# On the largest principal angle between random subspaces

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

Preprint
8 December 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 $\delta$, 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 $\gamma$ and $\alpha$ theorems [DPM03] provide values for $\delta$.

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 $\mathrm{dist}(\mathcal{Y}, \mathcal{S})$; see, e.g., [EAS98, p. 337], [QZL04]. The most important one in engineering applications is arguably the projection 2-norm

$$\mathrm{dist}_{p2}(\mathcal{Y}, \mathcal{S}) = \|P_{\mathcal{Y}} - P_{\mathcal{S}}\|_2,$$

[†]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 supported in part by the School of Computational Science of Florida State University through a postdoctoral fellowship, and by Microsoft Research through a Research Fellowship at Peterhouse, Cambridge.

where $P_{\mathcal{Y}}$ and $P_{\mathcal{S}}$ are the orthogonal projections onto $\mathcal{Y}$ and $\mathcal{S}$, respectively, and $\|\cdot\|_2$ denotes the matrix 2-norm. The projection 2-norm is related to the largest canonical angle. The *canonical angles* (or *principal angles*) $\theta_1, \ldots, \theta_p \in [0, \pi/2]$ between two $p$-dimensional subspaces $\mathcal{Y}$ and $\mathcal{S}$ are defined recursively by

$$\cos(\theta_k) = \max_{y \in \mathcal{Y}} \max_{s \in \mathcal{S}} y^T s = y_k^T s_k$$

subject to $\|y\| = \|s\| = 1$, $y^T y_i = 0$, $s^T s_i = 0$, $i = 1, \ldots, k-1$. Alternatively, if the columns of $Y$ and $S$ define orthonormal bases of $\mathcal{Y}$ and $\mathcal{S}$, respectively, then the cosines of the canonical angles are the singular values of $Y^T S$. The relation between the projection 2-norm and the canonical angles is that

$$\mathrm{dist}_{p2}(\mathcal{Y}, \mathcal{S}) = \sin(\theta_p(\mathcal{Y}, \mathcal{S})),$$

where $\theta_p(\mathcal{Y}, \mathcal{S})$ is the largest canonical angle between $\mathcal{Y}$ and $\mathcal{S}$; see, e.g., [GV96, Drm00] for details. Geometrically, the largest canonical angle is the distance induced by the Finsler metric defined by the operator 2-norm [Wei00].

In this paper, we give a formula for the probability distribution function $\Pr(\theta_p(\mathcal{S}, \mathcal{Y}) < \hat{\theta})$, $\hat{\theta} \in [0, \pi/2]$, where $\mathcal{S}$ and $\mathcal{Y}$ are chosen from the uniform distribution on the Grassmann manifold $\mathrm{Grass}(p, n)$ of $p$-planes in $\mathbb{R}^n$ (the *uniform distribution* on $\mathrm{Grass}(p, n)$ is the distribution with probability measure invariant under the transformations of $\mathrm{Grass}(p, n)$ induced by orthogonal transformations of $\mathbb{R}^n$; see [Jam54, §4.6]). We also derive a formula for the probability density function of $\theta_p(\mathcal{S}, \mathcal{Y})$. The formulas, given in Theorem 1, involve the gamma function (which extends the factorial over the reals) and the hypergeometric function of a matrix argument. Their derivation strongly relies on the material in Muirhead [Mui82].

It is direct, using the formulas of this paper, to obtain the probability distributions of distances based on the largest principal angle, like the projection 2-norm and the chordal 2-norm (see [EAS98]). An open problem is to obtain an expression for the probability distribution of the geodesic distance between two random subspaces, where the geodesic distance is given by the Euclidean norm of the vector of canonical angles, i.e., $d = \sqrt{\sum \theta_i^2}$ (see, e.g., [EAS98] or [AMS04, Section 3.8]).

## 2 Joint distribution of the canonical angles

Let $\mathcal{Y}$ and $\mathcal{S}$ be two independent observations from the uniform distribution on $\mathrm{Grass}(p, n)$, with $n \geq 2p$. Let $\theta_1 \leq \ldots \leq \theta_p$ be the canonical angles (see, e.g., [GV96]) between $\mathcal{Y}$ and $\mathcal{S}$. In this section, we give an expression for the joint probability density function (p.d.f.) of the $\theta$'s.

To this end we can assume that $\mathcal{S}$ is fixed to

$$\mathcal{S} = \mathrm{span}\,(S)\,, \qquad S = \begin{bmatrix} I_p \\ 0_{(n-p)\times p} \end{bmatrix}$$

and that

$$\mathcal{Y} = \mathrm{span}\,(Y)\,, \qquad Y = \begin{bmatrix} A \\ B \end{bmatrix},$$

where $A \in \mathbb{R}^{n_1 \times p}$ with $n_1 = p$ and $B \in \mathbb{R}^{n_2 \times p}$ with $n_2 = n - p$ are independent identically distributed Gaussian matrices.[1] This is because the Gaussian distribution is invariant under the orthogonal group of transformations; we refer to [Jam54, §6.2 and §7] for details.

To obtain the canonical angles between $\mathcal{Y}$ and $\mathcal{S}$, first orthonormalize $Y$ to obtain

$$\hat{Y} = \begin{bmatrix} A \\ B \end{bmatrix} (A^T A + B^T B)^{-1/2}.$$

Then the cosines of the canonical angles between $\mathcal{Y}$ and $\mathcal{S}$ are the singular values of $S^T \hat{Y} = A(A^T A + B^T B)^{-1/2}$. Consequently, the values $\cos^2 \theta_i$ are the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ of the matrix $(A^T A+B^T B)^{-1/2} A^T A (A^T A+B^T B)^{-1/2}$ or equivalently of $(A^T A+B^T B)^{-1} A^T A$. If $A^T A+B^T B = T^T T$ is a Cholesky decomposition (i.e., $T$ is upper triangular with positive diagonal elements), then the $\lambda$'s are the eigenvalues of the matrix $U = T^{-T} A^T A T^{-1}$. The matrix $U$ has the multivariate beta distribution $\mathrm{Beta}_p(\frac{1}{2}n_1, \frac{1}{2}n_2)$ [Mui82, Def. 3.3.2]. The joint p.d.f. of the $\lambda$'s is thus given by [Mui82, Th. 3.3.4]

$$\mathrm{dens}\,(\lambda_1, \ldots, \lambda_p) = c_{p,n_1,n_2} \prod_{i<j} |\lambda_i - \lambda_j| \prod_{i=1}^{p} \lambda_i^a (1-\lambda_i)^b \quad (1 \geq \lambda_1 \geq \ldots \geq \lambda_p \geq 0), \tag{1}$$

where $a = \frac{1}{2}(n_1 - p - 1) = -1/2$, $b = \frac{1}{2}(n_2 - p - 1)$, and

$$c_{p,n_1,n_2} = \frac{\pi^{p^2/2}}{\Gamma_p(p/2)} \cdot \frac{\Gamma_p((n_1+n_2)/2)}{\Gamma_p(n_1/2)\Gamma_p(n_2/2)}. \tag{2}$$

Observe that the determinant of the Jacobian of the change of variables between $\lambda$ and $\theta$ is $\prod_{i=1}^{p} 2 \sin\theta_i \cos\theta_i$, to obtain

$$\mathrm{dens}\,(\theta_1, \ldots, \theta_p) = 2c_{p,n_1,n_2} \prod_{i<j} |\cos^2\theta_i - \cos^2\theta_j|^\beta \prod_{i=1}^{p} \cos^{2a+1}\theta_i \sin^{2b+1}\theta_i. \tag{3}$$

Note as an aside that $\tan\theta_i$, $i = 1, \ldots, p$, can also be expressed as the singular values of $BA^{-1}$ (see, e.g., [AMSV02, Th. 3.1]). Hence $\tan^{-2}\theta_i$ is the $i$th largest eigenvalue of $(A^T A)(B^T B)^{-1}$. These eigenvalues are distributed as the latent roots in MANOVA [Mui82, Cor. 10.4.3], which again yields (3). Formulas for the p.d.f. of the largest latent root in MANOVA, which corresponds to the smallest $\theta$, are given in [Mui82, Section 10.6.4] for the case where $\frac{1}{2}(n-2p-1)$ is an integer. In this paper, we are interested in the probability distribution function of the largest canonical angle, which corresponds to the *smallest* latent root in MANOVA.

# 3 Distribution of the largest canonical angle

In this section, we derive formulas for the p.d.f. and the probability distribution function of the largest canonical angle $\theta_p$ between $\mathcal{S}$ and the random subspace $\mathcal{Y}$. The formulas are given in Theorem 1.

[1] We define a *Gaussian matrix* (over $\mathbb{R}$) to be a random matrix whose elements are generated by independent, normally distributed Gaussian random processes.

Let $x = \lambda_p$. We first calculate dens $(x)$, the p.d.f. of $x$. We start from

$$\text{dens}\,(x)dx = \int_{\{\lambda:1\geq\lambda_1\geq\ldots\geq\lambda_p\geq 0,\lambda_p\in(x,x+dx)\}} \text{dens}\,(\lambda_1,\ldots,\lambda_p)d\lambda_1\ldots d\lambda_p,$$

which yields

$$\begin{aligned}\text{dens}\,(x) &= \int_{1\geq\lambda_1\geq\ldots\geq\lambda_{p-1}\geq x} \text{dens}\,(\lambda_1,\ldots,\lambda_{p-1},x)d\lambda_1\ldots d\lambda_{p-1}\\ &= c_{p,n_1,n_2}x^a(1-x)^b\int_{1\geq\lambda_1\geq\ldots\geq\lambda_{p-1}\geq x}\prod_{i<j<p}|\lambda_i-\lambda_j|\prod_{i=1}^{p-1}|\lambda_i-x|\lambda_i^a(1-\lambda_i)^b d\lambda_1\ldots d\lambda_{p-1}.\end{aligned} \tag{4}$$

The change of variables $\lambda_i = (1-x)t_i + x$ gives

$$\begin{aligned}\text{dens}\,(x) &= c_{p,n_1,n_2}x^a(1-x)^b(1-x)^{(p-1)(p-2)/2}(1-x)^{p-1}x^{a(p-1)}(1-x)^{b(p-1)}(1-x)^{p-1}\\ &\quad\times\int_{1\geq t_1\geq\ldots\geq t_{p-1}\geq 0}\prod_{i<j<p}|t_i-t_j|\prod_{i=1}^{p-1}t_i\left(1+\tfrac{1-x}{x}t_i\right)^a(1-t_i)^b dt_1\ldots dt_{p-1}.\end{aligned} \tag{5}$$

Then

$$\begin{aligned}\text{dens}\,(x) &= c_{p,n_1,n_2}x^a(1-x)^b(1-x)^{(p-1)(p-2)/2}(1-x)^{p-1}x^{a(p-1)}(1-x)^{b(p-1)}(1-x)^{p-1}\\ &\quad\times\frac{2^{p-1}}{V(O_{p-1})}\int_{0<Y<I_{p-1}}\det(Y)\left(\det\left(I+\tfrac{1-x}{x}IY\right)\right)^a(\det(I-Y))^b dY,\end{aligned} \tag{6}$$

where $V(O_q) = \int_{O_q} Q^T dQ = \frac{2^q\pi^{q^2/2}}{\Gamma_q(\frac{q}{2})}$ is the volume of the orthogonal group $O_q$ [Mui82, p. 71]. Equation (6) comes from (5) because $dY = \prod|t_i - t_j|dT(Q^TdQ)$, where $Y = QTQ^T$ is an eigendecomposition with eigenvalues sorted in nonincreasing order, and $Q$ cancels out everywhere in the integrand. The factor $2^{p-1}$ appears because the eigendecomposition is defined up to the choice of the direction of the eigenvectors.

Combining [Mui82, Th. 7.4.2] and [Mui82, Th. 7.4.3] yields the identity

$$\begin{aligned}&\int_{0<Y<I_q}\det(I-XY)^{-f}(\det Y)^{e-(q+1)/2}\det(I-Y)^{g-e-(q+1)/2}dY\\ &\qquad = \frac{\Gamma_q(e)\Gamma_q(g-e)}{\Gamma_q(g)}\det(I-X)^{-f}\,{}_2F_1\left(g-e,f;g;-X(I-X)^{-1}\right)\end{aligned} \tag{7}$$

where $q$ is the size of $X$, valid for

$$\text{Re}(X) < I,\ \text{Re}(e) > \tfrac{1}{2}(q-1),\ \text{Re}(g-e) > \tfrac{1}{2}(q-1). \tag{8}$$

We make $q = p-1$, $-f = a$, $e-(q+1)/2 = 1$, $g-e-(q+1)/2 = b$ and $X = -(1-x)/x\, I_{p-1}$. From this we find $f = 1/2$, $e = 1+p/2$ and $g = (n+1)/2$. Conditions (8) yield $x > 0$ and $p < \frac{n+1}{2}$, and (6) becomes

$$\begin{aligned}\text{dens}\,(x) &= \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})}\cdot\frac{\Gamma_p(\frac{n}{2})}{\Gamma_p(\frac{p}{2})\Gamma_p(\frac{n-p}{2})}\cdot\frac{\Gamma_{p-1}(\frac{p-1}{2})}{\pi^{\frac{(p-1)^2}{2}}}\cdot\frac{\Gamma_{p-1}(1+\frac{p}{2})\Gamma_{p-1}(\frac{n-p-1}{2})}{\Gamma_{p-1}(\frac{n+1}{2})}\\ &\quad\times x^{\frac{-1}{2}}(1-x)^{\frac{p(n-p)-2}{2}}\,{}_2F_1(\tfrac{n-p-1}{2},\tfrac{1}{2};\tfrac{n+1}{2};(1-x)I_{p-1}).\end{aligned} \tag{9}$$

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

# 5 Numerical experiments

Formulas in Theorem 1 require an algorithm that evaluates the hypergeometric 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 $\mathrm{span}(X)$, with $n \times p$ matrix $X$ chosen from a Gaussian distribution. The principal angles between $\mathrm{span}(X)$ and the fixed subspace $\mathrm{span}(I_{n\times p})$ 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.

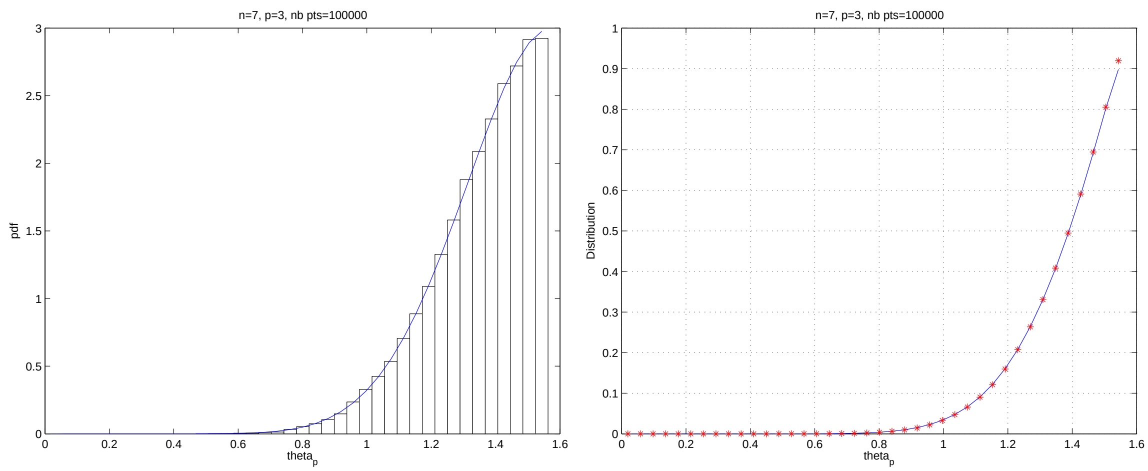


Figure 1: The solid curves correspond to the probability density function (11)—left-hand plot—and the probability distribution function (12)—right-hand plot—evaluated using `mhgi` for the case $n = 7$, $p = 3$. The histogram and the stars show approximations evaluated from a sample of $10^5$ three-dimensional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

[EAS98] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), no. 2, 303–353.

[GV96] G. H. Golub and C. F. Van Loan, *Matrix computations, third edition*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, 1996.

[Jam54] A. T. James, *Normal multivariate analysis and the orthogonal group*, Ann. Math. Statistics **25** (1954), 40–75.

[KE05] P. Koev and A. Edelman, *The efficient evaluation of the hypergeometric function of a matrix argument*, to appear in Math. Comp., 2005.

[Mat92] The MathWorks, Inc., Natick, MA, *MATLAB reference guide*, 1992.

[Mui82] R. J. Muirhead, *Aspects of multivariate statistical theory*, John Wiley & Sons, New York, 1982.

[QZL04] Li Qiu, Yanxia Zhang, and Chi-Kwong Li, *Unitarily invariant metric on the Grassmann space*, submitted, 2004.

[Wei00] Alan Weinstein, *Almost invariant submanifolds for compact group actions*, J. Eur. Math. Soc. (JEMS) **2** (2000), no. 1, 53–86.