is straightforward; include this extra ambiguity when searching for the minimum in (3). In practice, this extra level of difficulty is not warranted. First, if the A_i are noise-corrupted versions of the matrices C_i, then typically the probability that an A_i has a multiple eigenvalue is zero. More importantly, the joint diagonalisation problem is ill-conditioned[1] if an A_i has a multiple eigenvalue. For example, if $A_1 = I$ and A_2 is arbitrary, arbitrarily small perturbations of A_1 can lead to its eigenvectors being far away from the eigenvectors of A_2.

Thus, a joint diagonalisation algorithm detecting an A_i with two close eigenvalues should signal that the problem is ill-conditioned. Furthermore, all joint diagonalisation routines should be used cautiously in applications where the noise free matrices C_i may contain closely spaced eigenvalues.

This line of thinking motivates several new geometric problems which should be of interest to the signal processing community. The first is to find a joint diagonaliser X where the order of the columns of X matters.

Definition 2. Given symmetric $A_i \in \mathbb{R}^{n\times n}$ for $i = 1,\ldots,p$, let $X_i \in \mathcal{SO}_n$ be such that the jth column of X_i is the eigenvector associated with the jth largest eigenvalue of A_i. An ordered joint approximate diagonaliser of the A_i is any X minimising (1), with d defined in (3) but where $\mathcal{G}_n$ is instead the sub-group of all diagonal matrices in $\mathcal{SO}_n$.

[1] Prompted by a reviewer, we point out that ill-conditioned is different from ill-defined; our claim does not contradict Theorem 3 of [2].

If the C_i have well separated eigenvalues and $A_i = C_i + E_i$ with the E_i sufficiently small, then an ordered joint approximate diagonaliser (Definition 2) is also a joint approximate diagonaliser (Definition 1). Moreover, in certain cases, omitting the permutation matrices in $\mathcal{G}_n$ simplifies the joint diagonalisation algorithm. This is discussed in Section 4.

Finally, in some applications, each C_i might have some eigenvalues well separated but others closely spaced. For instance, the several smallest eigenvalues of C_i (the "noise subspace") might be closely spaced. Then, it may be of interest to find a rectangular matrix X which approximately diagonalises the principal parts of the C_i. Mathematically, let the columns of X_i be the principal eigenvectors of A_i and let X be the centroid of the X_i after appropriate permutation and sign change of the columns of the X_i. This differs from the geometric joint diagonalisation problem only in that the X and X_i are rectangular, hence belong to the Stiefel manifold rather than $\mathcal{SO}_n$. Similarly, if an approximately common principal subspace is sought, the X and X_i can be considered elements of the Grassmann manifold instead. Algorithms for solving these problems and applications thereof will be reported elsewhere.

3. Theoretical results

Let $\mathcal{J}$ denote an arbitrary joint approximate diagonalisation algorithm. For the algorithm to be sensible, it must satisfy the following requirements. In R1 and R4 below, Λ is the set of diagonal matrices with distinct eigenvalues.

R1. If $D_1, \ldots, D_p \in \Lambda$, then $\mathcal{J}(D_1, \ldots, D_p) = I$.
R2. For any permutation π, $\mathcal{J}(A_1, \ldots, A_p) = \mathcal{J}(A_{\pi(1)}, \ldots, A_{\pi(p)})$.
R3. For any orthonormal U, $\mathcal{J}(UA_1U^T, \ldots, UA_pU^T) = U\mathcal{J}(A_1, \ldots, A_p)$.
R4. If $D_1, \ldots, D_p \in \Lambda$, then $\mathcal{J}(D_1 + N_1, \ldots, D_p + N_p)$ is Lipschitz continuous for sufficiently small N_i.

The geometric joint diagonaliser is shown below to exist and be unique. Presupposing this, it follows that it satisfies R1–R4. Indeed, R1 and R2 are clear, while R3 follows from the left-invariance property of the distance function (3), namely $d(UX, UY) = d(X, Y)$ if $U \in \mathcal{SO}_n$. To prove R4, first note that if N_i is sufficiently small then the X_i, in Definition 1 will vary smoothly with N_i and be close to the identity matrix. If X and X_i are sufficiently close to the identity then it is a standard result that $d^2(X, X_i)$ is smooth, hence a small change in the X_i will induce only a small change in X, defined in (1), implying R4.

The remainder of this section is devoted to explaining the underlying geometry and proving uniqueness of the geometric joint diagonaliser. Referring to (1) and (3), although the functions f and d are defined on $\mathcal{SO}_n$, they should be thought of as being defined on the quotient space $\mathcal{SO}_n/\mathcal{G}_n$. Indeed, d is a true distance function on $\mathcal{SO}_n/\mathcal{G}_n$. Specifically, $\mathcal{SO}_n/\mathcal{G}_n$ is a manifold (in fact, a homogeneous space) and d is the distance function induced from $d_{\mathcal{SO}_n}$ by the natural projection $\pi\colon \mathcal{SO}_n \to \mathcal{SO}_n/\mathcal{G}_n$. (The situation is straightforward because $\mathcal{G}_n$ is a finite group. For complex-valued matrices, $\mathcal{G}_n$ would be an infinite group. This can be handled by the standard theory of reductive homogeneous spaces.)

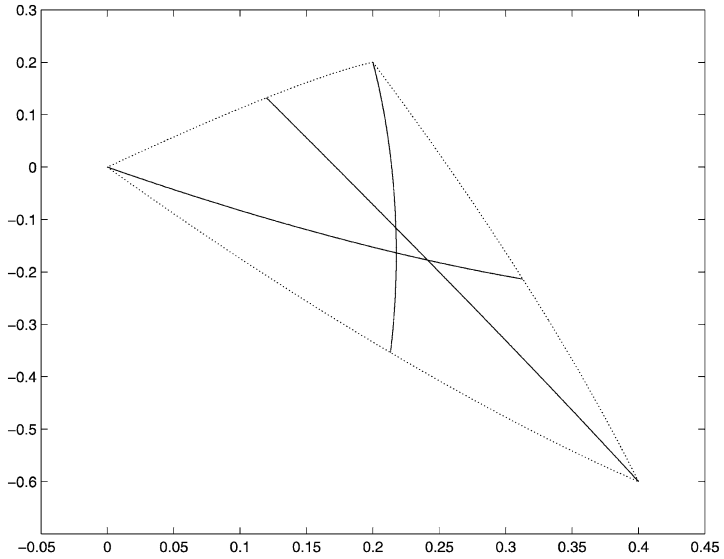

Fig. 3. The dotted lines are a geodesic triangle connecting three points (projected onto a two-dimensional plane) in $\mathcal{SO}_3$. The solid lines are geodesics joining the midpoint of each side of the triangle to the opposite corner. The curvature of $\mathcal{SO}_3$ prevents the three solid lines meeting at a unique point.

Although the projection $\pi\colon \mathcal{SO}_n \to \mathcal{SO}_n/\mathcal{G}_n$ is many-to-one, locally it is one-to-one. That is, the geometry (curvature, geodesics, etc.) of $\mathcal{SO}_n/\mathcal{G}_n$ coincides locally with the geometry of $\mathcal{SO}_n$, and the latter is well understood: $\mathcal{SO}_n$ equipt with its bi-invariant metric is a compact Riemannian manifold with non-negative curvature. Geodesics on $\mathcal{SO}_n$ are of the form $\gamma(t) = Xe^{At}$, where $X \in \mathcal{SO}_n$ and $A \in \mathfrak{so}_n$, where $\mathfrak{so}_n$ is the set of skew-symmetric matrices.

Since $\mathcal{SO}_n$, and hence $\mathcal{SO}_n/\mathcal{G}_n$ has positive curvature, if the X_i are too far apart the centroid need not be unique. To visualise this, consider the north and south poles on a sphere and note any point on the equator is equidistant from them. In practice, this means that if too much noise is present, so the A_i are far from being jointly diagonalisable, then the joint approximate diagonaliser is ill-defined. Since all joint diagonalisation algorithms minimise a cost function on $\mathcal{SO}_n$ (or a quotient space thereof), this is an inherent feature of the joint approximate diagonalisation problem and not of any specific definition or implementation.

It is remarked that although there are several different definitions of centroid which turn out to be equivalent in Euclidean space, on $\mathcal{SO}_n$ this is no longer true. See, for example, Fig. 3. The definition in (1) of a centroid is taken from [11] and has been studied in the special case of $\mathcal{SO}_n$ in [12]. The globally convergent algorithm in [13] for computing this centroid forms part of the algorithm in Section 4. Further details of the geometry of $\mathcal{SO}_n$ can be found in these papers.

The largest domain on which π is injective is the fundamental domain

$$\mathcal{F} = \bigl\{X \in \mathcal{SO}_n \mid \forall Q \in \mathcal{G}_n - \{I\},\ d_{\mathcal{SO}_n}(X, I) < d_{\mathcal{SO}_n}(X, Q)\bigr\}. \tag{4}$$

Denote the open ball in $\mathcal{SO}_n$ by $B(X;\rho) = \{Y \in \mathcal{SO}_n \mid d_{\mathcal{SO}_n}(X, Y) < \rho\}$.

Lemma 3. *Assume $n \geqslant 2$. The open ball $B(I;\rho)$ is contained in the fundamental domain $\mathcal{F}$ if and only if $\rho \leqslant \pi/4$.*

Proof. Define $\bar{\rho} = \frac{1}{2}\min_{Q\in\mathcal{G}_n - \{I\}} d_{\mathcal{SO}_n}(Q, I)$. Clearly $B(I;\rho) \subset \mathcal{F}$ if and only if $\rho \leqslant \bar{\rho}$. Let $Q \in \mathcal{G}_n$ be the matrix with all elements zero except for $Q_{12} = 1$, $Q_{21} = -1$, $Q_{ii} = 1$ for $i \geqslant 3$. Then $d_{\mathcal{SO}_n}(Q, I) = \pi/2$, proving $\bar{\rho} \leqslant \pi/4$. The proof that $\bar{\rho} \geqslant \pi/4$ is tedious and hence omitted. It is intuitive though that Q is the element in $\mathcal{G}_n - \{I\}$ closest to I because it differs from I in only two columns. □

Theorem 4. *Assume $n \geqslant 2$. Referring to* (1)*, if there exists a $Y \in \mathcal{SO}_n$ such that $X_i \in B(Y;\pi/8)$ then there is precisely one minimum of $f(X)$ in the region $B(Y;\pi/8)$. Moreover, f restricted to $B(Y;\pi/8)$ is strictly convex, and $\pi/8$ is the largest radius for which this is true.*

Proof. Define $\tilde{f}(X) = \sum_{i=1}^{p} d^2_{\mathcal{SO}_n}(X, X_i)$. If $X_i \in B(Y;\pi/8)$ then it is proved in [11,12] that $\tilde{f}$ restricted to $B(Y;\pi/8)$ is strictly convex and hence has a unique minimum $\tilde{X} \in B(Y;\pi/8)$, which is precisely the Karcher mean of the points X_i on $\mathcal{SO}_n$. (In fact, $\tilde{f}(X)$ is convex on a much larger ball.) Now, Lemma 3 implies that if $d_{\mathcal{SO}_n}(X, X_i) < \pi/4$ then $d(X, X_i) = d_{\mathcal{SO}_n}(X, X_i)$, so if $X, X_i \in B(Y;\pi/8)$ then $d(X, X_i) < \pi/4$, hence $\tilde{f}$ and f agree on $B(Y;\pi/8)$, thus proving f restricted to $B(Y;\pi/8)$ is strictly convex with minimum $\tilde{X}$.

To prove $\pi/8$ is the largest radius, define $Q(\theta) \in \mathcal{SO}_n$ to be the identity matrix except for the top left 2×2 block which is $[\cos\theta, -\sin\theta; \sin\theta, \cos\theta]$, a rotation matrix. Note $Q(\pi/2) \in \mathcal{G}_n$. Let $X_1 = Q(\pi/8)$ and $X_2 = Q(-\pi/8)$. Since $Q(-\pi/8)$ is equivalent to $Q(-\pi/8)Q(\pi/2) = Q(3\pi/8)$ in $\mathcal{SO}_n/\mathcal{G}_n$,

$$f\big(Q(\theta)\big) = \begin{cases} (\theta - \pi/8)^2 + (\theta + \pi/8)^2, & \theta \in [-\pi/8, \pi/8], \\ (\theta - \pi/8)^2 + (\theta - 3\pi/8)^2, & \theta \in [\pi/8, 3\pi/8] \end{cases} \tag{5}$$

and clearly $f(X)$ is not convex at $X = Q(\pi/8)$. □

Admittedly, the radius $\pi/8$ is small; if a moderate amount of noise is present, it might not be possible to choose $Q_i \in \mathcal{G}_n$ and $Y \in \mathcal{SO}_n$ such that $X_i Q_i \in B(Y;\pi/8)$, even though the geometric joint diagonaliser is well defined and correctly computed by the algorithm in Section 4. The problem is primarily caused by the volume of $B(I;\pi/4)$ being considerably smaller than the volume of $\mathcal{F}$ even though $\mathcal{F}$ contains no larger ball. The way to overcome this is to use a different metric for defining open balls. This is currently under investigation.

4. The algorithm, a numerical example, and a discussion

An algorithm for computing the geometric joint diagonaliser (Definition 1) is now described. Given symmetric $A_i \in \mathbb{R}^{n\times n}$ for $i = 1, \ldots, p$, find, by eigen-decomposition, matrices $X_i \in \mathbb{R}^{n\times n}$ such that the $X_i^T A_i X_i$ are diagonal. If any of the A_i have multiple eigenvalues, warn about ill-conditioning. Multiply the first column of each X_i by -1 if necessary, so $\det X_i = 1$. Recalling the definition of $\mathcal{G}_n$ in (3), compute $Q =$

$\arg\min_{Q\in\mathcal{G}_n} d_{\mathcal{SO}_n}(X_1, X_2Q)$ by exhaustive search. (The cardinality of $\mathcal{G}_n$ is $n!2^{n-1}$, so this is only practical for small n. Faster solutions not requiring an exhaustive search are currently under investigation. See too the note below.) Set $X := X_1 e^{(1/2)\log(X_1^T X_2)}$. If $p = 2$ then return X as the joint diagonaliser. Otherwise, for each i, compute $Q = \arg\min_{Q\in\mathcal{G}_n} d_{\mathcal{SO}_n}(X, X_iQ)$ and set $X_i := X_iQ$. (This ensures the X_i are clustered around X.) If there exists an i such that $d_{\mathcal{SO}_n}(X_i, X) \geqslant \pi/8$, warn that reliability cannot be guaranteed. Repeat the steps:

(1) Set $S := \frac{1}{p}\sum_{i=1}^{p} \log(X^T X_i)$. To reduce numerical roundoff errors, set $S := (S - S^T)/2$.
(2) If $\|S\| < 10^{-6}$ then return X as the joint diagonaliser.
(3) Set $X := Xe^S$ and go to step 1. (These three steps implement the steepest descent algorithm proposed in [13].)

Note that the exhaustive search is not necessary if the X_i are sufficiently close together on $\mathcal{SO}_n/\mathcal{G}_n$, such as when the A_i are close to being exactly diagonalisable and have well separated eigenvalues. In this case, if the jth column of X_i is the eigenvector of A_i associated with the jth largest eigenvalue, then $\min_{Q\in\mathcal{G}_n} d_{\mathcal{SO}_n}(X_1, X_2Q)$ achieves its minimum when Q is diagonal with $Q_{jj} = 1$ if the inner product of the jth columns of X_1 and X_2 is positive, otherwise $Q_{jj} = -1$. Similarly, computing the ordered joint diagonaliser (Definition 2) eliminates the need to search through all $n!$ permutation matrices.

The algorithm was implemented and compared with Cardoso's well-known Jacobi method for computing joint approximate diagonalisers [7]. Since Cardoso's method minimises a different cost function (namely, the sum of the squares of the off diagonal elements), whether Cardoso's method or the proposed method performs better should depend on the noise model.

Two different noise models are used. The first is $A_i = UDU^T + N_i$ for $i = 1, \ldots, 4$, where U is an arbitrary orthonormal matrix, $D = \text{diag}\{1, \ldots, 5\}$ and $N_i = (\sigma^2/2)(E_i + E_i^T)$ where the elements of E_i have a Gaussian $N(0, 1)$ distribution. It is anticipated Cardoso's method will outperform the geometric joint diagonaliser because the latter essentially assumes each eigenvector of A_i is perturbed by approximately the same amount from the corresponding eigenvector of UDU^T, whereas the additive noise N_i will perturb some eigenvectors more than others because the eigenvalues range from 1 to 5. Figure 4 illustrates this when $\sigma^2 = 0.5$. Here, if V is the joint approximate diagonaliser, the error is graphed as $d(U, V)$. Note that $\sigma^2 = 0.5$ is large enough for the $\pi/8$ distance rule to be violated frequently, yet the algorithm still works well. Note too that results not presented indicate that if $\sigma^2 < 0.2$ then the performance difference between the two methods is negligible. (When simulating this and the following model, 50 trials were performed, with the matrix U generated at random for each trial. Each point in the figures represents one trial.)

The second model is $A_i = Ue^{S_i}D(e^{S_i})^T U^T$ for $i = 1, \ldots, 4$, where U is an arbitrary orthonormal matrix, $D = \text{diag}\{1, \ldots, 5\}$ and $S_i = (\sigma^2/2)(E_i - E_i^T)$ where the elements of E_i have a Gaussian $N(0, 1)$ distribution. Since e^{S_i} represents a uniform random perturbation of the columns of U, it is anticipated the geometric joint diagonaliser will outperform Cardoso's method. Figure 5 shows that when $\sigma^2 = 0.5$, this is indeed the case; the geo-

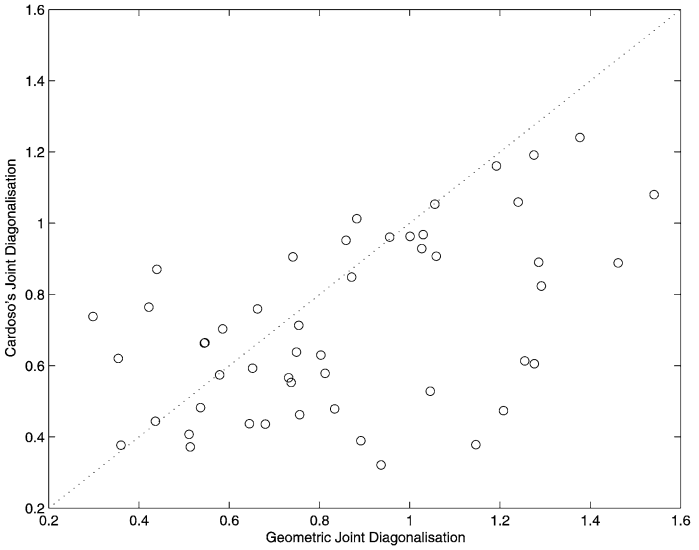

Fig. 4. For noise model one, Cardoso's method has a mean error of 0.71 and a standard deviation of 0.25 while the geometric joint diagonaliser has a mean error of 0.86 and a standard deviation of 0.32.

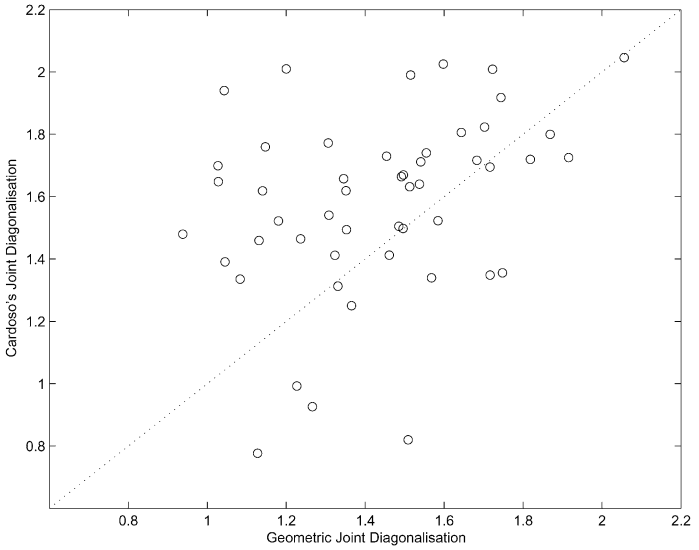

Fig. 5. For noise model two, Cardoso's method has a mean error of 1.58 and a standard deviation of 0.30 while the geometric joint diagonaliser has a mean error of 1.43 and a standard deviation of 0.26.

metric joint diagonaliser has not only a smaller mean but also a smaller standard deviation, implying it is more robust to outliers.

To make the geometric joint diagonaliser perform better for noise model one, the distance function (3) needs to be modified. Intuitively, changing an eigenvector associated

with a larger eigenvalue should incur a larger penalty since additive noise is less likely to affect these eigenvectors. However, an ad-hoc change to (3) can easily destroy the convexity of the cost function $f(X)$ in (1). In its full generality, the key idea in this paper is to choose $d(\cdot,\cdot)$ to be a distance function on a Riemannian manifold, since this will ensure $f(X)$ is convex provided the X_i are sufficiently close together. The only other restriction is d be left-invariant, that is $d(X,Y) = d(UX, UY)$ for all $U \in \mathcal{SO}_n$, since this ensures the geometric joint diagonaliser is equivariant to orthonormal changes of basis of the A_i (requirement R3 in Section 3).

Let $\mathfrak{so}_n$ denote the tangent space at the identity of $\mathcal{SO}_n$. (It is the Lie algebra associated with $\mathcal{SO}_n$ and consists of skew-symmetric matrices.) A left-invariant distance function on $\mathcal{SO}_n$ is found by assigning an inner product to $\mathfrak{so}_n$ and then extending it to the whole tangent bundle of $\mathcal{SO}_n$ by left translation. For noise model one, a suitable choice of inner product is found as follows. If $A \in \mathfrak{so}_n$ (so that $A^T = -A$) then $e^A \in \mathcal{SO}_n$ and

$$\left(e^A\right)^T D e^A = D + [D, A] + O\left(\|A\|^2\right), \tag{6}$$

where $[D, A] = DA - AD$ and D is diagonal. Provided D has distinct eigenvalues, given a symmetric N there exists an A such that $[D, A] = N$. Thus, (6) says to first order that a diagonal matrix D perturbed by N has its eigenvectors perturbed from I to e^A, where $[D, A] = N$. A least-squares estimator seeks to minimise $\mathrm{tr}\{N^T N\} = \mathrm{tr}\{[D, A]^2\}$. This suggests defining the norm of $A \in \mathfrak{so}_n$ to be $\|A\|^2 = \mathrm{tr}\{[D, A]^2\}$, and in fact, this norm comes from the inner product

$$\langle A, B\rangle = \mathrm{tr}\big\{[D, A][D, B]\big\}, \quad A, B \in \mathfrak{so}_n. \tag{7}$$

As already mentioned, (7) induces a left-invariant distance function, and geometric joint diagonalisation with respect to this distance function is expected to perform well for noise model one. Before this can be verified though, the geometry induced by (7) must be derived, and this is beyond the scope of the present paper.

5. Conclusion

A joint approximate diagonalisation algorithm should fulfill requirements R1–R4 in Section 3 and give reliable results. To achieve this, this paper proposed to factor out the difficult matrix eigenvalue problem from the joint diagonalisation problem and then use geometry to construct a convex cost function on a Riemannian manifold whose minimum defines the joint approximate diagonaliser. As such, this is the first joint diagonalisation algorithm to use a convex cost function and hence not suffer from convergence to a local minimum.

Acknowledgments

The author acknowledges the support of the Australian Research Council and the ARC Special Research Centre for Ultra-Broadband Information Networks (CUBIN).

References

[1] A.J. Van der Veen, A. Paulraj, An analytical constant modulus algorithm, IEEE Trans. Signal Process. 44 (5) (1996) 1136–1155.
[2] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, E. Moulines, A blind source separation technique using second-order statistics, IEEE Trans. Signal Process. 45 (2) (1997) 434–444.
[3] M. Wax, J. Sheinvald, A least-squares approach to joint diagonalization, IEEE Signal Process. Lett. 4 (2) (1997) 52–53.
[4] O.J. Micka, A.J. Weiss, Estimating frequencies of exponentials in noise using joint diagonalization, IEEE Trans. Signal Process. 47 (2) (1999) 341–348.
[5] K. Rahbar, J.P. Reilly, J.H. Manton, Blind identification of MIMO FIR systems driven by quasi-stationary sources using second order statistics: A frequency domain approach, IEEE Trans. Signal Process. 52 (2) (2004) 406–417.
[6] J. Cardoso, A.S. Soulomiac, Blind beamforming for non-Gaussian signals, Proc. IEE-F 140 (1993) 362–370.
[7] J. Cardoso, A.S. Soulomiac, Jacobi angles for simultaneous diagonalization, SIAM J. Matrix Anal. Applicat. 17 (1) (1996) 161–164.
[8] F.W. Warner, Foundations of Differentiable Manifolds and Lie Groups, Springer-Verlag, 1983.
[9] I. Chavel, Riemannian Geometry: A Modern Introduction, Cambridge Univ. Press, 1996.
[10] S. Helgason, Differential Geometry, Lie Groups, and Symmetric Spaces, American Mathematical Society, Providence, RI, 2001.
[11] H. Karcher, Riemannian center of mass and mollifier smoothing, Commun. Pure Appl. Math. 30 (1977) 509–541.
[12] J.H. Manton, K. Hüper, The Karcher mean of points on the special orthogonal group, in preparation.
[13] J.H. Manton, A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups, in: Eighth International Conference on Control, Automation, Robotics and Vision, Kunming, China, 2004 (on CD-ROM).