Introduction

For the immediate future my plan is to develop

New algorithms for computational multivariate statistical analysis,

as described in section 1.

I also plan on continuing work on

- Stable $\mathcal{O}(n^2)$ algorithm for the nonsymmetric tridiagonal eigenvalue problem;
- Accurate computations with totally nonnegative (TN) matrices;
- Research in matrix theory related to oscillating systems of vectors;
- Continuing collaborations on practical applications,

as described in section 2.

1. New algorithms for computational multivariate statistical analysis

The main goal of this project is to devise stable and efficient algorithms for computing the density, distribution, and quantile functions of select eigenvalues (and functions thereof) of the classical random matrix ensembles—Wishart, Jacobi, and Laguerre.

These functions are a fundamental tool in many multivariate statistical methods, such as hypothesis testing, principal component analysis, canonical correlation analysis, multivariate analysis of variance, etc. [38]. This type of analysis is an integral part of many practical applications where multiple signal sources/receivers are present and thus random covariance matrices naturally occur. Applications include telecommunications and wireless networks [6, 16, 18, 27, 36, 37, 39, 40, 48], image and signal processing [47], military applications (automatic target recognition and classification) [7, 23, 34], etc.[1]

Many explicit formulas for the distributions of the eigenvalues of the classical random matrix ensembles have been known for over 40 years (section 1.2). Unfortunately, most of these formulas are in terms of the hypergeometric function of a matrix argument—an extremely slowly converging series of Jack functions.

The hypergeometric function of a matrix argument has been incredibly difficult to compute even in the simplest cases (matrix argument of size 3 or 4); the development of efficient algorithms for its computation has been identified as a central open research problem in a large number of recent publications, e.g., [6, §5A and §5B], [7, §IV], [16, §III], [18, Appendix I], [23], [34, §III.B], [36, §I], [37, §III], [39, §III], [40, §III], [47, §II.C], [48, §5.2].

In our 2005 paper [33] Edelman and I developed the first practical algorithm for computing the hypergeometric function of a matrix argument (section 1.3.1). This new algorithm is very efficient for matrix arguments of size up to 10 (takes at most a few seconds time) and is exponentially faster than the previous best algorithm [21]: on the same 5×5 example our algorithm takes less than $\frac{1}{100}$ of a second as opposed to 8 days for the algorithm of [21].

The plan now is to develop new, faster algorithms for the computation of the hypergeometric function of a matrix argument, prove new theoretical results in random matrix theory, and ultimately develop reliable, accurate, and efficient algorithms for the computation of the density, distribution, and quantile functions of the extreme eigenvalues of the classical random matrix ensembles. In particular:

- I plan on developing new FFT-like algorithm for the hypergeometric function of a matrix argument (section 1.5). This algorithm, whose idea is analogous to that of Cooley and Tukey [11], would make it possible to increase the size of the random matrices whose eigenvalue distributions one can compute to 50 or even 100, up from 10;

[1] The above cited IEEE papers are available online from `http://ieeexplore.ieee.org`.

- I plan on proving new results in random matrix theory: formulas for the distributions of the extreme eigenvalues of the complex Jacobi ensemble (section 1.6);
- I plan on performing detailed convergence and stability analysis for each density, distribution, and quantile function for the extreme eigenvalues of each classical random matrix ensemble (section 1.8);
- I plan on implementing the above algorithms efficiently in MATLAB, R, and SAS;
- I plan on incorporating results of several authors whose work provides marginal information about the distributions of interest to us in cases when the explicit formulas and expressions become numerically infeasible (sections 1.3 and 1.7).

The rest of this section is organized as follows. In section 1.1 I give definitions of the basic objects in random matrix theory and multivariate analysis. In section 1.2 I list known formulas for the distributions of the eigenvalues of random matrices that I intend to use in my algorithms. In section 1.3 I describe the work of other authors that may have marginal, but important impact in the development of the algorithms. In section 1.4 I outline the work I plan on performing and outline reasons I believe it will be successful.

1.1. **Basic definitions.** In this section I present the definitions of the classical random matrix ensembles as well as the definitions of several objects commonly encountered in multivariate analysis—the multivariate Gamma function, the generalized Pochhammer symbol, and the hypergeometric function of a matrix argument.

1.1.1. *Classical random matrix ensembles.* The following definitions are taken directly from [38, Def. 3.1.3, p. 82], [45], [13, p. 24], and [38, section 3.3, p. 109].

Definition 1 (Real (complex) Wishart ensemble $W_m(n, \Sigma)$ ($\mathcal{C}W_m(n, \Sigma)$))**.** *Let the $n \times m$ real (complex) Gaussian random matrix Z be distributed as $N(0, I_n \otimes \Sigma)$ ($\mathcal{C}N(0, I_n \otimes \Sigma)$). The matrix $A = Z^T Z$ is called $m \times m$* real (complex) Wishart matrix *with n degrees of freedom and covariance matrix Σ.*

Definition 2 (Laguerre ensemble)**.** *The $m \times m$ β-*Laguerre matrix *is defined as*

$$L \equiv BB^T, \quad \textit{where} \quad B = \begin{bmatrix} \chi_{2a} & & & \\ \chi_{\beta(m-1)} & \chi_{2a-\beta} & & \\ & \ddots & \ddots & \\ & & \chi_{\beta} & \chi_{2a-\beta(m-1)} \end{bmatrix}, \quad a > \frac{\beta}{2}(m-1).$$

Definition 3 (Jacobi ensemble)**.** *Let $A \sim W_m(n_1, \Sigma)$ and $B \sim W_m(n_2, \Sigma)$, where $n_1 \geq m$ and $n_2 \geq m$, be independently distributed Wishart matrices. Then the matrix $C \equiv A(A+B)^{-1}$ has* real Jacobi *distribution. When $A \sim \mathcal{C}W_m(n_1, \Sigma)$ and $B \sim \mathcal{C}W_m(n_2, \Sigma)$, then C has* complex Jacobi *distribution.*

The Jacobi distribution is sometimes called multivariate beta *distribution and is closely related to the MANOVA ensemble* [38] *(which studies AB^{-1} instead of $A(A+B)^{-1}$).*

1.1.2. *Other common objects in multivariate analysis.* The raising factorial, the Gamma function, and the (univariate) hypergeometric function have natural multivariate analogues.

A *partition* κ of an integer $k \geq 0$ (denoted $\kappa \vdash k$) is a sequence $\kappa = (\kappa_1, \kappa_2, \ldots)$ such that $\kappa_1 \geq \kappa_2 \geq \cdots \geq 0$ are integers and $|\kappa| \equiv \kappa_1 + \kappa_2 + \cdots = k$.

The *multivariate Gamma function* of parameter α is defined as

$$\Gamma_m^{(\alpha)}(c) \equiv \pi^{\frac{m(m-1)}{2\alpha}} \prod_{i=1}^{m} \Gamma\left(c - \frac{i-1}{\alpha}\right) \text{ for } \Re(c) > \frac{m-1}{\alpha}.$$

The *generalized Pochhammer symbol* is defined as

$$(a)_\kappa^{(\alpha)} \equiv \prod_{(i,j)\in\kappa} \left(a - \frac{i-1}{\alpha} + j - 1\right).$$

The *hypergeometric function of a matrix argument and parameter* $\alpha > 0$ is defined as

$$(1) \qquad {}_pF_q^{(\alpha)}(a_1,\ldots,a_p;b_1,\ldots,b_q;X) \equiv \sum_{k=0}^{\infty}\sum_{\kappa\vdash k} \frac{(a_1)_\kappa^{(\alpha)}\cdots(a_p)_\kappa^{(\alpha)}}{k!(b_1)_\kappa^{(\alpha)}\cdots(b_q)_\kappa^{(\alpha)}}\cdot C_\kappa^{(\alpha)}(X).$$

where $p \ge 0$ and $q \ge 0$ are integers, X is an $m \times m$ complex symmetric matrix, and $C_\kappa^{(\alpha)}(X)$ is the Jack function (also sometimes known as *zonal polynomial* or *generalized Schur function*)—a symmetric, homogeneous polynomial of degree $|\kappa|$ in the eigenvalues x_1, x_2, …, x_m of X. Its formal definition can be found in Stanley's landmark paper [50].

1.2. **Distributions of the eigenvalues of the classical ensembles.** In this section I present explicit formulas for the densities and/or distributions of the extreme eigenvalues of the classical random matrix ensembles. These are typically expressed in terms of the hypergeometric function of a matrix argument.

1.2.1. *Wishart.* Let $A \sim W_m(n,\Sigma)$ or $A \sim \mathcal{C}W_m(n,\Sigma)$ be a real or a complex Wishart matrix. The values of $\alpha = 2$ and $\alpha = 1$ correspond to the real and complex cases, respectively. For the distributions of the extreme eigenvalues of A we have

$$(2) \qquad P(\lambda_{\max}(A) < x) = \frac{\Gamma_m^{(\alpha)}\left(\frac{m-1}{\alpha}+1\right)}{\Gamma_m^{(\alpha)}\left(\frac{n+m-1}{\alpha}+1\right)}\left(\det\left(\tfrac{x}{\alpha}\Sigma^{-1}\right)\right)^{\frac{n}{\alpha}} {}_1F_1^{(\alpha)}\left(\tfrac{n}{\alpha};\tfrac{n+m-1}{\alpha}+1;-\tfrac{x}{\alpha}\Sigma^{-1}\right);$$

$$(3) \qquad P(\lambda_{\min}(A) < x) = 1 - e^{\operatorname{tr}\left(-\frac{x}{\alpha}\Sigma^{-1}\right)}\sum_{k=0}^{mt}\sum_{\kappa\vdash k,\,\kappa_1\le t}\frac{C_\kappa^{(\alpha)}\left(\frac{x}{\alpha}\Sigma^{-1}\right)}{k!},$$

where (3) is only valid when $t \equiv \frac{n-m+1}{\alpha} - 1$ is a nonnegative integer (see [38, Thm. 9.7.1 and Cor. 9.7.4] for the real case and [45, Cor. 3.3 and 3.5] for the complex case).

In the real case, the density of $\operatorname{tr}(A)$ is [38, p. 341]

$$f(x) = \det(z^{-1}\Sigma)^{-\frac{n}{2}}\sum_{k=0}^{\infty}\frac{g_{\frac{nm}{2}+k,2z}(x)}{k!}\sum_{\kappa\vdash k}\left(\tfrac{n}{2}\right)_\kappa^{(2)}\cdot C_\kappa^{(2)}(I - z\Sigma^{-1}),$$

where z is arbitrary and $g_{r,2z}(x) = e^{-x/2z}x^{r-1}/\big((2z)^r\Gamma(r)\big)$, $x > 0$.

1.2.2. *β-Laguerre.* Let L be an $m \times m$ β-Laguerre matrix and let $\alpha = 2/\beta$. Then

$$P(\lambda_{\max}(L) < x) = \frac{\Gamma_m^{(\alpha)}\left(\frac{m-1}{\alpha}+1\right)}{\Gamma_m^{(\alpha)}\left(a+\frac{m-1}{\alpha}+1\right)}\left(\frac{x}{2}\right)^{am} {}_1F_1^{(\alpha)}\left(a;a+\tfrac{m-1}{\alpha}+1;-\tfrac{x}{2}I\right).$$

If $t = a - \frac{\beta}{2}(m-1) - 1$ is a nonnegative integer, then the density of $\lambda_{\min}(L)$ is proportional to (see, e.g., [13, Thm. 10.1.1, p. 146])

$$f(x) = x^{tm}\cdot e^{-\frac{mx}{2}}\cdot\ {}_2F_0^{(\alpha)}\left(-t,\beta\tfrac{m}{2}+1;-\tfrac{2}{x}I_{m-1}\right).$$

1.2.3. *Jacobi.* Let C have a real or a complex Jacobi distribution (see Definition 3). Then the distribution of $\lambda_{\max}(C)$ is known in the real ($\alpha = 2$) case from [10, eq. (61)]. I intend to prove that this formula is true in the complex ($\alpha = 1$) case (see section 1.6).

$$(4)\quad P(\lambda_{\max}(C) < x) = \frac{\Gamma_m^{(\alpha)}(\frac{n_1+n_2}{\alpha})\Gamma_m^{(\alpha)}(\frac{m-1}{\alpha}+1)}{\Gamma_m^{(\alpha)}(\frac{n_1+m-1}{\alpha}+1)\Gamma_m^{(\alpha)}(\frac{n_2}{\alpha})} \cdot x^{\frac{mn_1}{\alpha}} \cdot {}_2F_1^{(\alpha)}(\tfrac{n_1}{\alpha}, \tfrac{-n_2+m-1}{\alpha}+1; \tfrac{n_1+m-1}{\alpha}+1; xI).$$

The distributions of $\lambda_{\max}(C)$ and $\lambda_{\min}(C)$ are closely related:

$$(5)\qquad P(\lambda_{\min}(C) < x) = 1 - P(\lambda_{\min}(C) > x) = 1 - P(B(A+B)^{-1} < (1-x)I),$$

which is immediately evaluable using (4).

1.3. **Previous algorithms and related results.** In this section I survey existing algorithms for computing the hypergeometric function of a matrix argument (section 1.3.1) as well as relevant results which provide marginal information about the distribution of the eigenvalues of the random matrix ensembles. The latter may be useful when the formulas of section 1.2 are numerically infeasible and include include the work of Tracy–Widom and Iain Johnstone (section 1.3.2), Gross–Richards (section 1.3.4), Butler–Wood (section 1.3.3), and William Chen (section 1.3.5). The plan for exploiting these results is in section 1.7.

1.3.1. *Previous algorithms.* Edelman and I recently presented new algorithms [33] for approximating the hypergeometric function of a matrix argument. We exploited the combinatorial properties of the Jack function [50] and derived an algorithm for computing the truncation

$$(6)\qquad {}_pF_q^{(\alpha)}(a_1, \ldots, a_p; b_1, \ldots, b_q; X) \equiv \sum_{k=0}^{N} \sum_{\kappa \vdash k} \frac{(a_1)_\kappa^{(\alpha)} \cdots (a_p)_\kappa^{(\alpha)}}{k!(b_1)_\kappa^{(\alpha)} \cdots (b_q)_\kappa^{(\alpha)}} \cdot C_\kappa^{(\alpha)}(X)$$

in time that grows only linearly with the size m of the matrix argument X. As a function of N, our algorithm's cost grows as $\mathcal{O}(M^2)$, where $M \equiv \{\#\kappa \,|\, |\kappa| \le N\}$ is the number of terms in (6). Our algorithms efficiently (i.e., in less than 5 seconds) compute the distributions of the extreme eigenvalues of a random matrix of size up to 10.

Our algorithms in [33] are exponentially faster[2] than the previous best algorithms [21, 46]. In section 1.5 I intend to develop a new FFT-like algorithm with complexity (essentially) $\mathcal{O}(M \log M)$, that will likely allow us to efficiently compute distributions with random matrices of size up to 50 or 100. The work of Johnstone [25] and Tracy–Widom [52] may then allow us to compute distributions of the eigenvalues of random matrices of size larger than 100 in certain cases (sections 1.3.2 and 1.7).

1.3.2. *Tracy–Widom limits.* As the size of a Wishart matrix tends to infinity, its extreme eigenvalues converge in distribution to the so-called "*Tracy–Widom*" limits [25, 52]. This property is universal among many matrix ensembles [49].

For example, for the real Wishart ensemble, if $A \sim W_m(n, I)$, we define constants μ and σ

$$\mu = \left(\sqrt{n-\tfrac{1}{2}} + \sqrt{m-\tfrac{1}{2}}\right)^2, \quad \sigma = \left(\sqrt{n-\tfrac{1}{2}} + \sqrt{m-\tfrac{1}{2}}\right)\left(\frac{1}{\sqrt{n-\frac{1}{2}}} + \frac{1}{\sqrt{m-\frac{1}{2}}}\right)^{\frac{1}{3}}.$$

If $(m, n) \to \infty$ in such a way that $n/m \to \gamma \ge 1$, then [25, Thm. 1.1]

$$\frac{\lambda_{\max}(A) - \mu}{\sigma} \xrightarrow{\mathcal{D}} W_1,$$

[2] See [33, section 2] for a detailed analysis of the cost of the algorithms in [21, 46].

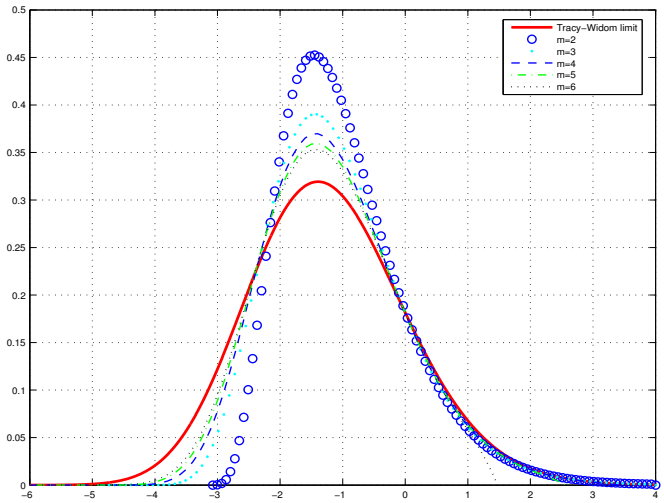

FIGURE 1. Convergence of the largest eigenvalue of the real Wishart matrix to the Tracy–Widom limit of order 1; m = size fo the matrix; $n = 4m$.

where W_1 is the Tracy–Widom law of order 1. Its density is

$$F_1(s) = \exp\left(-\frac{1}{2}\int_s^\infty q(x) + (x-s)^2 q^2(x) dx\right), \qquad s \in \mathbb{R},$$

where q solves the (nonlinear) Painlevé II differential equation

$$q''(x) = xq(x) + 2q^3(x), \qquad q(x) \to \mathrm{Ai}(x) \text{ as } x \to +\infty,$$

and $\mathrm{Ai}(x)$ denotes the Airy function.

The function F_1 (as well as the other Tracy–Widom limits, which are analogous) are readily computable numerically [42].

In Figure 1 we plot the Tracy–Widom limit against the formula (2) for $2 \le m \le 6$, and $n = 4m$.

Analogous Tracy–Widom limits apply to for the extreme eigenvalues of the complex Wishart and Jacobi ensembles; see [25, sec. 1.5], as well as [2, 24, 26, 49] for details.

1.3.3. *Laplace approximations.* These were used by Butler and Wood [5] to approximate

$$\begin{aligned} {}_1F_1^{(2)}(a;c;X) &= B \cdot \int_{0<Y<I} e^{\mathrm{tr}\,(XY)} (\det Y)^{a-\frac{n+1}{2}} \det(I-Y)^{c-a-\frac{n+1}{2}} (dY), \\ {}_2F_1^{(2)}(a,b;c;X) &= B \cdot \int_{0<Y<I} \det(I-XY)^{-b} (\det Y)^{a-\frac{n+1}{2}} \det(I-Y)^{c-a-\frac{n+1}{2}} (dY), \end{aligned}$$

where $B = \Gamma_n^{(2)}(c)/\big(\Gamma_n^{(2)}(a)\Gamma_n^{(2)}(c-a)\big)$, in order to approximate the value of ${}_pF_q^{(\alpha)}$ for $p = 1, 2$, $q = 1$, and $\alpha = 2$ quite accurately in many cases.

1.3.4. *The work of Gross and Richards.* A very elegant formula is presented in [19] for the hypergeometric function of *two* matrix arguments in the complex case ($\alpha = 1$) as a (scaled) determinant of classical hypergeometric functions:

$$\begin{aligned} {}_pF_q^{(1)}(a_{1:p}; b_{1:q}; X, Y) &\equiv \sum_{k=0}^\infty \sum_{\kappa \vdash k} \frac{(a_1)_\kappa^{(1)} \cdots (a_p)_\kappa^{(1)}}{k! (b_1)_\kappa^{(1)} \cdots (b_q)_\kappa^{(1)}} \cdot \frac{C_\kappa^{(1)}(X) C_\kappa^{(1)}(Y)}{C_\kappa^{(1)}(I)} \\ &= B \cdot \frac{\det\left({}_pF_q(a'_{1:p}; b'_{1:q}; x_i y_j)\right)_{i,j=1}^m}{V(X)V(Y)}, \end{aligned} \tag{7}$$

where $a_i' = a_i - m + 1,\ i = 1, 2, \ldots, p,\ b_i' = b_i - m + 1,\ i = 1, 2, \ldots, q,\ V(Z) = \prod_{i>j}(z_i - z_j)$, and B is a scalar constant. Unfortunately, this formula applies only to the complex case ($\alpha = 1$) and only for *two* matrix arguments, none of which have double eigenvalues. Thus (7) is inapplicable directly to any of the formulas in section 1.2 (since those involve the hypergeometric function of *one* matrix argument), but see section 1.7.

1.3.5. *The work of William Chen.* Given a matrix $C = A(A + B)^{-1}$ with real Jacobi distribution ($A \sim W_m(2n_1 + m + 1, \Sigma)$, $B \sim W_m(2n_2 + m + 1, \Sigma)$), William Chen [8, 9] presented tables for the upper percentage points (in the 0.8–0.99 range) for $m \le 6, n_1 \le 7$ and $n_2 \le 1000$. While these results are easily reproducible for $m \le 6$, $n_1 \le 7$, and $n_2 \le 30$ using the formulas from section 1.2, this approach seems attractive for values of n_2 between 30 and 1000.

1.4. **Research plan.** In this section I outline the results and algorithms I intend to prove and develop as a part of this project. They include:

(1) Implementation of the density, distribution, and quantile functions for the extreme eigenvalues, the trace, etc., of the classical random matrix ensembles (section 1.4.1);
(2) New Cooley–Tukey type algorithm for computing the hypergeometric function of a matrix argument (section 1.5);
(3) New formulas for the extreme eigenvalues of the complex Jacobi ensemble (section 1.6);
(4) Incorporation of the results described in section 1.3 in the limited circumstances in which they apply (section 1.7);
(5) Convergence and stability analysis for the hypergeometric function of a matrix argument (section 1.8);
(6) Efficient empirical sampling algorithms of eigenvalues of random matrices (section 1.9).

1.4.1. *Algorithms for applied multivariate statistical analysis.* I intend to design algorithms for computing the density, distribution, and quantile functions of the extreme eigenvalues and the trace of the Wishart, Jacobi, and Laguerre ensembles using the formulas from section 1.2 as well as the ones from section 1.6 below.

1.5. **New Cooley–Tukey type algorithm for computing the hypergeometric function of a matrix argument.** I intend to develop a new algorithm for approximating the hypergeometric function of a matrix argument (1). The idea, which I describe below, is analogous to that of Cooley and Tukey [11], and was inspired by the work of Püschel and Moura [44].

The idea is to compute a finite truncation of the series (1). If M is the number of terms in this truncation, this new algorithm would cost $O(M \log M)$ (as opposed to our current algorithm [30, 33] which costs, essentially, $\mathcal{O}(M^2)$).

The Cooley–Tukey FFT algorithm [11] is a fast $\mathcal{O}(n \log n)$ algorithm for matrix-vector multiplication by the DFT matrix D_n. The matrix D_n is the character table of the cyclic group. When the size n of the group factors as $n = km$, the cyclic group structure implies [43, eq. (10)]:

$$D_n = L_m^n (I_k \otimes D_m) T_m^n (D_k \otimes I_m),$$

where L_m^n is the stride permutation matrix and T_m^n is a diagonal matrix whose exact form does not concern us. The matrix D_n is thus decomposed as a product of simpler matrices. By exploiting this recursive structure, the FFT ends up taking $\mathcal{O}(n \log n)$ time instead of $\mathcal{O}(n^2)$, which is what a conventional matrix-vector multiplication algorithm would cost.

I intend to use the exact same idea for computing Jack functions, and illustrate this idea in the case $\alpha = 1$ (when the Jack function $C_\kappa^{(\alpha)}$ is the (scaled) Schur function s_λ).

The Schur functions describe the characters of $GL_n(\mathbb{C})$, thus the connection with the Cooley–Tukey idea. My algorithms for computing the Jack function [12, 33] use the identity [35, p. 72]

$$s_\lambda(x_1, \ldots, x_n) = \sum_{\mu \le \lambda} s_\mu(x_1, \ldots, x_{n-1}) x_n^{|\lambda/\mu|}, \tag{8}$$

which can be written in matrix form

$$S_n(x_1, \ldots, x_n) = S_{n-1}(x_1, \ldots, x_{n-1}) Y_n(x_n),$$

where S_n and S_{n-1} are vectors of Schur functions of n and $n-1$ variables respectively. Thus the problem of computing Schur (Jack) functions is once again a structured matrix-vector multiplication problem.

The matrix $Y_n(x_n)$ is dense, thus the cost of this evaluation is $\mathcal{O}(M^2)$, where M is the number of Schur functions we want to compute (i.e., the length of the vector S_n).

It turns out that the matrix Y_n is highly structured. For example

$$Y_2(x) = \left[\begin{array}{cccc|cccc|cccc|cccc} 1 & x & x^2 & x^3 & x & x^2 & x^3 & x^4 & x^2 & x^3 & x^4 & x^5 & x^3 & x^4 & x^5 & x^6 \\ & 1 & x & x^2 & & x & x^2 & x^3 & & x & x^2 & x^3 & & & & \\ & & 1 & x & & & x & x^2 & & & x & x^2 & & & & \\ & & & 1 & & & & & & & & & & & & \end{array}\right]$$

and (8) implies

$$S_2(x_1, x_2) = S_1(x_1) Y_2(x_2),$$

where $S_1(x_1) \equiv (s_{(0)}, s_{(1)}, s_{(2)}, s_{(3)}) = (1, x_1, x_1^2, x_1^3)$, and

$$\begin{aligned} S_2(x_1, x_2) \quad \equiv \quad & (s_{(0)}, s_{(1)}, s_{(2)}, s_{(3)}, s_{(1,1)}, s_{(2,1)}, s_{(3,1)}, s_{(4,1)}, s_{(2,2)}, \\ & s_{(3,2)}, s_{(4,2)}, s_{(5,2)}, s_{(3,3)}, s_{(4,3)}, s_{(5,3)}, s_{(6,3)}). \end{aligned}$$

Define

$$C(x) \equiv \left[\begin{array}{cccc} 1 & x & x^2 & x^3 \\ & 1 & x & x^2 \\ & & 1 & x \\ & & & 1 \end{array}\right], \quad \text{and} \quad B \equiv \left[\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right].$$

Thus

$$Y_2(x_2) = \left[\begin{array}{cccc} I_4 & x_2 B & x_2^2 B^2 & x_2^3 B^3 \end{array}\right] \cdot (I_4 \otimes C(x_2)).$$

This sparse structure is resemblant to the one in the FFT algorithm of Cooley and Tukey. We can now multiply by Y_2 in linear ($\mathcal{O}(M)$) time by observing that C^{-1} is bidiagonal

$$C^{-1} = \left[\begin{array}{cccc} 1 & -x & & \\ & 1 & -x & \\ & & 1 & -x \\ & & & 1 \end{array}\right].$$

When evaluating Jack functions for $\alpha \neq 1$, the only difference is that the matrix C is now Toeplitz

$$C = \left[\begin{array}{cccc} 1 & x & x^2(\alpha+1) & x^3(\alpha+1)(2\alpha+1) \\ & 1 & x & x^2(\alpha+1) \\ & & 1 & x \\ & & & 1 \end{array}\right],$$

but C^{-1} is no longer bidiagonal in general. However, we can still multiply by C in $\mathcal{O}(n \log n)$ time by exploiting its Toeplitz structure [17, p. 193].

Therefore I expect to derive an algorithm for computing the truncated hypergeometric function of a matrix argument in $\mathcal{O}(M \log M)$ time.

1.6. **The extreme eigenvalues of the complex Jacobi ensemble.** I intend to establish the following new result for the distribution of largest eigenvalue of the complex Jacobi ensemble. A formula for the smallest eigenvalue would then follow automatically using (5).

Theorem 1. *Let $C = A(A+B)^{-1}$ have a* complex Jacobi distribution*, where $A \sim \mathcal{C}W_m(n_1, \Sigma)$ and $B \sim \mathcal{C}W_m(n_2, \Sigma)$ are independently distributed Wishart matrices. Then* (4) *is true for $\alpha = 1$. Namely*

$$P(\lambda_{\max}(C) < x) = \frac{\Gamma_m^{(1)}(n_1+n_2)\Gamma_m^{(1)}(m)}{\Gamma_m^{(1)}(n_1+m)\Gamma_m^{(1)}(n_2)} \cdot x^{mn_1} \cdot {}_2F_1^{(1)}(n_1, m-n_2; n_1+m; xI). \tag{9}$$

Proof. I intend to prove this result as follows.

(1) Without loss of generality we can assume that $\Sigma = I$ (see Definition 3);

(2) The density of A is [22, section 8]

$$\frac{1}{\Gamma_m^{(1)}(n_1)} \cdot (\det A)^{n_1-m} \cdot e^{\operatorname{tr}(-A)}(dA);$$

(3) The joint density of A and B is

$$\frac{1}{\Gamma_m^{(1)}(n_1)\Gamma_m^{(1)}(n_2)} \cdot (\det A)^{n_1-m} \cdot (\det B)^{n_2-m} \cdot e^{\operatorname{tr}(-A-B)}(dA)(dB);$$

(4) Let $U \equiv L^{-1}AL^{-T}$, where L is the Cholesky factor of $A+B$. The matrices U and C thus have the same eigenvalues. By repeating, verbatim, the argument of Theorem 3.3.1 in Muirhead [38, p. 109] we conclude that the density function of U is

$$\frac{\Gamma_m^{(1)}(n_1+n_2)}{\Gamma_m^{(1)}(n_1)\Gamma_m^{(1)}(n_2)} \cdot (\det U)^{n_1-m} \cdot \det(I-U)^{n_2-m}; \tag{10}$$

(5) To compute the distribution $P(\lambda_{\max}(U) < t)$ of the largest eigenvalue of U, we integrate the density (10) from 0 to tI:

$$P(U < tI) = \frac{\Gamma_m^{(1)}(n_1+n_2)}{\Gamma_m^{(1)}(n_1)\Gamma_m^{(1)}(n_2)} \int_{0<U<tI} (\det U)^{n_1-m} \cdot \det(I-U)^{n_2-m}(dU), \tag{11}$$

where $B < C$ means that $C - B$ is Hermitian positive definite;

(6) To evaluate the integral in (11), we establish the formula

$${}_2F_1^{(1)}(a, b; c; X) = \frac{\Gamma_m^{(1)}(c)}{\Gamma_m^{(1)}(a)\Gamma_m^{(1)}(c-a)} \int_{0<Y<I} \det(I-XY)^{-b}(\det Y)^{a-m} \det(I-Y)^{c-a-m}(dY),$$

valid for $\|X\| < 1$, $\Re(a) > m-1$, $\Re(c-a) > m-1$ (this is the complex analogue of the integral formula in Muirhead [38, eq. (5), p. 264]). The integration is over all Hermitian positive definite matrices Y whose eigenvalues do not exceed 1. This formula is proved analogously to Theorem 7.4.2 in Muirhead [38, p. 264] using Proposition 2.5 from [45] and the identity (see, e.g., James [22, (90)])

$$\det(I-X)^{-b} = {}_1F_0^{(1)}(b; X) = \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} \frac{(b)_\kappa^{(1)}}{k!} C_\kappa^{(1)}(X).$$

Therefore

$${}_2F_1^{(1)}(n_1, m-n_2; n_1+m; xI) = \frac{\Gamma_m^{(1)}(n_1+m)}{\Gamma_m^{(1)}(n_1)\Gamma_m^{(1)}(m)} \int_{0<Y<I} \det(I-xY)^{n_2-m}(\det Y)^{n_1-m}(dY). \tag{12}$$

To complete the proof, we need to make an m^2-dimensional change of variables $xY = U$ in (12) and plug into (11). □

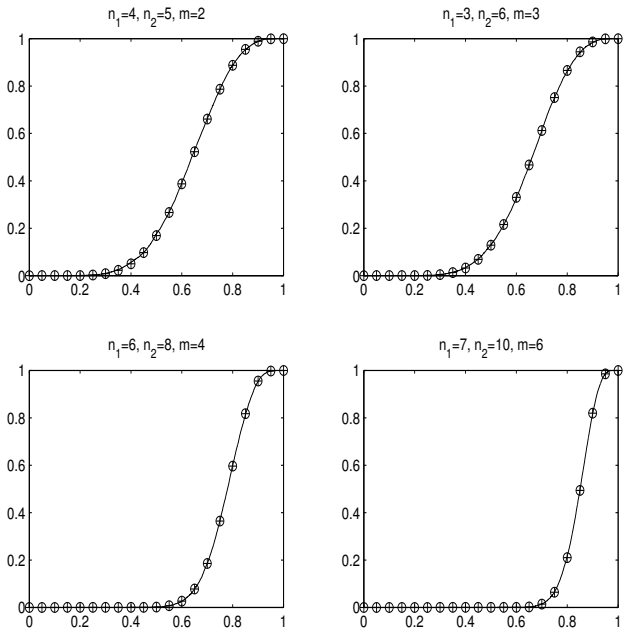

FIGURE 2. The distribution of the largest eigenvalue of the complex Jacobi ensemble. The solid line represents the result of a Monte–Carlo experiment with $10,000$ replications, while "$\oplus$" represents (9).

Remark 1. *To support the claim of Theorem 1, I performed a Monte–Carlo experiment and compared the results with those of* (9)*. The results in Figure 2 are a perfect visual match.*

1.7. **Using the results from section 1.3.** I expect to derive algorithms for computing the functions described in section 1.2 beyond the ranges of numerical feasibility of these formulas.

For example, Johnstone [25, section 1.4] showed that when $\Sigma = I$, the Tracy–Widom limits imply reasonable approximations to the distributions of the largest eigenvalue of the finite Wishart matrices. The recent work of El Karoui [28] may extend this implication beyond $\Sigma = I$.

The Butler–Wood approximations may be used in cases when the reported relative accuracy is sufficient in applications.

I intend to investigate the applicability of my TN algorithms [32] to the computation of the totally positive determinant in (7) and the possible removal of the singularity when $Y \to I$.

I expect to utilize the methods of William Chen for the computation of the ${}_2F_1$ function in those circumstances that makes his approach feasible.

1.8. **Stability and convergence analysis.** I intend to perform detailed stability and convergence analysis for the algorithms for each of the density, distribution, and quantile functions for the extreme eigenvalues of the classical random matrix ensembles. In particular, relationships such as the Kummer identity [38, p. 265, eq. (6)]

$$ {}_1F_1^{(\alpha)}(a;c;X) = e^{\operatorname{tr} X} \cdot {}_1F_1^{(\alpha)}(c-a;c;-X) \tag{13}$$

can greatly extend the region of numerical stability for the formula (2) in the presence of roundoff errors (see Figure 3).

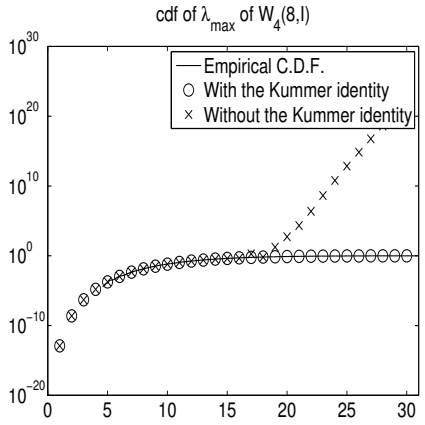

FIGURE 3. Formula (2) for the distribution of the largest eigenvalue of the Wishart ensemble is unstable for $x > 16$ (the "$\times$" marks). Transforming (2) using the Kummer identity (13) makes it stable (the "$\circ$" marks). For reference, the result of a Monte–Carlo experiment with $10,000$ replications is given in a solid line.

In addition to (7) and (13), I have collected a number of other identities involving the hypergeometric function of a matrix argument [3, 15, 19, 20]:

$$\begin{aligned}
{}_1F_1^{(\alpha)}(\tfrac{1}{\alpha};\tfrac{m}{\alpha};X) &= {}_1F_1^{(\alpha)}(\tfrac{1}{\alpha};\tfrac{m}{\alpha};X-tI)\cdot e^t, \text{ if } m=3,\ \alpha=2;\\
{}_2F_1^{(\alpha)}(a,b;c;I) &= \frac{\Gamma_m^{(\alpha)}(c)\Gamma_m^{(\alpha)}(c-a-b)}{\Gamma_m^{(\alpha)}(c-a)\Gamma_m^{(\alpha)}(c-b)};\\
{}_2F_1^{(2)}(a,b;c;xI) &= \mathrm{Pf}(A) \quad \text{(see [20] for the definition of } A);\\
{}_2F_1^{(\alpha)}(a,b;c;X) &= {}_2F_1^{(\alpha)}\left(c-a,b;c;-X(I-X)^{-1}\right)\cdot\det(I-X)^{-b}\\
&= {}_2F_1^{(\alpha)}\left(c-a,c-b;c;X\right)\cdot\det(I-X)^{c-a-b};\\
{}_2F_1^{(\alpha)}(a,b;c;X) &= \frac{\Gamma_m^{(\alpha)}(c)\Gamma_m^{(\alpha)}(c-a-b)}{\Gamma_m^{(\alpha)}(c-a)\Gamma_m^{(\alpha)}(c-b)}\cdot{}_2F_1^{(\alpha)}\left(a,b;a+b+1+\tfrac{m-1}{\alpha}-c;I-X\right),\\
&\quad \text{if } a-\tfrac{i-1}{\alpha}\in\mathbb{Z}_{\le 0} \text{ for some } i=1,2,\ldots,m;
\end{aligned}$$

These provide alternative ways of computing the expressions in section 1.2. Albeit mathematically equivalent, their numerical behavior can differ dramatically.

As far as the convergence of the series (1) is concerned, at this time I do not know how to optimally truncate it nor do I know how to automatically detect convergence. While the κ-term does approach zero as $|\kappa|\to\infty$, it need not monotonically decrease; it is unclear how to tell when convergence sets in.

Whereas I do not expect to find an universal answer to these problems (for every p and q and every value of α), I intend to analyze every formula in section 1.2 individually.

1.9. **Efficient sampling of eigenvalues of random matrices.** Say one wanted N samples of the largest (or smallest) eigenvalue of a Wishart (with identity covariance), Laguerre, or Jacobi matrix of size m. A direct and naive approach costs $\mathcal{O}(Nm^3)$. I intend to develop algorithms that cost only $\mathcal{O}(Nm)$ as follows.

The tridiagonal (scaled) β-Laguerre matrix L/β has the same eigenvalue distribution as a Wishart matrix with identity covariance. Thus sampling the largest eigenvalue of L will cost $O(m)$ and N samples will cost $O(Nm)$.

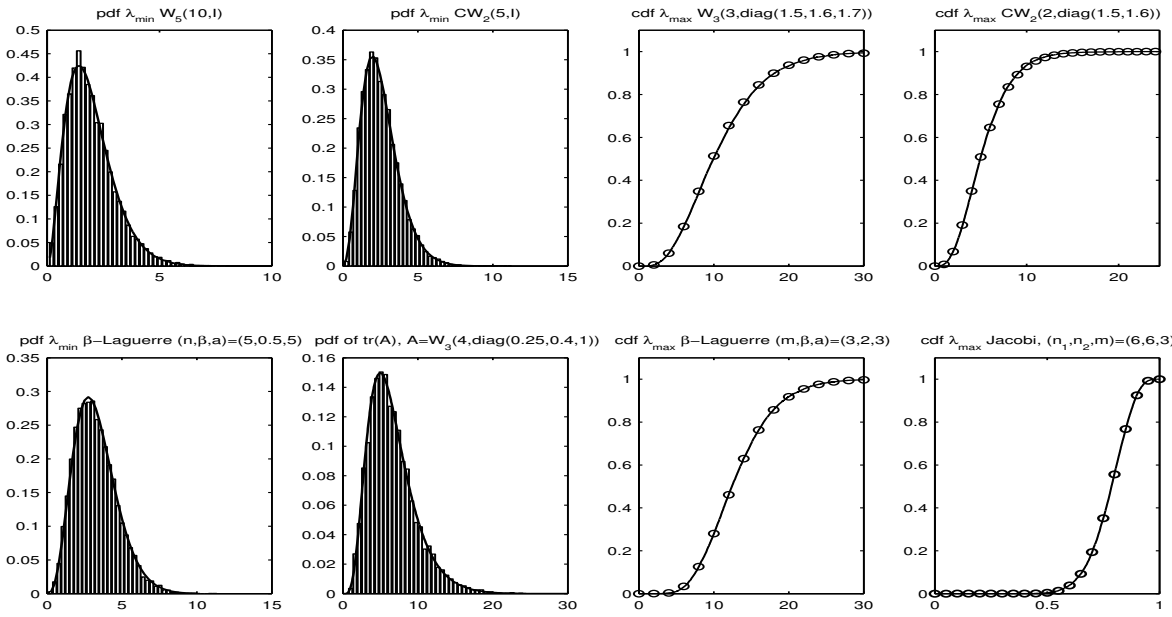

FIGURE 4. Densities and distributions of various eigenvalues of small random matrices computed using current techniques. On the left side, the empirical histogrammed p.d.f.'s are compared with the theoretical predictions in solid line. On the right side, the empirical c.d.f.'s (in solid line) are compared with the theoretical predictions (in "∘" marks). If the research in this proposal is successful, we will have algorithms to produce similar results for matrices of much larger size.

In the Jacobi case, we utilize the very elegant real tridiagonal β-Jacobi matrix C constructed by Sutton [51]. We have $C \equiv Z^T Z$, where

$$C \equiv \begin{bmatrix} c_n & -s_n c'_{n-1} & & \\ & c_{n-1}s'_{n-1} & \ddots & \\ & & \ddots & -s_2 c'_1 \\ & & & c_1 s'_1 \end{bmatrix},$$

with

$$\begin{aligned} c_k &\sim \sqrt{\mathrm{Beta}\left(\tfrac{n_1-m+k}{\alpha}, \tfrac{n_2-m+k}{\alpha}\right)}, \quad s_k = \sqrt{1-c_k^2}; \\ c'_k &\sim \sqrt{\mathrm{Beta}\left(\tfrac{k}{\alpha}, \tfrac{n_1+n_2-2m+k+1}{\alpha}\right)}, \quad s'_k = \sqrt{1-c_k'^2}. \end{aligned}$$

The β-Jacobi matrix has the same eigenvalue distribution as the Jacobi matrix, thus sampling N largest (or smallest) eigenvalues again only costs $O(Nm)$.

1.10. **Possible impact of this research.** The development of the algorithms described here has been identified as a central research problem in a number of practical applications. I expect such algorithms will have an extensive impact in each of those applications in areas ranging from telecommunications to wireless networks to military applications [6, 7, 16, 18, 23, 27, 34, 36, 37, 39, 40, 47, 48].

Our recent algorithms from [33] have already had a practical impact in the area of automatic target classification [23] and have been instrumental in the publications [1, 14].

Our current techniques allow us to compute certain densities and distribution of the extreme eigenvalues (as well as the trace) of small random matrices, see Figure 4. If the research described

here is successful, one would be able to perform the same computations with much larger random matrices. In certain cases, utilizing the Tracy–Widom limits, I expect to derive algorithms that can accurate compute the distributions for *any size* random matrix.

Overall, I plan on developing and analyzing of over 40 different algorithms for computing density, distribution, quantile, and sampling functions for the extreme eigenvalues and the trace of each classical random matrix ensemble in both the real and the complex cases.

2. Other projects

I also plan on continuing work in the following areas:

- *Stable $\mathcal{O}(n^2)$ algorithm for the nonsymmetric tridiagonal eigenvalue problem.*
  The development of such an algorithm is a major topic of research in numerical linear algebra [4]. I am planning on using the LR algorithm [41] with various shifting techniques as means of stabilizing it.
- *Accurate computations with totally nonnegative (TN) matrices.*
  I expect to extend my algorithms [32, 31] to TN matrices of less than full rank. If successful, this will lead to the first algorithm for computing nontrivial Jordan structures accurately. algorithm for computing nontrivial Jordan structures accurately.
- *Research in matrix theory.*
  I plan on establishing a characterization, in the language of bidiagonal decompositions, of what constitutes an *oscillating* system of vectors (i.e., a system of vectors such that jth one has $j-1$ sign changes). These vectors often describe the modes of vibration of mechanical systems. In particular I expect to obtain simple proofs of the Schoenberg's theorem for the variation diminishing property of totally positive matrices, and prove that the columns of the matrix Q of the QR decomposition of a totally positive matrix has the same sign-oscillating properties as the eigenvector matrix of a TP matrix.
- *Collaborations on practical problems.* I plan on continuing collaborations with Vladimir Druskin of Schlumburger Oilfield Research, Inc. in Boston in resolving accuracy issues in their totally positive matrix problems related to circular resistor networks.

References

[1] P.-A. Absil, A. Edelman, and P. Koev. On the largest principal angle between random subspaces. *Linear Algebra Appl.*, 414:288–294, 2006.

[2] T. H. Baker, P. J. Forrester, and P. A. Pearce. Random matrix ensembles with an effective extensive external charge. *J. Phys. A*, 31(29):6087–6101, 1998.

[3] Christopher Bingham. An antipodally symmetric distribution on the sphere. *Ann. Statist.*, 2:1201–1225, 1974.

[4] D. Bini, L. Gemignani, and F. Tisseur. The Ehrlich-Aberth method for the nonsymmetric tridiagonal eigenvalue problem. *SIAM J. Matrix Anal. Appl.*, 27(1):153–175, 2005.

[5] R. W. Butler and A. T. A. Wood. Laplace approximations for hypergeometric functions with matrix argument. *Ann. Statist.*, 30(4):1155–1177, 2002.

[6] G.J. Byers and F. Takawira. Spatially and temporally correlated MIMO channels: modeling and capacity analysis. *IEEE Transactions on Vehicular Technology*, 53:634–643, May 2004.

[7] Pinyuen Chen, G.J. Genello, and M.C. Wicks. Estimating the number of signals in presence of colored noise. In *Radar Conference 2004. Proceedings of the IEEE*, pages 432–437, 26–29 April 2004.

[8] William R. Chen. Table for upper percentage points of the largest root of a determinantal equation with five roots.

[9] William R. Chen. Some new tables of the largest root of a matrix in multivariate analysis: A computer approach from 2 to 6, 2002. Presented at the 2002 American Statistical Association.

[10] A. G. Constantine. Some non-central distribution problems in multivariate analysis. *Ann. Math. Statist.*, 34:1270–1285, 1963.

[11] James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, 19:297–301, 1965.

[12] J. Demmel and P. Koev. Accurate and efficient evaluation of Schur and Jack functions. *Math. Comp.*, 75:223–239, 2006.

[13] I. Dumitriu. *Eigenvalue Statistics for the Beta-Ensembles*. PhD thesis, Massachusetts Institute of Technology, 2003.
[14] Alan Edelman and Brian Sutton. Tails of condition number distributions. SIAM J. Matrix Anal. Appl., accepted for publication, 2005.
[15] P. Forrester. Log-gases and random matrices, 2005. preprint.
[16] H. Gao, P.J. Smith, and M.V. Clark. Theoretical reliability of MMSE linear diversity combining in Rayleigh-fading additive interference channels. *IEEE Transactions on Communications*, 46:666–672, May 1998.
[17] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.
[18] A.J. Grant. Performance analysis of transmit beamforming. *IEEE Transactions on Communications*, 53:738–744, April 2005.
[19] K. I. Gross and D. St. P. Richards. Total positivity, spherical series, and hypergeometric functions of matrix argument. *J. Approx. Theory*, 59(2):224–246, 1989.
[20] Rameshwar D. Gupta and Donald St. P. Richards. Hypergeometric functions of scalar matrix argument are expressible in terms of classical hypergeometric functions. *SIAM J. Math. Anal.*, 16(4):852–858, 1985.
[21] R. Gutiérrez, J. Rodriguez, and A. J. Sáez. Approximation of hypergeometric functions with matricial argument through their development in series of zonal polynomials. *Electron. Trans. Numer. Anal.*, 11:121–130, 2000.
[22] Alan T. James. Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.*, 35:475–501, 1964.
[23] Michael Jeffris. MITRE Corporation. Private communication, 2004.
[24] Kurt Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000.
[25] Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.
[26] Iain M. Johnstone. Notes on the largest canonical correlation, 2004. Work in progress.
[27] M. Kang and M.-S. Alouini. Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems. *IEEE Journal on Selected Areas in Communications*, 21(3):418–431, 4 2003.
[28] Noureddine El Karoui. Tracy–Widom limit for the largest eigenvalue of a large class of complex Wishart matrices. `http://www.arxiv.org/pdf/math.PR/0503109`, 2005.
[29] P. Koev. `http://www-math.mit.edu/~plamen`.
[30] P. Koev. Software for computing the hypergeometric function of a matrix argument. available from [29].
[31] P. Koev. Accurate eigenvalues and SVDs of totally nonnegative matrices. *SIAM J. Matrix Anal. Appl.*, 27(1):1–23, 2005.
[32] P. Koev. Accurate computations with totally nonnegative matrices. submitted to SIAM J. Matrix Anal. Appl., available from [29], 2006.
[33] P. Koev and A. Edelman. The efficient evaluation of the hypergeometric function of a matrix argument. Math. Comp., to appear, available from [29].
[34] C. Lopez-Martinez, E. Pottier, and S.R. Cloude. Statistical assessment of eigenvector-based target decomposition theorems in radar polarimetry. *IEEE Transactions on Geoscience and Remote Sensing*, 43:2058–2074, 2005.
[35] I. G. Macdonald. *Symmetric functions and Hall polynomials*. Oxford University Press, New York, Second edition, 1995.
[36] M.R. McKay and I.B. Collings. Capacity bounds for correlated rician MIMO channels. In *2005 IEEE International Conference on Communications. ICC 2005.*, volume 2, pages 772–776, 16-20 May 2005.
[37] M.R. McKay and I.B. Collings. General capacity bounds for spatially correlated Rician MIMO channels. *IEEE Transactions on Information Theory*, 51:3121–3145, September 2005.
[38] R. J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York, 1982.
[39] A. Ozyildirim and Y. Tanik. Outage probability analysis of a CDMA system with antenna arrays in a correlated Rayleigh environment. In *IEEE Military Communications Conference Proceedings, 1999. MILCOM 1999*, volume 2, pages 939–943, 31 Oct.–3 Nov. 1999.
[40] A. Ozyildirim and Y. Tanik. SIR statistics in antenna arrays in the presence of correlated Rayleigh fading. In *IEEE VTS 50th Vehicular Technology Conference, 1999. VTC 1999 - Fall*, volume 1, pages 67–71, 19-22 September 1999.
[41] B. Parlett. The development and use of methods of LR type. *SIAM Rev.*, 6:275–295, 1964.
[42] Per-Olof Persson. Numerical methods for random matrices, class report for 18.325. Available from `http://math.mit.edu/~persson`, 2002.
[43] Markus Püschel. Cooley-Tukey FFT like algorithms for the DCT. In *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 501–504, 2003.
[44] Markus Püschel and José M. F. Moura. The algebraic approach to the discrete cosine and sine transforms and their fast algorithms. *SIAM J. Comput.*, 32(5):1280–1316, 2003.

[45] T. Ratnarajah, R. Vaillancourt, and M. Alvo. Eigenvalues and condition numbers of complex random matrices. *SIAM J. Matrix Anal. Appl.*, 26(2):441–456, 2004/05.

[46] A. J. Sáez. Software for calculus of zonal polynomials. `http://estio.ujaen.es/Profesores/ajsaez/software.html`, 2004.

[47] Hyundong Shin and Jae Hong Lee. Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole. *IEEE Transactions on Information Theory*, 49:2636–2647, October 2003.

[48] V. Smidl and A. Quinn. Fast variational PCA for functional analysis of dynamic image sequences. In *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003.*, volume 1, pages 555–560, 18-20 September 2003.

[49] Alexander Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108(5-6):1033–1056, 2002.

[50] R. Stanley. Some combinatorial properties of Jack symmetric functions. *Adv. Math.*, 77(1):76–115, 1989.

[51] Brian Sutton. *The stochastic operator approach to random matrix theory*. PhD thesis, Massachusetts Institute of Technology, 2005.

[52] Craig A. Tracy and Harold Widom. The distribution of the largest eigenvalue in the Gaussian ensembles: $\beta = 1, 2, 4$. In *Calogero-Moser-Sutherland models (Montréal, QC, 1997)*, CRM Ser. Math. Phys., pages 461–472. Springer, New York, 2000.