Jacobi Ensemble (MANOVA matrix, Multivariate Beta)

$A = randn(n_1, p)$
$B = randn(n_2, p)$ — See Bottom of pg

$A^TA) / (A^TA + B^TB) \sim M = (Y_1^TY_1)$ (all $p \times p$)

Joint Density of M is

$$\frac{\Gamma_p\left(\frac{1}{2}(n_1+n_2)\right)}{\Gamma_p\left(\frac{1}{2}n_1\right)\Gamma_p\left(\frac{1}{2}n_2\right)} |M|^{(n_1-p-1)/2} |I-M|^{(n_2-p-1)/2}$$

on $0 \leq M \leq I_p$ — M is pos def & $I-M$ is pos def or eigs in $[0,1]$

QR factor (MATLAB $[Y_1, \sim] = qr([A;B], 0)$)

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} R \qquad Y_1^TY_1 + Y_2^TY_2 = I_p \leftarrow \text{orthogonality}$$

$A^TA = R^T(Y_1^TY_1)R$
$A^TA + B^TB = R^T(Y_1^TY_1 + Y_2^TY_2)R = R^TR$

$(A^TA) / (A^TA + B^TB) = R^T(Y_1^TY_1)R^{-T} \sim Y_1^TY_1$

Recall

$$\Gamma_p(x) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left(x - \tfrac{1}{2}(i-1)\right)$$

Angles Between Subspaces

Scalar

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} r$$

$$\frac{a^2}{a^2+b^2} = \cos^2\theta \quad (y_1^2)$$

= cos squared of angle with x-axis

Analog: Take a p-dimensional subspace of $\mathbb{R}^n$ & ask for angles with standard n_1 dimensional subspace.

eig $Y_1^T Y_1$ --- $\cos^2$ angles
Svd Y_1 --- cos angles

Geometrically: What are angles between subspaces?

$\theta_i \in [0, \frac{\pi}{2}]$
Peel off maximum angle, then only consider vectors orthogonal & recur.

Sometimes: only largest angle is needed

Example: $\mathbb{R}^3$ is really small

line + plane: Angle between the + its projection

plane + plane: Angle between vectors $\perp$ line of intersection

$\theta = 0$ usually considered "extra" or accidental alignment

In general, if
W_1 & W_2 are orthogonal bases
$n \times n_1$ $n \times n_2$ for two spaces,

$\text{svd}(W_1^T W_2)$ gives the cosines.

The GSVD and the CS Decomposition

The Generalized Singular Value Decomposition

$[U, V, X, C, S] = \text{gsvd}(A, B)$

(same # of columns)

$$\begin{matrix} m \\ n \end{matrix} \begin{bmatrix} A \\ B \end{bmatrix}_{p} = \begin{bmatrix} U & \\ & V \end{bmatrix} \begin{bmatrix} C \\ S \end{bmatrix} X^T$$

$$C = \begin{bmatrix} \cos\theta_1 & & \\ & \ddots & \\ & & \end{bmatrix}$$

$$S = \begin{bmatrix} \sin\theta_1 & & \\ & \ddots & \\ & & \end{bmatrix}$$

$\text{gsvd}(A, B)$ its gives the "cotans" $= \frac{\cos\theta}{\sin\theta}$

square case: $gsvd(A,B) = svd(A/B)$
in analogy of $eig(A,B) = eig(A/B)$
known as the generalized
eigenvalue problem $0 = det(B - \lambda A)$

The CS Decomposition

Essentially the same idea of the GSVD is in the CS decomposition. It says that if you split up an orthogonal matrix the svd's are in "cosine-sine" pairs.

$$\left[\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array}\right]$$

Try $svd(Q_{11})$
$svd(Q_{22})$
$svd(Q_{12})$
$svd(Q_{21})$

There are always "Filler" 0's + 1's

$$Q = \begin{matrix} n_1 \\ n_2 \end{matrix}\left[\begin{array}{c|c} & \\ & \end{array}\right] \quad \begin{matrix} p & n-p \end{matrix}$$

$$C = \begin{bmatrix} \cos\theta_1 & & \\ & \ddots & \\ & & \cos\theta_p \end{bmatrix} \quad S = \begin{bmatrix} \sin\theta_1 & & \\ & \ddots & \\ & & \sin\theta_p \end{bmatrix}$$

$$= \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix} \left[\begin{array}{cc|ccc} C & & S & 0 & 0 \\ & & 0 & 0 & I_{n-p} \\ \hline -S & & C & 0 & 0 \\ 0 & & 0 & I_{n_2-p} & \end{array}\right] \begin{bmatrix} V_1 & \\ & V_2 \end{bmatrix}^T$$

Application:

Random p-dim subspace of $\mathbb{R}^n$ $\longleftrightarrow$ span(randn(n, p))

Fancy Math Term $\equiv$ Random point in Grassmann Manifold $G(n,p)$

What is the distribution of the largest principal angle between two randomly chosen p-dim subspaces of $\mathbb{R}^n$?

Answer: Smallest $\cos^2$ from Jacobi density with $n_1 = p$

$$\text{Density} = \underbrace{\int \cdots \int}_{\substack{p-1 \text{ angles} \\ \text{that are} \\ \text{smaller}}}$$

$$= \frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{n-p+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \beta(n-p)(\sin\theta)^{p(n-p-1)} \, {}_2F_1\left(\frac{n-p-1}{2}, \frac{1}{2}; \frac{n+1}{2}; \sin^2\theta I_{p-1}\right)$$

What is ${}_2F_1$?

$$ {}_2F_1(a,b;c;X) = \frac{\Gamma_m(c)}{\Gamma_n(a)\Gamma_n(b)} \times $$

$$ \int_{0<Y<I_n} |I-XY|^{-b} |Y|^{a-\frac{n+1}{2}} |I-Y|^{c-a-\frac{(n+1)}{2}} (dY) $$

(dY) ↑ Sym

$$ {}_2F_1(a,b;c;X) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{K+k} \frac{(a)_{1a}(b)_{1a}}{(c)_{1a}} C_K(X) $$

$$ (a)_{1a} = \prod_{i=1}^{1} \left(a - \tfrac{1}{2}(i-1)\right)_{K_i} $$

$$ (a)_k = a(a+1)\cdots(a+k-1) $$

= rising factorial

or Pochammer Symbol

(7)

Joint Density of Jacobi in
Muirhead p109 Theorem 3.3.1

Might be easier to just
get GSVD and "back out"
the Jacobian

From Dumitriu's Thesis should
be possible to use

$$C \prod |\sin(\theta_i - \theta_j)| \; |\sin(\theta_i + \theta_j)| \; \left| \prod \sin\theta_i \, (V_1^* dV_1) \right|^{\wedge}$$

$$= \prod \left(\sin^2\theta_i \cos^2\theta_j - \cos^2\theta_i \sin^2\theta_j \right)$$

details need to be derived
Help Welcome ☺

McKay
Steven Miller

gsvd

Generalized singular value decomposition

Syntax

```
[U,V,X,C,S] = gsvd(A,B)
sigma = gsvd(A,B)
```

Description

`[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
```

`A` and `B` must have the same number of columns, but may have different numbers of rows. If `A` is `m`-by-`p` and `B` is `n`-by-`p`, then `U` is `m`-by-`m`, `V` is `n`-by-`n` and `X` is `p`-by-`q` where `q = min(m+n,p)`.

`sigma = gsvd(A,B)` returns the vector of generalized singular values, `sqrt(diag(C'*C)./diag(S'*S))`.

The nonzero elements of `S` are always on its main diagonal. If `m >= p` the nonzero elements of `C` are also on its main diagonal. But if `m < p`, the nonzero diagonal of `C` is `diag(C,p-m)`. This allows the diagonal elements to be ordered so that the generalized singular values are nondecreasing.

`gsvd(A,B,0)`, with three input arguments and either `m` or `n >= p`, produces the "economy-sized"decomposition where the resulting `U` and `V` have at most `p` columns, and `C` and `S` have at most `p` rows. The generalized singular values are `diag(C)./diag(S)`.

When `B` is square and nonsingular, the generalized singular values, `gsvd(A,B)`, are equal to the ordinary singular values, `svd(A/B)`, but they are sorted in the opposite order. Their reciprocals are `gsvd(B,A)`.

In this formulation of the `gsvd`, no assumptions are made about the individual ranks of `A` or `B`. The matrix `X` has full rank if and only if the matrix `[A;B]` has full rank. In fact, `svd(X)` and `cond(X)` are equal to `svd([A;B])` and `cond([A;B])`. Other formulations, eg. G. Golub and C. Van Loan [1], require that `null(A)` and `null(B)` do not overlap and replace `X` by `inv(X)` or `inv(X')`.

Note, however, that when `null(A)` and `null(B)` do overlap, the nonzero elements of `C` and `S` are not uniquely determined.

Examples

Example 1

The matrices have at least as many rows as columns.

```
A = reshape(1:15,5,3)
B = magic(3)
A =
     1     6    11
     2     7    12
     3     8    13
     4     9    14
     5    10    15
B =
     8     1     6
     3     5     7
     4     9     2
```

The statement

```
[U,V,X,C,S] = gsvd(A,B)
```

produces a 5-by-5 orthogonal `U`, a 3-by-3 orthogonal `V`, a 3-by-3 nonsingular `X`,

```
X =
    2.8284   -9.3761   -6.9346
   -5.6569   -8.3071  -18.3301
    2.8284   -7.2381  -29.7256
```

and

```
C =
    0.0000         0         0
         0    0.3155         0
         0         0    0.9807
         0         0         0
         0         0         0
S =
    1.0000         0         0
         0    0.9489         0
         0         0    0.1957
```

Since `A` is rank deficient, the first diagonal element of `C` is zero.

The economy sized decomposition,

```
[U,V,X,C,S] = gsvd(A,B,0)
```

produces a 5-by-3 matrix `U` and a 3-by-3 matrix `C`.

```
U =
    0.5700   -0.6457   -0.4279
   -0.7455   -0.3296   -0.4375
   -0.1702   -0.0135   -0.4470
    0.2966    0.3026   -0.4566
    0.0490    0.6187   -0.4661

C =
    0.0000         0         0
         0    0.3155         0
         0         0    0.9807
```

The other three matrices, `V`, `X`, and `S` are the same as those obtained with the full decomposition.

The generalized singular values are the ratios of the diagonal elements of `C` and `S`.

```
sigma = gsvd(A,B)
sigma =
    0.0000
    0.3325
    5.0123
```

These values are a reordering of the ordinary singular values

```
svd(A/B)
ans =
    5.0123
    0.3325
    0.0000
```

Example 2

The matrices have at least as many columns as rows.

```
A = reshape(1:15,3,5)
B = magic(5)
A =
     1     4     7    10    13
     2     5     8    11    14
     3     6     9    12    15
B =
    17    24     1     8    15
    23     5     7    14    16
     4     6    13    20    22
    10    12    19    21     3
    11    18    25     2     9
```

The statement

```
[U,V,X,C,S] = gsvd(A,B)
```

produces a 3-by-3 orthogonal `U`, a 5-by-5 orthogonal `V`, a 5-by-5 nonsingular `X` and

```
C =
         0         0    0.0000         0         0
         0         0         0    0.0439         0
         0         0         0         0    0.7432
S =
    1.0000         0         0         0         0
         0    1.0000         0         0         0
         0         0    1.0000         0         0
         0         0         0    0.9990         0
         0         0         0         0    0.6690
```

In this situation, the nonzero diagonal of `C` is `diag(C,2)`. The generalized singular values include three zeros.

```
sigma = gsvd(A,B)
sigma =
         0
         0
    0.0000
    0.0439
    1.1109
```

Reversing the roles of `A` and `B` reciprocates these values, producing two infinities.

```
gsvd(B,A)
ans =
   1.0e+016 *

    0.0000
    0.0000
```

```
    4.4126
       Inf
       Inf
```

Algorithms

The generalized singular value decomposition uses the C-S decomposition described in [1], as well as the built-in `svd` and `qr` functions. The C-S decomposition is implemented in a subfunction in the `gsvd` program file.

Diagnostics

The only warning or error message produced by `gsvd` itself occurs when the two input arguments do not have the same number of columns.

References

[1] Golub, Gene H. and Charles Van Loan, *Matrix Computations*, Third Edition, Johns Hopkins University Press, Baltimore, 1996

See Also

`qr` | `svd`

On the largest principal angle between random subspaces

P.-A. Absil[†‡] A. Edelman[*] P. Koev[*]

Manuscript re-accepted by Dr H. Schneider on 22 Oct 2005

LAA 8707
LAA # 0409-473S, RFP # 6646/R
(This is an update of the version accepted on 5 Oct 2005.)

Abstract

Formulas are derived for the probability density function and the probability distribution function of the largest canonical angle between two p-dimensional subspaces of $\mathbb{R}^n$ chosen from the uniform distribution on the Grassmann manifold (which is the unique distribution invariant by orthogonal transformations of $\mathbb{R}^n$). The formulas involve the gamma function and the hypergeometric function of a matrix argument.

AMS subject classifications. 15A51 Stochastic matrices; 15A52 Random matrices; 33C05 Classical hypergeometric functions, ${}_2F_1$; 33C45 Orthogonal polynomials and functions of hypergeometric type (Jacobi, Laguerre, Hermite, Askey scheme, etc.); 62H10 Distribution of statistics

Key words. Largest principal angle, largest canonical angle, projection 2-norm, Grassmann manifold, random matrices, gamma function, hypergeometric function of matrix argument.

1 Introduction

Several numerical algorithms on Grassmann manifolds (i.e., sets of fixed-dimensional subspaces of a Euclidean space) display a convergence property of the following type: if the distance $\mathrm{dist}(\mathcal{Y}, \mathcal{S})$ between the initial point (i.e., subspace) $\mathcal{Y}$ and the solution point $\mathcal{S}$ is smaller than some given number δ, then the sequence of iterates generated by the algorithm is guaranteed to converge to the solution $\mathcal{S}$. An example is Newton's method on Riemannian manifolds (Grassmann manifolds are particular cases of Riemannian manifolds) for which γ and α theorems [DPM03] provide values for δ.

Then the question naturally arises to determine the probability that a randomly chosen initial subspace $\mathcal{Y}$ satisfies the distance condition $\mathrm{dist}(\mathcal{Y}, \mathcal{S}) < \delta$. There are several definitions for

[†]Département d'ingénierie mathématique, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium, and Peterhouse, University of Cambridge, Cambridge CB2 1RD, UK. URL: http://www.inma.ucl.ac.be/~absil/

[*]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

[‡]The work of this author was done in part while the author was a Postdoctoral Associate with the School of Computational Science of Florida State University.

Now remember that $x = \cos^2\theta_p$. The (one-dimensional) Jacobian of this change of variables is $2\sin\theta_p\cos\theta_p$ and we obtain

$$\operatorname{dens}(\theta_p) = 2\frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})}\cdot\frac{\Gamma_p(\frac{n}{2})}{\Gamma_p(\frac{p}{2})\Gamma_p(\frac{n-p}{2})}\cdot\frac{\Gamma_{p-1}(\frac{p-1}{2})}{\pi^{\frac{(p-1)^2}{2}}}\cdot\frac{\Gamma_{p-1}(1+\frac{p}{2})\Gamma_{p-1}(\frac{n-p-1}{2})}{\Gamma_{p-1}(\frac{n+1}{2})}$$
$$\times(\sin\theta_p)^{p(n-p)-1}\,{}_2F_1(\tfrac{n-p-1}{2},\tfrac{1}{2};\tfrac{n+1}{2};\sin^2\theta_p I_{p-1}) \quad (10)$$

for $\theta_p \in [0, \frac{\pi}{2})$ and $p < \frac{n+1}{2}$. Formula (10) can be simplified using the formula for the multivariate Γ function [Mui82, p. 62]

$$\Gamma_m(y) = \pi^{m(m-1)/4}\prod_{i=1}^{m}\Gamma\left(y-\tfrac{i-1}{2}\right) = \pi^{m(m-1)/4}\prod_{i=1}^{m}\Gamma\left(y-\tfrac{m}{2}+\tfrac{i}{2}\right), \quad \operatorname{Re}(y) > \frac{1}{2}(m-1).$$

along with the identities $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(m+1) = m\Gamma(m)$. This yields the following result.

Theorem 1 *The probability density function of the largest canonical angle θ_p between two subspaces chosen from the uniform distribution on the Grassmann manifold of p-planes in $\mathbb{R}^n$ ($p < \frac{n+1}{2}$) endowed with its canonical metric, is given for $\theta_p \in [0, \frac{\pi}{2})$ by*

$$\operatorname{dens}(\theta_p) = p(n-p)\frac{\Gamma(\frac{p+1}{2})\Gamma(\frac{n-p+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n+1}{2})}(\sin\theta_p)^{p(n-p)-1}\,{}_2F_1(\tfrac{n-p-1}{2},\tfrac{1}{2};\tfrac{n+1}{2};\sin^2\theta_p I_{p-1}), \quad (11)$$

where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt$ is the classical gamma function, related to the factorial by $\Gamma(n) = (n-1)!$, and ${}_2F_1$ is the Gaussian hypergeometric function of matrix argument (see [Mui82, Def. 7.3.1]). The corresponding probability distribution function is

$$\Pr(\theta_p < \hat{\theta}_p) = \frac{\Gamma(\frac{p+1}{2})\Gamma(\frac{n-p+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n+1}{2})}(\sin\hat{\theta}_p)^{p(n-p)}\,{}_2F_1(\tfrac{n-p}{2},\tfrac{1}{2};\tfrac{n+1}{2};\sin^2\hat{\theta}_p I_p). \quad (12)$$

The probability distribution function (12) is obtained from

$$\Pr(\lambda_p > x) = \int_{1\geq\lambda_1\geq\ldots\geq\lambda_p>x}\operatorname{dens}(\lambda_1,\ldots,\lambda_p)d\lambda_1\ldots d\lambda_p,$$

using the same techniques as above.

4 The case $p = 1$

The case $p = 1$ is much simpler as there is only one canonical angle. Starting from (1), one obtains

$$\operatorname{dens}(x) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})}x^{-1/2}(1-x)^{(n-3)/2}$$

and since $x = \cos^2\theta$

$$\operatorname{dens}(\theta) = 2\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})}(\sin\theta)^{n-2}.$$

For $n = 3$, using $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(1) = 1$, we obtain

$$\operatorname{dens}(\theta) = \sin\theta.$$

For $n = 2$, $\operatorname{dens}(\theta) = 2/\pi$.

Computing the Complete CS Decomposition

Brian D. Sutton*

May 19, 2008

Abstract

An algorithm is developed to compute the complete CS decomposition (CSD) of a partitioned unitary matrix. Although the existence of the CSD has been recognized since 1977, prior algorithms compute only a reduced version (the 2-by-1 CSD) that is equivalent to two simultaneous singular value decompositions. The algorithm presented here computes the complete 2-by-2 CSD, which requires the simultaneous diagonalization of all four blocks of a unitary matrix partitioned into a 2-by-2 block structure. The algorithm appears to be the only fully specified algorithm available. The computation occurs in two phases. In the first phase, the unitary matrix is reduced to bidiagonal block form, as described by Sutton and Edelman. In the second phase, the blocks are simultaneously diagonalized using techniques from bidiagonal SVD algorithms of Golub, Kahan, and Demmel. The algorithm has a number of desirable numerical features.

1 Introduction

The *complete CS decomposition* (CSD) applies to any m-by-m matrix X from the unitary group $U(m)$, viewed as a 2-by-2 block matrix,

$$X = \begin{array}{c} \\ p \\ m-p \end{array} \begin{array}{c} \begin{array}{cc} q & m-q \end{array} \\ \left[\begin{array}{c|c} X_{11} & X_{12} \\ \hline X_{21} & X_{22} \end{array}\right] \end{array}.$$

For convenience, we assume $q \leq p$ and $p + q \leq m$. A complete CS decomposition has the form

$$X = \left[\begin{array}{cc} U_1 & \\ & U_2 \end{array}\right] \left[\begin{array}{c|c|c|c} C & S & 0 & 0 \\ \hline 0 & 0 & I_{p-q} & 0 \\ \hline -S & C & 0 & 0 \\ \hline 0 & 0 & 0 & I_{m-p-q} \end{array}\right] \left[\begin{array}{cc} V_1 & \\ & V_2 \end{array}\right]^*, \qquad (1.1)$$

$$C = \operatorname{diag}(\cos(\theta_1), \ldots, \cos(\theta_q)), \quad S = \operatorname{diag}(\sin(\theta_1), \ldots, \sin(\theta_q)),$$

in which $\theta_1, \ldots, \theta_q \in [0, \frac{\pi}{2}]$, $U_1 \in U(p)$, $U_2 \in U(m-p)$, $V_1 \in U(q)$, and $V_2 \in U(m-q)$. The letters *CS* in the term *CS decomposition* come from *cosine-sine*.

The major contribution of this paper is an algorithm for computing (1.1). We believe this to be the only fully specified algorithm available for computing the complete CS decomposition. Earlier algorithms compute only a reduced form, the "2-by-1" CSD, which is defined in the next section. The algorithm developed in this article is based

*Department of Mathematics, Randolph-Macon College, P.O. Box 5005, Ashland, VA 23005 USA. email: bsutton@rmc.edu. Supplementary material is available at the author's home page.

on the SVD algorithm of Golub and Kahan and has a number of desirable numerical properties.

The algorithm proceeds in two phases.

1. Phase I: Bidiagonalization. In the special case $p = q = \frac{m}{2}$, the decomposition is

$$X = \begin{bmatrix} P_1 & \\ & P_2 \end{bmatrix} \begin{bmatrix} B_{11}^{(0)} & B_{12}^{(0)} \\ B_{21}^{(0)} & B_{22}^{(0)} \end{bmatrix} \begin{bmatrix} Q_1 & \\ & Q_2 \end{bmatrix}^*, \tag{1.2}$$

   in which $B_{11}^{(0)}$ and $B_{21}^{(0)}$ are upper bidiagonal, $B_{12}^{(0)}$ and $B_{22}^{(0)}$ are lower bidiagonal, and P_1, P_2, Q_1, and Q_2 are q-by-q unitary. We say that the middle factor is a real orthogonal matrix in *bidiagonal block form*. (See Definition 1.1.)

2. Phase II: Diagonalization. The CSD of $\begin{bmatrix} B_{11}^{(0)} & B_{12}^{(0)} \\ B_{21}^{(0)} & B_{22}^{(0)} \end{bmatrix}$ is computed,

$$\begin{bmatrix} B_{11}^{(0)} & B_{12}^{(0)} \\ B_{21}^{(0)} & B_{22}^{(0)} \end{bmatrix} = \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix} \begin{bmatrix} C & S \\ -S & C \end{bmatrix} \begin{bmatrix} V_1 & \\ & V_2 \end{bmatrix}^*.$$

Combining the factorizations gives the CSD of X,

$$X = \begin{bmatrix} P_1U_1 & \\ & P_2U_2 \end{bmatrix} \begin{bmatrix} C & S \\ -S & C \end{bmatrix} \begin{bmatrix} Q_1V_1 & \\ & Q_2V_2 \end{bmatrix}^*. \tag{1.3}$$

Phase I is a finite-time procedure first described in [18], and Phase II is an iterative procedure based on ideas from bidiagonal SVD algorithms [6, 8].

Some of the earliest work related to the CSD was completed by Jordan, Davis, and Kahan [4, 5, 12]. The CSD as we know it today and the term *CS decomposition* first appeared in a pair of articles by Stewart [16, 17]. Computational aspects of the 2-by-1 CSD are considered in [3, 13, 14, 17, 19] and later articles. A "sketch" of an algorithm for the complete CSD can be found in a paper by Hari [11], but few details are provided. For general information and more references, see [2, 10, 15].

1.1 Complete versus 2-by-1 CS decomposition

Most commonly available CSD algorithms compute what we call the *2-by-1 CS decomposition* of a matrix $\hat{X}$ with orthonormal columns partitioned into a 2-by-1 block structure. In the special case $p = q = \frac{m}{2}$, $\hat{X}$ has the form

$$\hat{X} = \begin{matrix} q \\ q \end{matrix} \begin{matrix} \overset{q}{\overset{\downarrow}{\phantom{\hat{X}_{11}}}} \\ \begin{bmatrix} \hat{X}_{11} \\ \hat{X}_{21} \end{bmatrix} \end{matrix},$$

and the CSD is

$$\hat{X} = \begin{bmatrix} U_1 & \\ & U_2 \end{bmatrix} \begin{bmatrix} C \\ -S \end{bmatrix} V_1^*.$$

A naive algorithm for computing the 2-by-1 CSD is to compute two SVD's,

$$\begin{cases} \hat{X}_{11} = U_1 C V_1^* \\ \hat{X}_{21} = (-U_2) S V_1^*, \end{cases}$$

reordering rows and columns and adjusting signs as necessary to make sure that the two occurrences of V_1^* are identical and that $C^2 + S^2 = I$. This works in theory if no two singular values of $\hat{X}_{11}$ are repeated, but in practice it works poorly when there are

5 Numerical experiments

Formulas in Theorem 1 require an algorithm that evaluates the hyperbolic function ${}_2F_1$ with matrix argument. Recently, new algorithms were proposed that efficiently approximate the hypergeometric function of matrix argument through its expansion as a series of Jack functions [KE05], and MATLAB [Mat92] implementations of these algorithms were made available. In numerical experiments, we used the MATLAB script `mghi`—dedicated to special case when the matrix argument is a multiple of the identity—to evaluate formulas (11) and (12).

For the purpose of checking the results given by the formulas evaluated with the MATLAB script, comparisons were made with the probability functions estimated from samples of p-dimensional subspaces of $\mathbb{R}^n$. The subspaces were selected as span(X), with $n \times p$ matrix X chosen from a Gaussian distribution. The principal angles between span(X) and the fixed subspace span($I_{n\times p}$) were computed using the MATLAB expression `asin(svd(X(p+1:n,:)))`. As an illustration, the case $n = 7$, $p = 3$ is shown on Figure 1. An excellent agreement is observed between the computed probability functions and the ones estimated from the sample.

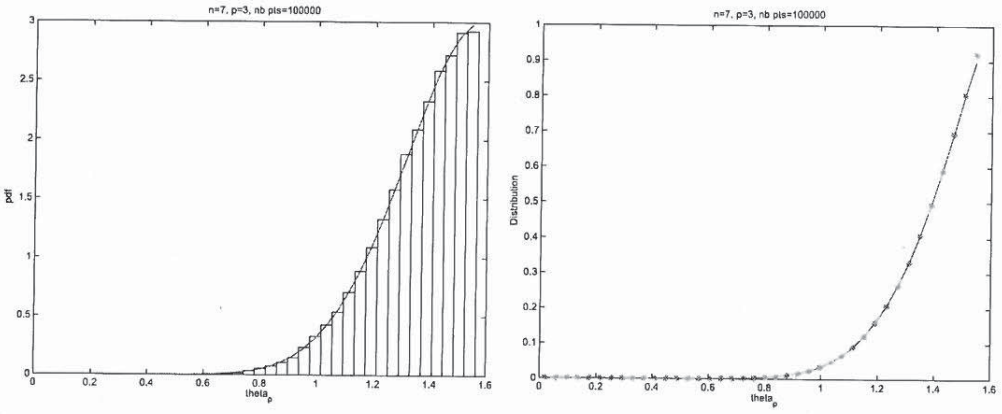

Figure 1: The solid curves correspond to the probability density function (11)—left-hand plot—and the probability distribution function (12)—right-hand plot—evaluated using `mhgi` for the case $n = 7$, $p = 3$. The histogram and the stars show approximations evaluated from a sample of 10^5 three-dimensional subspaces in $\mathbb{R}^7$ selected from the uniform distribution on the Grassmannian.

References

[AMS04] P.-A. Absil, R. Mahony, and R. Sepulchre, *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*, Acta Appl. Math. **80** (2004), no. 2, 199–220.

[AMSV02] P.-A. Absil, R. Mahony, R. Sepulchre, and P. Van Dooren, *A Grassmann-Rayleigh quotient iteration for computing invariant subspaces*, SIAM Rev. **44** (2002), no. 1, 57–73 (electronic).

[DPM03] Jean-Pierre Dedieu, Pierre Priouret, and Gregorio Malajovich, *Newton's method on Riemannian manifolds: covariant alpha theory*, IMA J. Numer. Anal. **23** (2003), no. 3, 395–419.

[Drm00] Zlatko Drmač, *On principal angles between subspaces of Euclidean space*, SIAM J. Matrix Anal. Appl. **22** (2000), no. 1, 173–194 (electronic).

where σ_i are the singular values of A and the eigenvalues of S.

The proof follows from the fact that, if $A = U\Sigma V'$ is the singular value decomposition for A, then $Q = UV'$ and $S = V\Sigma V'$, which also explains why the Jacobian is written in terms of the singular values σ_i of A.

- **CS decomposition** (see [35]).

Valid for all three cases ($\beta = 1, 2, 4$). Q is $n \times n$ orthogonal/unitary/symplectic. Then for any $k + j = n$, $p = k - j \geq 0$, the decomposition is

$$Q = \begin{pmatrix} U_{11} & U_{12} & 0 \\ U_{21} & U_{22} & 0 \\ 0 & 0 & U_2 \end{pmatrix} \begin{pmatrix} I_p & 0 & 0 \\ 0 & C & S \\ 0 & S & -C \end{pmatrix} \begin{pmatrix} V_{11}' & V_{12}' & 0 \\ V_{21}' & V_{22}' & 0 \\ 0 & 0 & V_2' \end{pmatrix},$$

such that U_2, V_2 are $j \times j$ orthogonal/unitary/symplectic, $\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$ and $\begin{pmatrix} V_{11}' & V_{12}' \\ V_{21}' & V_{22}' \end{pmatrix}$ are $k \times k$ orthogonal/unitary/symplectic, with U_{11} and V_{11} being $p \times p$, and C and S are $j \times j$ real, positive, and diagonal, and $C^2 + S^2 = I_j$.

Parameter count:

$$\begin{aligned} \beta\frac{n(n+1)}{2} - n &= \Big(\beta j(j+1) - (\beta-1)j\Big) + j \; + \\ &\quad + \Big(\beta k(k+1) - k - \beta\frac{p(p+1)}{2} + p\Big) \end{aligned}$$

This count is a little special, and so we will give it in detail. The first part in the right hand side counts the number of parameters in U_2 and V_2 (which are independent), and accounts for the choice of phases, since $U_2 C V_2'$ is the SVD of the bottom $j \times j$ part of Q. The second term accounts for the j parameters of C (which are the same in S).

Finally, the last parenthesis counts the number of parameters in $\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$ and $\begin{pmatrix} V_{11}' & V_{12}' \\ V_{21}' & V_{22}' \end{pmatrix}$, and accounts for the fact that $\begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} \begin{pmatrix} V_{11}' & V_{12}' \end{pmatrix}$ is

determined by the other matrices in the decomposition and thus, one has to subtract the missing number of parameters. Since the number of parameters in either $\begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix}$ or $\begin{pmatrix} V_{11}' & V_{12}' \end{pmatrix}$, given the rest of the orthogonal matrices, is $\beta p(p+1)/2 - p$, the count follows.

Note that the choice of phases in U_2, V_2 determines all phases in the decomposition.

Now let $\theta_i \in (0, \frac{\pi}{2})$. $q \leq i \leq j$ be the angles such that C =diag(cos(θ_1),...,cos(θ_j)), and S =diag(sin(θ_1),...,sin(θ_j)). To ensure uniqueness of the decomposition we order the angles, $\theta_i \geq \theta_j$, for all $i \leq j$.

For notational purposes, let $U_1 = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$, and $V_1' = \begin{pmatrix} V_{21}' & V_{22}' \end{pmatrix}$.

Jacobian:

$$(Q'dQ) = \prod_{i<j} \sin(\theta_i - \theta_j)^\beta \ \sin(\theta_i + \theta_j)^\beta \ \prod_{i=1}^{j} \cos(\theta_i)^{\beta-1} \ \sin(\theta_i) \ d\theta \ \times$$
$$\times \ (U_1'dU_1) \ (U_2'dU_2) \ (V_1'dV_1) \ (V_2'dV_2) \ .$$

The proof follows from the decomposition itself. Note that $V_1 \in V_{j,k}^{\beta}$.

- **Tridiagonal QΛQ′ (eigenvalue) decomposition** ($T = Q\Lambda Q'$).

  Valid for real matrices. T is an $n \times n$ tridiagonal symmetric matrix as in (3.3), Q is an orthogonal $n \times n$ matrix, and Λ is diagonal. To make the factorization unique, we impose the condition that the first row of Q is all positive. The number of independent parameters in Q is $n-1$ and the can be seen as being all in the first row q of Q. The rest of Q can be determined from the orthogonality constraints, the tridiagonal symmetric constraints on A, and from Λ. Parameter count:

$$2n - 1 = n - 1 + n \ .$$

  Jacobian:

$$dT = \frac{\prod_{i=1}^{n-1} b_i}{\prod_{i=1}^{n} q_i} \ (dq) \ d\Lambda.$$