Making Matrices Better:

Geometry and Topology of Singular Value and Polar Decomposition

Dennis DeTurck, Amora Elsaify, Herman Gluck, Benjamin Grossmann, Joseph Ansel Hoisington, Anusha M. Krishnan, and Jianru Zhang

ABSTRACT. Our goal is to see the space of matrices of a given size from a geometric and topological perspective, with emphasis on the families of various ranks and how they fit together. We pay special attention to the nearest orthogonal neighbor and nearest singular neighbor of a given matrix, since both play central roles in matrix decompositions, and then against this visual backdrop examine the polar and singular value decompositions and some of their applications.

Figure 1 is the kind of picture we have in mind, in which we focus on 3×3 matrices, view them as points in Euclidean 9-space $\mathbb{R}^9$, ignore the zero matrix at the origin, and scale the rest to lie on the round 8-sphere $S^8(\sqrt{3})$ of radius $\sqrt{3}$, so as to include the orthogonal group $O(3)$.

The orthogonal group $O(3)$ has two components: $SO(3)$, where the determinant is $+1$, and $O^-(3)$, where the determinant is -1. The first, $SO(3)$, is a real projective 3-space $\mathbb{R}P^3$ at the core of its open neighborhood N of nonsingular matrices of positive determinant. The second, $O^-(3)$, is another $\mathbb{R}P^3$ at the core of its open neighborhood N' of nonsingular matrices of negative determinant. The common boundary of these two neighborhoods on the 8-sphere is the 7-dimensional algebraic variety V^7 of singular matrices, that is, of determinant 0 and rank 1 or 2.

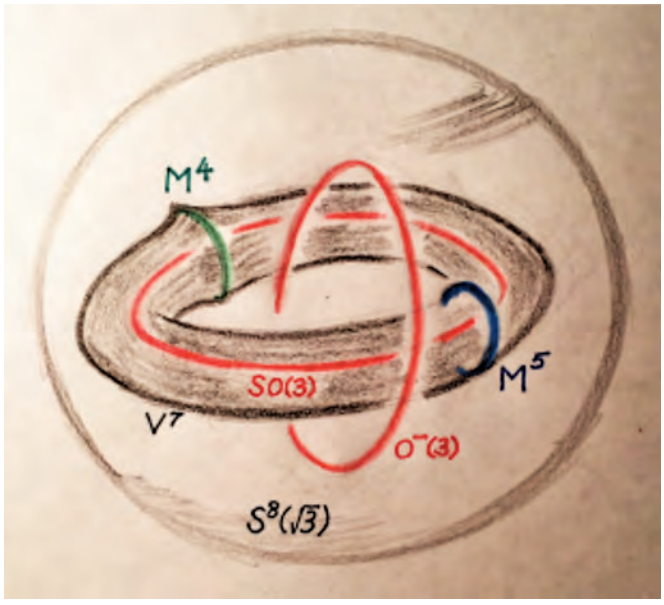

Figure 1. A view of 3×3 matrices.

This variety of singular matrices fails to be a submanifold precisely along the 4-manifold M^4 of matrices of rank 1. The rest of the matrices on V^7 have rank 2, and at their core is the 5-manifold M^5 consisting of the "best matrices of rank 2," namely those that up to scale are orthogonal on a 2-plane through the origin and zero on its orthogonal complement.

In Figure 2, we start with a 3×3 matrix A on $S^8(\sqrt{3})$ that has positive determinant, and is therefore inside the neighborhood N of $SO(3)$. We show it lying on an open 5-cell D^5, which is a portion of a great 5-sphere intersecting $SO(3)$ orthogonally, and also show the corresponding 5-cell centered at the identity I. The neighborhood N is

Dennis DeTurck is professor of mathematics at the University of Pennsylvania. His e-mail address is deturck@math.upenn.edu.

Amora Elsaify is a PhD student in finance at the University of Pennsylvania. His e-mail address is aelsaify@wharton.upenn.edu.

Herman Gluck is professor of mathematics at the University of Pennsylvania. His e-mail address is gluck@math.upenn.edu.

Benjamin Grossmann is a PhD student in mathematics at Drexel University. His e-mail address is bwg25@drexel.edu.

Joseph Hoisington is a PhD student in mathematics at the University of Pennsylvania. His e-mail address is jhois@math.upenn.edu.

Anusha Krishnan is a PhD student in mathematics at the University of Pennsylvania. Her e-mail address is anushakr@math.upenn.edu.

Jianru Zhang is a PhD student in mathematics at the University of Pennsylvania. Her e-mail address is jianruzh@math.upenn.edu.

DOI: http://dx.doi.org/10.1090/noti1571

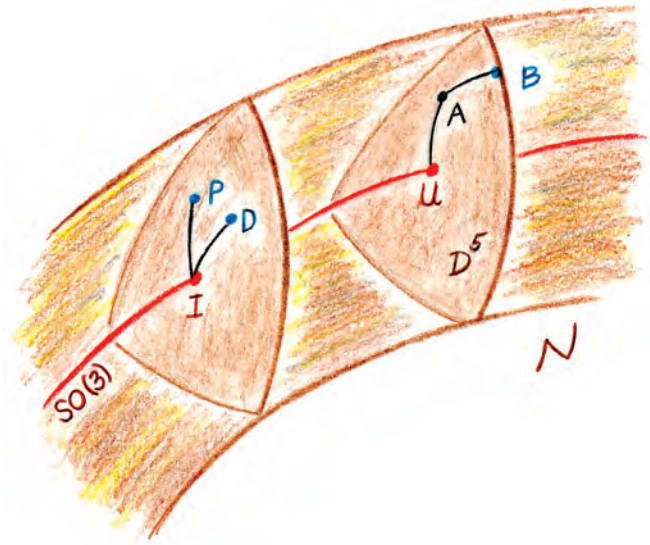

Figure 2. The ingredients for polar and singular value decomposition of A.

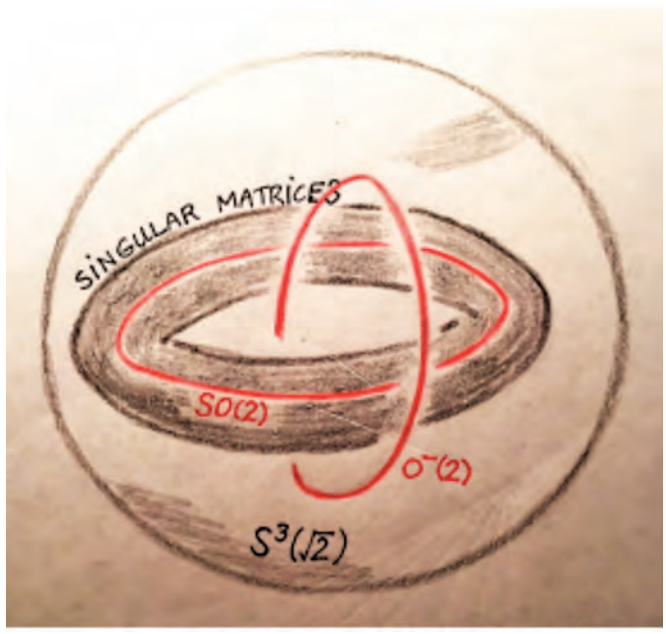

Figure 3. A view of 2×2 matrices.

filled with such non-intersecting 5-cells, and thus fibred by them.

The ***nearest orthogonal neighbor*** U to A is at the center of the 5-cell D^5, while the ***nearest singular neighbor*** B to A lies on its boundary. U is uniquely determined by A, but B may not be.

These two nearest neighbors play a central role in the applications.

For the positive definite symmetric matrix $P = \sqrt{A^T A} = U^{-1}A$, the identity I is the nearest orthogonal matrix. The same is true for its orthogonal diagonalization $D = V^{-1}PV$. The nearest singular neighbor B to A fails to be unique precisely when D has two equal smallest eigenvalues.

We have two matrix decompositions

$$A = UP \qquad \text{(polar decomposition)}$$
$$= U(VDV^{-1}) = (UV)DV^{-1} = WDV^{-1}$$
$$\text{(singular value decomposition)}$$

These decompositions have a wealth of applications, many of which fall into two different types: least squares estimate of satellite attitude as well as computational comparative anatomy (both instances of nearest orthogonal neighbor, and known as the ***Orthogonal Procrustes Problem***); and facial recognition via eigenfaces as well as interest rate term structure for US treasury bonds (both instances of nearest singular neighbor, and known as ***Principal Component Analysis***).

In the first part of this paper, we focus on the surprisingly rich and beautiful geometry of 3×3 matrices, which we view as the gateway to higher dimensions. In the second part, we consider matrices of arbitrary size and shape, as we focus on their singular value and polar decompositions, and applications of these.

As usual, figures depicting higher-dimensional phenomena are at best artful lies, emphasizing some features and distorting others, and need to be viewed charitably and cooperatively by the reader.

An expanded version of this article, with more mathematical details, and with historical comments about the origin, evolution, and numerical implementation of the various matrix decompositions, as well as fuller references, appears on the arXiv [2017].

Geometry and topology of spaces of matrices

Although we restrict ourselves to the 3×3 case, we invite the reader to warm up with nonzero 2×2 matrices, normalized to lie on the round 3-sphere of radius $\sqrt{2}$ in $\mathbb{R}^4$, and to obtain Figure 3, in which the two components of $O(2)$ appear as linked orthogonal great circles, while the singular matrices appear as the Clifford torus halfway between them.

In the expanded arXiv version of this paper [2017], the reader can see how we did this ourselves, and also see the additional details that caught our attention in this case.

We turn now to 3×3 matrices, and ask the reader to once again look at Figure 1. Contrary to appearances there, the two components of $O(3)$ are too low-dimensional to be linked in the 8-sphere. The neighborhoods N of $SO(3)$ and N' of $O^-(3)$ have nearest neighbor projections to their cores, whose point inverse images are the 5-dimensional cells lying on great 5-spheres that meet the cores orthogonally, as shown in Figure 2. The subspaces V^7, M^4, and M^5, defined earlier, will be examined in detail as we proceed.

The key to everything lies in the diagonal 3×3 matrices.

(1) A 2-sphere's worth of diagonal 3×3 matrices. In Figure 4, the diagonal matrix $\text{diag}(x, y, z)$ is located at the point (x, y, z), and indicated "distances" are really angular separations.

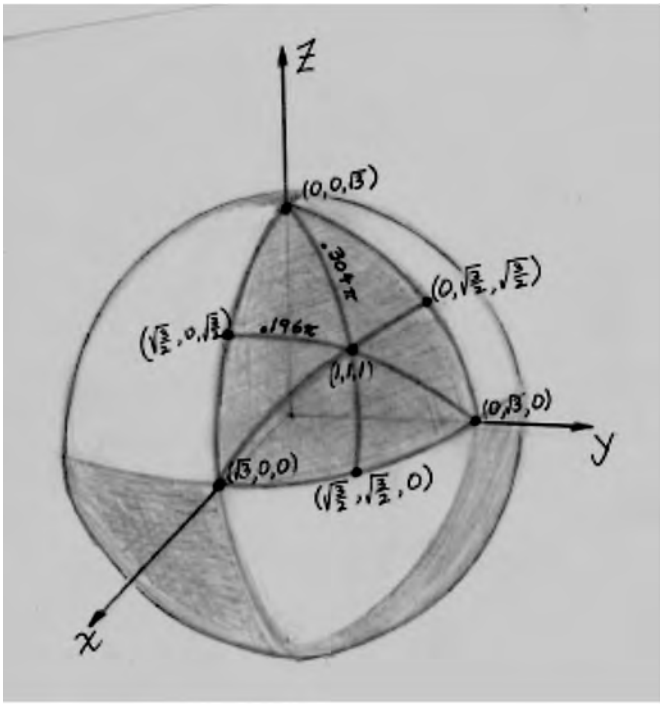

Figure 4. Diagonal matrices in $S^8(\sqrt{3})$.

This 2-sphere is divided into eight spherical triangles, with the shaded ones centered at the points $(1,1,1)$, $(-1,-1,1)$, $(1,-1,-1)$, and $(-1,1,-1)$ of $SO(3)$, and the unshaded ones, of negative determinant, centered at points of $O^-(3)$.

The interiors of the shaded triangles lie in the open neighborhood N of $SO(3)$ on $S^8(\sqrt{3})$, the interiors of the unshaded triangles lie in the open neighborhood N' of $O^-(3)$, while the shared boundaries lie on the variety V^7 of singular matrices of determinant 0, with the vertices of rank 1, the open edges of rank 2, and the centers of the edges "best of rank 2."

The shaded triangle centered at the identity $(1,1,1)$ lends its shape to that of the 5-cell fibre of N centered there. We can see in this triangle where the nearest singular neighbor fails to be unique.

(2) Symmetries. We have $O(3)\times O(3)$ acting as a group of isometries of our space $\mathbb{R}^9$ of all 3×3 matrices, and hence of the normalized ones on $S^8(\sqrt{3})$, via the map

$$(U,V) * A = UAV^{-1}.$$

This action is a rigid motion of the 8-sphere that takes $O(3) = SO(3) \cup O^-(3)$ to itself, possibly interchanging these two components, and takes the variety V^7 of singular matrices separating them to itself as well. In most cases, we can restrict our attention to the corresponding $SO(3)\times SO(3)$ action.

For example, conjugation by elements of $SO(3)$, when applied to the spherical triangle in Figure 4 centered at $(1,1,1)$, "fattens" this triangle into the 5-cell fibre of N centered there.

"Natural geometric constructions" for 3×3 matrices are those that are equivariant with respect to this action of $O(3)\times O(3)$.

(3) The determinant function on $S^8(\sqrt{3})$ takes its maximum value of $+1$ on $SO(3)$, its minimum value of -1 on $O^-(3)$, and its intermediate value of 0 on V^7.

(4) The tangent space to $S^8(\sqrt{3})$ at the identity matrix consists of all matrices orthogonal to I, that is, all matrices of trace 0. It decomposes orthogonally into the three-dimensional space of skew-symmetric matrices tangent to $SO(3)$, and the five-dimensional space of traceless symmetric matrices tangent to the great 5-sphere of symmetric matrices. Within the traceless symmetric matrices is the two-dimensional space of traceless diagonal matrices, tangent to the great 2-sphere of diagonal matrices in $S^8(\sqrt{3})$.

(5) The 7-dimensional variety V^7 of singular matrices on $S^8(\sqrt{3})$. The singular 3×3 matrices A on $S^8(\sqrt{3})$ fill out a 7-dimensional algebraic variety V^7 defined by the equations $\|A\|^2 = 3$ and $\det A = 0$, by far the most interesting part of our picture.

What portion of V^7 is a manifold?

Let $A = (a_{rs})$ be a given 3×3 matrix. One easily computes the gradient at A of the determinant function to be

$$(\nabla \det)_A = \sum_{r,s} A_{rs} \frac{\partial}{\partial a_{rs}},$$

where A_{rs} is the cofactor of a_{rs} in A.

Thus $(\nabla \det)_A$ vanishes if and only if all the 2×2 cofactors of A vanish, which happens only when A has rank ≤ 1.

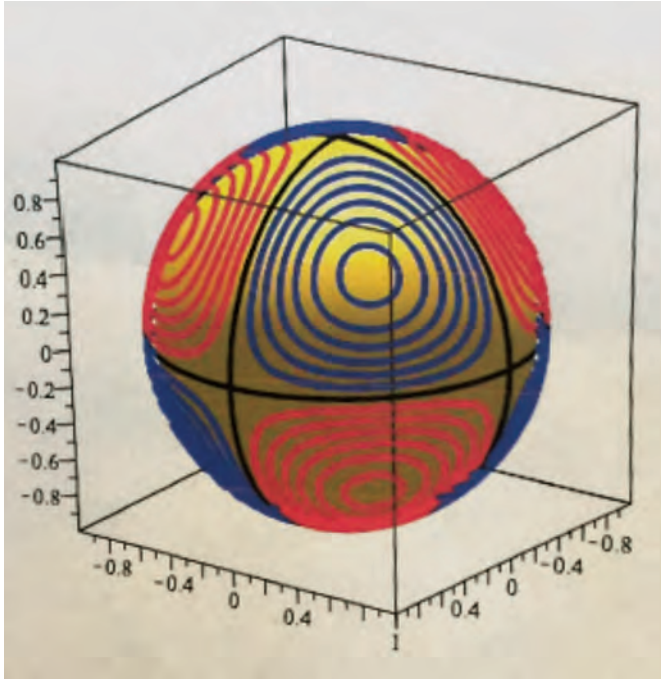

Figure 5. Level curves of det on the 2-sphere of diagonal matrices.

The subvariety V^7 of $S^8(\sqrt{3})$ consisting of the singular matrices is the zero set of the determinant function $\det\colon S^8(\sqrt{3}) \to \mathbb{R}$.

If A is a matrix of rank 2 on V^7 then the gradient vector $(\nabla \det)_A$ is nonzero there, when det is considered as a function from $\mathbb{R}^9 \to \mathbb{R}$. But if $\det A = 0$, then also $\det(tA) = 0$ for all real numbers t, hence the nonzero vector $(\nabla \det)_A$ must be orthogonal to the ray through A, and therefore tangent to $S^8(\sqrt{3})$. This means that $(\nabla \det)_A$ is also nonzero when det is considered as a function from $S^8(\sqrt{3}) \to \mathbb{R}$. It follows that V^7 is a submanifold of $S^8(\sqrt{3})$ at all its points A of rank 2.

But V^7 fails to be a manifold at all its points of rank 1.

(6) The submanifold M^4 of matrices of rank 1. First we identify M^4 as a manifold. Define $S^2 \otimes S^2$ to be the quotient of $S^2 \times S^2$ by the equivalence relation $(x, y) \sim (-x, -y)$, a space that is (coincidentally) also homeomorphic to the Grassmann manifold of unoriented 2-planes through the origin in real 4-space. To see that M^4 is homeomorphic to $S^2 \otimes S^2$, define a map $f\colon S^2 \times S^2 \to M^4$ by sending the pair of points $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$ on $S^2 \times S^2$ to the 3×3 matrix $(x_r y_s)$, scaled up to lie on $S^8(\sqrt{3})$. Then check that this map is onto, and that the only duplication is that $(\mathbf{x}, \mathbf{y})$ and $(-\mathbf{x}, -\mathbf{y})$ go to the same matrix.

M^4 is an ***orientable*** manifold because the involution $(\mathbf{x}, \mathbf{y}) \to (-\mathbf{x}, -\mathbf{y})$ of $S^2 \times S^2$ is orientation preserving, and it is a single orbit of the $O(3) \times O(3)$ action.

(7) Tangent and normal vectors to M^4. At the point $P = \operatorname{diag}(\sqrt{3}, 0, 0)$, the tangent and normal spaces to M^4 within $S^8(\sqrt{3})$ are

$$T_P M^4 = \left\{ \begin{bmatrix} 0 & a & b \\ c & 0 & 0 \\ d & 0 & 0 \end{bmatrix} \; a, b, c, d \in \mathbb{R} \right\}$$

and

$$(T_P M^4)^\perp = \left\{ \begin{bmatrix} 0 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{bmatrix} \; a, b, c, d \in \mathbb{R} \right\}.$$

We leave this to the interested reader to confirm.

(8) The submanifold M^5 = {Best of rank 2}. Recall that the "best" 3×3 matrices of rank 2 are those that, up to scale, are orthogonal on a 2-plane through the origin, and zero on its orthogonal complement.

An example of such a matrix is $P = \operatorname{diag}(1, 1, 0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, representing orthogonal projection of xyz-space to the xy-plane.

We let M^5 denote the set of best 3×3 matrices of rank 2, scaled up to lie on $S^8(\sqrt{3})$. This set is a single orbit of the $O(3) \times O(3)$ action on $\mathbb{R}^9$.

Claim: *M^5 is homeomorphic to $\mathbb{R}P^2 \times \mathbb{R}P^3$.*

Proof. Let T be one of these best 3×3 matrices of rank 2 . Then the kernel of T is some unoriented line through the origin in $\mathbb{R}^3$, hence an element of $\mathbb{R}P^2$.

An orthogonal transformation of $(\ker T)^\perp$ to a 2-plane through the origin in $\mathbb{R}^3$ can be uniquely extended to an orientation-preserving orthogonal transformation A_T of $\mathbb{R}^3$ to itself, hence an element of $SO(3)$.

Then the correspondence $T \to (\ker T, A_T)$ gives the homeomorphism of M^5 with $\mathbb{R}P^2 \times SO(3)$, equivalently, with $\mathbb{R}P^2 \times \mathbb{R}P^3$.

M^5 is non-orientable, and is a single orbit of the $O(3) \times O(3)$ action. □

(9) Tangent and normal vectors to M^5. At the point $P = \operatorname{diag}\left(\sqrt{\tfrac{3}{2}}, \sqrt{\tfrac{3}{2}}, 0\right)$, the tangent and normal spaces to M^5 within V^7 are

$$T_P M^5 = \left\{ \begin{bmatrix} 0 & -a & b \\ a & 0 & c \\ d & e & 0 \end{bmatrix} \; a, b, c, d, e \in \mathbb{R} \right\}$$

and

$$(T_P M^5)^\perp = \left\{ \begin{bmatrix} a & b & 0 \\ b & -a & 0 \\ 0 & 0 & 0 \end{bmatrix} \; a, b \in \mathbb{R} \right\},$$

and $(T_P V^7)^\perp \subset T_P S^8(\sqrt{3})$ is spanned by $\operatorname{diag}(0, 0, 1)$, as the reader can confirm.

(10) The wedge norm on V^7. To help us understand the detailed structure of V^7, we seek a real-valued function there whose level sets fill the space between M^4 and M^5, and to this end turn to the wedge norm $\|A \wedge A\|$, defined as follows.

If $A\colon V \to W$ is a linear map between the real vector spaces V and W, then the induced linear map $A \wedge A\colon \wedge^2 V \to \wedge^2 W$ between spaces of 2-vectors is defined by

$$(A \wedge A)(\mathbf{v}_1 \wedge \mathbf{v}_2) = A(\mathbf{v}_1) \wedge A(\mathbf{v}_2),$$

with extension by linearity. If $V = W = \mathbb{R}^2$, then the space $\wedge^2 \mathbb{R}^2$ is one-dimensional, and $A \wedge A$ is simply multiplication by $\det A$, while if $V = W = \mathbb{R}^3$, then the space $\wedge^2 \mathbb{R}^3$ is three-dimensional, and $A \wedge A$ coincides with the matrix of cofactors of A.

The wedge norm is defined by $\|A \wedge A\|^2 = \sum_{i,j} (A \wedge A)_{ij}^2$, and is easily seen to be $O(3) \times O(3)$-invariant, and thus constant along the orbits of this action.

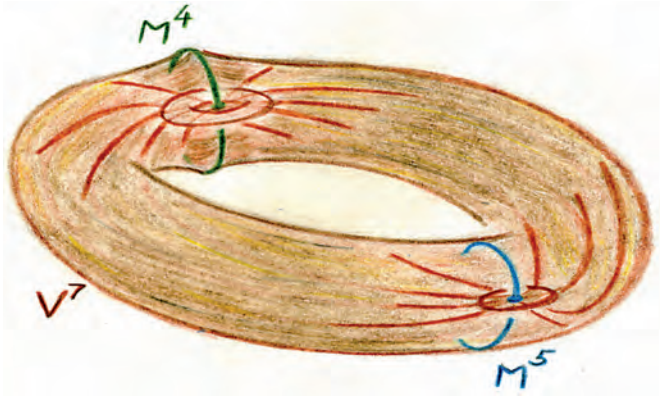

Figure 6. The variety V^7 of singular 3×3 matrices.

It has the following properties:

(1) On V^7 the wedge norm takes its maximum value of $3/2$ on M^5 and its minimum value of 0 on M^4.

(2) The level sets between these two extreme values are 6-dimensional submanifolds that are principal orbits of the $O(3) \times O(3)$ action.
(3) The orthogonal trajectories of these level sets are geodesic arcs, each an eighth of a great circle, meeting both M^4 and M^5 orthogonally, and filling the space between them without overlap along their interiors.
(4) At the upper left in Figure 6 is the 4-manifold M^4 of matrices of rank 1, along which V^7 fails to be a manifold, and at the lower right is the 5-manifold M^5 of best matrices of rank 2. The little torus linking M^4 signals that a torus's worth of geodesics on V^7 shoot out orthogonally from each of its points, while the little circle linking M^5 signals that a circle's worth of geodesics on V^7 shoot out orthogonally from each of its points. These are exactly the great circle arcs mentioned in (3) above.

(11) What else? In the expanded arXiv version of this article, the reader can find proofs of the assertions above, as well as

- A detailed description of the neighborhoods N of $SO(3)$ and N' of $O^-(3)$ on the 8-sphere $S^8(\sqrt{3})$ and of their 5-cell fibres, and a proof that these fibres meet the variety V^7 orthogonally;
- A detailed description of the singularity of V^7 along M^4;
- A description of $V^7 - M^4$ as a bundle over M^5 with round 2-cell fibres;
- A description of concrete cycles that generate the 4-dimensional homology $H_4(V^7;\mathbb{Z}) \cong \mathbb{Z} + \mathbb{Z}$ of the variety V^7 of singular matrices. These cycles are 4-spheres on which there is a well-known cohomogeneity-one action of $SO(3)$ by conjugation, with two singular $\mathbb{R}P^2$ orbits, one in M^4 and the other in M^5.

These details are not needed here, but may be useful to the reader who is aiming to generalize our description of the space of 3×3 matrices to larger real and even complex rectangular matrices.

Matrix Decompositions

Singular Value Decomposition

Let A be an $n \times k$ matrix, thus representing a linear map $A\colon \mathbb{R}^k \to \mathbb{R}^n$. We seek a matrix decomposition of A,

$$A = WDV^{-1},$$

where V is a $k \times k$ orthogonal matrix, where D is an $n \times k$ diagonal matrix,

$$D = \operatorname{diag}(d_1, d_2, \dots, d_r),$$

with $d_1 \geq d_2 \geq \cdots \geq d_r \geq 0$ with $r = \min(k, n)$, and where W is an $n \times n$ orthogonal matrix.

The message of this decomposition is that A takes some right angled k-dimensional box in $\mathbb{R}^k$ to some right angled box of dimension $\leq k$ in $\mathbb{R}^n$, with the columns of the orthogonal matrices V and W serving to locate the edges of the domain and image boxes, and the diagonal matrix D reporting expansion and compression of these edges (Figure 7). See Horn and Johnson [1991] for derivation of this singular value decomposition.

Remarks. (1) Consider the map $A^TA\colon \mathbb{R}^k \to \mathbb{R}^k$, and note that

$$A^TA = (VDW^{-1})(WDV^{-1}) = VD^2V^{-1},$$

with eigenvalues $d_1^2, d_2^2, \dots, d_r^2$ and if $r = n < k$, then also with $k - n$ zero eigenvalues. The orthonormal columns $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ of V are the corresponding eigenvectors of A^TA, since for example

$$\begin{aligned} A^TA(\mathbf{v}_1) = VD^2V^{-1}(\mathbf{v}_1) &= VD^2(1, 0, \dots, 0) \\ &= V(d_1^2, 0, \dots, 0) = d_1^2\mathbf{v}_1, \end{aligned}$$

and likewise for $\mathbf{v}_2, \dots, \mathbf{v}_k$.

(2) In similar fashion, consider the map $AA^T\colon \mathbb{R}^n \to \mathbb{R}^n$, note that

$$AA^T = (WDV^{-1})(VDW^{-1}) = WD^2W^{-1},$$

with eigenvalues $d_1^2, d_2^2, \dots, d_r^2$, and if $r = k < n$, then also with $n - k$ zero eigenvalues. The orthonormal columns $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$ of W are the corresponding eigenvectors of AA^T.

Polar Decomposition

The ***polar decomposition*** of an $n \times n$ matrix A is the factoring

$$A = UP,$$

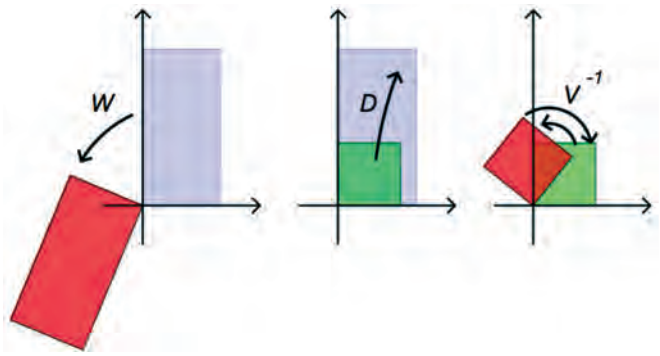

Figure 7. Singular value decomposition: $A = WDV^{-1}$.

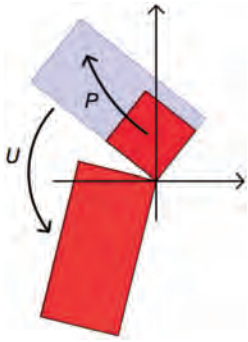

Figure 8. Polar decomposition: $A = UP$.

where U is orthogonal and P is symmetric positive semi-definite.

The message of this decomposition is that P takes some right angled n-dimensional box in $\mathbb{R}^n$ to itself, edge by edge, expanding and compressing some while perhaps sending others to zero, after which U moves the image box rigidly to another position (Figure 8).

See Horn and Johnson [1991] for derivation of this polar decomposition, and for problems to help develop expertise.

Remarks. (1) Existence of the polar decomposition follows immediately from the singular value decomposition for A:

$$A = WDV^{-1} = (WV^{-1})(VDV^{-1}) = UP.$$

Furthermore, if $A = UP$, then $A^T = P^T U^T = PU^{-1}$, and hence

$$A^T A = (PU^{-1})(UP) = P^2.$$

Now the symmetric matrix $A^T A$ is positive semi-definite, and has a unique symmetric positive semi-definite square root $P = \sqrt{A^T A}$.

(2) In the polar decomposition $A = UP$, the factor P is uniquely determined by A, while the factor U is uniquely determined by A if A is nonsingular, but not in general if A is singular.

(3) If $n = 3$ and A is nonsingular, with polar decomposition $A = UP$, and if we scale A to lie on $S^8(\sqrt{3})$, then P will also lie on that sphere, and the polar decomposition of A is just the product coordinatization of the open tubular neighborhoods N and N' of $SO(3)$ and $O^-(3)$.

(4) An $n \times n$ matrix A of rank r has a factorization $A = UP$, with U best of rank r and P symmetric positive semi-definite, and with both factors U and P uniquely determined by A and having the same rank r as A.

(5) Let A be a real nonsingular $n \times n$ matrix, and let $A = UP$ be its polar decomposition. Then U is the nearest orthogonal matrix to A, in the sense of minimizing the norm $\|A - V\|$ over all orthogonal matrices V.

(6) Let $A = UP$ be an $n \times n$ matrix of rank r with U best of rank r and P symmetric positive semi-definite. Then U is the nearest best of rank r matrix to A, in the sense of minimizing the norm $\|A - V\|$ over all best of rank r matrices V.

(7) The decomposition $A = UP$ is called ***right polar decomposition***, to distinguish it from the ***left polar decomposition*** $A = P'U'$. Given the right polar decomposition $A = UP$, we can write $A = UP = (UPU^{-1})U = P'U$ to get the left polar decomposition. If A is nonsingular, then the unique orthogonal factor U is the same for both right and left polar decompositions, but the symmetric positive semi-definite factors P and P' are not. The orthogonal factor must be the same since in either case it is the unique element of the orthogonal group $O(n)$ closest to A.

The Nearest Singular Matrix

Theorem (Eckart and Young, 1936)**.** *Let A be an $n \times k$ matrix of rank r, with singular value decomposition $A = WDV^{-1}$, where V is a $k \times k$ orthogonal matrix, where D is an $n \times k$ diagonal matrix,*

$$D = \mathrm{diag}(d_1, d_2, \dots, d_r, 0, \dots, 0), \text{ with } d_1 \geq d_2 \geq \dots d_r > 0,$$

and where W is an $n \times n$ orthogonal matrix.

Then the nearest $n \times k$ matrix A' of rank $\leq r' < r$ is given by $A' = WD'V^{-1}$, with W and V as above, and with

$$D' = \mathrm{diag}(d_1, d_2, \dots, d_{r'}, 0, \dots, 0).$$

If any two of the diagonal entries $d_{r'+1}, d_{r'+2}, \dots, d_r$ are equal to one another, then A' is not unique, but this is hidden from immediate view by the convention of listing diagonal entries in decreasing order.

Principal component analysis

Consider the singular value decomposition $A = WDV^{-1}$ of an $n \times k$ matrix A, where V is a $k \times k$ orthogonal matrix, where D is an $n \times k$ diagonal matrix,

$$D = \mathrm{diag}(d_1, d_2, \dots, d_r), \quad \text{with } d_1 \geq d_2 \geq \dots \geq d_r \geq 0,$$

with $r = \min(k, n)$, and where W is an $n \times n$ orthogonal matrix.

Suppose that the rank of A is $s \leq r = \min(k, n)$, and that $s' < s$. Then from the Eckart-Young theorem, we know that the nearest $n \times k$ matrix A' of rank $\leq s' < s$ is given by $A' = WD'V^{-1}$, with W and V as above, and with

$$D' = \mathrm{diag}(d_1, d_2, \dots, d_{s'}).$$

The image of A' has the orthonormal basis $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{s'}\}$, which are the first s' columns of the matrix W.

The columns of W are the vectors $\mathbf{w}_1, \mathbf{w}_2, \dots$, and are known as the ***principal components*** of the matrix A , and the first s' of them span the image of the best rank s' approximation to A.

If the matrix A is used to collect a family of data points, and these data points are listed as the columns of A, then the orthonormal columns of W are regarded as the principal components of this family of data points.

But if the data points are listed as the rows of A, then it is the orthonormal columns of V that serve as the principal components.

Remark. Principal Component Analysis began with Karl Pearson in 1901. He wanted to find the line or plane of closest fit to a system of points in space, in which the measurement of the locations of the points are subject to errors in any direction.

His key observation was that to achieve this, one should seek to minimize the sum of the squares of the perpendicular distances from all the points to the proposed line or plane of best fit. The best fitting line is what we now view as the first principal component, described earlier (Figure 9).

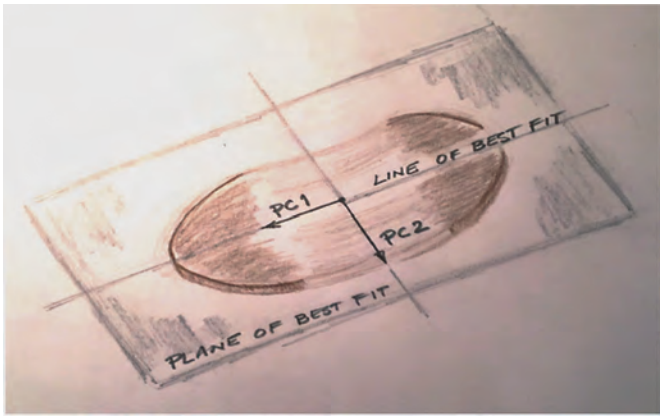

Figure 9. Principal components 1 and 2.

Applications of Nearest Orthogonal Neighbor

The Orthogonal Procrustes Problem

Let $P = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k\}$ and $Q = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k\}$ be two ordered sets of points in Euclidean n-space $\mathbb{R}^n$. We seek a rigid motion U of n-space that moves P as close as possible to Q, in the sense of minimizing the ***disparity*** $d_1^2 + d_2^2 + \cdots + d_k^2$ between $U(P)$ and Q, where $d_i = \|U(\mathbf{p}_i) - \mathbf{q}_i\|$.

It is easy to check that if we first translate the sets P and Q to put their centroids at the origin, then this will guarantee that the desired rigid motion U also fixes the origin and so lies in $O(n)$. We assume this has been done, so that the sets P and Q have their centroids at the origin.

Then we form the $n \times k$ matrices A and B whose columns are the vectors $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k$ and $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k$, and we seek the matrix U in $O(n)$ that minimizes the disparity $d_1^2 + d_2^2 + \cdots + d_k^2 = \|UA - B\|^2$ between $U(P)$ and Q.

We start by expanding

$$\langle UA - B\,,\, UA - B\rangle = \langle UA\,,\, UA\rangle - 2\langle UA\,,\, B\rangle + \langle B, B\rangle.$$

Now $\langle UA\,,\, UA\rangle = \langle A, A\rangle$, which is fixed, and likewise $\langle B, B\rangle$ is fixed, so we want to *maximize* the inner product $\langle UA\,,\, B\rangle$ by appropriate choice of U in $O(n)$. But

$$\langle UA\,,\, B\rangle = \langle U\,,\, BA^T\rangle,$$

and so, reversing the above steps, we want to *minimize* the inner product

$$\langle U - BA^T\,,\, U - BA^T\rangle,$$

which means that we are seeking the orthogonal transformation U that is closest to BA^T in the space of $n \times n$ matrices.

The above argument was given by Peter Schönemann in his 1996 PhD thesis at the University of North Carolina.

When $n \geq 3$, we don't have a simple explicit formula for U, but it is the orthogonal factor in the polar decomposition

$$BA^T = UP = P'U.$$

Visually speaking, if we scale BA^T to lie on the round $n^2 - 1$ sphere of radius $\sqrt{n}$ in n^2-dimensional Euclidean space $\mathbb{R}^{n^2}$, then U is at the center of the cross-sectional cell in the tubular neighborhood of $O(n)$ that contains BA^T and is unique if $\det(BA^T) \neq 0$.

A Least Squares Estimate of Satellite Attitude

Let $P = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k\}$ be unit vectors in 3-space that represent the direction cosines of k objects observed in an earthbound fixed frame of reference, and $Q = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k\}$ the direction cosines of the same k objects as observed in a satellite fixed frame of reference. Then the element U in $SO(3)$ that minimizes the disparity between $U(P)$ and Q is a least squares estimate of the rotation matrix that carries the known frame of reference into the satellite fixed frame at any given time.

Errors incurred in computation of U can result in a loss of orthogonality, and be compensated for by moving the computed U to its nearest orthogonal neighbor.

Procrustes Best Fit of Anatomical Objects

The challenge is to compare two similar anatomical objects: two skulls, two teeth, two brains, two kidneys, and so forth.

Anatomically corresponding points (*landmarks*) are chosen on the two objects, say the ordered set of points $P = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k\}$ on the first object, and the ordered set of points $Q = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k\}$ on the second object. They are translated to place their centroids at the origin, and then the Procrustes procedure is applied by seeking a rigid motion U of 3-space so as to minimize the disparity $d_1^2 + d_2^2 + \cdots + d_k^2$ between $U(P)$ and Q , where $d_i = \|U(\mathbf{p}_i) - \mathbf{q}_i\|$.

If size is not important in the comparison of two shapes, then it can be factored out by scaling the two sets of landmarks, $P = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k\}$ and $Q = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k\}$, so that $\|\mathbf{p}_1\|^2 + \cdots + \|\mathbf{p}_k\|^2 = \|\mathbf{q}_1\|^2 + \cdots + \|\mathbf{q}_k\|^2$.

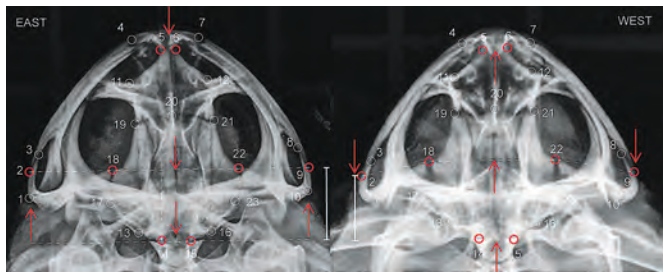

Figure 10. Dorsal x-ray images of Rhinella marina skulls, one eastern and one western, with corresponding landmarks.

Rhinella marina is a tropical toad found in the western hemisphere that has toxic effects on frog-eating predators. Previous studies suggested two genetically distinct species of this toad, one east of and one west of the Andes. Procrustes best fit of these two x-ray images is said to support the hypothesis of two separate evolutionary lineages, which have significant differences in skull shape (Figure 10).

Figure 11. Sample face and caricature.

Figure 12. First eight eigenfaces.

Applications of nearest singular neighbor

Facial Recognition and Eigenfaces

We follow Sirovich and Kirby [1987] in which the principal components of the data base matrix of facial pictures are suggestively called ***eigenpictures***.

The authors and their team assembled a file of 115 pictures of undergraduate students at Brown University. Aiming for a relatively homogeneous population, these students were all smooth-skinned Caucasian males. The faces were lined up so that the same vertical line passed through the symmetry line of each face, and the same horizontal line through the pupils of the eyes. Size was normalized so that facial width was the same for all images.

Each picture contained $128 \times 128 = 2^{14} = 16,384$ pixels, with a grey scale determined at each pixel. So each picture was regarded as a single vector $\varphi^{(n)}$, $n = 1, 2, \ldots, 115$, called a ***face***, in a vector space of dimension 2^{14}.

The challenge was to find a low-dimensional subspace of best fit to these 115 faces, so that a person could be sensibly recognized by the projection of his picture into this subspace.

To make sure that the subspace passes through the origin (i.e., is a linear rather than affine subspace), the data is adjusted so that its average is zero, as follows.

Let $\langle\varphi\rangle = (1/M)\sum_{n=1}^{M} \varphi^{(n)}$ be the average face, where $M = 115$, and then let $\phi^{(n)} = \varphi^{(n)} - \langle\varphi\rangle$ be the deviation of each face from the average. The authors refer to each such deviation ϕ as a ***caricature***. Figure 11 shows a sample face and its caricature.

The collection of caricatures $\phi^{(n)}$, $n = 1, 2, \ldots, 115$ was then regarded as a $2^{14} \times 115$ matrix A, with each caricature appearing as a column of A.

Let the singular value decomposition of A be $A = WDV^{-1}$, with W a $2^{14} \times 2^{14}$ orthogonal matrix,

$$D = \operatorname{diag}(d_1, d_2, \ldots, d_{115}) \quad \text{with } d_1 \geq d_2 \geq \cdots \geq d_{115} \geq 0$$

a $2^{14} \times 115$ diagonal matrix, and V a 115×115 orthogonal matrix. The orthonormal columns $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{2^{14}}$ of W are the principal components of the matrix A.

It was found that the first 100 principal components of A span a subspace sufficiently large to recognize any of the faces $\varphi^{(n)}$ by projecting its caricature into this

subspace and then adding back the average face:

$$\varphi^{(n)} \sim \langle \varphi \rangle + \sum_{k=1}^{100} \langle \phi^{(n)}, \mathbf{w}_k \rangle \mathbf{w}_k.$$

Figure 12 shows the first eight eigenpictures starting at the upper left, moving to the right, and ending at the lower right, in which each picture is cropped to focus on the eyes and nose. Since the eigenpictures can have negative entries, a constant was added to all the entries to make them positive for the purpose of viewing.

Figure 13. Cropped sample face.

Figure 13 shows a sample face, correspondingly cropped, and Figure 14 shows the approximations to that sample face, using 10, 20, 30, and 40 eigenpictures.

Figure 14. Approximations to sample face.

After working with the initial group of 115 male students, the authors tried out the recognition procedure on one more male student and two females, using 40 eigenpictures, with errors of 7.8%, 3.9%, and 2.4% in these three cases.

Principal Component Analysis Applied to Interest Rate Term Structure

How does the interest rate of a bond vary with respect to its *term*, meaning time to maturity? The answer involves one of the oldest and best known applications of Principal Components Analysis (PCA) to the field of economics and finance, originating in the 1991 work of Litterman and Scheinkman.

To begin, economists plot the interest rate for a given bond against a variety of different maturities, and call this a ***yield curve***. Figure 15 shows such a curve for US Treasury bonds from an earlier date, when interest rates were higher than they are now.

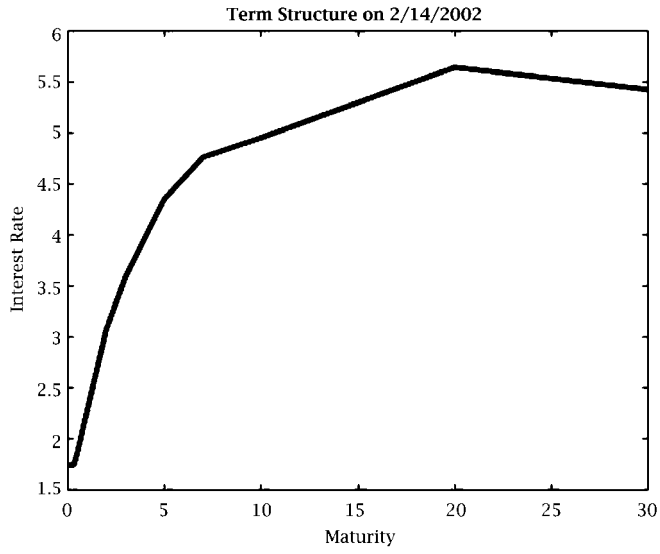

Figure 15. Yield curve.

Predicting the relation shown by such a curve can be crucial for investors trying to determine which assets to invest in, and for government officials who wish to determine the best mix of Treasury maturities to auction on any given day. For this reason, a number of investigators have tried to understand whether there are common factors embedded in the term structure. In particular, identifying whether there are factors that affect all interest rates equally, or that affect interest rates for bonds of certain maturities but not of others, is important for understanding how the term structure behaves.

To help understand how these questions are answered, we replicated the methodology in the Litterman and Scheinkman paper, using a newer data set that gives the daily interest rate term structure for US Treasury bonds over a long span of time, 2,751 days between 2001 and 2016. For each of these days, we recorded the interest rates for bonds of 11 different maturities: 1, 3, and 6 months, and 1, 2, 3, 5, 7, 10, 20, and 30 years. Each data vector is an 11-tuple of interest rates, which we collected as the rows of a $2{,}751 \times 11$ matrix.

The average of the rows is depicted graphically in Figure 16. We subtracted this average from each of the rows, and called the resulting matrix A. The rows of A are our *adjusted data vectors*, which now add up to zero.

Let $A = WDV^{-1}$ be the singular value decomposition of A, where V is an 11×11 orthogonal matrix, D is a

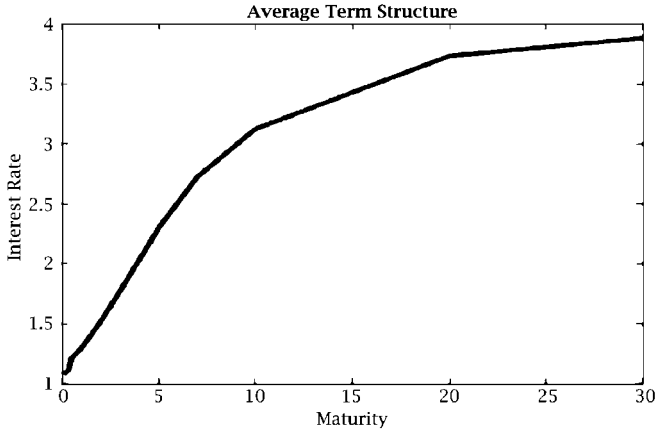

Figure 16. Average yield curve.

$2{,}751 \times 11$ diagonal matrix, and W is a $2{,}751 \times 2{,}751$ orthogonal matrix. Since the data points are the *rows* of A, the principal components are the 11 orthonormal columns of V.

These principal components reveal the line of best fit, the plane of best fit, the 3-space of best fit, and so forth for our 2,751 data points. They were obtained using the PCA package of MATLAB. The first three principal components are shown graphically in Figure 17.

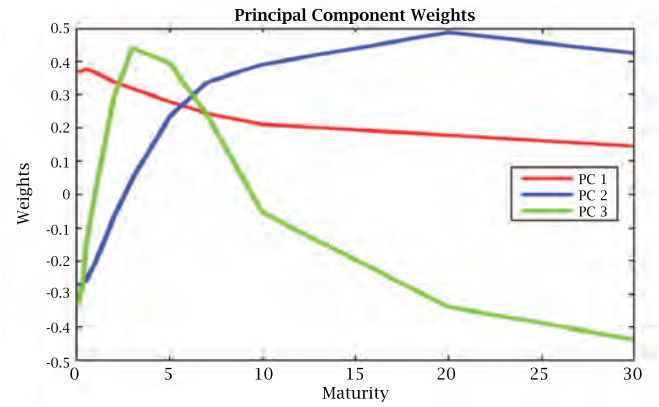

Figure 17. Principal components of A.

The first principal component is more constant than the other two, and captures the fact that most of the variation in term structures comes from changes that affect the levels of all yields. The second most important source of variation in term structure comes from the second principal component, which reflects changes that most affect yields on bonds of longer maturities, while the third principal component reflects changes that affect medium term yields the most. These features of the first three principal components were called *level*, *steepness*, and *curvature* in the foundational paper by Litterman and Scheinkman.

In Figure 18, the black curve is the term structure on 2/14/2002, duplicating the first figure in this section. We subtract the average term structure from this particular one, project the difference onto the one-dimensional

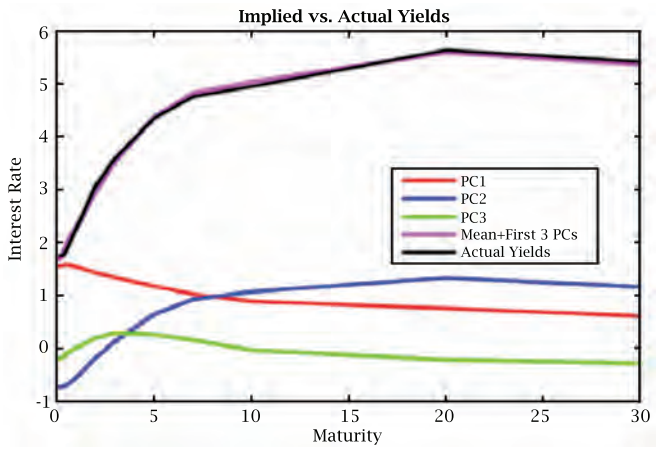

Figure 18. Approximation of a yield curve by its first three principal components.

subspaces spanned in turn by the first three principal components, and show these projections above in red, blue, and green. Finally, we sum up these three projections, add back the average term structure, show the result in purple, and see how closely this purple curve approximates the black curve we started with.

References

[1987] L. Sirovich and M. Kirby, Low-dimensional procedure for the characterization of human faces, *J. Optical Society of America* **4** (1987), no. 3, 519–524.

[1991] Roger Horn and Charles Johnson, *Topics in Matrix Analysis*, Cambridge University Press.

[2016] Aldemar A. Acevedo, Margarita Lampo, and Roberto Cipriani, The cane or marine toad, *Rhinella marina* (Anura, Bufonidae): Two genetically and morphologically distinct species, *Zootaxa* 4103 (6), 574–586.

[2017] D. DeTurck, A. Elsaify, H. Gluck, B. Grossmann, J. Hoisington, A. M. Krishnan and J. Zhang, Making matrices better: Geometry and topology of polar and singular value decomposition, (arXiv 1702.02131)

Acknowledgments

We are grateful to our friends Christopher Catone, Joanne Darken, Ellen Gasparovic, Chris Hays, Kostis Karatapanis, Rob Kusner, and Jerry Porter for their help with this paper.

Special thanks to Lawrence Sirovich for permission to use the pictures from his 1987 paper with M. Kirby on facial recognition, and to Aldemar Acevedo for permission to use the picture from his 2016 paper with M. Lampo and R. Cipriani on the marine toad.

Photo Credits

Figures 1–9, Figures 15–18, and all headshots courtesy of the authors.

Figure 10 reprinted with permission from Acevedo, Lampo, and Cipriani [2016].

Figures 11, 12, 13, and 14 reprinted with permission from Sirovich and Kirby [1987].


Dennis DeTurck

ABOUT THE AUTHORS

Dennis DeTurck is Dean of the College and faculty director of Riepe College House, where he enjoys hanging out with about 500 of his closest friends.


Amora Elsaify

Amora Elsaify's research interests include asset pricing, macroeconomics, differential geometry, and applications of nonparametric statistics and machine learning. Outside of his studies, Amora enjoys watching and playing soccer and following national politics.


Herman Gluck

Herman Gluck sees himself as a 4 channel receiver in a 198 channel world: family and friends; mathematics (geometry and topology with applications to molecular biology, hydrodynamics, and plasma physics); tennis (he is playing captain on two teams in USTA league competition); and music (where in recent years he has sung roles in *The Barber of Seville*, *Fledermaus*, and *The Merry Wives of Windsor* with the Delaware Valley Opera Company).


Benjamin Grossman

Benjamin Grossman's research interests include rank minimization in matrix patterns, quantum game theory, positive maps on operator algebras, and the combinatorics of matrices over finite fields. He likes to sing and go rock climbing, but not usually at the same time.


Joseph Ansel Hoisington

Joseph Ansel Hoisington is a mathematics PhD student at the University of Pennsylvania. Before graduate school, he worked for the Institute of Health Metrics and Evaluation at the University of Washington on quantitative public health research. He enjoys running, listening to (and occasionally making) music, and reading.


Anusha M. Krishnan

Anusha M. Krishnan is a graduate student in mathematics at the University of Pennsylvania.


Jianru Zhang

Jianru Zhang is a Chinese female graduate student in the Penn math department. She worked on this project in the summer of her first year with some other graduate students and gave a short talk on it in a sectional meeting of MAA the following October with Anusha Krishnan.