To appear in *Linear and Multilinear Algebra*
Vol. 00, No. 00, Month 20XX, 1–14

On Procrustes matching of non-negative matrices and an application to random tomography

Christian Rau[a*]

[a] *Department of Mathematics, Shantou University, Shantou Guangdong 515063, P.R. China*

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

We consider a Procrustes matching problem for non-negative matrices that arose in random tomography. As an alternative to the Frobenius distance, we propose an alternative non-symmetric distance using generalized inverses. Among its advantages is that it leads to a relatively simple quadratic function that can be optimized with least-square methods on manifolds.

Keywords: Brockett cost function; generalized inverses; least-square methods on manifolds; Procrustes methods; random tomography

AMS Subject Classification: 65F20; 65C60

1. Introduction

Procrustes problems have, over the last few decades, attained considerable importance in the literature on multivariate statistical analysis, and shape analysis in particular; see [1, Ch. 14]. Problems of this sort have been known to give rise to

*Email: rau@stu.edu.cn, christianrau080@gmail.com

a variety of matrix decomposition problems and factorizations such as the polar factorization, see [2, pp. 12–13] and the references cited therein. An important yet considerably less well studied problem is the choice of distance used in the Procrustes step, and in more general matching problems. The Frobenius norm $\|\cdot\|_F$ defined by $\|A\|_F^2 = \langle A, A\rangle = \sum_{i,j} a_{ij}^2$, for $A = (a_{ij})_{i,j=1}^n \in \mathbb{R}^{n\times n}$, seems a natural choice to define distance, and has been widely used. A naive view is that this distance would still be a good choice if the matrices to be matched are often not positive definite but merely non-negative (which throughout this paper means positive semidefinite). However, even if one assumes that the matrices involved have fixed rank, there exists no clearly distinguished, or we may say 'canonical', metric; it is known that the Frobenius norm does not provide a complete metric space in this case, see e.g. [3]. A further complicating factor is that the matrices that are to be matched, are oftentimes not equally 'reliable', in that some non-negative matrices (or the images which they derive from, for example as Gram matrices; see Section 2) are of higher 'quality' than others. It is this last feature that may prompt us to no longer require that the distance be symmetric.

We shall mainly be concerned with a problem addressed in [4, 5] concerning random tomography with cryo-electron microscopy, in particular the reconstruction of a three-dimensional particle such as a protein fragment—construed as a three-dimensional compactly supported probability density—from its two-dimensional projections at random and, importantly, unknown angles. We refer to Section 2 for relevant details. Properties and theoretical results pertaining to this model, including a shape inversion formula that connects the three-dimensional image with its projections, were provided in [4], and related geometric problems and numerical implementation were investigated in [5], in particular Section 5 and 6 therein. Our problem concerns a single but central step in the tomographic reconstruction procedure. A non-symmetric distance is suggested here, and we argue why it may be better suited to the problem. Our reasons for using an alternative distance are given later in this introduction, after some more motivating material is given. Here we merely note that our distance uses generalized inverses.

A further issue that we address in this paper relates to the space over which the optimization is carried out. The problem that we are focusing on here is a matching problem for certain 'landmarks' of the image (see Subsection 2.2), which gives rise to a combinatorial optimization problem. In [8], a main focus was to investigate when such problems may be solved by *relaxing* the combinatorial problem to an embedding continuous group. That is, the problem is relaxed from the permutation group $\mathfrak{S}_m$ on $\{1,\dots,m\}$, $m \geq 2$, to the orthogonal group $O(m)$. Conditions in corresponding results, such as in [8, Thm. 5, p. 776], are quite stringent and probably not justified in problems such as [4, 5] where, in any case, only a rough estimate for

reconstruction is sought. A more pertinent point for our purposes is the focus on a certain term that is 'mixed' in that it depends on both of the sets to be matched. To explain this point in a simple setting, consider the example from [8, pp. 773–774] of matching two sets of $m \geq 2$ real numbers, say $x_1, \ldots, x_m$ and $y_1, \ldots, y_m$. That is, we choose $\pi \in \mathfrak{S}_m$ such that the sum

$$\begin{aligned} \sum_{i=1}^{m} (x_{\pi(i)} - y_i)^2 &= \sum_{i=1}^{m} (x_{\pi(i)}^2 + y_i^2) - 2 \sum_{i=1}^{m} x_{\pi(i)} y_i \\ &= \sum_{i=1}^{m} (x_{\pi(i)}^2 + y_i^2) - 2 \operatorname{tr}\big(\Pi^{\mathsf{T}} \operatorname{diag}(x_1, \ldots, x_m)\, \Pi \operatorname{diag}(y_1, \ldots, y_m) \big) \end{aligned} \tag{1}$$

is minimised, where tr denotes trace, Π is the permutation matrix representing π in the sense that $\Pi \mathrm{e}_i = \mathrm{e}_{\pi(i)}$ for all $i = 1, \ldots, m$, and e_i is the i'th column of I_m, the $m \times m$ identity matrix. But since the first sum on the right in (1) is independent of π, one may as well maximize the stated trace; and one may then go further by relaxing to the maximization of the function $Q \mapsto \operatorname{tr}\big(Q^{\mathsf{T}} \operatorname{diag}(x_1, \ldots, x_m)\, Q \operatorname{diag}(y_1, \ldots, y_m) \big)$, for $Q \in O(m)$, as analyzed in [8]. Brockett also showed that even if the x_i and y_i are vectors, there are situations where the extremum is attained at a permutation. The trace expression in (1) was further generalized from the orthogonal group to the Stiefel manifold, and termed "Brockett cost function", in [2, Sec. 4.8.1, pp. 80ff].

We are now in a position to give the two reasons for why we are looking for a replacement for distance induced by the Frobenius norm (the analogue of the Euclidean 2-norm for matrices) in the matching problem at hand. First, the images to be matched (strictly speaking, their average Gram matrices) are of differing quality, so a non-symmetric distance, in the spirit of a divergence, seems preferable. Second, there is not only a permutational or combinatorial aspect, as described in the previous paragraph in a simple setting; there is also the aspect of how to reduce from rank three to the correct rank two (since the landmarks, as well as the Gram matrices, are observed in their projections into the two-dimensional imaging plane—see Section 2) by an additional $\mathbb{S}^2$-valued variable. (However, that variable should not be confused with a viewing angle, since it does not directly relate to physical space.) This reduction to rank two introduces a certain 'obliqueness' into the problem, which manifests itself in the fact that a certain important expression only simplifies in a special case, see Remark 4. Our replacement for the Frobenius norm will exhibit a simple form for the impact of the $\mathbb{S}^2$-valued variable. While on the surface it may appear that the remaining mixed term (the analogue of the last term on the right in the first line of (1)) is more complicated than in the Frobenius norm case, we shall see that this there is, to the contrary, a considerable simplification, insofar as the Brockett cost function may be regarded as a 'square', thus lending itself to least-squares type methods on manifolds, after a relaxation is car-

ried out. Along the way, we provide a number of results and insights that we hope make the problem more accessible to future research.

The paper is organised as follows. In Section 2, after giving notation, we recall the setting from [5] (separately discussing the geometric and linear algebraic aspects in Subsections 2.2 and 2.3). Section 3 gives our proposal of alternative distance measure, and main results. Section 4 gives concluding remarks. Ancillary material about parametrization with Givens rotations is given in an appendix.

2. Background

2.1. Notation

For a matrix $W \in \mathbb{R}^{3\times K}$, its associated Gram matrix is $\mathrm{Gram}(W) = W^{\mathsf{T}}W$. All non-negative matrices, that is symmetric matrices A with $x^{\mathsf{T}}Ax \geq 0$ for all x, can be written as Gram matrices, and all matrices in the orbit $\{QW, Q \in O(3)\}$ have the same Gram matrix. Always assuming $K \geq 3$, the noncompact Stiefel manifold $\mathrm{ST}(3,K)$ consists of the matrices in $\mathbb{R}^{K\times 3}$ of full (column) rank, while the compact Stiefel manifold is $\mathrm{St}(3,K) = \{Q \in \mathbb{R}^{K\times 3} : Q^{\mathsf{T}}Q = I_3\}$, where I_p is the $p \times p$ identity matrix; we write $I = I_K$. The tangent space of $Q \in O(m)$ at I_m may be identified with the space $\mathrm{Skew}(m)$ of skew-symmetric matrices, which has dimension $m(m-1)/2$.

Two actions are defined on $\mathrm{ST}(3,K)$: matrix multiplication from the left by $O(K)$, and from the right by $O(3)$. The former action is transitive while the latter is not (as it preserves the column space), and they both have a role to play in the present paper. The non-negative matrices form a homogeneous space of $GL(K)$ under the action

$$G \mapsto \Gamma_W(G) = WGW^{\mathsf{T}}, \qquad W \in GL(K). \tag{2}$$

If $\operatorname{rank} G = 3$, then $G = YY^{\mathsf{T}}$ with $Y \in \mathrm{ST}(3,K)$ (Y is not unique), and the action (2) when restricted to $O(K)$ is the analogue of the action from the left on $\mathrm{ST}(3,K)$.

2.2. *Random tomography model*

The random tomography model on which the present paper is based was introduced in [4]. The particle to be reconstructed was modelled as a three-dimensional probability density $\rho(x_1, x_2, x_3)$, which is assumed to be compactly supported. An observed image of ρ is a (discretized to a regular grid in practice) projection of ρ, given by the random field (bivariate $\mathbb{R}$-valued stochastic process)

$$\check{\rho}(x_1, x_2) = \int_{-\infty}^{+\infty} \rho(U^{-1}x)\, \mathrm{d}x_3\,, \tag{3}$$

which can be thought of the intensity values on the projection plane. Equation (3) exhibits the action of $O(3)$ on three-dimensional densities, which is defined as $(Q\rho)x = \rho(Q^{-1}x)$; and U, which is uniform (Haar) distributed (meaning that UQ, and also QU, has the same distribution as U, for any fixed $Q \in O(3)$) defines the random orientation at which ρ is viewed. The stochastic Radon transform of length $N \geq 1$ is a sample $\{\check{\rho}_1, \ldots, \check{\rho}_N\}$ of N independent and identically distributed (i.i.d.) copies of $\check{\rho}$, generated by using a sample $\{U_1, \ldots, U_N\}$ of N i.i.d. copies of U.

In order to arrive at a tractable model, in [4, 5] the density ρ was modelled as a mixture of translates $\phi(\cdot - y)$ of a spherical (isotropic) Gaussian density ϕ with fixed unknown covariance matrix σI_3, where $\sigma > 0$:

$$\rho(x) = \sum_{k=1}^{K} q_k\, \phi(x - \mu_k) \qquad \text{with } q_k > 0 \text{ for } k = 1, \ldots, K \text{ and } \quad \sum_{k=1}^{K} q_k = 1.$$

The landmarks $\{\mu_k\} \subset \mathbb{R}^3$ may be written in the columns of $W \in \mathbb{R}^{3\times K}$. Due to the ill-posed nature of the tomographic reconstruction problem, the density ρ itself, and in particular the absolute locations of the landmarks in $\mathbb{R}^3$, cannot be recovered. As is known [4, Thm. 3.1, p. 3278], the shape of ρ, that is its equivalence class $\{Q\rho\,, Q \in O(3)\}$, can be recovered. As the associated Gram matrix $\mathrm{Gram}(W) = W^{\mathsf{T}}W$ has the invariance property $\mathrm{Gram}(QW) = \mathrm{Gram}(W)$ for all $Q \in O(3)$, it is a suitable rudimentary descriptor for shape. A 'shape inversion' formula for the expected value of the Gram matrix was given in [4, Thm. 4.1, p. 3285], for any dimension. Moreover, if three orthogonal perfect (noise-free and unblurred) views of an object in $\mathbb{R}^3$ are available, then the object may be reconstructed from these [5, Lemma 5.1, p. 2588].

2.3. Procrustes procedure

We recall the Procrustes procedure in [5, Sec. 6], with one changed definition because the discretisation carried out in [5, p. 2595], which pertains to the u variable appearing in (8) below, will be a focus of attention here. Alongside we will introduce the necessary notation, mostly taken from [5].

We simplify the setting of [5, Sec. 6] by assuming that there are only two classes (Class 1 and Class 2) of images to be considered, distinguished by the observed number of landmarks. The Procrustes procedure in [5, Sec. 6] combines the information from the two classes into a single (albeit rough) starting estimate for tomographic reconstruction. By definition, images in Class 1 are those with $K \geq 3$ landmarks, while images in Class 2 have exactly $K-1$ identified landmarks; also — an assumption which is implicit in [5] — for Class 2 it is always the same landmark which is missing, perhaps the faintest among them. Assume also that there are n_i images available in Class i for $i = 1, 2$. (We will not consider asymptotics in n_1, n_2.)

Write $\hat{\mu}_k^{(i\cdot j)}$, $k = 1, \ldots, K_i$, $j = 1, \ldots, n_i$, $i = 1, 2$, for the k'th landmark in the j'th image of Class i, and define $\hat{\mu}^{(i\cdot j)} \in \mathbb{R}^{3\times K_i}$ as the matrix whose k'th column is $\hat{\mu}_k^{(i\cdot j)}$, with $K_1 = K$ and $K_2 = K - 1$. The first step is to obtain the Gram matrix $\hat{G}_1$ for the Class 1 images. Using the spectral decomposition of $\tilde{G}_1$,

$$\tilde{G}_1 = \frac{3}{2 \cdot n_1} \sum_{j=1}^{n_1} \mathrm{Gram}\big(\hat{\mu}^{(1\cdot j)}\big) = U_1 D_1 U_1^{\mathsf{T}}$$

with $U_1 \in O(K)$ and a diagonal matrix $D_1 = \mathrm{diag}(\sigma_i)$, $\sigma_1 \geq \ldots \geq \sigma_K \geq 0$, let

$$\hat{G}_1 = U_1' D_1' \big(U_1'\big)^{\mathsf{T}}, \tag{4}$$

where $U_1' \in \mathrm{St}(3, K)$ consists of the first three columns of U_1, and $D_1' = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$ contains the first three singular values of $\tilde{G}_1$. (The factor $3/2$ in the definition of $\tilde{G}_1$ comes from [5, Lemma 5.1, p. 2588].) For later purposes we also define $U_1'' \in \mathrm{St}(K, 2)$ as the first two columns of U_1, and $D_1'' = \mathrm{diag}(\sigma_1, \sigma_2)$.

By the Eckart-Young theorem, $\hat{G}_1$ is the nearest rank-three approximation of $\tilde{G}_1$ in the Frobenius norm. As noted in [5, p. 2594], the slight variations of orientations of images in what we call Class 1 encodes *some* three-dimensional information; this is why one does not reduce to rank two from the beginning by considering the rank-two matrix $U_1'' D_1'' \big(U_1''\big)^{\mathsf{T}}$, but reduces instead to the rank-three matrix $\hat{G}_1$. We shall assume that $\sigma_2 > \sigma_3 > \sigma_4 > 0$, thus making the column spaces of U_1' and U_1'' well-defined. (In [5, p. 2594] a matrix V_1' is defined in the context of (4), which

however has a different meaning than in the present paper, and will not be needed.)

We follow [5] in that the missing landmark in Class 2 is construed as being made invisible by another (brighter) landmark, and consider ensembles of means of the form $\big[\hat\mu^{(2\cdot j)}\ \hat\mu_k^{(2\cdot j)}\big]$ for $k = 1,\dots,K-1$ and $j = 1,\dots,n_2$. Let $\{P_\ell, \ell \in \mathcal{L}\}$ denote that subset of $\mathfrak{S}_K$ that has 1 as a fixed point, with $\mathcal{L}$ some index set, and P_ℓ^0 denote the restriction of P_ℓ to $\{2,\dots,K\}$. Let further (with the conjugation map as in (2))

$$G_{\ell k} = \frac{3}{2\cdot n_2}\sum_{j=1}^{n_2} \mathrm{Gram}\big(\big[\hat\mu^{(2\cdot j)}\ \hat\mu_k^{(2\cdot j)}\big]P_\ell\big) = \Gamma_{\Pi_{\ell k}}(\tilde G_2) = \Gamma_{F_\ell}(\tilde G_{2,k})\,, \tag{5}$$

where

$$\tilde G_2 = \frac{3}{2\cdot n_2}\sum_{j=1}^{n_2} \mathrm{Gram}\big(\hat\mu^{(2\cdot j)}\big)\,,\quad \tilde G_{2,k} = \Gamma_{E_k}(\tilde G_2)\,,$$
$$\Pi_{\ell k} = \underbrace{\begin{pmatrix} 1 & 0_{1\times(K-1)} \\ 0_{(K-1)\times 1} & P_\ell^0 \end{pmatrix}}_{=:F_\ell} \begin{pmatrix} I_{K-1} \\ \mathrm{e}_k \end{pmatrix} \in \mathbb{R}^{K\times(K-1)}\,.$$

Here P_ℓ^0 is identified with a permutation in the same manner as for P_ℓ.

Remark 1 *As in [5, p. 2594], the reason for assuming that landmark 1 is a fixed point of the permutation is because the brightest component can be identified reliably in all images; this assumption is not essential to our methodology, but does have a role to play later, see the discussion around* (15) *below.*

Note that

$$\mathrm{tr}(G_{\ell k}) = \frac{3}{2\cdot n_2}\sum_{j=1}^{n_2}\left(\sum_{m=1}^{K-1}\left\|\hat\mu_m^{(2\cdot j)}\right\|^2 + \left\|\hat\mu_k^{(2\cdot j)}\right\|^2\right) \tag{6}$$

depends on k but not on ℓ.

Let $(V_1')^{\mathsf T} \in \mathrm{ST}(3,K)$ be such that $(V_1')^{\mathsf T}V_1' = \hat G_1$. The choice of V_1' fixes the coordinate system for $u \in \mathbb{S}^2$ in what follows; a natural choice is, using notation from around (4),

$$V_1' = \Big(U_1'\big(D_1'\big)^{1/2}\Big)^{\mathsf T} = \big(U_1'\,\mathrm{diag}\big(\sqrt{\sigma_1},\sqrt{\sigma_2},\sqrt{\sigma_3}\big)\big)^{\mathsf T}\,, \tag{7}$$

where $A^{1/2}$ is the symmetric square root of the non-negative matrix A. (This exhibits the polar decomposition of $(V_1')^{\mathsf T}$, see [6, p. 540]. We stick with the transpose in

the definition of V_1' to align with notation in [5].) For $u \in \mathbb{S}^2$, define (correcting transpose sign placements in [5])

$$S(u) = (V_1')^{\mathsf{T}}(I - uu^{\mathsf{T}})V_1' \,. \tag{8}$$

In [5], the (square) Frobenius norm distance was used in the Procrustes procedure:

$$d(G_{\ell k}) = \min_{u\in\mathbb{S}^2} \|G_{\ell k} - S(u)\|_F^2 = \|S(u)\|_F^2 + \|G_{\ell k}\|_F^2 - 2\,\mathrm{tr}\big(G_{\ell k}S(u)\big)\,. \tag{9}$$

The reason for expanding (9) in three terms is to have a comparison with the case of the alternative distance introduced in Section 3. Here we merely observe that no simplification as was seen in (1) is possible.

For $\{u_n\} \subset \mathbb{S}^2$, the matrix S_n in [5, p. 2595] equals our $S(u_n)$. With $(\tilde\ell, \tilde k)$ such that $d(G_{\tilde\ell\,\tilde k}) \le d(G_{\ell k})$ for all (ℓ, k), Panaretos and Konis updated $\hat G_1$ by first computing

$$\tilde G_{12} = \frac{3}{2\cdot(n_1+n_2)}\left[\sum_{j=1}^{n_1} \mathrm{Gram}\big(\hat\mu^{(1\cdot j)}\big) + \sum_{j=1}^{n_2} \mathrm{Gram}\big(\big[\hat\mu^{(2\cdot j)}\ \hat\mu_{\tilde k}^{(2\cdot j)}\big]P_{\tilde\ell}\big)\right], \tag{10}$$

and then again using spectral decomposition and rank-three reduction to obtain the updated Gram matrix $\hat G_{12}$.

In the case of the Frobenius distance at hand, just as with our own distance proposed in Section 3, it is of interest to replace the combinatorial optimization over $\{P_\ell, \ell \in \mathcal{L}\}$ by a continuous one over $O(K)$. That is, we replace $G_{\ell k} = \Gamma_{F_\ell}(\tilde G_{2,k})$ by an arbitrary conjugation $\Gamma_Q(\tilde G_{2,k})$ with $Q \in O(K)$. By [10, Thm. VI.4.3, p. 167] (which, as noted there, does cover the case of Hermitian and hence of real symmetric matrices; recall that the latter matrix space may be identified with its own tangent space at I), for fixed k, any (even local) minimum of (9) in the orbit $\{\Gamma_Q(\tilde G_{2,k}) : Q \in O(K)\}$ will be attained at a matrix which commutes with $S(u)$. Any permutation that is applied to both that minimum and $S(u)$ will not change the minimum value, and so we have the following.

Proposition 2 *The number of permutations ℓ that are components of a local minimum $(\tilde u, \tilde k, \ell)$ of (9) is equal to $K!/6$, the cardinality of $\mathfrak{S}_K/\mathfrak{S}_3$.*

Proposition 2 casts doubts on whether, further to the basic task of updating the Gram matrix via equation (10), the harder task of finding correspondences between individual landmarks (the "by-product" in [5, p. 2596]) may be accomplished.

3. Main results: A divergence-type distance and orthogonal relaxation

In order to motivate our definition of alternative distance to replace $\|\cdot\|_F$, assume at first that A, B are (symmetric and) positive definite, and let

$$\begin{aligned} D(A;B) &= \|A^{1/2} - BA^{-1/2}\|_F^2 = \big\|A^{1/2}(I - A^{-1/2}BA^{-1/2})\big\|_F^2 \\ &= \operatorname{tr}\Big\{\big(I - A^{-1/2}BA^{-1/2}\big)^{\mathsf{T}} A \big(I - A^{-1/2}BA^{-1/2}\big)\Big\} \\ &= \operatorname{tr}\Big\{A - A^{-1/2}BA^{1/2} - A^{1/2}BA^{-1/2} + A^{-1/2}B^2A^{-1/2}\Big\} \\ &= \operatorname{tr}(A) - 2\operatorname{tr}(B) + \operatorname{tr}\big(B^2A^{-1}\big)\,. \end{aligned} \tag{11}$$

In [7, Eqn. (58), p. 560], the distance D appeared as a (non-stochastic) time derivative of an L^2 distance between two means in what that author called a stochastic localization scheme, where A and B are subject to a certain coupled system of stochastic differential equations. The reason for using the derivative, rather than the function itself, is that the (degenerate) densities whose covariance matrices are given by A and B already have mean zero, by the definition of the Gram matrices. An alternative interpretation of D is given in Section 4.

In our case, the participating matrices $A = S(u)$ and $B = G_{\ell k}$ are only non-negative, and thus the first problem is to find a replacement for the non-existent ordinary inverse of A. Recall e.g. from [9] that there are several notions of generalized inverse, several of which are defined through the four Penrose equations

$$AXA = A\,, \quad XAX = X\,, \quad (AX)^{\mathsf{T}} = AX\,, \quad (XA)^{\mathsf{T}} = XA\,. \tag{12}$$

For $M \subseteq \{1,\dots,4\}$, one says that X is an M-inverse of A if X satisfies the r'th equation in (12), for every $r \in M$, and writes $X = A^{(\cdots)}$ with the elements of M listed inside the $(\cdots)$ parentheses. The commonly used Moore-Penrose (MP) inverse or pseudo-inverse $X = A^{\dagger}$, which corresponds to $M = \{1,\dots,4\}$, is uniquely determined. However, in most of what follows we shall use the specific $\{1\}$-inverse (here again $u \in \mathbb{S}^2$)

$$\begin{aligned} S^{(1)}(u) \equiv \big(S(u)\big)^{(1)} &= U_1'(D_1')^{-1/2}(I - uu^{\mathsf{T}})(D_1')^{-1/2}(U_1')^{\mathsf{T}} = C_u C_u^{\mathsf{T}}\,, \\ C_u &= U_1'(D_1')^{-1/2}(I - uu^{\mathsf{T}})\,. \end{aligned} \tag{13}$$

This inverse is completely determined given the decomposition of $S(u)$ as in (7) and (8). It is readily checked that (13) does not commute with $S(u)$, and thus is not a $\{1,3\}$-inverse of $S(u)$. Note that the set of all $\{1,3\}$-inverses of $S(u)$ is given by all solutions of X of $S(u)X = S(u)S^{(1)}(u)$ [9, Thm. 3, p. 48], which includes

non-symmetric matrices. However, restricting attention to symmetric X has the major advantage that we have an explicit quadratic expression available for the most complicated term. Indeed, (11) becomes

$$\begin{aligned} D(S(u);G_{\ell k}) &= \operatorname{tr}(S(u)) - 2\operatorname{tr}(\tilde{G}_{2,k}) + \operatorname{tr}\big(G_{\ell k}^2 S^{(1)}(u)\big) \qquad (14)\\ &= \operatorname{trace}\big(D_1'(I-uu^{\mathsf{T}})\big) - 2\cdot(\text{a sum independent of } \ell) + \|G_{\ell k}C_u\|_F^2 \\ &= \sum_{i=1}^{3}(1-u_i^2)\sigma_i \ -2\cdot(\text{a sum independent of } \ell) + \|\tilde{G}_{2,k}F_\ell^{\mathsf{T}}C_u\|_F^2\,, \end{aligned}$$

where the "sum" term is given in (6).

We now replace the combinatorial problem of finding the optimal $\tilde{\ell}\in\mathcal{L}$ by the calculus problem of finding the optimal $Q\in O(K)$ in the expression $D(S(u);\Gamma_Q(\tilde{G}_{2,k}))$; recall the definitions in (2) and (5). There is still a combinatorial problem left in the k variable, but the computational complexity of this scales linearly in K and is thus not a serious numerical problem. Note that only the term $\|\tilde{G}_{2,k}Q^{\mathsf{T}}C_u\|_F^2$ depends on $Q\in O(K)$. The term $\sum(1-u_i^2)\sigma_i$ is a penalty that favors the deviation from the non-Procrustes approach of just taking $u=\mathrm{e}_3$, for then this term is even maximal.

Methodology for least-squares methods of the above kind is described in [2, Sec. 8.4, pp. 184ff]. This requires that the number of (independent) equations be at least the same as the number of (manifold) 'variables'. The arguments in this paragraph apply to all distances, whether $\|\cdot\|_F^2$, D, or other. As the factor $I-uu^{\mathsf{T}}$ means that the rank of $\tilde{G}_{2,k}Q^{\mathsf{T}}C_u\in\mathbb{R}^{K\times 3}$ is deficient and $\mathbb{S}^2$ contributes two scalar variables, this requirement translates to

$$\operatorname{card}(\mathcal{L}) + 2 \overset{!}{\leq} 2K\,. \qquad (15)$$

Since the permutations in $\mathcal{P}_\ell$ fix a landmark, and interchanging two others has no effect, those permutations, viewed as a subgroup of $\mathfrak{S}_K$, have as smallest continuous embedding group the quotient group $O(K-1)/O(2)$, which has dimension $K(K-3)/2$. (In the appendix we suggest a convenient parametrization of this group.) Hence (15) becomes

$$\frac{K(K-3)}{2} + 2 \overset{!}{\leq} 2K \quad\Longleftrightarrow\quad K^2+4\leq 7K\,. \qquad (16)$$

This is satisfied for $K\leq 6$, making the analysis given in [5, Sec. 6] as well as that with our distance D more valid from the theoretical viewpoint. We formulate a stronger statement as the following conjecture.

CONJECTURE 3 *In the class of all possible landmark configurations (modulo a set of measure zero), the reconstruction problem exhibits a 'phase transition' to being more severely ill-posed, starting from the value* $K = 7$.

REMARK 4 *In the case* $u = \mathrm{e}_3$, *where Procrustes is not used, the MP inverse in* (14) *may be explicitly computed, by McDuffee's theorem [9, Thm. 5, p. 42]; this is also the so-called Drazin inverse [9, Thm. 8, p. 148]:*

$$S^\dagger(\mathrm{e}_3) = U_1''(D_1'')^{-1}(U_1'')^\mathsf{T} = CC^\mathsf{T}\,, \quad \textit{where} \;\; C = U_1''(D_1'')^{-1/2}(U_1'')^\mathsf{T}. \tag{17}$$

Thus

$$\begin{aligned} D(S(e_3); G_{\ell k}) &= \operatorname{tr}(S(e_3)) - 2\operatorname{tr}(G_{\ell k}) + \operatorname{tr}\big(G_{\ell k} S^\dagger(e_3) G_{\ell k}\big) \\ &= \sigma_1 + \sigma_2 - 2 \cdot \textit{(a sum independent of } \ell\textit{)} + \|G_{\ell k} C\|_F^2 \\ &= \sigma_1 + \sigma_2 - 2 \cdot \textit{(a sum independent of } \ell\textit{)} + \|\tilde{G}_{2,k} F_\ell^\mathsf{T} C\|_F^2\,, \end{aligned}$$

where the "sum" term is again given in (6).

As our last result, we prove that D is essentially the only distance that enjoys basic desirable properties encoded in the following decomposition:

$$D(S(u); G) = f(U_1', D_1', u, G) = \varphi(D_1', u) + \psi(D_2') + \|G_{\ell k} \cdot \theta(U_1', D_1', u)\|_F^2\,,$$

where φ and ψ are $\mathbb{R}$-valued functions on their respective domains, with $\varphi(D_1', \mathrm{e}_3) = \varphi_0(D_1'')$ for some function φ_0; θ is a 'simple' and matrix-valued function; (U_1, D_1') is the polar decomposition of the natural Stiefel factor (computed as in (7)) of the initial Gram matrix (the matrix $\hat{G}_1$ from earlier); u is a unit vector (which, as we noted before (7), is well-defined); and G is a non-negative matrix added in during the matching or Procrustes step (the matrix $G_{\ell k}$ from earlier), with eigenvalues encoded in D_2'. In order to circumvent problems that arise from non-uniqueness, we use the Moore-Penrose inverse instead of the $\{1\}$-inverse. From MacDuffee's theorem, as well as the property $(Q^\mathsf{T} A Q)^\dagger = Q^\mathsf{T} A^\dagger Q$ for $Q \in O(K)$ (see [9, Exercise 25, p. 43]), one has

$$D(A; B) = D(\Gamma_X(A); \Gamma_X(B)) \tag{18}$$

for any $X \in [0, \infty) \times O(K)$, the group of scalings, rotations and reflections. (The element (c, Q) in this group is identified with cQ.) For the general linear group $GL(K)$, equation (18) would be false. A similar restriction was also found to be a necessary compromise in [11] in the context of axiomatically defining a sensible mean of non-negative matrices. As the last preliminary observation, we note that (18) expresses

a kind of 'point-pair invariance' of D when regarded as a kernel function on the product of non-negative matrices, analogous to [12, p. 29].

PROPOSITION 5 *Consider (possibly asymmetric) distances of the form* $d(A;B) = \|g(A,B)\|_F^2$, *where* $A = A(u)$ *and* $B = B(Q)$ *with* $Q \in O(K)$ *and* $u \in \mathbb{S}^2$, *and* g *is a 'simple' bivariate power function of two non-negative matrices as exhibited on the left-hand side in* (19) *below. Up to symmetries, the only distance which has only one term depending on* $Q \in O(K)$, *and simple explicit dependence on* u *in the 'non-mixed' term involving only* A, *is* D.

Proof. For $\alpha, \beta, \gamma, \delta \in \mathbb{R}$, we have

$$\|A^\alpha - B^\beta A^\gamma B^\delta\|_F^2 = \operatorname{tr}\big(A^{2\alpha}\big) - 2\operatorname{tr}\big(A^\alpha B^\beta A^\gamma B^\delta\big) - \operatorname{tr}\big(B^{2\beta} A^\gamma B^{2\delta} A^\gamma\big)\,. \tag{19}$$

First, (18) implies $\alpha = \beta + \gamma + \delta$. Now, if $\gamma = 0$, then we have a power version of the distance from [5], which does not lend itself to further simplification. There remain two other, essentially symmetric, cases where the last term on the right in (19) simplifies, namely $\beta = 0$ (and then necessarily $\delta > 0$), and $\delta = 0$ (and then necessarily $\beta > 0$). If $\alpha = 0$ then in the first case, the second and third terms on the right have sum $\|B^{\delta/2} A^{\gamma/2}\|_F^2 + \|B^\delta A^\gamma\|_F^2$, and similar in the second case; such two mixed terms are highly inconvenient from both the theoretical as well as the numerical viewpoint. If $\alpha = 1/2$, then the cases $\beta = 0$ and $\delta = 0$ again exhibit a symmetry. The latter case yields

$$\operatorname{tr}(A) - 2\operatorname{tr}\big(A^{\gamma+1/2} B^\beta\big) - \|B^\beta A^\gamma\|_F^2\,.$$

The second term simplifies according to the stipulations in the formulation of the proposition if, and only if, $\beta = 1$ and $\gamma = -1/2$, which yields $d = D$. □

4. Conclusion

This paper, besides drawing attention to a possible better alternative to the simple Frobenius norm, intended to lay groundwork for a systematic study of the Procrustes procedure of [5], by formulating it as an optimization problem on a manifold. As noted in Section 1, one of the aims of [8] was to give assumptions that allow one to replace a combinatorial optimisation over $\mathfrak{S}_K$ by a continuous optimisation on the embedding manifold and continuous Lie group $O(K)$; that is to relax the problem. What lends weight to the use of the relaxation in our case is the presence of the additional u variable. Indeed, a small change in u may be seen as a small change

in the rotation in $O(K)$, acting from the left side of $(V_1')^{\mathsf{T}}$, though the required rotation in $O(K)$ in general neither needs to exist nor needs to be unique.

Here is an alternative way of understanding the distance D.

REMARK 6 *In the terminology of [13], the block matrix* $M = \begin{pmatrix} A & B \\ B & BA^{(1)}B \end{pmatrix}$ *is a* Schur compression *of the* (*generally indefinite*) *matrix* $\begin{pmatrix} A & B \\ B & A \end{pmatrix}$. *If trace is viewed as 'energy', then our* D *is the energy of the difference* $M - \mathrm{diag}\left(\begin{pmatrix} B & 0 \\ 0 & B \end{pmatrix}\right)$, *where* $\mathrm{diag}(X)$ *is the matrix which replaces off-diagonal elements of* X *by zeros.*

A main idea in the present paper was to write the Brockett cost function (see Section 1) as a function $\|g(u,Q)\|_F^2$ for $(u,Q) \in \mathbb{S}^2 \times O(K)$, where g takes values in $\mathrm{ST}(3,K)$. The Frobenius norm squared is of course just about the simplest function one can combine with g. We hope that our ideas may find applications in problems featuring a more complicated function of g. There may even be applications that just use the compact Stiefel manifold $\mathrm{St}(3,K)$. Such applications arise for example in astronomy; see [1, pp. 285] for an exposition from a statistical viewpoint.

References

[1] Mardia, K.V. and Jupp, P.E. Directional Statistics. Wiley Series in Probability and Statistics, Wiley, Chichester, 2000.

[2] Absil, P.-A., Mahony, R., Sepulchre, R. Optimization algorithms on matrix manifolds. Princeton University Press. Princeton, NJ, 2008.

[3] Vandereycken, B., Absil, P.-A., and Vandewalle, S. A Riemannian geometry with complete geodesics for the set of positive semidefinite matrices of fixed rank. *IMA J. Numer. Anal.* **33** (2013), 481–514.

[4] Panaretos, V. On random tomography with unobservable projection angles. *Ann. Statist.* **37** (2009), 3272–3306.

[5] Panaretos, V. and Konis, K. Sparse approximations of protein structure from noisy random projections. *Ann. Appl. Statist.* **5** (2011), 2572–2602.

[6] Golub, G., van Loan, C.F. Matrix computations. Fourth edition. The Johns Hopkins University Press, Baltimore, 2013.

[7] Eldan, R. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geom. Funct. Anal.* **23** (2013), 532–569.

[8] Brockett, R.W. Least squares matching problems. *Linear Algebra Appl.* **122/123/124** (1989), 761–777.

[9] Ben-Israel, A., Greville, T. Generalized inverses. Theory and applications. Second edi-

tion. CMS Books in Mathematics, vol. 15. Springer, NY, 2003.
[10] Bhatia, R., Matrix analysis. Graduate Texts in Mathematics, vol. 169. Springer, NY, 1997.
[11] Bonnabel, S., Collard, A., and Sepulchre, R. Rank-preserving geometric means of positive semi-definite matrices. *Linear Algebra Appl.* **438** (2013), 3202–3216.
[12] Terras, A. Harmonic analysis on symmetric spaces—higher rank spaces, positive definite matrix space and generalizations. Second edition. Springer, NY, 2016.
[13] Ando, T. Generalized Schur complements. *Linear Algebra Appl.* **27** (1979), 173–186.

Appendix A. Parametrization with Givens rotations

In this appendix we suggest a parametrization of $O(K)$ that seems to combine best with the relaxation of the problem at hand. Choosing a parametrization is important for the algorithms discussed in [2], see in particular Algorithm 14, p. 186 for least-squares, as needed in the present paper. Roughly speaking, a retraction is a modification of Newton or other line-search algorithm to project the endpoint of a segment, which in general is outside the manifold of interest, back to the manifold. See [2, Sec. 4.1, pp. 54ff] for a precise description and details.

We suggest to use the following retraction defined in [2, pp. 58–59]. For $1 \le i < j \le K$ and $\theta \in [0, 2\pi)$, the Givens rotation $G(i,j,\theta)$ in $\mathbb{R}^K$ is the identity matrix with the substitutions $\mathrm{e}_i^\mathsf{T} G(i,j,\theta)\mathrm{e}_i = \mathrm{e}_j^\mathsf{T} G(i,j,\theta)\mathrm{e}_j = \cos\theta$ and $\mathrm{e}_i^\mathsf{T} G(i,j,\theta)\mathrm{e}_j = -\mathrm{e}_j^\mathsf{T} G(i,j,\theta)\mathrm{e}_i = \sin\theta$, so that right-multiplication by $G(i,j,\theta)$ is counter-clockwise rotation by the angle θ in the plane spanned by the ordered pair $\{\mathrm{e}_i, \mathrm{e}_j\}$. Define

$$\mathrm{Giv}_{-k}(\Omega) = \prod_{\substack{2 \le i < j \le K \\ (i,j) \ne (k,K)}} G(i,j,\Omega_{ij})\,, \qquad k = 1, \dots, K-1\,, \tag{A1}$$

where $\Omega = (\Omega_{ij}) \in \mathrm{Skew}(m)$, and the order of multiplication is any fixed order, and define $\hat\nu_m^{(1)}$ as the m'th column of V_1'; we may call $\hat\nu_m^{(1)}$ the m'th 'pseudo-landmark' (this being a landmark that is derived from an averaging process). Note that the exclusion of the (k,K) factor in (A1) is because we do not need to relax to the continuum the permutation which just interchanges the two pseudo-landmarks $\hat\nu_k^{(1)} = \hat\nu_K^{(1)}$. The retraction R_Q at $Q \in O(K)$ is now defined as

$$R_Q(Q\,O(K)) = Q\mathrm{Giv}_{-k}(O(K))\,,$$

for $Q \in O(K-1)$ canonically embedded into $O(K)$ as the replacement for the right-hand corner of I.