THE GSVD: WHERE ARE THE ELLIPSES?, MATRIX TRIGONOMETRY, AND MORE

ALAN EDELMAN* AND YUYANG WANG†

Abstract.
This paper provides an advanced mathematical theory of the Generalized Singular Value Decomposition (GSVD) and its applications. We explore the geometry of the GSVD which provides a long sought for ellipse picture which includes a horizontal and a vertical multiaxis. We further propose that the GSVD provides natural coordinates for the Grassmann manifold. This paper proves a theorem showing how the finite generalized singular values do or do not relate to the singular values of $AB^\dagger$.

We then turn to the applications arguing that this geometrical theory is natural for understanding existing applications and recognizing opportunities for new applications. In particular the generalized singular vectors play a direct and as natural a mathematical role for certain applications as the singular vectors do for the SVD. In the same way that experts on the SVD often prefer not to cast SVD problems as eigenproblems, we propose that the GSVD, often cast as a generalized eigenproblem, is rather best cast in its natural setting.

We illustrate this theoretical approach and the natural multiaxes (with labels from technical domains) in the context of applications where the GSVD arises: Tikhonov regularization (unregularized vs regularization), Genome Reconstruction (humans vs yeast), Signal Processing (signal vs noise), and statical analysis such as ANOVA and discriminant analysis (between clusters vs within clusters.) With the aid of our ellipse figure, we encourage in the future the labelling of the natural multiaxes in any GSVD problem.

Key words. GSVD, SVD, ellipse, CS Decomposition, Tikhonov Regularization

AMS subject classifications. 65F22, 15A18, 15A23

1. Introduction.

1.1. Prelude. If $a \in \mathbb{R}^{m_1}$ and $b \in \mathbb{R}^{m_2}$ are two vectors, then the block vector equation in $\mathbb{R}^{m_1+m_2}$:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix}$$

may be thought of geometrically as a hypotenuse vector decomposed as the sum of two legs of a right triangle. If $h = \sqrt{\|a\|^2 + \|b\|^2} \neq 0$ is the length of this hypotenuse and $u = a/\|a\|, v = b/\|v\|$ are the unit direction vectors for a, b then we can write

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} uc \\ vs \end{bmatrix} h,$$

where c and s are the cosine and sine of the corresponding angles, namely $c = \|a\|/h$ and $s = \|b\|/h$. This is ordinary planar trigonometry of a right triangle.

For notational convenience, we will sometimes use a semicolon (";") to denote the stacking (or vertical concatenation) of vectors and matrices, so that

$$[a; b] = [a; 0] + [0; b].$$

We note that $[uc; vs]$ is a unit vector in the direction $[a; b]$. The cotangent $\sigma = c/s$ is a slope which provides a measure of whether the vector is primarily in the "a" (or top) direction, or the "b", or a mix depending on whether σ is large, small, or in between.

The GSVD extends the above ideas to matrices.

*Department of Mathematics, MIT, Cambridge, MA (edelman@math.mit.edu).
†AWS AI Labs, East Palo Alto, CA (yuyawang@amazon.com). Work done prior joined Amazon.

1.2. The GSVD. This paper provides a new approach and understanding of the generalized SVD (GSVD) [25, 32] of two matrices $A \in \mathbb{R}^{m_1 \times n}, B \in \mathbb{R}^{m_2 \times n}$. Generalizing the introductory paragraphs, the GSVD may be understood in the context of a generalized Pythagorean theorem on

$$[A; B] = [A; 0] + [0; B].$$

Let r denote rank$([A; B])$.

We take as our definition of a GSVD, a decomposition of $[A; B]$ in the form

$$\left[\frac{A}{B}\right] = \left[\frac{UC}{VS}\right] H, \tag{1.1}$$

where U, V are square orthogonal in $\mathbb{R}^{m_1 \times m_1}$. $\mathbb{R}^{m_2 \times m_2}$; C, S are 1-diagonal such that $C'C + S'S = I_r$, and H has full row rank r. The remaining dimensions are implied, namely C, S are in $\mathbb{R}^{m_1 \times r}$, $\mathbb{R}^{m_2 \times r}$, and H is in $\mathbb{R}^{r \times n}$,

The SVD is so widely used, applications need not be listed. Historically this was not always the case. Fields such as biology, economics, and computer science could be watched learning about the svd one-by-one with great impact. Perhaps a kind of folklore notion is that the SVD applies any time an array A needed to be quickly compressed to get the main information out, or whenever AA^T was lurking. We would love to foster a world where the GSVD finds applications one-by-one in many fields. Perhaps the new folklore is that the GSVD applies when two arrays with a common dimension need to be quickly compressed or whenever two matrices AA^T and BB^T are lurking. Of course both the SVD and GSVD underly more.

Some selected applications of the GSVD include oriented energy analysis [5, 6, 7, 8, 9, 33], (here the GSVD is sometimes called by the more descriptive name QSVD for "quotient" SVD), Tykhonov regularization [18, 12], Linear Discriminant Analysis [26, 20], and more recently in microarray analysis [2]. A review from 1992 and discussion of algorithms may be found in [4].

As a point of mathematical taste, many textbooks today still treat SVDs as a byproduct of exposition on eigenvalues. This is unfortunate, as most of the time considerations of AA^T or A^TA create unnecessary mathematical baggage, best abandoned. The SVD is mature enough to live its own life separate from the symmetric eigenvalue problem. Taking this notion one step further, the GSVD deserves to live separately from generalized eigenvalue problems or the SVD. When a GSVD lurks, it is recommended to abandon old fashioned language and see the true GSVD construction in full mature light. We take this approach in a number of examples in this paper.

1.3. A "GH" decomposition. To clarify and streamline our view of the roles of the pieces of the GSVD, we propose that the GSVD be considered a GH decomposition:

$$\left[\frac{A}{B}\right] = GH.$$

where $G = [UC; VS]$ (for Grassmann or geometric) denotes the information in the r-dimensional hyperplane representing the column space of $[A; B]$. Specifically the columns of G are a natural orthonormal basis for that hyperplane in $\mathbb{R}^{m_1+m_2}$, and the columns of H represent the columns of $[A; B]$ in that basis. Of course the QR decomposition of $[A; B]$ has exactly the same properties, with one important difference:

the Q is not uniquely defined by the hyperplane, while in the GSVD, the choice is more or less canonical.

We further feel that the factorization into the two matrices G and H emphasizes the outer product rank r form:

$$\left[\frac{A}{B}\right] = \sum_{i=1}^{r} (i^{\text{th}} \text{ column of } [UC; VS])(i^{\text{th}} \text{ row of } H), \tag{1.2}$$

which can be readily missed in the long form.

In analog with the SVD or NMF decompositions, one might consider a simultaneous rank reducing method where only the k rows of H with largest norm are kept.

In particular if we multiply $[A; B]$ on the right by $H^\dagger I_{r,k} I_{r,k}^T H$, where $I_{r,k}$ is the first k columns of the $r \times r$ identity, we obtain a rank reduced $[A; B]$:

$$[A; B] \approx \left([UC; VS] I_{r,k}\right) \left(I_{r,k}^T H\right) = \sum_{i=1}^{k} (i^{\text{th}} \text{ column of } [UC; VS])(i^{\text{th}} \text{ row of } H).$$

We remark that $H^\dagger I_{r,k} I_{r,k}^T H$ is an oblique projector when H is square non-singular, and an orthogonal projector when H is orthogonal.

1.4. More details about U, V, C, S, H. The matrices U, V, C, S, H deserve more detailed discussion, though we recommend the reader only skim these sections on a first read lest losing the forest for the trees.

To help guide the reader, we make a table of bases for the fundamental subspaces that appear in the GSVD. It is helpful to keep in mind that the columns of C and S are leftward looking towards the orthogonal U and V matrices in the GSVD factorization, while the rows of C and S are rightward looking towards the full row rank H in the GSVD factorization.

Fundamental Spaces	Basis (with Link to C, S)
Column Spaces of A, B:	Columns of U, V corresponding to non-zero cols of C, S
Left-Null Spaces of A, B:	Columns of U, V corresponding to zero cols of C, S
Row Space of $[A; B]$:	Rows of H
Row Spaces of A, B:	Rows of H corresponding to non-zero rows of C, S
Null Spaces of A, B:	Columns of $H^\dagger$ corresponding to zero columns of C, S + common null space (if $r < n$)
Gen Eigenvector Spaces:	Columns of $H^\dagger$ (for the problem $\det(A'A - \lambda B'B) = 0$)
Common Null Space:	Null space of H (Also see 1.4.6 for an RQ drilldown)

It is useful to point out that the common nullspace of A and B is killed by H, i.e., if $Ax = 0$ and $Bx = 0$ then $Hx = 0$. A vector that is in only one of the nullspaces is not killed by H, but Hx is killed by 0 columns in C or S respectively.

1.4.1. The square orthogonal matrices U and V. The U and V matrices represent orthogonal bases for $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$ respectively.

One obtains an orthogonal basis for the column space of A (B) by taking the columns of U (V) corresponding to the non-zero rows of C (S). The remaining columns are an orthogonal basis for the left nullspace. (Recall, in the ordinary SVD, the "U" can be chosen to be square, and one can divide "U" into the column space/left nullspace through "Σ" in the analagous way.)

We will see in Section 2 that the columns of U and the columns of V may be thought of as semi-axes of ellipses (with the possibility of degenerate axes.)

Notation: For $i = 1, \ldots, r$, let u_i denote the normalized ith column of UC if $c_i \neq 0$, or else define $u_i = 0$. Similarly, let v_i denote the normalized ith column of VS if $s_i \neq 0$, or else define $v_i = 0$. This notation conveniently avoids issues of different sizes and conventions. For example, U or V may have fewer than r columns. Details of the placement of the c_i and s_i appear in Section 1.4.2. Suffice it to say for now that u_i is the *ith* column of the U matrix when $c_i > 0$, and v_i may be found in the kth column of the V matrix when $S_{ki} = s_i > 0$. The indirection in V is admittedly unfortunate, but in all cases, the non-zero v_i by convention are left to right contiguous columns of V that may either start from the left, or end at the right, but in many situations v_i is not in the ith column.

1.4.2. The diagonal cosine and one-diagonal sine matrices: C **and** S**.** The cosines $1 \geq c_1 \geq \ldots \geq c_r \geq 0$ and sines $0 \leq s_1 \leq \ldots \leq s_r \leq 1$ satisfy $c_i^2 + s_i^2 = 1$. They represent the lengths of the semi-axes of our two ellipses. The generalized singular values are the cotangents $\sigma_i = c_i/s_i$ which may be 0 or infinite. When $0 < \sigma_i < \infty$, we say that σ_i is finite.

The cosine matrix $C \in \mathbb{R}^{m_1 \times r}$ matrix puts the c_i on the diagonal starting with c_1 in the (1,1) position. If we run out of room, by not having enough rows, we drop some of the 0 cosines.

The sine matrix $S \in \mathbb{R}^{m_2 \times r}$ puts the s_i on *some* diagonal, and again if we run out of room, by not having enough rows, we drop some of the 0 sines. One convention (used by LAPACK [3]) puts all the positive s_i in the top rows by putting the positive diagonal in the top right corner. Another [25, Eq. 2.3] puts them in the bottom rows which as Paige and Saunders remark (and we agree) creates [25, p.401 top]: "as easy way to remember the symmetry."

Either way $C'C$ and $S'S$ are square $r \times r$ diagonal with the c_i^2 and s_i^2 on the main diagonal and $C'C + S'S = I_r$ The only difference between the two conventions is where the orthogonal basis for the column space of B ends up in the columns of V, i.e., the left or right side. (It is always those columns of V that correspond to the rows where an $s_i > 0$.) When the significant elements of S are on top, the column space basis is on the left like it is with U. When it is on the bottom, one feels that the B is being treated as something of a "mirror image" of A with the column space basis on the right of V, and S being something of a 180 degree rotation (in structure) from C.

Let $r_a = \text{rank}(A), r_b = \text{rank}(B), r = \text{rank}[A; B]$. Table 1 shows the structure of C and S:

PROPERTY OF C AND S	C	S
total # columns	r	
# zero columns in S (left columns):	$r - r_b \quad = \#\ \{c_i = 1\} = \#\{s_i = 0\}$	
# non-zero columns (middle columns):	$r_a + r_b - r = \#\{0 < c_i, s_i < 1\}$	
# zero columns in C (right columns):	$r - r_a \quad = \#\ \{c_i = 0\} = \#\{s_i = 1\}$	
total # rows	m_1= # rows A	m_2= # rows B
# non-zero rows	$r_a \leq m_1$	$r_b \leq m_2$
# zero rows	$m_1 - r_a$	$m_2 - r_b$

TABLE 1

The C and S matrices are naturally simultaneously partitioned into three block columns such that the number of columns $r = (r - r_b) + (r_a + r_b - r) + (r - r_a)$ in left to right order. The row sizes conform to A and B which means that we add rows of zeros to C, S or possibly delete some of the zero cosines/sines to achieve a row count of $m_1, m2$. The number of non-degenerate angles (not 0 nor $\pi/2$) is the middle number $(r_a + r_b - r)$.

A very common case has $r = n$ in which case the sizes of C, S match that of A, B.

1.4.3. The matrix H that has no orthogonality or diagonal properties. On a first glance, no self-respecting decomposition in the SVD family should be neither diagonal nor orthogonal. Nonetheless, all we can say about H is that $H \in \mathbb{R}^{r \times n}$ is a full row rank matrix whose rowspace is that of $[A; B]$. A very common case is $r = n$ in which case H is square non-singular.

In the same way that a vector is specified by its direction and length, we think of $[A; B]$ as being specified by its column space and the rest of the information. The matrix H specifies the rest of the information.

Rowspace information is available in H. The first r_a rows of H form a basis for the rowspace of A. The last r_b rows of H form a basis for the rowspace of B. The nullspaces are not as immediately available due to the non-orthogonality of H. The nullspace of H is the common nullspace of A and B. Of course one can use a QR decomposition.

1.4.4. Compact Formats. One can optionally delete all the zero rows of C or S and the corresponding columns of U and V. This kills the left nullspace basis vectors, but preserves the column space vectors.

1.4.5. Expanded Format. When $r < n$ one can add $n - r$ zero columns to both C and S and expand H to a full square non-singular matrix by adding any $n-r$ rows to H that would make it invertible.

1.4.6. Further reduction to Orthogonal and Triangular. The Expanded H matrix can be written $[0\ R]Q'$, where R is triangular $\mathbb{R}^{r,r}$, and Q is square orthogonal $\mathbb{R}^{n,n}$. In this case the initial $n-r$ columns of Q are an orthogonal basis for the common nullspace of A and B.

1.5. Summary. This paper contains a number of insights and results about the GSVD:

- The GSVD generalizes planar trigonometry to matrix trigonometry.
- We present an ellipse picture of the GSVD, which requires four dimensions to get a good feel for the general case.
- We consider $[UC; VS]$ as natural coordinates for r dimensional hyperplanes (the Grassmann manifold) in $\mathbb{R}^m$ given that $m = m_1 + m_2$.

- We use the Grassmann manifold coordinates to clarifiy the link between the CS decomposition and the GSVD (other authors have observed vaguely that they are closely related).
- We view the H matrix as the change of coordinates from canonical coordinates $[UC; VS]$ to the specifics of $[A; B]$.
- We prove a theorem relating $\mathrm{gsvd}(A, B)$ and $\mathrm{svd}(AB^{\dagger})$. They are not generally identical.
- We revisit Tikhonov regularization in the geometric context.
- We interpret the GSVD as a multi-dimensional slope and connect applications.

2. Where are The Ellipses?. The SVD ellipse picture for a matrix A (Figure 1) is a very familiar visual for the action of A on the unit ball. We are not aware of any ellipse pictures in the literature nor even a notion that a natural ellipse picture exists for the GSVD or even the CSD (CS Decomposition) [17]. We believe that the lack of a geometric view of the GSVD is part of the reason that the GSVD is not as widely understood or as widely used as it should be.

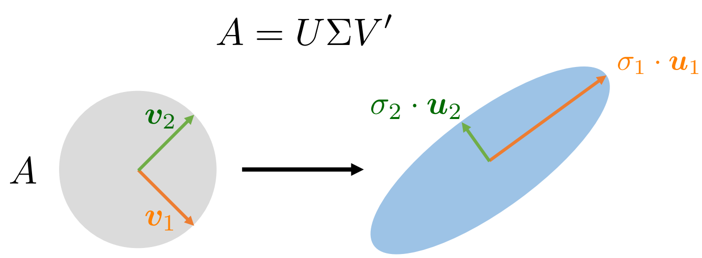


FIG. 1. *Familiar SVD visual showing singular vectors and singular values of a matrix A through the action of A on the unit ball.*

Regarding an ellipse picture one might blame some sort of human inability to perceive higher dimensions as a complication, but we show that this is not really the case in Figure 2.

The gap in understanding is underscored by the curiosity expressed online, but without answer, on such sites as MATLAB Central [11] (reproduced here) and a similar request on the question-and-answer site Quora [29] (not reproduced here).

```
Subject: Generalized SVD geometry?
From: Bob Dyas
Date: 29 Feb, 2000 15:31:31
Message: 1 of 1 ⟵ indicates no answer in 19 years!
Is there a geometric interpretation of the generalized singular
value decomposition? I'm looking for something comparable to
the geometry associated with the standard SVD. I understand how
U, V and the singular values of the SVD relate to the geometry
of the input matrix but I don't have an intuitive feel for how
U, V, X and the generalized singular values relate to the
geometry of the two input matrices of the GSVD.

Any help would be appreciated.
--
Bob Dyas Email: dyas@ccrl.mot.com
```

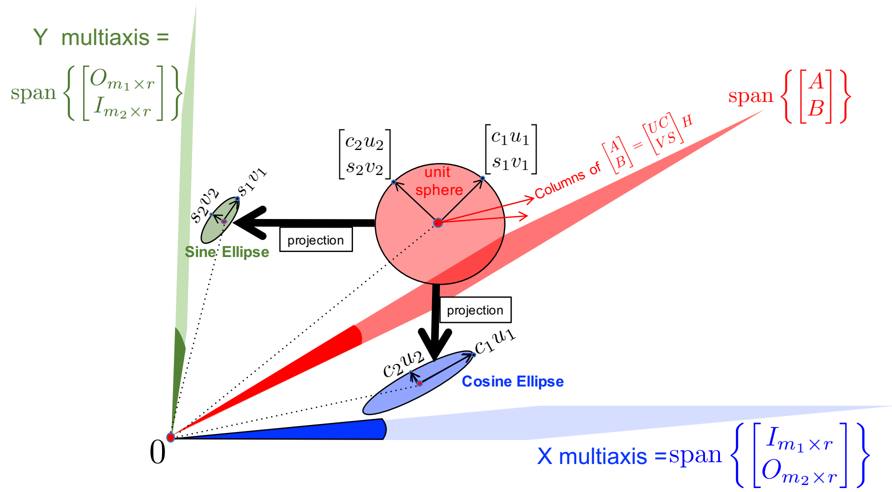


Fig. 2. *Ellipse picture for the GSVD (illustrated generically in four dimensions with $m_1 = m_2 = r = r_a = r_b = 2$): containing a green plane (span of $[A; B]$), and two gray planes (the X and Y multiaxes). Centered at the origin is a unit sphere (light green) and two ellipses (light orange) shown in exploded view format. The ellipses, which may be named the cosine and sine ellipses are "horizontal" and "vertical" shadows of the unit sphere.*

Color Coding (consistent for all figures in this paper):
shade of RED=Span([A;B]), shade of BLUE=X Multiaxis, shade of GREEN=Y Multiaxis.

Motorola Labs - IL02 Phone: 847-576-8702
1301 E. Algonquin Road, Room 2724 Fax: 847-538-4593
Schaumburg, IL 60196

2.1. Understanding the Ellipse Picture for the GSVD. Figure 2, portrayed in four dimensional space, generically serves to illustrate the GSVD in any dimensions.

Given $A \in \mathbb{R}^{m_1,n}, B \in \mathbb{R}^{m_2,n}$, we consider the unit sphere (shown in exploded form in Figure 2 as a green cirlce) in the span of $[A; B]$ (shown as a green plane). In orange we have the ellipses that show the "downward" and "leftward" projections of these ellipses onto the multiaxes X and Y defined as those vectors whose first m_1 or last m_2 coordinates may not vanish. (For example if $m_1 = m_2$ in $\mathbb{R}^4$, then the X multiaxis consists of vectors of the form $(x_1, x_2, 0, 0)$ and the Y multiaxis consists of vectors of the form $(0, 0, x_3, x_4)$.

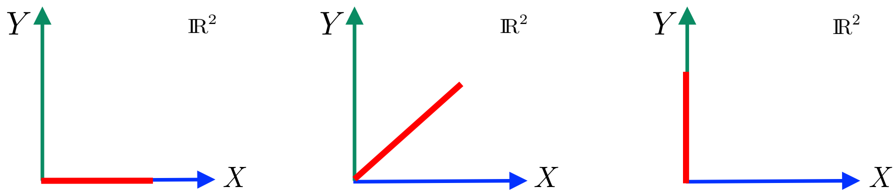
The u_i, v_i are semi-axes of these ellipses, with lengths c_i, s_i. The vector $[u_i c_i; v_i s_i]$ is on the (green) unit sphere in the span of $[A; B]$.

Since we have the equality $[A; B]x = [UC; VS]Hx$, we see that H is the change of coordinates from the columns of $[A; B]$ to the orthonormal columns of $[UC; VS]$, and $H^\dagger$ goes the other way.

2.2. An in depth look at small dimensional special cases.

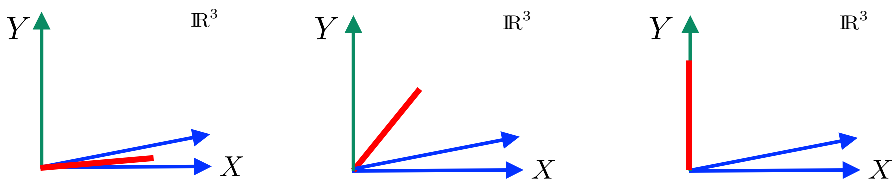
2.2.1. A red line in $\mathbb{R}^2$, X=the x-axis, Y=the y-axis. ($m_1 = m_2 = n = r = 1$)

Below we show the possibilities for $[C; S]$ for a line in $\mathbb{R}^2$ (drawn in red as the span of $[a, b]$ where a and b are $\in \mathbb{R}^1$) which may be horizontal $a \neq 0, b = 0$, general position $a \neq 0, b \neq 0,$, or vertical $a = 0, b \neq 0$. In any event the c and s are the cosine and sine of the angle with the horizontal.



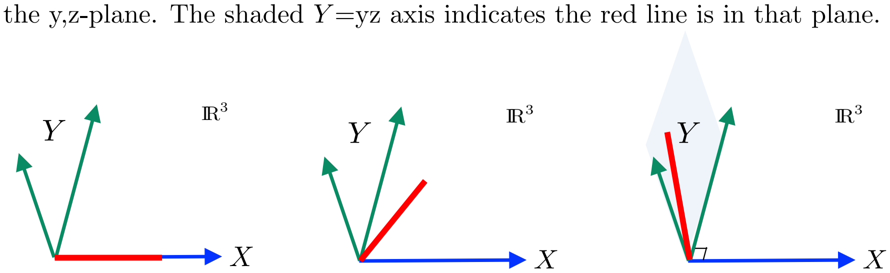
$$\left[\begin{array}{c} \mathbf{1} \\ \hline \mathbf{0} \end{array}\right] \qquad \left[\begin{array}{c} \boldsymbol{c} \\ \hline \boldsymbol{s} \end{array}\right] \qquad \left[\begin{array}{c} \mathbf{0} \\ \hline \mathbf{1} \end{array}\right]$$

2.2.2. A red line in $\mathbb{R}^3$, X=the x, y plane, Y=the z axis. ($m_1 = 2, m_2 = n = r = 1$) Below we show the possibilities for $[C; S]$ for a line in $\mathbb{R}^3$ (drawn in red as the span of $[a, b]$, where $a \in \mathbb{R}^2$, $b \in \mathbb{R}^1$). The X multiaxis is traditionally labeled the x, y plane, and the Y is the z-axis. A line can be in the x, y plane, in general position, or along the z-axis. The corresponding [C:S] matrix is illustrated. The c is the angle between the red line and the x, y plane, while the s is the angle of the red line and the z-axis.



$$\left[\begin{array}{c} \mathbf{1} \\ \mathbf{0} \\ \hline \mathbf{0} \end{array}\right] \qquad \left[\begin{array}{c} \boldsymbol{c} \\ \mathbf{0} \\ \hline \boldsymbol{s} \end{array}\right] \qquad \left[\begin{array}{c} \mathbf{0} \\ \mathbf{0} \\ \hline \mathbf{1} \end{array}\right]$$

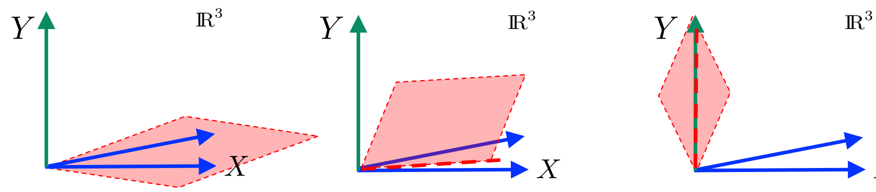
2.2.3. A red line in $\mathbb{R}^3$, X=the axis, Y=the y,z plane. ($m_1 = 2, m_2 = n = r = 1$) Below we show the possibilities for $[C; S]$ for a line in $\mathbb{R}^3$ (drawn in red as the span of $[a, b]$, where $a \in \mathbb{R}^1$, $b \in \mathbb{R}^2$). A line can be along the x-axis, in general position, or in the y, z plane. The corresponding [C:S] matrix is illustrated. The c is the angle between the red line and the $x$axis, while the s is the angle of the red line and the y,z-plane. The shaded Y=yz axis indicates the red line is in that plane.



$$\left[\begin{array}{c} \mathbf{1} \\ \hline \mathbf{0} \\ \mathbf{0} \end{array}\right] \qquad \left[\begin{array}{c} \boldsymbol{c} \\ \hline \boldsymbol{s} \\ \mathbf{0} \end{array}\right] \qquad \left[\begin{array}{c} \mathbf{0} \\ \hline \mathbf{1} \\ \mathbf{0} \end{array}\right]$$

2.2.4. A red plane in $\mathbb{R}^3$, X=the x,y plane, Y=the z axis. ($m_1 = 2, m_2 = 1, n = r = 2$) Below we show the possibilities for $[C; S]$ for a plane in $\mathbb{R}^3$ (drawn in red as the span of $[A, B]$, where $A \in \mathbb{R}^{2,2}$, $B \in \mathbb{R}^{1,2}$). A plane can be the x-y plane. A plane in general position in $\mathbb{R}^3$ intersects the x-y plane in a line (shown as a dashed red line) but does not include the z-axis. A final possibility for a plane is that it includes the z-axis (broken red/green line.)

The corresponding $[C; S]$ matrix is illustrated. We have $c_1 = 1$ corresponding to the 0 degree angle from a line in the red plane and the x-y axis. We have c_2 which is the cosine of the angle formed from a line at right angles from the aforementioned line and the x-y axis. Note that $s_1 = 0$ is not found in the S matrix, since there is room for only one row which contains s_2.



$$\left[\begin{array}{cc} \mathbf{1} & \\ & \mathbf{1} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \qquad \left[\begin{array}{cc} \mathbf{1} & \\ & \boldsymbol{c} \\ \hline & \boldsymbol{s} \end{array}\right] \qquad \left[\begin{array}{cc} \mathbf{1} & \\ & \mathbf{0} \\ \hline & \mathbf{1} \end{array}\right]$$

Figure 3 below is the ellipse picture in 3 dimensions, which admittedly has too few dimensions to understand the general picture. Nevertheless, one can clearly see the unit circle in the sphere being projected down to an ellipse on the x-y axis. We see the $c_1 = 1$ and $c_2 = \cos\theta$ as the lengths of the semi-axis of the ellipse. The u_1 direction is where the plane representing span$([A; B])$ intersects the xy plane. The u_2 direction is orthogonal to u_1 and also in the span$([A; B])$ plane. The u_2 direction is the maximum slope off the xy plane. $s_2 = \sin\theta$ is the length of the projection of the unit circle onto the z-axis. The orthogonal direction projects to 0 giving the $s_1 = 0$.

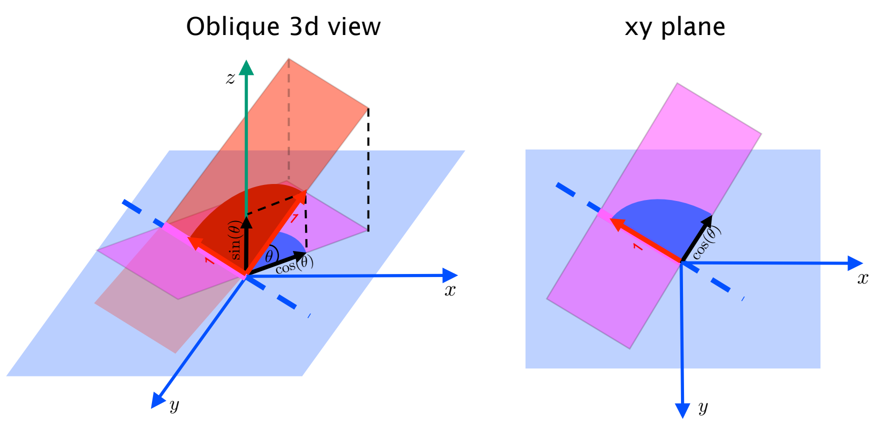


Fig. 3. *GSVD in 3d is a bit cramped: Oblique 3d view (left) and x-y plane (right). Generically a hyperplane will intersect the x-y plane in a line (blue dashed line) which will contain simultaneously the major axis of the blue (cosine) ellipse and a diameter of the red circle. In 3d, we have* $c_1 = 1, c_2 = \cos\theta$ *to indicate the intersection and the angle* θ *with the x-y plane, respectively. We also have* $s_1 = 0, s_2 = \sin\theta$ *which indicates that with respect to the z-axis, the red hyperplane has one vacuous direction (the red arrow in the x-y plane) and the orthogonal direction (other red arrow in the red hyperplane) makes an angle of* $\pi/2 - \theta$*. In summary, the blue (cosine) ellipse has semi-axes* 1 *and* $\cos\theta$*, the green (sine) ellipse is confined to 1d and has an unseen 0 and* $\sin\theta$*, while of course the unit circle has radius* 1.

2.3. On infinite generalized singular values and horizontal directions. As may become clear upon inspection of the small dimensional cases, it is very possible that we have some $c_i = 1$ and $s_i = 0$ so that the generalized singular value c_i/s_i is infinite. These infinite singular values are associated with horizontal directions $[u_i; 0]$ in the "red" hyperplane, i.e. $[u_i; 0] \in span([A; B])$. They arise when our hyperplane intersects our X multiaxis in any non-zero direction.

The situation in Section 2.2.4 illustrates that this is typical when we consider a plane in $\mathbb{R}^3$ and X is the xy plane. (A is 2×2 and B is 1×2.) The unit circle in the plane has a vector of length 1, $[u_1; 0]$, that lives on the horizontal x-y plane. The orthogonal direction, $[c_2u_2, s_2]$ has a projection $[c_2u_2; 0]$ on the x-y plane that is generically shorter than a unit vector, but still orthogonal to $[u_1; 0]$.

In Section 4.2 we will be interested in the projection P onto all directions $[u_i; 0]$ where $c_i < 1$.

3. Matrix Trigonometry. We claim that the GSVD is the natural generalization of high school trigonometry to what we might call "matrix trigonometry."

3.1. The Main Idea. There is so much in Figure 4 that we are all familiar with. There is all of **trigonometry,** and in particular there is $\tan\theta$ which has a special role because B/A is the **slope** of the line. If $|B|$ is small relative to $|A|$, we have a shallow slope, and vice versa. The only hint that there is some directionality is the possibility of a $\pm$ sign. To specify directions we sometimes would write a hypotenuse vector in **component form**: $A\mathbf{i} + B\mathbf{j}$. If we take the components of a unit vector in the direction of the hypotenuse, then the components form a **cosine-sine** pair: $\cos\theta\mathbf{i} + \sin\theta\mathbf{j}$.

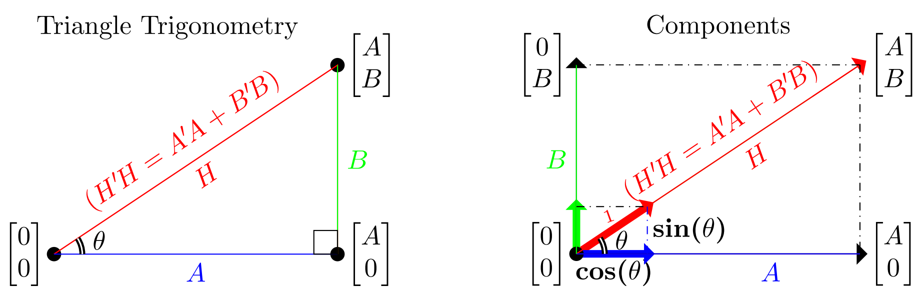


FIG. 4. *The GSVD is the generalization of the trigonometry picture (left) or the components picture (right) to higher dimensions. When A and B are 1×1 these pictures specialize to familiar grade school trigonometry (the 2d case).*

As a portrayer of higher dimensions, line segments represent hyperplanes, and the desired ellipses are hiding inside the subspaces as the thick unit vector along the hypotenuse (unit sphere in higher dimensions), and the thick components in the cosine-sine pair (horizontal and vertical ellipses in higher dimensions).

Notice that the generalized hypotenuse H is not the matrix square root but does satisfy $H'H = A'A + B'B$ (The reason a simple matrix sqrt does not work is that we must denote the direction of every component in higher dimensions). The cosine form of the gsvd denotes the singular values of A/H, and the sine form denotes the singular values of B/H.

The ideas of trigonometry, slope, component form and cosine-sine pairs extend to higher dimensions through the GSVD. Instead of one triangle, there are n triangles. Instead of one vector $\mathbf{i}$, there are n vectors in the columns of U. Instead of one vector $\mathbf{j}$, there are n vectors in the columns of V. Instead of a unit length hypotenuse there are n unit length hypotenuses, which can be written in the component form

$$\cos\theta_k \begin{bmatrix} u_k \\ 0 \end{bmatrix} + \sin\theta_k \begin{bmatrix} 0 \\ v_k \end{bmatrix}, \qquad k = 1, 2, \ldots, n.$$

The n hypotenuses, as we show in Figure 2, live on a unit sphere that projects nicely "down"ward and "left"ward. The $\cos\theta_k u_k$ are semi-axes of the downward ellipse; and the $\sin\theta_k v_k$ on the leftward ellipse.

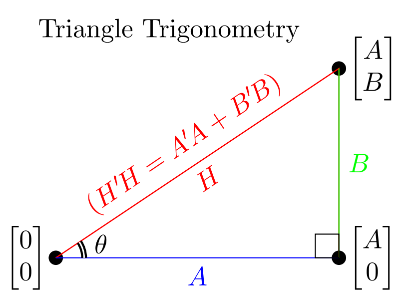
Just as b/a tells you how small or big b is relative to a, the gsvd tells you how small or big B is relative to A, but now it is in n natural directions. Thus B can be larger than A in some directions, and smaller in others.

There is some temptation to try to say that the GSVD is related to the principle angles of the column space of A and the column space of B. This of course makes no more sense than looking for anything other than right angles between the x-axis and the y-axis in 2d. The interesting angles are between the span of the column space of $\begin{bmatrix}A\\B\end{bmatrix}$. and the canonical axes $\begin{bmatrix}I_1\\0\end{bmatrix}$. More details can be found in Section 5.

One quick algebraic way to define the singular values of an $m\times n$ matrix A is to find the diagonal matrix with non-negative entries in the set $\{UAV'\}$ where U is $m\times m$ orthogonal and V is $n\times n$ orthogonal. Similarly one can define the generalized singular values of a pair of matrices (A,B) with the same number of columns. The "cosine-sine" format, is the pair of diagonal matrices (C,S) with non-negative entries in the set of matrix pairs $\{(UAH^{-1}, VBH^{-1}) : U, V \text{ orthogonal}, H \text{ non-singular}\}$. Often the gsvd is given in "cotangent" format, which is the ratio of cosines to sines.

We summarize the GSVD properties with the following table.

$$\left[\frac{A}{B}\right] = \left[\frac{UC}{VS}\right] H$$

C,S		θ: Principle angle between span$\{\begin{bmatrix}A\\B\end{bmatrix}\}$ and span $\{\begin{bmatrix}\mathbb{I}_n\\0\end{bmatrix}\}$ $\sin\theta$: svd$(BH^\dagger)$ $\cos\theta$: svd$(AH^\dagger)$ $\tan\theta$: svd$(BA^\dagger)$ if $r=r_a$ $\boldsymbol{\cot\theta}$: svd$(AB^\dagger)$ if $r=r_b$
U	left singular vectors of $AH^\dagger$ (or $AB^\dagger$ if $r=r_b$)	
V	left singular vectors of $BH^\dagger$ (or $BA^\dagger$ if $r=r_a$)	

3.2. The relationship between the GSVD and the CS Decomposition. It is often written [17] that the GSVD and the CS Decomposition are closely related. The geometric viewpoint highlights the GSVD and the CS decomposition as rooted in representations of points in the Grassmann manifold (linear hyperplanes through the origin) in an $m = m_1 + m_2$ dimensional space using $[UC; VS]$ as natural coordinates.

The simple notion is that the information may be thought of as

$$\begin{bmatrix}A\\B\end{bmatrix} = \underbrace{\begin{bmatrix}UC\\VS\end{bmatrix}}_{\substack{\text{column space as a hyperplane}\\\text{(a canonical basis!)}}} \times \underbrace{H}_{\substack{\text{Coordinates of }[A;B]\\\text{in the }[UC;VS]\text{ basis.}}}$$

This connection is rooted ultimately in the Cartan decomposition of the Grassmann manifold, one of the finitely many classes of symmetric spaces [19]. The idea is that certain matrix spaces have a "KAK" or compact/abelian/compact decomposition. The SVD is one example as it is orthogonal/diagonal/orthogonal. The CS decomposition is another. This observation may be found in a numerical linear algebra conference presentation [13] and in the quantum computing literature [31].

To be sure if $[A;B]$ is already orthogonal then so is H. This constitutes the "left half" of the complete CS decomposition. Thus a GSVD is a left-half of a CS, when $[A;B]$ are orthogonal, and the left-half of a CS is a GSVD. One can also have a basis for the orthogonal complement of span($[A;B]$) to get the "right half." This captures the isomorphism between $G_{m,n}$ and $G_{m,n-m}$. Thus if one takes the combined svd's of orthogonal matrices whose spans are orthogonal complements, one has the CS decomposition and vice versa.

Any which way, the mathematical idea underlying all is that there is a fairly canonical representation for generic elements of the Grassmann manifold and a matrix connecting back to an orthogonal or arbitrary basis which has a further symmetry property when taking both the span of $[A;B]$ and its orthogonal complement in conjunction in that transposing a full orthogonal matrix reverses the roles canonical coordinates and basis converter.

3.2.1. Parameter Count. There has been a longstanding tradition in numerical linear algebra to overwrite matrix inputs with the parameters from the factored form. Thus if A is $n \times n$, the LU factorization has the $n(n-1)/2$ parameters from L and te $n(n+1)/2$ parameters from U. Similarly if $A = QR$, the Q while appearing naively as an $n \times n$ matrix, actually only has $n(n-1)/2$ parameters, which is exactly what is computed in software.

Given an $m \times n$ matrix $[A;B]$ of rank r, and a decomposition of m as $m = m_1 + m_2$, we can count parameters on both the left and right sides of $[A;B] = [UC;VS]H$. While tricky, the only facts needed are:

1. Rank Codimension: The codimension of the rank r matrices of size $m \times n$ is $(m-r)(n-r)$ [10, Lemma 3.3].
2. Stiefel Manifold Dimension: The dimension of the Stiefel manifold $V_{m,n}$ of n ordered orthonormal directions in $\mathbb{R}^m$ is $n(m-n) + n(n-1)/2$ [15, Section 2.2].
3. Grassmann Manifold Dimension: The dimension of the Grassmann manifold $G_{m,n}$ of n-dimensional subspaces in $\mathbb{R}^m$ is $n(m-n)$ [15, Section 2.5].

	$r \le m_1 \le m_2$	$m_1 \le r \le m_2$	$m_1 \le m_2 \le r$
rank r codim	$(m-r)(n-r)$	$(m-r)(n-r)$	$(m-r)(n-r)$
H $(r \times n)$	rn	rn	rn
$0 < \theta_i < \pi/2$	r	m_1	$m-r$
U Stiefel	$(m_1-r)r$ $+r(r-1)/2$	$m_1(m_1-1)/2$	$(r-m_2)(m-r)$ $+(m-r)(m-r-1)/2$
V Stiefel	$(m_2-r)r$ $+r(r-1)/2$	$(m_2-m_1)m_1$ $+m_1(m_1-1)/2$	$(r-m_1)(m-r)$ $+(m-r)(m-r-1)/2$
V Grassmann	0	$(r-m_1)(m_2-r)$	0
Total	mn	mn	mn

To understand the parameter count we begin with the simple observation that $r_a = \min(r, m_1)$ generically and $r_b = \min(r, m_2)$, from which we can derive the number of θ_i that are strictly between 0 and $\pi/2$ as $r_a + r_b - r$. The relevant Stiefel manifolds are V_{m_1, r_a+r_b-r} and V_{m_2, r_a+r_b-r}. These correspond exactly to choosing the directions for the axes of the ellipses. Also one must consider $G_{m_i-(r_a+r_b-r), r-r_a}$ for $i = 1, 2$ as this is the dimension divide between the 0 degree angles and the $\pi/2$ angles when this has content. This data is summarized below:

	$r \le m_1 \le m_2$	$m_1 \le r \le m_2$	$m_1 \le m_2 \le r$
r_a	r	m_1	m_1
r_b	r	r	m_2
$r_a + r_b - r$	r	m_1	$m - r$
Stiefel Manifolds	$V_{m_1,r}, V_{m_2,r}$	V_{m_1,m_1}, V_{m_2,m_1}	$V_{m_1,m-r}, V_{m_2,m-r}$
Grassmann Manifold		$G_{m_2-m_1,r-m_1}$	

We remark that further fine grain detailed parameter counts are possible including lower rank A and B, but we content ourselves with the table above.

4. On the gsvd(A, B) and the svd$(AB^\dagger)$. In this section we relate the finite part (nonzero, noninfinite) of the generalized singular values of (A, B) to the singular values of $AB^\dagger$ where $B^\dagger$ is the pseudoinverse of B. We may use the notation A/B for $AB^\dagger$. An issue arises that may surprise some readers.

4.1. Why there is an issue. One may expect that there may always be a relation between the gsvd of A, B and the svd of $AB^\dagger$. For example, in the matlab documentation[1] it is stated that the generalized singular values are the ratios of the diagonal elements of C and S in a given example. One might infer from the documentation that this is always the case.

However it is not generally true when there are infinite singular values, i.e, when $r_b < r$.

A simple example takes A to be a non-singular $n \times n$ matrix, and B to be a nonzero $1 \times n$ matrix. In this case $r_b = 1, r = n$. The gsvd of A, B is readily verified to have $n - 1$ infinite singular values, and the finite value

$$\sigma_{\text{gsvd}} = 1/\|B/A\|.$$

The svd of $AB^\dagger$ is just the length of $AB^\dagger = AB'/\|B\|^2$ or

$$\sigma_{\text{svd}} = \|BA'\|/\|B\|^2.$$

When $n = 1, A = a, B = b$, both of these expressions are equal to the absolute ratio $|a/b|$, ($r = r_b = 1$ after all) but for larger n the two matrix expressions are not equal.

An extremely simple special case takes $A = \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \end{pmatrix}$. The two values are $\sigma_{\text{gsvd}} = 2.4$ and $\sigma_{\text{svd}} = 2.5$ exactly.

The issue arises exactly when there are infinite σ. If there are no infinite σ, S has no 0 columns, and we can write

$$AB^\dagger = (UCH)(VSH)^\dagger = UCHH^\dagger S^\dagger V' = U(C/S)V',$$

which is a singular value decomposition of A/B. (We use the property that H has full row rank to conclude $HH^\dagger = I_r$ and that C/S is an $m_1 \times m_2$ matrix with c_i/s_i on the main diagonal.)

[1] https://www.mathworks.com/help/matlab/ref/gsvd.html

The problem that arises when some $\sigma = \infty$ is that $B^\dagger = (VSH)^\dagger = (SH)^\dagger V'$ does not equal $H^\dagger S^\dagger V'$ when S has any zero columns.

4.2. The significance of horizontal directions and their orthogonal complement in X. In Section 2.3, we considered the intersection of span($[A; B]$) with the X multiaxis. An orthogonal basis for this intersection is $[u_1; 0], \ldots, [u_{r-r_b}; 0]$ which correspond exactly to the $c_i = 1$.

Working entirely in X as an m_1 dimensional space, we are interested in the $m_1 \times m_1$ projection matrix P that kills the directions of intersection. Precisely we define P on the orthogonal basis for $\mathbb{R}^{m_1}$:

$$Pu_i = \begin{cases} u_i & \text{if } c_i < 1 \\ 0 & \text{if } c_i = 1 \end{cases}$$

Suppose N is a matrix whose columns are a basis for the null space of B. If we consider AN then the span of the columns of AN is the intersection we are discussing, i.e., the intersection of X with span($[A; B]$). To be sure either the column of N is in the common null space of A and B, so that the corresponding column of AN is 0, or else if one follows through the first $r - r_b$ columns of $H^\dagger$ in $A = UCH^\dagger$, one sees that we will hit the "$c_i = 1$" columns in C only, hence we will emerge a linear combination of $u_1, \ldots, u_{r-r_b}$.

We can thus describe P as the orthogonal projection onto the left nullspace of AN which is the orthogonal complement of the column space of AN.

4.3. The correct modified theorem requires PA/B . We remind the reader of the usual definition of the matrix pseudoinverse in terms of the singular value decomposition:

$$A^\dagger = V\Sigma^\dagger U^T, \tag{4.1}$$

where $\Sigma^\dagger$ means taking the inverse of the finite entries in Σ. When A has full column rank and B has full row rank, we have $(AB)^\dagger = B^\dagger A^\dagger$. It is easy to see that $[\mathbf{0}\ B]^\dagger = [\mathbf{0}; B^\dagger]$.

Theorem 4.1. *Let N be a matrix whose columns are a basis for the nullspace of B, and P be the orthogonal projection onto the left nullspace of AN. The finite non-zero generalized singular values of (A, B) are the same as the non-zero singular values of $PAB^\dagger$.*

Proof. Setting notation, we have

$$A = U \left[\begin{array}{c|c} \begin{smallmatrix} 1 & & \\ & 1 & \\ & & \ddots & \\ & & & 1 \end{smallmatrix} & \overbrace{\quad 0 \quad}^{C_*} \\ \hline & \begin{smallmatrix} c_{r-r_b+1} & \\ & \ddots \end{smallmatrix} \end{array}\right] H$$

$$B = V \left[\begin{array}{c|c} \begin{smallmatrix} 0 & & \\ & 0 & \\ & & \ddots & \\ & & & 0 \end{smallmatrix} & \overbrace{\quad 0 \quad}^{S_*} \\ \hline & \begin{smallmatrix} s_{r-r_b+1} & \\ & \ddots \end{smallmatrix} \end{array}\right] H = V \overbrace{\left[\begin{array}{c} 0 \\ \hline \begin{smallmatrix} s_{r-r_b+1} & \\ & \ddots \end{smallmatrix} \end{array}\right]}^{S_*} H_* = VS_*H_*. \qquad \square$$

so that $B = VS_*H_*$, where S_* are the rightmost r_b non-zero columns of S (indexed by $i = r - r_b + 1, ..., r$) and H_* are the corresponding rows (the bottom r_b) of H.

(To see this note that $B = V[0\ S_*][?; H_*]$ where the "?" denotes rows that hit the 0 columns in S so we do not care what they are.) We point out that H_* has full row rank as the rows of H_* are a subset of the full row rank matrix H. We immediately conclude that

$$B^\dagger = H_*^\dagger S_*^\dagger V'.$$

We further claim that

$$PA = UC_*H_*,$$

where C_* are the exact corresponding columns of C (the rightmost r_b indexed by $i = r - r_b + 1, ..., r$), which are the $c_i < 1$. To see this, first observe that the definition of P as described in Section 4.2. is $PU = U[0\ \ I_*]$ where I_* are the rightmost r_b columns of the identity indexed by $i = r - r_b + 1, ..., r$. Thus $PA = U[0\ C_*][?; H_*] = UC_*H_*$ the 0 indicating the columns of U killed by P.

Now that we have compressed out the immaterial columns, and knowing that $H_*H_*^\dagger = I_{r_b}$ by the full row rank condition, we can compute

$$PAB^\dagger = UC_*H_*H_*^\dagger S_*^\dagger V' = UC_*/S_*V'.$$

This is a singular value decomposition of $PAB^\dagger$, with $\Sigma = C_*/S_*$ an $m_1 \times m_2$ diagonal matrix, with the c_i/s_i in decreasing order on the diagonal and no $s_i = 0$.

Corollary 4.2. *If B has full column rank ($r_b = n$) or if the weaker condition holds that $r = rank([A; B]) = r_b = rank(B)$, then P is not needed, i.e. the finite non-zero generalized singular values of (A, B) are the same as the non-zero singular values of $AB^\dagger$.*

Proof. If $r_b = n$, then B has nothing in the nullspace, N has no columns, and P is obviously I. More generally, if $r_b = r$, then B has nothing in its nullspace that is not also in the nullspace of A, so if AN has any columns at all, it is the zero matrix, so again projection onto the left nullspace is $P = I$.

4.4. Blame the pseudoinverse not the GSVD. The difficulty with $AB^\dagger$ may seem like an unfortunate consequence of infinite singular values, but in point of fact, it is related to the discontinuity in the definition of the pseudoinverse. If one takes a bigger picture viewpoint, it is easy to see that infinite singular values are natural limits of finite singular values.

The only truly natural discontinuity in the GSVD is the reduction of rank of $[A; B]$ which reduces the dimensionality of the hyperplane (and the rank of H.)

We mention some limit type results which help understand the nature of the infinite generalized singular values:

Theorem 4.3. *If rank($[A; B]$)=r, and $m_2 \geq r$, then we can define a continuous curve of matrices $[A_\epsilon, B_\epsilon]$ of the same shape as $[A; B]$ without infinite generalized singular values when $\epsilon > 0$ is small but whose limit as $\epsilon \to 0$ continuously converges to the generalized singular values of $[A, B]$, finite or infinite.*

Proof. Take

$$[A_\epsilon; B_\epsilon] = [UC(\epsilon); VS(\epsilon)]H,$$

where

$$c_i(\epsilon) = \begin{cases} c_i & s_i > 0 \\ \cos(\epsilon) & s_i = 0 \end{cases} \quad \text{and} \quad s_i(\epsilon) = \begin{cases} s_i & s_i > 0 \\ \sin(\epsilon) & s_i = 0 \end{cases}.$$

Corollary 4.4. *If rank($[A;B]$)=r, and $m_2 < r$, then we can define a continuous curve of matrices $[A_\epsilon, B_\epsilon]$ without infinite generalized singular values when $\epsilon > 0$ is small but whose limit as $\epsilon \to 0$ continuously converges to the generalized singular values of $[A, B]$ by row augmenting B_ϵ to contain r rows.*

Proof. Simply add $r - m_2$ rows of zeros to the bottom of B. This does not change the generalized singular values of $[A;B]$ or U,C or H. S is augmented with $r - m_2$ rows of zeros and V is augmented with $r - m_2$ rows and columns with an identity matrix. Apply the construction in Theorem 4.3 to complete the proof. □

Example: Consider that

$$\text{gsvd}\left(\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 1 \end{bmatrix}\right) = \ 2.4 \text{ and } \infty.$$

One might seek nearby matrices with no infinite generalized singular values. This is impossible if we insist that B remain 1×2 but is possible if we augment B with one row, which in this case we can simply take

$$\text{gsvd}\left(\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & \epsilon \end{bmatrix}\right) = \ 2.4 + O(\epsilon^2) \text{ and } 5/\epsilon + O(\epsilon).$$

Corollary 4.5. *Suppose $[A_\epsilon, B_\epsilon]$ has rank r for $0 \le \epsilon < \epsilon_0$ is a continuous curve, where B_ϵ has rank r for $\epsilon > 0$ but may drop rank at $\epsilon = 0$. We then have that the generalized singular values are a continuous function of $[A_\epsilon, B_\epsilon]$ as $\epsilon \to 0$.*

Proof. The only true discontinuity in the gsvd is the potential for a drop in rank of $[A;B]$. This is avoided in the statement by keeping $[A_\epsilon, B_\epsilon]$ rank r. Thus the limit of the column space is the column space of the limit. □

We do remark on the other hand that if $[A_\epsilon, B_\epsilon]$ drops rank, then we can only say that the limit of the column space contains the column space of the limit, which can lead to all kind of discontinuities in the generalized singular values.

5. Principal angles between subspaces.

It is well known that if one has two matrices A_1 and A_2 with the same number of rows then one can compute the principal angles and principal vectors [17] by computing orthogonal bases encoded in Q_1 and Q_2 for $A_1 = Q_1 R_1$ and $A_2 = Q_2 R_2$ using QR, or rank revealing QR, or the SVD. Then one computes the svd of $Q_1^T Q_2$, which yields the cosines of the principal angles.

One can obtain the same information from the GSVD. Merely one has to realize that one has to rotate the coordinate axes to the span of A_2.

Thus, let Q_1 be the product of Householder matrices that encode the span of A_2, presumably stored in compact format so that it also encodes an orthogonal basis for the subspace perpendicular to the span of A_2. Then one merely needs to compute $Q^T A_1$ and separate this matrix according to the rank of A_2 and take the gsvd. This can be made explicit by writing $Q = [Y | Y^\perp]$, but remember that in compact format, one does not ever explicitly store $Y^\perp$. The GSVD is performed on $Y^T A_1$ and $(Y^\perp)^T A_1$.

Note that $Y^T A_1$ amounts to $(A_1^T Y Y^T A_1)^{1/2}$ or $(A_1^T P A_1)^{1/2}$, where P is the projection onto the span of A_1, and similarly $(Y^\perp)^T A_1$ amounts to $(A_1^T P^\perp A_1)^{1/2}$ where $P^\perp$ is the projection onto the orthogonal complement of the span of A_1, but this would be not the best numerical procedure.

6. Geometry of Tikhonov Regularization.

6.1. The two cosine damping. We show how geometry can add insight to our understanding of Tikhonov Regularization:

$$\min_x \{\|Ax - b\| + \lambda \cdot \|Lx\|\} \tag{6.1}$$

by providing a two cosines view of damping. Specifically, the way Tikhonov regularization reduces the solution or "weights," is usually understood algebraically in terms of adding a regularizer term that moves the original problem away from some kind of ill-conditioned setting. We will show that one cosine comes from the projection from the horizontal (blue) plane to the span of $[A; \lambda L]$ red plane. The other cosine comes from the non-canonical basis of the plane: the columns of $[A; \lambda L]$ which enlongate with λ, hence the coordinates shrink.

While the "calming influence" [17, Section 6.1.26], [4, Section 4.4],[18] of the regularization parameter λ has been well studied algebraically, we identify geometrically in (6.2) the influence as a factor of $\cos^2\theta_\lambda$ where $\tan\theta_\lambda = \lambda\tan\theta_1$ so that $\cos^2\theta_\lambda = 1/(1+\lambda^2\tan^2\theta_1)$, where θ_1 is the angle that corresponds to $\lambda = 1$. We will compare the $\cos^2$ formulation with previous formulations explaining why we find that this formulation feels somewhat more insightful.

Before we start, let us recap Tikhonov regularization. Suppose we have a matrix A, which we will assume has full column rank. The $\lambda = 0$ problem (standard least squares) is the computation of $x_0 = A^\dagger b = (A^TA)^{-1}A^Tb$, the standard solution to the normal equations $A^TAx = A^Tb$. To regularize we pick a suitable matrix L, and a "regularization parameter" λ, and then solve instead $(A^TA + \lambda^2L^TL)x = b$, which is equivalent to computing

$$x_\lambda = \begin{bmatrix} A \\ \lambda L \end{bmatrix}^\dagger \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

From the geometrical point of view, we believe the reformulation in Theorem 6.1 below is more revealing of the "calming effect." Figure 5 demonstrates the hyperplane onto which $[b; 0]$ gets projected for varying λ.

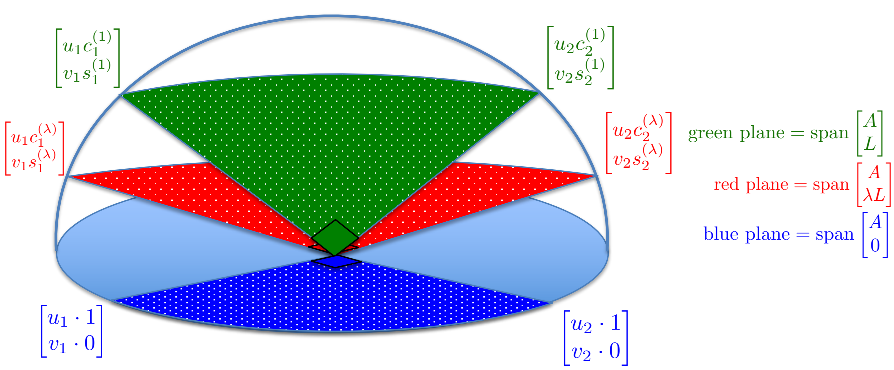


FIG. 5. *This n=2 Tikhonov regularization picture in the four dimensional hypersphere illustrates the hyperplanes onto which* $[b; 0]$ *gets projected for varying* λ. *The projection gives one cosine, while the representation (not pictured) in ever elongating bases gives the second cosine. Portrayed is the unique hypersphere containing the four mutually orthogonal vectors in four dimensions:* $[u_1, 0], [u_2, 0], [0, v_1], [0, v_2]$, *While tempting to see this as a 3d object, as* $\lambda \to \infty$ *the wedge drawn does not shrink but remains a quarter circle wedge.*

For every λ, we obtain the GSVD as a continuous function of λ:

$$[A; \lambda L] = [UC_\lambda, VS_\lambda]H_\lambda,$$

where it is easy to check that H_λ is square non-singular. It is convenient to use the compact format described in Section 1.4.4 here. Thus we take U to be $m_1 \times n$, C and S to be square diagonal $n \times n$. The exact values in C and S come from the trigonometry with unit hypotenuse, fixed base, and sliding height of a $c, s, 1$ triangle at $\lambda = 1$, as shown in the left side of Figure 6. Namely

$$C_\lambda = \frac{C_1}{\sqrt{C_1^2 + \lambda^2 S_1^2}} \quad \text{and} \quad S_\lambda = \frac{\lambda S_1}{\sqrt{C_1^2 + \lambda^2 S_1^2}},$$

where the operations happen on the diagonal. It also follows that

$$H_0 = C_\lambda H_\lambda, \text{ and, } A = UH_0 = UC_\lambda H_\lambda, \ \forall \lambda \geq 0.$$

The equation $H_0 = C_\lambda H_\lambda$ has a nice trigonometric interpretation. As the column vectors of $[A; \lambda L]$ grow in length (these lengths are encoded in H_λ). the cosines in C_λ relate back to the $[A; 0]$ columns which are shorter in length. This is depicted in Figure 6.

Theorem 6.1. *The solution x_λ to the Tikhonov Regularization problem can be written as*

$$x_\lambda = \left(H_0^{-1} C_\lambda^2 \ H_0\right) x_0, \tag{6.2}$$

where x_0 is the least squares solution to $Ax = b$ and $A = UH_0$, where $[A; \lambda L] = [UC_\lambda; VS_\lambda] H_\lambda$.

Proof. Since

$$x_\lambda = \begin{bmatrix} A \\ \lambda L \end{bmatrix}^\dagger \begin{bmatrix} A \\ 0 \end{bmatrix} x_0,$$

we can calculate

$$x_\lambda = H_\lambda^{-1} C_\lambda U^T U H_0 x_0 = H_\lambda^{-1} C_\lambda H_0 x_0$$

and use the relation $H_\lambda^{-1} = H_0^{-1} C_\lambda$ to complete the proof.

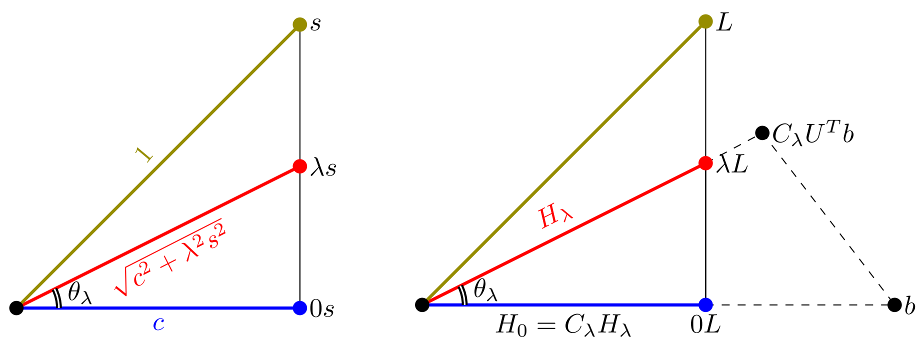


Fig. 6. *The "two cosine" Geometric interpretation of Tikhonov regularization: Single u-v plane (left) vs. general (right). The green, red and blue lines represent the span of $[A; L]$ (green) , $[A; \lambda L]$ (red) and $[A; 0L]$ (blue) respectively. Our "two cosines" view of regularization is that one cosine dampening comes from the projection of b from the blue plane to the red plane, and the second cosine comes from the extended basis $H_\lambda = C_\lambda^{-1} H_0$ which gets divided. Note that the value of λ may be greater than 1 (not shown).*

6.2. Comparison and Discussion. The standard application of the GSVD to Tikhonov relates x_λ to b and thus gives formulas involving the non-physical, non-homogeneous factor of $c/(c^2+\lambda^2 s^2)$ rather than the homogeneous $c_\lambda^2 = c^2/(c^2+\lambda^2 s^2)$.

The formulation in Theorem 6.1 diagonalizes the operator that relates x_λ to x_0. We understand that when x are the coordinates of a linear combination of the columns of $[A;B]$, we have that $H_0 x$ are the coordinates of that same vector in the natural basis. Thus the interpretation of $H_0^{-1} C_\lambda^2 H_0$ simply is:

1. Write the vector in the natural coordinate system
2. Multiply by a cosine squared in every natural direction
3. Return to the original coordinate sytem.

7. Comparative Data Modelling. In a series of beautiful applications of the GSVD, Alter, et.al. [2, 27, 28, 30, 1] propose an approach towards data reconstruction and classification. In their case [2], the A and B are two DNA microarrays, one from humans and the other from yeast. The rows of A and B live in $\mathbb{R}^n$ or gene space. The rows of H form a basis for this row (or gene) space, and are denoted genelets. A natural question is whether the genelet is primarily human, primarily yeast, or a mixture. In general, given two matrices with equal columns, one wants to classify the basis vectors in the rows of H according to its source.

The GSVD provides a natural solution by creating a single coherent model from the two datasets recording different aspects of interrelated phenomena by simultaneously identifying the similar and dissimilar between the two corresponding column-matched but row-independent matrices. For each of the r rows, we have that θ_i denotes the angle towards A. In Figure 7, we portray this. We note that [2] displays the angles from $-\pi/4$ to $\pi/4$, but we will stick with the 0 to $\pi/2$ convention. It is convenient that the rows of H are already sorted from “most A”, to “most B”.

Our ellipse picture Figure 2 reveals the geometry readily. The $[u_i c_i; v_i s_i]$ all appear on the unit ball.

The comparative Data Reconstruction equation is

$$\begin{pmatrix} A \\ B \end{pmatrix} = \sum_{i=1}^{r} \begin{pmatrix} u_i c_i \\ v_i s_i \end{pmatrix} h_i^T,$$

where h_i^T is the ith row of H. (This is exactly Equation 1.2.) One can preprocess H so that each row is of unit direction as it is only the ratio of c_i to s_i that matters. Any ill-conditioning of H could be worrisome.

8. The Lemniscate Plots from Leuven, Belgium. In a series of early papers most of which date back to the 1980s [5, 6, 7, 8, 9, 33], energy portraits that relate to the SVD and GSVD of a matrix or a pair of matrices are discussed with applications.

The definition of an energy portrait of a single matrix is

$$\text{Energy}(A) = \{e\|Ae\|^2 : \|e\| = 1\} \subset \mathbb{R}^n,\ (A \in \mathbb{R}^{m,n})$$

and for a pair of matrices with the same number of columns

$$\text{Energy}(A,B) = \{e\frac{\|Ae\|^2}{\|Be\|^2} : \|e\| = 1\} \subset \mathbb{R}^n,\ (A \in \mathbb{R}^{m_1,n}, B \in \mathbb{R}^{m_2,n})$$

It is important to point out that these are not ellipse pictures but rather lemniscate like portraits. They do not even live in the same spaces as the ellipse pictures.

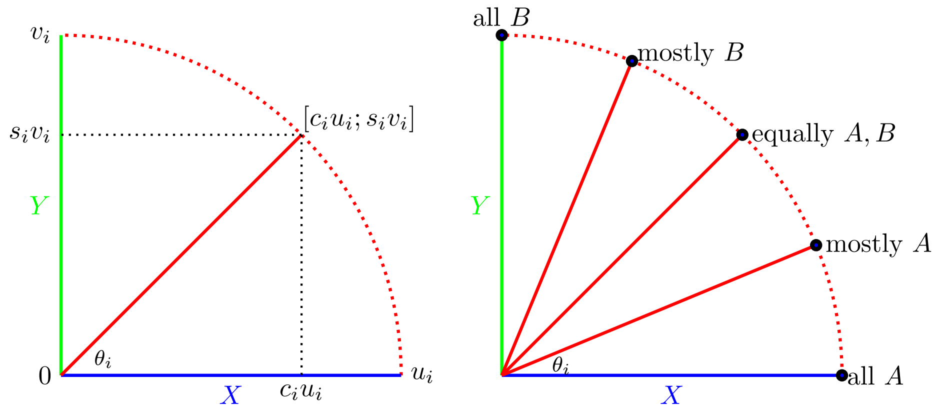


FIG. 7. *Are genomes human, yeast, or a combination? (Application from Alter [2]) Left: $[c_iu_i; s_iv_i]$ makes an angle θ_i with the X multiaxis. Right: Depending on the angle we apportion the ith row of H (a basis element for the rowspaces of A and B) as being attributable to A or B.*

The standard SVD ellipse lives in $\mathbb{R}^m$ and the GSVD picture in this paper lives in $\mathbb{R}^{m_1+m_2}$. By contrast the energy portraits from Leuven live in $\mathbb{R}^n$.

For completeness, we thought we would take a closer look at these older plots. To explain in what sense the curves are lemniscates, it is best to eliminate the "e" in the definition and rewrite the energy plots as the zero set of an algebraic equation, thereby connecting the portraits to the field of algebraic geometry.

THEOREM 8.1. *If $Vx \in$ Energy(A), then x satisfies the algebraic polynomial equation*

$$\left[\sum x_i^2\right]^3 = \left[\sum \sigma_i^2 x_i^2\right]^2,$$

where $A = U\Sigma V^T$. Further if $x \in$ Energy(A, B), then x satisfies the algebraic polynomial equation

$$\|x\|^2\|SHx\|^4 = \|CHx\|^4,$$

where $[A; B] = [UC; VS]H$.

Before proving the theorem we provide a historical analog. We might compare the solution set of $(\sum_{i=1}^n x_i^2)^3 = (\sum_{i=1}^n \sigma_i^2 x_i^2)^2$, with that of $(\sum_{i=1}^2 x_i^2) = (\sum_{i=1}^2 \sigma_i^2 x_i^2)$, which is the lemniscate of Booth whose study traces back to the 5th century Greek philosopher Proclus. The difference being that Booth specialized to $n = 2$ and only took first powers of the quantities, but in spirit it is a similar algebraic polynomial equation.

Proof. Taking $e = Vy$, we see that $e\|Ae\|^2 = Vy\|\Sigma y\|^2 = Vx$ where $x = y\|\Sigma y\|^2$. It is straighforward to check $\|x\|^6 = \|\Sigma x\|^4 = \|\Sigma y\|^{12}$, since $\|y\| = 1$ which is exactly the result for a single matrix.

For the two matrix case, where $A = UCH$ and $B = VSH$, if $x = e\|Ae\|^2/\|Be\|^2$, then

$$\|x\|^2 = \frac{\|CHe\|^4}{\|SHe\|^4}, \quad \text{and} \quad \frac{\|CHx\|}{\|SHx\|} = \frac{\|CHe\|}{\|SHe\|}.$$

9. A one matrix and one subspace view of the GSVD. The focus on two matrices with the same number of columns is not always the best view of the GSVD.

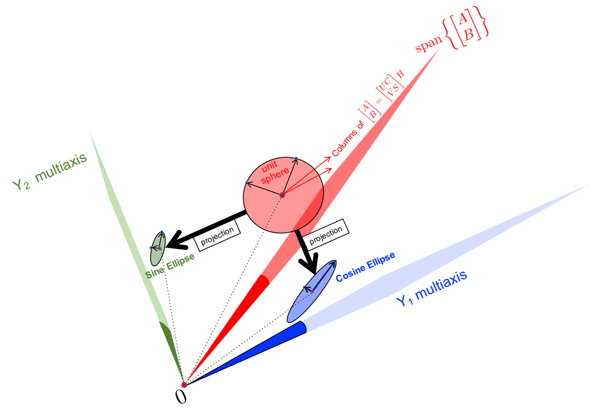


FIG. 8. *The ellipse picture in Figure 2 need not fundamentally line up with horizontal and vertical multiaxes. This rotated geometry underlies a signal processing application in [21, 22].*

One can take rather a single $m \times n$ matrix M and any m_1 dimensional reference subspace $\mathcal{S}$ of $\mathbb{R}^m$. We can then think of the GSVD as an additive decomposition:

$$M = P + Q,$$

where $P = Y_1UCH$ and $Q = Y_2VSH$, and the columns of Y_1, Y_2 are orthonormal bases for $\mathcal{S}$ and $\mathcal{S}^\perp$ respectively. Conversely, $[Y_1 \ Y_2]^T M = [Y_1^T M; Y_2^T M]$ is an ordinary GSVD.

By doing this we have a decomposition of $M = P + Q$ such that $P^TQ = Q^TP = 0_{n\times n}$. Geometrically, instead of decomposing into a "top half" and "bottom half", into a "horizontal" and "vertical" multiaxis subspace, we are rather allowing for general multiaxes subspaces. One might think of this as a rotated view of Figure 2. More specifically most of this paper would take $Y_1 = [I; 0]$ and $Y_2 = [0; I]$, but all that is required is that Y_1 and Y_2 are orthogonal complements.

This geometrical insight underlies an additive decomposition signal processing application found in [21, 22] where P and Q play the role of signal + noise.

10. Orthonormal Bases for $\{Ax : Bx = 0\}$ and Friends. The U matrix of the GSVD provides, in its columns, orthonormal bases for three mutually orthogonal subspaces that arise in many applications:

$$U = \left[\begin{array}{c|c|c} U_1 = & U_2 = & U_3 = \\ \text{orthonormal} & \text{completion to} & \text{orthonormal} \\ \text{basis for} & \text{all of} & \text{basis for} \\ \{Ax : Bx = 0\} & col(A) = \{Ax\} & col(A)^\perp \end{array}\right].$$

From the perspective of Figure 2, there are the horizontal directions in the red unit sphere, the generic directions, and the directions that are not present.

10.1. Clustering Matrices. An important example where the GSVD lurks implicitly or explicitly is clustering. We will consider an A matrix that indicates the clustering, and a B matrix that indicates equality between the clusters.

We consider data in $\mathbb{R}^p$ and assume a partitioning of $p = p_1 + \ldots + p_k$, into clusters. The indicator matrix corresponding to the partition of p is :

$$A = \text{Indicator}(p_1,p_2,\ldots,p_k) \;=\; \begin{array}{c} | \\ p_1 \\ | \\ | \\ p_2 \\ | \\ \vdots \\ | \\ p_k \\ | \end{array} \overbrace{\begin{bmatrix} 1 & & & \\ \vdots & & & \\ 1 & & & \\ & 1 & & \\ & \vdots & & \\ & 1 & & \\ & & \vdots & \\ & & & 1 \\ & & & \vdots \\ & & & 1 \end{bmatrix}}^{k},$$

which we can normalize by setting

$$Y_1 = \text{Indicator}(p_1,p_2,\ldots,p_k) \times \text{Diagonal}(\frac{1}{\sqrt{p_1}}, \frac{1}{\sqrt{p_2}}, \ldots, \frac{1}{\sqrt{p_k}}).$$

In the Julia computing language, the indicator matrix can be generated succintly with `A = cat(ones.(Int,partition)...,dims=1:2)`.

Another useful matrix in this context is the constraint matrix whose nullspace is the all ones vector:

$$B = \begin{array}{c} | \\ | \\ k-1 \\ \text{rows} \\ | \\ | \end{array} \overbrace{\begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & -1 \\ 0 & 1 & 0 & \ldots & 0 & 0 & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 & -1 \\ 0 & 0 & 0 & \ldots & 0 & 1 & -1 \end{bmatrix}}^{k \text{ columns}}.$$

In Julia, with the LinearAlgebra package, this may be written succintly as `B = [I -ones(k-1)]`. To obtain the U matrix for the GSVD, one can then set `U, = svd(A,B)`, where the comma indicates that you are requesting only the U matrix. The immediate interpretation in this special case of the GSVD U is

$$U = \begin{array}{c} | \\ p \\ \text{rows} \\ | \end{array} \left[\begin{array}{c|c|c} \overbrace{\begin{array}{c} 1/\sqrt{p} \\ 1/\sqrt{p} \\ \vdots \\ 1/\sqrt{p} \end{array}}^{1 \text{ col}} & \overbrace{\begin{array}{c} U_2 = \\ \text{completion to} \\ \text{“between” clusters} \end{array}}^{k-1 \text{ columns}} & \overbrace{\begin{array}{c} U_3 = \\ \text{completion to} \\ \text{“within” clusters} \end{array}}^{p-k \text{ columns}} \end{array} \right].$$

The “between” and “within” terms are statistics jargon. Given a data vector, the first column extracts the normalized mean. The next block gives a basis for clustered vectors that are mean-free which by removing the fine details within cluster provides a way to compare between clusters. The last block provides the within cluster details. The number of columns is the dimension of the space, and in statistics jargon is known as the "degrees of freedom." (See [24, Chap. 10].)

10.2. One Way ANOVA made simple. A commonly used statistics test is to decide whether a proposed clustering of a vector v is justified. The test takes the average (meaning divide by $k-1$) square component in the U_2 direction and divides it by the average (meaning divide by $p-k$) square component in the U_3 direction. The following Julia code shows how compactly one can reproduce an example from Wikipedia where one can quickly obtain the number computed in Step 5 of https://en.wikipedia.org/wiki/One-way_analysis_of_variance#Example.

```
using LinearAlgebra
v = [6,8,4,5,3,4,8,12,9,11,6,8,13,9,11,8,7,12]    # data vector
A = cat(ones.([6,6,6])...,dims=1:2)               # Indicator(6,6,6)
B = [1 0 -1; 0 1 -1]                              # Constraint matrix
U,= svd(A,B)                                      # gsvd
(norm(U[:,2:3]'v)/norm(U[:,4:18]'v))^2 * 15/2     # The F value
```

```
9.264705882352956
```

While for this problem the classic approach is fine as an algorithm, for general tests for being in the column space of A but orthogonal to $\{Ax : Bx = 0\}$, the GSVD is worth considering algorithmically and how we are projecting into the non-horizontal directions is worth understanding geometrically.

10.3. See a slope? Generalize to a GSVD. The innocent looking

`norm(U[:,2:3]'v)/norm(U[:,4:18]'v)`

for an orthogonal matrix U carries a message of generalization if you know how to read it. It is a ratio of components in two orthogonal directions. You can call it a slope, or a cotangent, or a tangent. What we called horizontal and vertical multiaxes in Figure 2 may now be labeled in this coordinate system: the between and within axes, following the aforementioned statistics nomenclature.

The generalization of the vector $v \in \mathbb{R}^p$ example of Section 2 is a $p \times n$ matrix M of data, each data item being one row of length n. It is therefore natural geometrically to consider and interpret

$$\operatorname{gsvd}(U_2^T M, U_3^T M) = [U_b C; U_w S]H.$$

The result is n canonical directions for considering between vs within as naturally as comparing human vs yeast, or signal vs noise as we have seen in previous applications. The multislope, i.e. the generalized singular values (or perhaps we can call this the ANOVA structure) is 0 in all but at most $k-1$ directions, owing to the number of columns in U_2.

10.4. Discriminant Analysis Dimension Reduction. Continuing with the idea in Section 10.3. we obsever that it is natural to reduce out all but the $k-1$ nonzero ANOVA directions by multiplying M on the right by $G = H^\dagger I_{r,k-1}$ or (for that matter any matrix whose columns span the same subspace of $\mathbb{R}^n$.).

The reduction to $k-1$ columns

$$[U_2^T M; U_3^T M] \approx_{\text{reduction}} [U_2^T M; U_3^T M]G,$$

can be rotated back to the standard coordinate system without any change to the nonzero generalized singular values (the ANOVA structure) to yield

$$[U_2\ U_3][U_2^T M; U_3^T M]G = \left(U_2 U_2^T M + U_3 U_3^T M\right) G = (I - U_1 U_1^T) MG,$$

since $UU^T = I$. We can reduce the mean also by adding back $U_1 U_1^T G$ producing our final reduction, MG.

Our simple summary is that for a data matrix M ANOVA measures the nonzero generalized singular values in $[U_2^T; U_3^T]M$, a rotated multiaxis system which gives the ratios of the between to the within, and these are the same as for the reduced data matrix MG because we are suppressing the directions with 0 generalized singular values.

This is a geometrical derivation of an idea and algorithm presented by Park and others [20] with a minimization approach. In their algorithm G can be derived efficiently as the first $k-1$ columns of the Q from the GSVD, and the authors point out that the GSVD idea is robust even in the case of too little data.

11. The Jacobi Ensemble from Random Matrix Theory is a GSVD. Classical random matrix theory centers are Hermite, Laguerre, and Jacobi ensembles. Historically, they are presented in eigenvalue format, but we have argued that the eigenvalue, svd, gsvd formats, respectively, are mathematically more natural providing simpler derviations and clearer insights. Suppose we have two Gaussian random matrices A $(m_1 \times n)$ and B $(m_2 \times n)$ with $m_1 \geqslant n$ and $m_2 \geqslant n$. For example, `A=randn(m1,n)` and `B=randn(m2,n)` using common technical computing notation. The so-called *MANOVA matrix* (Multivariate Analysis of Variance) is defined to be

$$(A'A + B'B)^{-1} A'A \tag{11.1}$$

or in the symmetric form $(A'A + B'B)^{-1/2} A'A (A'A + B'B)^{-1/2}$. The eigenvalues are the squares of the cosines (c_i^2) and are jointly distributed as [24]

$$c \cdot \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^{n} \lambda_i^{a_1 - p} (1 - \lambda_i)^{a_2 - p}, \tag{11.2}$$

where $a_1 = \frac{\beta}{2} m_1, a_2 = \frac{\beta}{2} m_2$ and $p = 1 + \frac{\beta}{2}(n-1)$,

$$c = \prod_{j=1}^{n} \frac{\Gamma(1 + \frac{\beta}{2})\Gamma(a_1 + a_2 - \frac{\beta}{2}(n-j))}{\Gamma(1 + \frac{\beta}{2} j)\Gamma(a_1 - \frac{\beta}{2}(n-j))\Gamma(a_2 - \frac{\beta}{2}(n-j))},$$

where $\beta = 1$ for real matrices, $\beta = 2$ for complex matrices, $\beta = 4$ for quaternion matrices, and general β is worth considering, as in [14] . The eigenvalue distribution is known as the *Jacobi ensemble*, which was first referred by name in [23]. We refer interested readers to [16], where the geometrical picture motivates a direct derivation of the joint density of the Jacobi ensemble.

12. Mathematical Software. Suppose one looks up the GSVD in the help pages of your favorite technical computing language, shown in Table 2 One gets lost in a sea of matrices whose meaning is very hard to fully appreciate. Surprisingly, we find no standard function for the GSVD in Python (NumPy and SciPy) though there is some discussion on stackoverflow [2] and Github Numpy issue #3475[3] and scipy issue

[2] https://stackoverflow.com/questions/37814024/gsvd-for-python-generalized-singular-value-decomposition
[3] https://github.com/numpy/numpy/issues/3475

Fig. 9. *The SVD was once an obscure theoretical tool, and now it is everywhere, in part due to the work of the late Gene Golub at Stanford University (Gene Golub's famous vanity license plate illustrated, photographed by Professor P. M. Kroonenberg of Leiden University.). It is time for the GSVD to undergo the same transformation.*

#743[4] and #1491[5].

13. Acknowledgments. We thank Orly Alter, Zhaojun Bai, Michael Kirby, Andreas Noack, Chris Paige, Haesun Park, Sri Priya Ponnapalli, Charlie van Loan, Sabine van Huffel, and Joos Vandewalle for interesting conversations about the GSVD theory, software, and feedback from lectures at the 2017 Householder Symposium, the 2018 SIAM Applied Linear Algebra meeting.

We also wish to acknowledge and remember the late Gene Golub, over a decade since his passing, who so effectively promoted the singular value decomposition. We remember a time, not so long ago, when the SVD was unheard of outside of numerical linear algebra circles, and eigenvalues were all that were known. Then like dominos falling, one field after another, biology, economics, fields of engineering, statistics, computer science, and yes pure mathematics learned about the value of the SVD as a tool, as an algorithm, and even as vocabulary for effective communication. Gene with his $\boxed{\text{PROF SVD}}$ and $\boxed{\text{DR SVD}}$ California vanity license plates seemed always nearby when a field was starting to catch on. Today the GSVD is as obscure as the SVD was in the early days. We feel that the GSVD's time has come. We would be very pleased if one by one other fields would catch on.

REFERENCES

[1] K. A. Aiello, S. P. Ponnapalli, and O. Alter, *Mathematically universal and biologically consistent astrocytoma genotype encodes for transformation and predicts survival phenotype*, APL bioengineering, 2 (2018), p. 031909.
[2] O. Alter, P. O. Brown, and D. Botstein, *Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms*, Proceedings of the National Academy of Sciences, 100 (2003), pp. 3351–3356.
[3] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' guide*, vol. 9, Siam, 1999, http://www.netlib.org/lapack/lug/node36.html.
[4] Z. Bai, *The csd, gsvd, their applications and computations*, (1992).
[5] D. Callaerts, *Signal separation methods based on singular value decomposition and their application to the real-time extraction of the fetal electrocardiogram from cutaneous recordings*, (1989).
[6] D. Callaerts, B. De Moor, J. Vandewalle, W. Sansen, G. Vantrappen, and J. Janssens, *Comparison of svd methods to extract the foetal electrocardiogram from cutaneous electrode signals*, Medical and Biological Engineering and Computing, 28 (1990), p. 217.
[7] D. Chu, L. De Lathauwer, and B. De Moor, *A qr-type reduction for computing the svd of a general matrix product/quotient*, Numerische Mathematik, 95 (2003), pp. 101–121.
[8] B. De Moor, *Mathematical concepts for modeling of static and dynamic systems*, (1988).

[4] https://github.com/scipy/scipy/issues/743
[5] https://github.com/scipy/scipy/issues/1491

[9] B. De Moor, J. Staar, and J. Vandewalle, *Oriented energy and oriented signal-to-signal ratio concepts in the analysis of vector sequences and time series*, SVD and signal processing, (1988), pp. 209–232.

[10] J. W. Demmel and A. Edelman, *The dimension of matrices (matrix pencils) with given jordan (kronecker) canonical forms*, Linear algebra and its applications, 230 (1995), pp. 61–87.

[11] B. Dyas, *Generalized svd geometry?*, (2000), http://www.mathworks.com/matlabcentral/newsreader/view_thread/15120.

[12] L. Dykes and L. Reichel, *Simplified gsvd computations for the solution of linear discrete ill-posed problems*, Journal of Computational and Applied Mathematics, 255 (2014), pp. 15–27.

[13] A. Edelman, *Applied geometrical matrix computations*, in Proceedings of the Householder Symposium, http://www-math.mit.edu/ edelman/homepage/talks/hh2002.ppt, 2002.

[14] A. Edelman, *The random matrix technique of ghosts and shadows*, Markov Processes and Related Fields, 16 (2010), pp. 783–790.

[15] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM journal on Matrix Analysis and Applications, 20 (1998), pp. 303–353.

[16] A. Edelman and Y. Wang, *Random hyperplanes, generalized singular values & whats my β?*, in 2018 IEEE Statistical Signal Processing Workshop (SSP), IEEE, 2018, pp. 458–462.

[17] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3, JHU Press, 2012.

[18] P. C. Hansen, *Regularization, gsvd and truncatedgsvd*, BIT numerical mathematics, 29 (1989), pp. 491–504.

[19] S. Helgason, *Differential geometry and symmetric spaces*, vol. 341, American Mathematical Soc., 2001.

[20] P. Howland, M. Jeon, and H. Park, *Structure preserving dimension reduction for clustered text data based on the generalized singular value decomposition*, SIAM Journal on Matrix Analysis and Applications, 25 (2003), pp. 165–179.

[21] D. Hundley, M. Kirby, and M. Anderle, *A solution procedure for blind signal separation using the maximum noise fraction approach: algorithms and examples*, in Proceedings of the Conference on Independent Component Analysis, 2001, pp. 337–342.

[22] D. R. Hundley, M. J. Kirby, and M. Anderle, *Blind source separation using the maximum signal fraction approach*, Signal processing, 82 (2002), pp. 1505–1508.

[23] H. S. Leff, *Class of ensembles in the statistical theory of energy-level spectra*, Journal of Mathematical Physics, 5 (1964), pp. 763–768.

[24] R. J. Muirhead, *Aspects of multivariate statistical theory*, vol. 197, Wiley-Interscience, 2005.

[25] C. C. Paige and M. A. Saunders, *Towards a generalized singular value decomposition*, SIAM Journal on Numerical Analysis, 18 (1981), pp. 398–405.

[26] C. H. Park and H. Park, *Nonlinear discriminant analysis using kernel functions and the generalized singular value decomposition*, SIAM journal on matrix analysis and applications, 27 (2005), pp. 87–102.

[27] S. P. Ponnapalli, G. H. Golub, and O. Alter, *A novel higher-order generalized singular value decomposition for comparative analysis of multiple genome-scale datasets*, Workshop on Algorithms for Modern Massive Datasets (MMDS), (2006), https://www.alterlab.org/publications/Ponnapalli_et_al_MMDS_2006_Abstract.pdf.

[28] S. P. Ponnapalli, M. A. Saunders, C. F. Van Loan, and O. Alter, *A higher-order generalized singular value decomposition for comparison of global mrna expression from multiple organisms*, PloS one, 6 (2011), p. e28072.

[29] Quora, *What is the geometric interpretation and significance of the generalized singular value decomposition?*, (2014), https://goo.gl/sLpVha.

[30] P. Sankaranarayanan, T. E. Schomay, K. A. Aiello, and O. Alter, *Tensor gsvd of patient- and platform-matched tumor and normal dna copy-number profiles uncovers chromosome arm-wide patterns of tumor-exclusive platform-consistent alterations encoding for cell transformation and predicting ovarian cancer survival*, PloS one, 10 (2015), p. e0121396.

[31] R. R. Tucci, *An introduction to cartan's kak decomposition for qc programmers*, arXiv preprint quant-ph/0507171, (2005).

[32] C. F. Van Loan, *Generalizing the singular value decomposition*, SIAM Journal on Numerical Analysis, 13 (1976), pp. 76–83.

[33] J. Vandewalle, L. De Lathauwer, and P. Comon, *The generalized higher order singular value decomposition and the oriented signal-to-signal ratios of pairs of signal tensors and their use in signal processing*, in Proc ECCTD, Citeseer, 2003.

LANGUAGE	GSVD DOCUMENTATION IN CORRESPONDING LANGUAGE

MATLAB (R2018b)	https://www.mathworks.com/help/matlab/ref/gsvd.html
	`[U,V,X,C,S] = gsvd(A,B)` returns unitary matrices U and V, a (usually) square matrix X, and nonnegative diagonal matrices C and S so that
	`A = U*C*X'`
	`B = V*S*X'`
	`C'*C + S'*S = I`

MATHEMATICA (11.3.0)	https://reference.wolfram.com/language/ref/SingularValueDecomposition.html `>Details and Options`. `SingularValueDecomposition[m,a]` gives a list of matrices `{{u,ua},{w,wa},v}` such that `m` can be written as `u.w.Conjugate[Transpose[v]]` and `a` can be written as `ua.wa.Conjugate[Transpose[v]]`.
R (geigen v2.2)	https://www.rdocumentation.org/packages/geigen/versions/2.2/topics/gsvd The matrix A is a m-by-n matrix and the matrix B is a p-by-n matrix. This function decomposes both matrices; if either one is complex than the other matrix is coerced to be complex. The Generalized Singular Value Decomposition of numeric matrices A and B is given as $A = UD_1[0\ R]Q^T, \quad \text{and} \quad B = VD_2[0\ R]Q^T,$ where U an $m \times m$ orthogonal matrix V an $p \times p$ orthogonal matrix Q an $n \times n$ orthogonal matrix R an r-by-r upper triangular non singular matrix and the matrix $[0\ R]$ is an r-by-n matrix. D_1, D_2 are quasi diagonal matrices and nonnegative and satisfy $D_1^T D_1 + D_2^T D_2 = I$. D_1 is an m-by-r matrix and D_2 is a p-by-r matrix. For details on this decomposition and the structure of the matrices D_1 and D_2. see http://www.netlib.org/lapack/lug/node36.html.

TABLE 2

The GSVD as portrayed in the documentation of most technical computing languages seems unlikely to inspire the user unfamiliar with the GSVD.

LANGUAGE	GSVD DOCUMENTATION IN CORRESPONDING LANGUAGE
Julia 1.x	SVD(A, B) -> GENERALIZEDSVD The generalized SVD is used in applications such as when one wants to compare how much belongs to A vs. how much belongs to B, as in human vs yeast genome, or signal vs noise, or between clusters vs within clusters. (See Edelman and Wang for discussion: https://arxiv.org/pdf/1901.00485.pdf.) It decomposes [A;B] into [UC;VS]H, where [UC;VS] is a natural orthogonal basis for the column space of [A;B], and H=RQ' is a natural non-orthogonal basis for the rowspace of [A;B], where the top rows are most closely attributed to the A matrix, and the bottom to the B matrix. The multi-cosine/sine matrices C and S provide a multi-measure of how much A vs how much B, and U and V provide directions in which these are measured. svd(A,B) computes the generalized SVD of ‘A‘ and ‘B‘, returning a ‘GeneralizedSVD‘ factorization object ‘F‘ such that ‘[A;B] = [F.U * F.D1; F.V * F.D2] * F.R0 * F.Q'‘

TABLE 3

Proposed Documentation in Julia 1.0 at time of writing in Julia pull request https://github.com/JuliaLang/julia/pull/30239.