For isotropic offset normal shape distributions, covariance distance is proportional to Procrustes distance

Fred L. Bookstein

University of Vienna, University of Washington

Earlier this year Philipp Mitteroecker and I published an application of one particularly natural distance function ρ between positive definite matrices, the formula $\rho^2(S_1, S_2) = \sum(\log \lambda_i)^2$ where the λ_i are the relative eigenvalues of S_2 with respect to S_1 (or vice-versa). Ian Dryden and friends (*Annals of Applied Statistics,* forthcoming; also, probably, the preceding talk in these Proceedings) usefully tabulate a variety of distance measures between covariance structures. The suggestion here is, in their notation, d_R, the "Riemannian" formula, as that is the only one that is affine-invariant, unchanged when matrices $S_1,\ S_2$ are replaced by $AS_1A^t,\ AS_2A^t$. For our applications, which lie in auxology (the study of organismal growth) and evolution, this was the principal property the metric needed to have. The requirement arises because we wished our analysis to place iterations of a constant biological growth programme at a constant spacing along straight lines in the metric: it must satisfy $\rho(I, A^m) = \rho(A^k, A^{k+m}) = m\rho(I, A)$ for any $k, m,$ and positive definite A. For scalars, this requirement is for a function f with $f(x^n) = nf(x)$, the familiar characterization of the logarithm, and in fact the formula $\rho^2(S_1, S_2) = \sum(\log \lambda_i)^2$ reduces to a multidimensional version of a classic loglinear machine, the slide rule, that may be familiar to those born before, say, 1955.

The difference between the Dryden metric and ours nicely exemplifies one general point in the philosophy of applied statistics that is not discussed at these LASR meetings as much as it should be: that the connection between the structures of statistics and the structures of a quantitative science is metaphorical, not itself necessarily either scientific or statistical. The word "distance" here is a metaphor standing for the scientific intuition of *dissimilarity.* I have criticized this metaphor at length in an earlier LASR presentation (Bookstein, 2007) that, alas, mystified many of its listeners. The $\sum(\log \lambda_i)^2$ metric arises naturally in a context where covariance matrices are operators as well as sample descriptions (the Lie algebra setting). This is the case in applications to biological form — the best factor models for growth can be notated in this way — but not to diffusion tensor analysis, as far as I can tell. The contrast of distance formulas between Dryden and myself is a consequence, then, of the difference in scientific contexts. The same distinction is reflected in the titles of our publication. For the Dryden trio, the topic is statistics for covariance matrices, with "applications" to diffusion tensor imaging; for Mitteroecker and me, the topic is organismal biology, with the notoriously model-free principal coordinates of the covariance metric playing the role of a sturdy pattern-recognition tool. (The principal coordinates introduced by Gower in 1966 are the most convenient visualization of some explanatory models — *stages of growth* — that arose at the dawn of morphometrics and that are known to be promising on biological grounds of long standing.)

Consider now one typical setting of Procrustes analysis, the study of Gaussian models $N(\mu, \Sigma)$ across two or more samples (species, ages, experimental conditions) of organisms. One common style of investigation is of the means μ_i of different groups; another is of the corresponding covariances Σ_i; yet a third is of the possible relationship between μ and Σ as we survey the range of samples. In the Procrustes setting, Σ is not a general covariance structure, but is deficient in rank by the 4 (for 2D data) or 7 (for 3D data) constraints of the linearized Procrustes formalism. Hence whenever comparisons of Σ are among the concerns, it would be worth knowing whether there is some intrinsic "uncertainty principle" according to which the covariance structures of Procrustes analyses at different mean forms are necessarily different themselves. Such a conjecture is reasonable, as the constraints on the null space of Σ are phrased precisely in terms of the Procrustes mean form μ itself (see the matrix J below). We did not expect the simple functional form of this dependence, the straightforward linear proportionality.

Following is a proof that on the offset isotropic normal model, in the limit of small digitizing variance and small Procrustes distance, a slight modification of the Mitteroecker-Bookstein covariance metric $\sqrt{\Sigma(\log \lambda_i)^2}$ is proportional to the Procrustes metric. The result means, in effect, that one cannot put the default (offset isotropic) Mardia–Dryden distributions at two mean shapes down any closer together than the Procrustes distance between those means — between interpretations of variance-covariance structures on two sets of shape data there is an irreducible incommensurability with magnitude equal to the magnitude of the shape difference between the means around which those variance-covariance structures are calibrated. Put even more simply, there is a limit on the precision with which one can test covariance matrices around different mean shapes for equality or proportionality, regardless of sample size.

To start, we need to extend the definition of the covariance metric to accommodate the intentionally reduced rank of the Procrustes shape coordinates. While the Helmertizing dimensions are the same everywhere, the dimensions for size standardization and rotation are different from mean form to mean form. In the *Evolution* publication there is reference to rank that is reduced by reason of sample size, but not by reason of a mean-dependent construction like this. That is in fact the algebraic annoyance that the present note sets out to solve.

Write $RI(M, N)$ for the relative eigenvalues of two symmetric square matrices M, N whenever N is nonsingular. ("RI" is intended to be read, in a multilingual pun, as "relative eigenvalues.") Write $RI(M) = RI(M, I)$ for the vector of the usual eigenvalues of M by itself. For any square symmetric M, N define $ERI(M, N, \alpha)$ as $RI(\alpha I + M, \alpha I + N)$. ("ERI" stands for "extended relative eigenvalues.") This is a ridge version of RI; it is sufficient for our purposes to leave α indeterminate, because ERI will turn out to be proportional to Procrustes distance for every feasible value of α. We will consider, among others, $\alpha = 1$ and $\alpha = (\sqrt{17} - 1)/2$.

Approximate the offset isotropic normal Procrustes coordinates by projecting out size and rotation instead of dividing or rotating. (The Helmertization step, which is not an approximation, is unchanged.) LASR regulars have seen this notation before. It invokes a **mean shape** μ on k landmarks $\mu_1 = (x_1, y_1), \mu_2 = (x_2, y_2), \ldots, \mu_k = (x_k, y_k)$ vectorized as $vec(\mu) = (x_1, y_1, x_2, y_2, \ldots, x_k, y_k)$ in two dimensions. μ is standardized as $\Sigma x_i = \Sigma y_i = \Sigma x_i y_i = 0$, $\Sigma(x_i^2 + y_i^2) = 1$. (Meaning: μ is centered, its Centroid Size is 1, and it has been rotated to principal axes horizontal and vertical.)

Consider the matrix

$$J = \begin{pmatrix} \delta & 0 & \delta & 0 & \dots & \delta & 0 \\ 0 & \delta & 0 & \delta & \dots & 0 & \delta \\ -y_1 & x_1 & -y_2 & x_2 & \dots & -y_k & x_k \\ x_1 & y_1 & x_2 & y_2 & \dots & x_k & y_k \end{pmatrix}$$

where $\delta^2 = 1/k$, and write $J^i, i = 1, \dots, 4$, for the i^{th} row of J.

- J is orthonormal.

Sample configurations C of k landmark locations from the distribution $N(\mu, \sigma^2 I)$, treated as column vectors, and consider the various derived configurations $C^j = C - \Sigma_{i=1}^{j} (J^i C)(J^i)^t, \quad j = 2, 3, 4$. Because J is orthonormal, each C^i is spherically symmetric within its subspace.

- C^2 is centered in both even-numbered and odd-numbered coordinates.
- C^3 is centered and has been rotated to a position of zero torque against μ.
- The fourth row of J scales each C to centroid size 1 (approximate, to second order in σ). Hence C^4 is, approximately, the vector of Procrustes shape coordinates for the distribution C.
- For small σ, because $\log(1+\epsilon) \sim \epsilon$, C^3 is a rotation of the matrix that augments C^4 by log Centroid Size; hence it is a representation of the size-shape of C.

The distribution C^4 has rank $2k - 4$. It approximates the Procrustes shape coordinates in the vicinity of the mean shape μ.

Write $V(\mu)$ for the covariance matrix $I_{2k} - (J^1)^t(J^1) - (J^2)^t(J^2) - (J^3)^t(J^3) - (J^4)^t(J^4)$ of the distribution C^4. This is the limiting covariance matrix for the complete set of Procrustes shape coordinates for the offset isotropic distribution of shape in the vicinity of any mean form μ. It too has rank $2k - 4$.

In this same notation, the squared Procrustes distance $|\mu - \nu|^2$ between two (unnormalized) samples of the offset isotropic normal is equal to the squared Euclidean distance between their vectorizations minus the sum of squared differences of projections on the J's. Again for the limiting case of sufficiently small σ^2, we have $|\mu - \nu|^2 = \|vec(\mu) - vec(\nu)\|^2 - (vec(\mu) - vec(\nu))^t \big(J^{1t}J^1 + J^{2t}J^2 + J^{3t}J^3 + J^{4t}J^4\big)(vec(\mu) - vec(\nu))$ where $|\cdot|$ is Procrustes distance and $\|\cdot\|$ is Euclidean distance.

Now $RI(I + \epsilon\ \nu\nu^t) = (1, 1, \dots, 1, 1 + \epsilon|\nu|^2)$. (The nontrivial eigenvector is in the direction of ν.) Also, if μ and ν are perpendicular, $RI(I + \epsilon\ \mu\mu^t + \eta\ \nu\nu^t) = (1, 1, \dots, 1, 1 + \epsilon|\mu|^2, 1 + \eta|\nu|^2)$. Furthermore, for sufficiently small ϵ, $RI(I, I + \epsilon M) \sim RI(I - \epsilon M, I)$ for any suitable symmetric M and ϵ small, and (as a corollary) for sufficiently small ϵ, $RI(I + \epsilon N, I + \epsilon M) \sim RI(I + \epsilon N - \epsilon M, I)$.

Combining these, we arrive at the **Helpful Approximation**

$$RI\big(I - \epsilon\, vec(\mu)vec(\mu)^t + \epsilon\, vec(\nu)vec(\nu)^t\big) \sim 1 \pm \epsilon|\mu - \nu|$$

for μ and ν registered Procrustes mean shapes with $|\mu - \nu| << 1$, and ϵ small. (Setting aside the term in I for the moment, $x^t\big(vec(\nu)vec(\nu)^t - vec(\mu)vec(\mu)^t\big)x$ is a simple difference of squared lengths of projections. As a function of unit vectors x, it is a maximum for x along one bisector of the angle bisectors of μ and ν and a minimum along the other bisector. These differences evaluate to $\pm|\mu - \nu|$. See the figure.)

Write $CD(M, N, \alpha) = \sum \log(ERI(M, N, \alpha))^2$ as in our *Evolution* paper. ("CD" is for "covariance distance" [squared].) For small ϵ and any symmetric M, using the familiar identity for the logarithm of quantities near 1, we have $CD(I + \epsilon M, I, 0) \sim \epsilon^2 \sum(RI(M) - 1)^2$.

As a final bit of notation, for any complex vector μ, write $\mu_\perp$ for the k-vector with entries equal to the entries of μ rotated by $\pi/2$, so that $vec(\mu_\perp) = (-y_1, x_1, -y_2, x_2, \ldots, -y_k, x_k) = J^3$ for $vec(\mu) = (x_1, y_1, x_2, y_2, \ldots, x_k, y_k) = J^4$ as above.

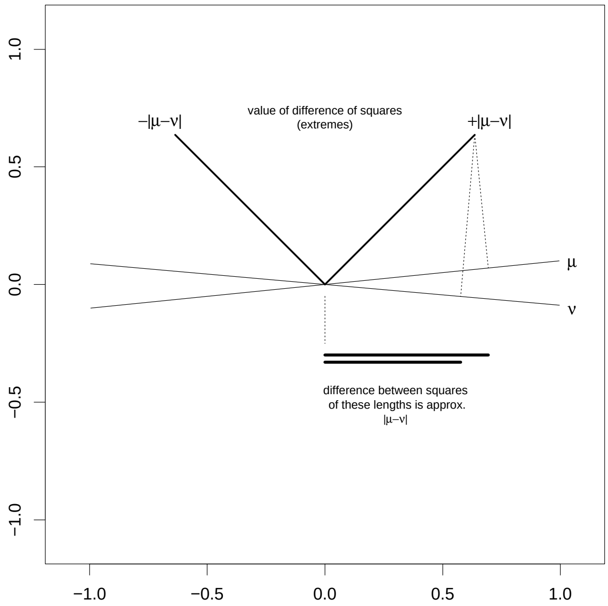
Theorem. On the offset isotropic normal model, in the limit of small digitizing variance σ^2 and for mean shapes μ and ν that are sufficiently close together, the covariance metric $CD(V(\mu), V(\nu), \alpha)$ is proportional to squared Procrustes distance $|\mu - \nu|$, and is equal to it for one particular value of α.

Proof. To illustrate the algebra we proceed for a while with $\alpha = 1$. We have $ERI(V(\mu), V(\nu), 1) = RI\big(2I_{2k} - (J^1)^t J^1 - (J^2)^t J^2 - vec(\mu)vec(\mu)^t - vec(\mu_\perp)vec(\mu_\perp)^t, 2I_{2k} - (J^1)^t J^1 - (J^2)^t J^2 - vec(\nu)vec(\nu)^t - vec(\nu_\perp)vec(\nu_\perp)^t\big) = RI\big(2I_{2k} - vec(\mu)vec(\mu)^t - vec(\mu_\perp)vec(\mu_\perp)^t, 2I_{2k} - vec(\nu)vec(\nu)^t - vec(\nu_\perp)vec(\nu_\perp)^t\big) \sim RI(2I - vec(\mu)vec(\mu)^t - vec(\mu_\perp)vec(\mu_\perp)^t + vec(\nu)vec(\nu)^t + vec(\nu_\perp)vec(\nu_\perp)^t, 2I)$. The last equality is because the eigenvalues along J^1 and J^2 (the dimensions for centering) are 1.0, of log 0.0, and thus do not contribute to the CD scoring, and the last approximation comes from transferring two rank-one terms from one side to the other.

The terms for μ and $\mu_\perp$ and for ν and $\nu_\perp$ are perpendicular, and so we apply our Helpful Approximation twice, once for μ, ν and once for their perpendiculars. There result two eigenvalues of $1 + |\mu - \nu|/\sqrt{2}$ and two of $1 - |\mu - \nu|/\sqrt{2}$. The term in $\sqrt{2}$ arises because in comparison to the gauge metric $2I$ the rank-1 matrices $\mu\mu^t$, etc., are no longer outer products of unit vectors.

Conversion to logarithmic form yields two copies each of $\pm|\mu - \nu|/\sqrt{2}$. Squaring, summing, and taking the square root gives us $\sqrt{2}|\mu - \nu|$, proportional to $|\mu - \nu|$, as was to be proved. Following through on the role of α, we see that a value of $\alpha = 3$ would result in a sum of four terms each $|\mu - \nu|^2/4$ in the formula, and thereby equality between $CD(V(\mu), V(\nu), 3)$ and $|\mu - \nu|^2$, as claimed.

The preceding applies approximations with ϵ effectively equal to 0.5, which is not actually a very small quantity. A more conscientious computation retains the symmetry of the geometry but adjusts the ratio between squared Procrustes distance and squared covariance distance from $4/(\alpha + 1)$ to $4(\alpha^{-1} - (\alpha + 1)^{-1})$. Equality is achieved, then, for $\alpha = \sqrt{17/4} - .5 \sim 1.56...$. This is easily confirmed numerically.



Acknowledgement

This is my first publication under my new Vienna title. I am grateful for the help of my collaborator Philipp Mitteroecker and for ten years of support from the Austrian Ministry of Science and Research via grants to Prof. Horst Seidler. I also acknowledge support from the 6th European Union Framework Programme under Marie Curie grant MRTMN-CT-2005-019564, EVAN (European Virtual Anthropology Network), to the University of Vienna and fourteen collaborating sites.

References

F. L. Bookstein, 2007. Shape coordinates or Procrustes distances? Pp. 41–45 in S. Barber, P.D. Baxter, and K.V.Mardia, eds., *Systems Biology and Statistical Bioinformatics*, Leeds University Press.

Dryden, I. L., A. Koloydenko, and D. Zhou, 2009. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics,* to appear.

P. Mitteroecker and F. L. Bookstein, 2009. The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. *Evolution* 63:727-37.