On the universality of the probability distribution of the product $B^{-1}X$ of random matrices

2004 J. Phys. A: Math. Gen. 37 6823

(http://iopscience.iop.org/0305-4470/37/26/015)

J. Phys. A: Math. Gen. **37** (2004) 6823–6835 PII: S0305-4470(04)76469-0

On the universality of the probability distribution of the product $B^{-1}X$ of random matrices

Joshua Feinberg

Physics Department, University of Haifa at Oranim, Tivon 36006, Israel
and
Physics Department, Technion, Israel Institute of Technology, Haifa 32000, Israel

E-mail: joshua@physics.technion.ac.il

Received 22 February 2004
Published 16 June 2004
Online at stacks.iop.org/JPhysA/37/6823
doi:10.1088/0305-4470/37/26/015

Abstract

Consider random matrices A, of dimension $m \times (m+n)$, drawn from an ensemble with probability density $f(\operatorname{tr} AA^\dagger)$, with $f(x)$ a given appropriate function. Break $A = (B, X)$ into an $m \times m$ block B and the complementary $m \times n$ block X, and define the random matrix $Z = B^{-1}X$. We calculate the probability density function $P(Z)$ of the random matrix Z and find that it is a universal function, independent of $f(x)$. The universal probability distribution $P(Z)$ is a spherically symmetric matrix-variate t-distribution. Universality of $P(Z)$ is, essentially, a consequence of rotational invariance of the probability ensembles we study. As an application, we study the distribution of solutions of systems of linear equations with random coefficients, and extend a classic result due to Girko.

PACS numbers: 02.50.Cw, 02.50.Sk
Mathematics Subject Classification: 15A52, 60E05, 62H10, 34F05

1. Introduction

In this paper we will address the issue of universality of the probability density function (PDF) of the product $B^{-1}X$ of real and complex random matrices.

In order to motivate our discussion, before delving into random matrix theory, let us discuss a simpler problem. Thus, consider the random variables x and y drawn from the normal distribution

$$G(x, y) = \frac{1}{2\pi\sigma^2}\, \mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{1.1}$$

Define the random variable $z = x/y$. Obviously, its PDF is independent of the width σ of $G(x, y)$, and it is a straightforward exercise to show that

$$P(z) = \frac{1}{\pi}\frac{1}{1+z^2} \tag{1.2}$$

i.e., the standard Cauchy distribution.

A slightly more interesting generalization of (1.1) is to consider the family of joint probability density (JPD) functions of the form

$$G(x, y) = f(x^2 + y^2) \tag{1.3}$$

where $f(u)$ is a given appropriate PDF, subject to the normalization condition

$$\int_0^\infty f(u)\,\mathrm{d}u = \frac{1}{\pi}. \tag{1.4}$$

A straightforward calculation of the PDF of $z = x/y$ leads again to (1.2). Thus, the random variable $z = x/y$ is distributed according to (1.2), *independently* of the function $f(u)$. In other words, (1.2) is a *universal* probability density function[1]. $P(z)$ is universal, essentially, due to rotational invariance of (1.3). More generally, $P(z)$ must be independent, of course, of any common scale of the distribution functions of x and y.

We will now show that an analogue of this universal behaviour exists in random matrix theory. Our interest in this problem stems from the recent application of random matrix theory made in [1] to calculate the complexity of an analogue computation process [2], which solves linear programming problems.

2. The universal probability distribution of the product $B^{-1}X$ of real random matrices

Consider a real $m \times (m+n)$ random matrix A with entries $A_{i\alpha}$ $(i = 1, \ldots m;\ \alpha = 1, \ldots m+n)$. We take the JPD for the $m(m+n)$ entries of A as

$$G(A) = f(\mathrm{tr}\, AA^T) = f\left(\sum_{i,\alpha} A_{i\alpha}^2\right) \tag{2.1}$$

with $f(u)$ a given appropriate PDF. From[2]

$$\int G(A)\,\mathrm{d}A = 1 \tag{2.2}$$

we see that $f(u)$ is subject to the normalization condition

$$\int_0^\infty u^{\frac{m(m+n)}{2}-1} f(u)\,\mathrm{d}u = \frac{2}{S_{m(m+n)}} \tag{2.3}$$

where

$$S_d = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \tag{2.4}$$

is the surface area of the unit sphere embedded in d dimensions. This implies, in particular, that $f(u)$ must decay *faster* than $u^{-m(m+n)/2}$ as $u \to \infty$, and also, that if $f(u)$ blows up

[1] We can generalize (1.3) somewhat further, by considering circularly asymmetric distributions $G(x, y) = f(ax^2 + by^2)$ (with $a, b > 0$ of course, and the r.h.s. of (1.4) changed to $\sqrt{ab}/\pi$), rendering (1.2) a Cauchy distribution of width $\sqrt{b/a}$, independently of the function $f(u)$.

[2] We use the ordinary Cartesian measure $\mathrm{d}A = \mathrm{d}^{m(m+n)}A = \prod_{i\alpha} \mathrm{d}A_{i\alpha}$. Similarly, $\mathrm{d}B = \mathrm{d}^{m^2}B$ and $\mathrm{d}X = \mathrm{d}^{mn}X$ for the matrices B and X in (2.6) and (2.18).

as $u \to 0+$, its singularity must be *weaker* than $u^{-m(m+n)/2}$. In other words, $f(u)$ must be subjected to the asymptotic behaviour

$$u^{m(m+n)/2} f(u) \to 0 \tag{2.5}$$

both as $u \to 0$ and $u \to \infty$.

We now choose m columns out of the $m+n$ columns of A, and pack them into an $m \times m$ matrix B (with entries B_{ij}). Similarly, we pack the remaining n columns of A into an $m \times n$ matrix X (with entries X_{ip}). This defines a partition

$$A \to (B, X) \tag{2.6}$$

of the columns of A.

The index conventions throughout this paper are such that indices

$$\begin{array}{lll} i, j, \ldots & \text{range over} & 1, 2, \ldots, m \\ p, q, \ldots & \text{range over} & 1, 2, \ldots, n \end{array} \tag{2.7}$$

and α ranges over $1, 2, \ldots, m+n$.

In this notation we have $\operatorname{tr} AA^T = \sum_{i,j} B_{ij}^2 + \sum_{i,p} X_{ip}^2 = \operatorname{tr} BB^T + \operatorname{tr} XX^T$, and thus (2.1) reads

$$G(B, X) = f(\operatorname{tr} BB^T + \operatorname{tr} XX^T). \tag{2.8}$$

We now define the random matrix $Z = B^{-1}X$. Our goal is to calculate the JPD $P(Z)$ for the mn entries of Z. $P(Z)$ is clearly independent of the particular partitioning (2.6) of A, since $G(B, X)$ is manifestly independent of that partitioning. The main result in this section is stated as follows:

Theorem 2.1. *The JPD for the mn entries of the real random matrix $Z = B^{-1}X$ is independent of the function $f(u)$ and is given by the universal function*

$$P(Z) = \frac{C}{[\det(\mathbb{1} + ZZ^T)]^{\frac{m+n}{2}}} \tag{2.9}$$

where C is a normalization constant.

Remark 2.1. The probability density function (2.9) is a special (spherically symmetric) case of the so-called[3] matrix variate t-distributions [3, 4]: the $m \times n$ random matrix Z is said to have a matrix variate t-distribution with parameters M, Σ, Ω and q (a fact we denote by $Z \sim T_{n,m}(q, M, \Sigma, \Omega)$) if its PDF is given by

$$D(\det\Sigma)^{-\frac{n}{2}}(\det\Omega)^{-\frac{m}{2}}[\det(\mathbb{1}_m + \Sigma^{-1}(Z-M)\Omega^{-1}(Z-M)^T)]^{-\frac{1}{2}(m+n+q-1)} \tag{2.10}$$

where M, Σ and Ω are fixed real matrices of dimensions $m \times n$, $m \times m$ and $n \times n$, respectively. Σ and Ω are positive definite, and $q > 0$. The normalization coefficient is

$$D = \frac{1}{\pi^{\frac{mn}{2}}} \frac{\prod_{j=1}^{n} \Gamma\left(\frac{m+n+q-j}{2}\right)}{\prod_{j=1}^{n} \Gamma\left(\frac{n+q-j}{2}\right)}. \tag{2.11}$$

It arises in the theory of matrix variate distributions as the PDF of a random matrix which is the product of the inverse square root of a certain Wishart-distributed matrix and a matrix taken from a normal distribution, and by shifting this product by M, as described in [3, 4]. Our universal distribution (2.9) corresponds to setting $M = 0$, $\Sigma = \mathbb{1}_m$, $\Omega = \mathbb{1}_n$ and $q = 1$ in (2.10) and (2.11).

[3] Our notation in remark (2.1) is slightly different from the notation used in [4]. In particular, we interchanged their Σ and Ω, and also denoted their $(T-M)^T$ by $Z-M$ here. Finally, we applied the identity $\det(\mathbb{1}+AB) = \det(\mathbb{1}+BA)$ to arrive after all these interchanges from their equation (4.2.1) at our equation (2.10).

Remark 2.2. It would be interesting to distort the parent JPD (2.1) into a non-isotropic distribution and see if the generic matrix variate t-distribution (2.10) arises as the corresponding universal probability distribution function in this case.

To prove theorem (2.1), we need

Lemma 2.1. *Given a function $f(u)$, subjected to (2.3), the integral*

$$I = \int \mathrm{d}B\, f(\mathrm{tr}\, BB^T)|\det B|^n \tag{2.12}$$

converges, and is independent of the particular function $f(u)$.

Remark 2.3. A qualitative and simple argument, showing the convergence of (2.12), is that the measure $\mathrm{d}\mu(B) = \mathrm{d}B|\det B|^n$ scales as $\mathrm{d}\mu(tB) = t^{m(m+n)}\,\mathrm{d}\mu(B)$, and thus has the same scaling property as $\mathrm{d}A$ in (2.2), indicating that the integral (2.12) converges, in view of (2.5). To see that I is independent of $f(u)$ one has to work harder.

Proof. We would like first to integrate over the rotational degrees of freedom in $\mathrm{d}B$. Any real $m \times m$ matrix B may be decomposed as [5, 6]

$$B = \mathcal{O}_1 \Omega \mathcal{O}_2 \tag{2.13}$$

where $\mathcal{O}_{1,2} \in \mathcal{O}(m)$, the group of $m \times m$ orthogonal matrices and $\Omega = \mathrm{Diag}(\omega_1, \ldots, \omega_m)$, where $\omega_1, \ldots, \omega_m$ are the singular values of B. Under this decomposition we may write the measure $\mathrm{d}B$ as [5, 6]

$$\mathrm{d}B = \mathrm{d}\mu(\mathcal{O}_1)\,\mathrm{d}\mu(\mathcal{O}_2) \prod_{i<j} \left|\omega_i^2 - \omega_j^2\right| \mathrm{d}^m\omega \tag{2.14}$$

where $\mathrm{d}\mu(\mathcal{O}_{1,2})$ are Haar measures over the $\mathcal{O}(m)$ group manifold. The measure $\mathrm{d}B$ is manifestly invariant under actions of the orthogonal group $\mathcal{O}(m)$

$$\mathrm{d}B = \mathrm{d}(B\mathcal{O}) = \mathrm{d}(\mathcal{O}'B) \qquad \mathcal{O}, \mathcal{O}' \in \mathcal{O}(m) \tag{2.15}$$

as should have been expected to begin with.

Remark 2.4. Note that the decomposition (2.13) is not unique, since $\mathcal{O}_1\mathcal{D}$ and $\mathcal{D}\mathcal{O}_2$, with $\mathcal{D}$ being any of the 2^m diagonal matrices $\mathrm{Diag}(\pm 1, \ldots, \pm 1)$, is an equally good pair of orthogonal matrices to be used in (2.13). Thus, as $\mathcal{O}_1$ and $\mathcal{O}_2$ sweep independently over the group $\mathcal{O}(m)$, the measure (2.14) over counts B matrices. This problem can be easily rectified by appropriately normalizing the volume $\mathcal{V}_m = \int \mathrm{d}\mu(\mathcal{O}_1)\,\mathrm{d}\mu(\mathcal{O}_2)$. One can show[4] that the correct normalization of the volume is

$$\mathcal{V}_m = \frac{\pi^{\frac{m(m+1)}{2}}}{2^m \prod_{j=1}^m \Gamma\left(1 + \frac{j}{2}\right)\Gamma\left(\frac{j}{2}\right)}. \tag{2.16}$$

Let us now turn to (2.12). The integrals over the orthogonal group in (2.12) clearly factor out, and we obtain

$$I = \mathcal{V}_m \int_{-\infty}^{\infty} \prod_{i=1}^m \mathrm{d}\omega_i \prod_{j<k} \left|\omega_j^2 - \omega_k^2\right| \left(\prod_{i=1}^m \omega_i\right)^n f\left(\sum_{i=1}^m \omega_i^2\right). \tag{2.17}$$

[4] One simple way to establish (2.16), is to calculate $\int \mathrm{d}B \exp -\frac{1}{2}\mathrm{tr}\, B^T B = (2\pi)^{\frac{m^2}{2}} = \mathcal{V}_m \int_{-\infty}^{\infty} \mathrm{d}^m\omega \prod_{i<j} |\omega_i^2 - \omega_j^2| \exp -\frac{1}{2}\sum_i \omega_i^2$. The last integral is a known Selberg type integral [7].

Finally, we change the integration variables in (2.17) to the polar coordinates associated with the ω_i. The angular part of that integral is fixed only by dimensionality and by the factor $\prod_{j<k} |\omega_j^2 - \omega_k^2| \left(\prod_{i=1}^m \omega_i\right)^n$, and thus is *independent* of the function $f(u)$.

To prove that $I < \infty$ we need only consider integration over the radius $r^2 = \sum_{i=1}^m \omega_i^2$, since integration over the angles obviously produces a finite result. Using (2.3), we find that the radial integral in question is

$$\int_0^\infty \mathrm{d}r\, r^{m^2+nm-1} f(r^2) = \frac{1}{2}\int_0^\infty u^{\frac{m(m+n)}{2}-1} f(u)\, \mathrm{d}u = \frac{2}{S_{m(m+n)}}$$

independently of $f(u)$. □

We are now ready to prove theorem (2.1):

Proof. By definition[5]

$$P(Z) = \int \mathrm{d}B\, \mathrm{d}X\, f(\operatorname{tr} BB^T + \operatorname{tr} XX^T)\delta(Z - B^{-1}X)$$
$$= \int \mathrm{d}B\, \mathrm{d}X\, f(\operatorname{tr} BB^T + \operatorname{tr} XX^T)|\det B|^n \delta(X - BZ). \tag{2.18}$$

Integration over X gives

$$P(Z) = \int \mathrm{d}B\, f(\operatorname{tr} BB^T + \operatorname{tr} BZZ^T B^T)|\det B|^n$$
$$= \int \mathrm{d}B\, f[\operatorname{tr} B(\mathbb{1} + ZZ^T)B^T]|\det B|^n. \tag{2.19}$$

The $m \times m$ symmetric matrix $\mathbb{1} + ZZ^T$ can be diagonalized as $\mathbb{1} + ZZ^T = \mathcal{O}\Lambda\mathcal{O}^T$, where $\mathcal{O}$ is an orthogonal matrix, and $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_m)$ is the corresponding diagonal form. Obviously, all $\lambda_i \geqslant 1$, since ZZ^T is positive definite. Substituting this diagonal form into (2.19) we obtain

$$P(Z) = \int \mathrm{d}B\, f(\operatorname{tr} B\mathcal{O}\Lambda\mathcal{O}^T B^T)|\det B|^n.$$

From the invariance of the determinant $|\det B\mathcal{O}| = |\det B|$ and of the volume element $\mathrm{d}(B\mathcal{O}) = \mathrm{d}B$ under orthogonal transformations we have

$$P(Z) = \int \mathrm{d}B\, f(\operatorname{tr} B\Lambda B^T)|\det B|^n. \tag{2.20}$$

Let us now rescale B as $\tilde{B} = B\sqrt{\Lambda}$. Thus,

$$\det \tilde{B} = \sqrt{\det \Lambda}\, \det B \qquad \text{and} \qquad \mathrm{d}\tilde{B} = (\det \Lambda)^{\frac{m}{2}}\, \mathrm{d}B. \tag{2.21}$$

Finally, substituting (2.21) in (2.20) we obtain

$$P(Z) = \int \frac{\mathrm{d}\tilde{B}}{(\det \Lambda)^{\frac{m}{2}}} f(\operatorname{tr} \tilde{B}\tilde{B}^T)\left(\frac{|\det \tilde{B}|}{(\det \Lambda)^{\frac{1}{2}}}\right)^n$$
$$= \frac{C}{(\det \Lambda)^{(m+n)/2}} = \frac{C}{[\det(\mathbb{1} + ZZ^T)]^{(m+n)/2}} \tag{2.22}$$

where C is the normalization constant

$$C = \int \mathrm{d}B\, f(\operatorname{tr} BB^T)|\det B|^n \tag{2.23}$$

[5] Our notation is such that $\delta(X) = \prod_{i=1}^m \prod_{p=1}^n \delta(X_{ip}) = \prod_{p=1}^n \delta^{(m)}(X_p)$, X_p being the pth column of X.

rendering

$$\int P(Z)\,\mathrm{d}Z = 1. \tag{2.24}$$

C is nothing but the integral (2.12). Thus, according to lemma (2.1), $C < \infty$ and is also independent of the function $f(u)$. □

Remark 2.5. The JPD $P(Z)$ in (2.9) is manifestly a symmetric function only of the eigenvalues of ZZ^T, and thus, a symmetric function only of the singular values of Z.

Remark 2.6. From the normalization condition (2.24) we obtain an alternative expression for the normalization constant (2.23) as

$$\frac{1}{C} = \frac{1}{\int \mathrm{d}B\, f(\mathrm{tr}\, BB^T)|\det B|^n} = \int \frac{\mathrm{d}Z}{[\det(\mathbb{1} + ZZ^T)]^{(m+n)/2}} \tag{2.25}$$

which is manifestly independent of the particular function $f(u)$, in accordance with lemma (2.1). The integral over the matrix Z can be reduced to a multiple integral of the Selberg type [5, 7] over the singular values of the matrix Z, which can be carried out explicitly

$$\int \frac{\mathrm{d}Z}{[\det(\mathbb{1} + ZZ^T)]^{(m+n)/2}} = \pi^{\frac{mn}{2}} \prod_{j=1}^{n} \frac{\Gamma\left(\frac{j}{2}\right)}{\Gamma\left(\frac{m+j}{2}\right)}. \tag{2.26}$$

For particular choices of the function $f(u)$, we can use (2.25) to derive explicit integration formulae. For example, the function

$$f(u) = \frac{\mathrm{e}^{-u}}{\pi^{m(m+n)/2}} \tag{2.27}$$

(i.e., the entries $A_{i\alpha}$ in (2.1) are IID according to a normal distribution of variance $1/2$) satisfies (2.3). Thus, we obtain from (2.25) that

$$\int \mathrm{d}B\, \mathrm{e}^{-\mathrm{tr}\, BB^T} |\det B|^n = \pi^{\frac{m^2}{2}} \prod_{j=1}^{n} \frac{\Gamma\left(\frac{m+j}{2}\right)}{\Gamma\left(\frac{j}{2}\right)}. \tag{2.28}$$

Note that the integral on the left-hand side of (2.28) can also be reduced to a multiple integral of the Selberg type (this time, over the singular values of B), which can be carried out explicitly. The result is

$$\pi^{\frac{m^2}{2}} \prod_{j=1}^{m} \frac{\Gamma\left(\frac{n+j}{2}\right)}{\Gamma\left(\frac{j}{2}\right)}.$$

Since this must coincide with (2.28), we obtain the identity

$$\prod_{j=1}^{m} \frac{\Gamma\left(\frac{n+j}{2}\right)}{\Gamma\left(\frac{j}{2}\right)} = \prod_{j=1}^{n} \frac{\Gamma\left(\frac{m+j}{2}\right)}{\Gamma\left(\frac{j}{2}\right)}. \tag{2.29}$$

Example 1. For $n = 1$, i.e., the case where X and Z are m-dimensional vectors, (2.9) simplifies into the m-dimensional Cauchy distribution

$$P(Z) = \frac{C}{(1 + Z^T Z)^{(m+1)/2}}. \tag{2.30}$$

This is so because for $n = 1$, the matrix ZZ^T has $m-1$ eigenvalues equal to 0, that correspond to the $(m-1)$-dimensional subspace of vectors orthogonal to Z, and one eigenvalue equal to $Z^T Z$. Thus, $\det(\mathbb{1} + ZZ^T) = 1 + Z^T Z$. Equation (2.30) then follows by substituting this determinant into (2.22).

3. The universal probability distribution of the product $B^{-1}X$ of complex random matrices

The results of the previous section are readily generalized to complex random matrices. One has only to count the number of independent real integration variables correctly. In what follows we will use the notation defined in the previous section (unless specified otherwise explicitly). Thus, consider a complex $m \times (m+n)$ random matrix A with entries $A_{i\alpha}$ $(i = 1, \dots m; \alpha = 1, \dots m+n)$. We take the JPD for the $m(m+n)$ entries of A as

$$G(A) = f(\operatorname{tr} AA^\dagger) = f\left(\sum_{i,\alpha} |A_{i\alpha}|^2\right) \tag{3.1}$$

with $f(u)$ a given appropriate PDF. From[6]

$$\int G(A)\, \mathrm{d}A = 1 \tag{3.2}$$

we see that $f(u)$ is subjected to the normalization condition

$$\int_0^\infty u^{m(m+n)-1} f(u)\, \mathrm{d}u = \frac{2}{S_{2m(m+n)}}. \tag{3.3}$$

This implies that $f(u)$ must be subjected to the asymptotic behaviour

$$u^{m(m+n)} f(u) \to 0 \tag{3.4}$$

both as $u \to 0$ and $u \to \infty$.

As in the previous section, we choose a partition

$$A \to (B, X) \tag{3.5}$$

of the columns of A. Thus, $\operatorname{tr} AA^\dagger = \sum_{i,j} |B_{ij}|^2 + \sum_{i,p} |X_{ip}|^2 = \operatorname{tr} BB^\dagger + \operatorname{tr} XX^\dagger$, and thus (3.1) reads

$$G(B, X) = f(\operatorname{tr} BB^\dagger + \operatorname{tr} XX^\dagger). \tag{3.6}$$

We now define the random matrix $Z = B^{-1}X$. Our goal is to calculate the JPD $P(Z)$ for the mn entries of Z. The main result in this section is stated as follows:

Theorem 3.1. *The JPD for the mn entries of the complex random matrix $Z = B^{-1}X$ is independent of $f(u)$ and is given by the universal function*

$$P(Z) = \frac{C}{[\det(1 + ZZ^\dagger)]^{m+n}} \tag{3.7}$$

where C is the normalization constant (3.12).

Proof. The proof proceeds in a similar manner to the proof of theorem (2.1). The only important difference is that now $\delta(X) = \prod_{i=1}^m \prod_{p=1}^n \delta(\operatorname{Re} X_{ip})\delta(\operatorname{Im} X_{ip}) = \prod_{p=1}^n \delta^{(2m)}(X_p)$, X_p being the pth column of X. One obtains

$$\begin{aligned} P(Z) &= \int \mathrm{d}B\, \mathrm{d}X\, f(\operatorname{tr} BB^\dagger + \operatorname{tr} XX^\dagger)\delta(Z - B^{-1}X) \\ &= \int \mathrm{d}B\, f[\operatorname{tr} B(\mathbb{1} + ZZ^\dagger)B^\dagger]|\det B|^{2n} \end{aligned} \tag{3.8}$$

where we have integrated over X.

[6] We use the Cartesian measure $\mathrm{d}A = \mathrm{d}^{2m(m+n)}A = \prod_{i\alpha} \mathrm{d}\operatorname{Re} A_{i\alpha}\, \mathrm{d}\operatorname{Im} A_{i\alpha}$, with analogous definitions for $\mathrm{d}B$ and $\mathrm{d}X$ below.

The $m \times m$ complex Hermitean matrix $\mathbb{1} + ZZ^\dagger$ can be diagonalized as $\mathbb{1} + ZZ^\dagger = \mathcal{U}\Lambda\mathcal{U}^\dagger$, where $\mathcal{U}$ is a unitary matrix, and $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_m)$ is the corresponding diagonal form. Obviously, all $\lambda_i \geqslant 1$, since $ZZ^\dagger$ is positive definite. Substituting this diagonal form into (3.8) we obtain

$$P(Z) = \int \mathrm{d}B\ f(\mathrm{tr}\, B\mathcal{U}\Lambda\mathcal{U}^\dagger B^\dagger)|\det B|^{2n} = \int \mathrm{d}B\ f(\mathrm{tr}\, B\Lambda B^\dagger)|\det B|^{2n} \tag{3.9}$$

where we used the invariance of the determinant $|\det B\mathcal{U}| = |\det B|$ and the invariance of the volume element $\mathrm{d}(B\mathcal{U}) = \mathrm{d}B$ under unitary transformations.

As in the previous section we now rescale B as $\tilde{B} = B\sqrt{\Lambda}$. Thus,

$$\det \tilde{B} = \sqrt{\det \Lambda}\, \det B \qquad \text{and} \qquad \mathrm{d}\tilde{B} = (\det \Lambda)^m\, \mathrm{d}B. \tag{3.10}$$

Finally, substituting (3.10) in (3.9) we obtain

$$\begin{aligned} P(Z) &= \int \frac{\mathrm{d}\tilde{B}}{(\det \Lambda)^m} f(\mathrm{tr}\, \tilde{B}\tilde{B}^\dagger) \left(\frac{|\det \tilde{B}|}{(\det \Lambda)^{\frac{1}{2}}}\right)^{2n} \\ &= \frac{C}{(\det \Lambda)^{m+n}} = \frac{C}{[\det(\mathbb{1} + ZZ^\dagger)]^{m+n}} \end{aligned} \tag{3.11}$$

where C is the normalization constant

$$C = \int \mathrm{d}B\ f(\mathrm{tr}\, BB^\dagger)|\det B|^{2n} \tag{3.12}$$

rendering

$$\int P(Z)\, \mathrm{d}Z = 1. \tag{3.13}$$

Finally, one can show, in a manner analogous to lemma (2.1), that $C < \infty$ and that it is independent of the particular function $f(u)$. □

There are obvious analogues to the remarks made in the previous section, which follow from theorem (3.1), which we will not write down explicitly.

4. More on the distribution of solutions of systems of linear equations with random coefficients: extension of a result due to Girko

The methods of the previous sections may be applied in studying the distribution of solutions of systems of linear equations with random coefficients. For concreteness, let us concentrate on real linear systems in real variables.

Consider a system of m real linear equations

$$\sum_{\alpha=1}^{m+n} A_{i\alpha}\xi_\alpha = b_i \qquad i = 1, \ldots m \tag{4.1}$$

in the $m + n$ real variables ξ_α. With no loss of generality, we will treat the first m components of the vector ξ as unknowns, and the remaining n components of ξ as given parameters. Thus, we split

$$\xi = \begin{pmatrix} z \\ u \end{pmatrix} \tag{4.2}$$

where z is the vector of unknowns $z_i = \xi_i$ $(i = 1, \ldots m)$ and u is the vector of parameters $u_p = \xi_{p+m}$ $(p = 1, \ldots n)$. Similarly, we split the matrix of coefficients

$$A = (B, X)$$

where the $m \times m$ matrix B (with entries B_{ij}) and the $m \times n$ matrix X (with entries X_{ip}) were defined in (2.6). Thus, we may rewrite (4.1) explicitly as a system for the z_i:

$$Bz = b - Xu. \tag{4.3}$$

If we consider an ensemble of systems (4.1), in which A and b are drawn according to some probability law, the unknowns z_i become random variables, which depend on the parameters u_p. Girko proved [8] the following theorem for a particular family of such ensembles:

Theorem 4.1. *(Girko) If the random variables $A_{i\alpha}$ and b_i $(i = 1, \ldots m; \alpha = 1, \ldots m + n)$ are independent, identically distributed variables, having a stable distribution law with the characteristic function*

$$g(t; \alpha, c) = \mathrm{e}^{-c|t|^\alpha} \qquad 0 < \alpha \leqslant 2; c > 0 \tag{4.4}$$

then the random variables z_i $(i = 1, \ldots m)$ are identically distributed with the probability density function

$$p(\zeta; \alpha, \beta) = \frac{2}{\beta}\int_0^\infty r\rho\left(\frac{r\zeta}{\beta}; \alpha\right)\rho(r; \alpha)\,\mathrm{d}r \tag{4.5}$$

where $\rho(r; \alpha)$ is the probability density of the postulated stable distribution, and

$$\beta = \left(1 + \sum_{p=1}^{n} |u_p|^\alpha\right)^{\frac{1}{\alpha}}. \tag{4.6}$$

The ratios z_i/z_j $(i \neq j; i, j = 1, \ldots m)$ have the density $p(r; \alpha, 1)$.

In the special case $\alpha = 2$ in (4.4), the random variables $A_{i\alpha}$ and b_i are normally distributed. For this case Girko obtained

Corollary 4.1. *If, under the conditions of theorem (4.1), $\alpha = 2$, then*

$$p(\zeta; 2, \beta) = \frac{\beta}{\pi(\zeta^2 + \beta^2)} \tag{4.7}$$

i.e., ζ follows a Cauchy distribution of width β.

When $\alpha = 2$, the JPD of the $A_{i\alpha}$ and the b_i is

$$G(A, b) = (2\pi\sigma^2)^{\frac{m(m+n+1)}{2}}\mathrm{e}^{-\frac{1}{2\sigma^2}(\mathrm{tr}\, A^T A + b^T b)} \tag{4.8}$$

which is a special case of the JPD we have discussed in the previous sections. Thus, in the spirit of the discussion in the previous sections, we will study systems of linear equations (4.1) with random coefficients A and inhomogeneous terms b with JPDs of the form

$$G(A, b) = f(\mathrm{tr}\, A^T A + b^T b) \tag{4.9}$$

with $f(u)$ a given appropriate PDF subjected to the normalization condition

$$\int_0^\infty u^{\frac{m(m+n+1)}{2}-1} f(u)\,\mathrm{d}u = \frac{2}{S_{m(m+n+1)}}. \tag{4.10}$$

Our goal is to calculate the JPD $P(z; u)$ for the m unknowns z_i. We summarize our main result in this section as

Theorem 4.2. *If the random variables $A_{i\alpha}$ and b_i $(i = 1, \ldots m; \alpha = 1, \ldots m + n)$ are distributed with a JPD given by (4.9), with $f(u)$ being any appropriate probability density* *function subjected to (4.10), then the random variables* z_i $(i = 1, \ldots m)$ *are distributed with the universal JPD function*

$$P(z; u) = C\frac{\beta}{(\beta^2 + z^T z)^{\frac{m+1}{2}}} \tag{4.11}$$

independently of the function $f(u)$*, where*

$$\beta = \sqrt{1 + u^T u} \tag{4.12}$$

and C is a normalization constant given by

$$C = \int \mathrm{d}A|\det B| f(\operatorname{tr} A^T A). \tag{4.13}$$

Remark 4.1. Note that theorem (4.2) generalizes the case $\alpha = 2$ of Girko's result, theorem (4.1), from the particular $f(u) \sim \mathrm{e}^{-u}$ to a whole class of probability densities $f(u)$, and moreover, it determines for this class of distributions the (universal) JPD of the z_i, and not only the distribution of a single component. Thus, it is an interesting question whether Girko's result could be generalized also to other ensembles of systems of linear equations as well.

Proof. By definition, from (4.3),

$$\begin{aligned} P(z; u) &= \int \mathrm{d}A\, \mathrm{d}b\, f(\operatorname{tr} A^T A + b^T b)\delta(z - B^{-1}(b - Xu)) \\ &= \int \mathrm{d}A\, \mathrm{d}b\, f(\operatorname{tr} A^T A + b^T b)|\det B|\delta(Bz + Xu - b) \\ &= \int \mathrm{d}A\, f[\operatorname{tr} A^T A + (A\xi)^T (A\xi)]|\det B| \end{aligned} \tag{4.14}$$

where in the last step we integrated over the m-dimensional vector b and used $Bz + Xu = A\xi$.

The last expression in (4.14) is manifestly invariant under $\mathcal{O}(m) \times \mathcal{O}(n)$ orthogonal transformations

$$P(\mathcal{O}_1 z; \mathcal{O}_2 u) = P(z; u) \qquad \mathcal{O}_1 \in \mathcal{O}(m), \mathcal{O}_2 \in \mathcal{O}(n) \tag{4.15}$$

due to the invariance of the measure $\mathrm{d}A = \mathrm{d}B\, \mathrm{d}X = \mathrm{d}(B\mathcal{O}_1)\mathrm{d}(X\mathcal{O}_2)$, the invariance of the determinant $|\det B| = |\det B\mathcal{O}_1|$, and the invariance of the trace $\operatorname{tr} A^T A = \operatorname{tr}(B\mathcal{O}_1)^T (B\mathcal{O}_1) + \operatorname{tr}(X\mathcal{O}_2)^T (X\mathcal{O}_2)$. With this symmetry at our disposal, we may simplify the calculation of $P(z; u)$ by rotating the vectors z and u into fixed convenient directions, e.g., into the directions in which only z_1 and u_1 do not vanish

$$z_i^{(0)} = z_0 \delta_{i1} \qquad u_p^{(0)} = u_0 \delta_{p1}. \tag{4.16}$$

with

$$z_0 = (z^T z)^{\frac{1}{2}} \qquad u_0 = (u^T u)^{\frac{1}{2}}. \tag{4.17}$$

Thus, we obtain

$$\begin{aligned} P(z; u) &= \int \mathrm{d}A\, f[\operatorname{tr} A^T A + (Bz^{(0)} + Xu^{(0)})^T (Bz^{(0)} + Xu^{(0)})]|\det B| \\ &= \int \mathrm{d}B\, \mathrm{d}X\, f(S)|\det B| \end{aligned} \tag{4.18}$$

where

$$S = \sum_{j=2}^{m} B_j^T B_j + \sum_{p=2}^{n} X_p^T X_p + \left(1 + z_0^2\right) B_1^T B_1 + 2u_0 z_0 B_1^T X_1 + \left(1 + u_0^2\right) X_1^T X_1 \tag{4.19}$$

in which B_i is the ith column of B, and X_p is the pth column of X.

The bilinear form involving B_1 and X_1 in (4.19) may be diagonalized as

$$(1 + \xi^T \xi) \left(\frac{z_0 B_1 + u_0 X_1}{\sqrt{\xi^T \xi}} \right)^2 + \left(\frac{u_0 B_1 - z_0 X_1}{\sqrt{\xi^T \xi}} \right)^2 \tag{4.20}$$

where we have used $z_0^2 + u_0^2 = \xi^T \xi$. We now perform a rotation in the B_1–X_1 plane, followed by a scale transformation of the first term in (4.20), thus defining

$$B_1' = (1 + \xi^T \xi)^{\frac{1}{2}} \frac{z_0 B_1 + u_0 X_1}{\sqrt{\xi^T \xi}} \qquad X_1' = \frac{u_0 B_1 - z_0 X_1}{\sqrt{\xi^T \xi}} \tag{4.21}$$

such that $\mathrm{d}^m B_1' \, \mathrm{d}^m X_1' = (1 + \xi^T \xi)^{\frac{m}{2}} \, \mathrm{d}^m B_1 \, \mathrm{d}^m X_1$. We will also need the inverse transformation for B_1

$$B_1(B_1', X_1') = \frac{1}{\sqrt{\xi^T \xi}} \left(\frac{z_0 B_1'}{(1 + \xi^T \xi)^{\frac{1}{2}}} + u_0 X_1' \right) \tag{4.22}$$

in order to express the matrix B in terms of the primed column vectors

$$\tilde{B} = (B_1(B_1', X_1'), B_2, \ldots, B_m). \tag{4.23}$$

Thus, using (4.20)–(4.23) and the trivial fact that $\mathrm{d}A = \prod_{i=1}^{m} \mathrm{d}^m B_i \prod_{i=p}^{n} \mathrm{d}^m X_p$, we obtain

$$P(z; u) = \frac{1}{(1 + \xi^T \xi)^{\frac{m}{2}}} \int \mathrm{d}A \, f(\operatorname{tr} A^T A) |\det \tilde{B}| \tag{4.24}$$

where we have removed the primes from the integration variables. We are not done yet, since $\tilde{B}_1$, the first column of $\tilde{B}$, depends on z_0 and u_0. To rectify this problem, we note from (4.22) that

$$\tilde{B}_1(B_1, X_1) = \sqrt{\frac{1 + u_0^2}{1 + \xi^T \xi}} \, (B_1 \cos\theta + X_1 \sin\theta) \qquad \cos\theta = \frac{z_0}{\sqrt{\xi^T \xi \left(1 + u_0^2\right)}}. \tag{4.25}$$

Thus, performing one final rotation by an angle θ in the B_1–X_1 plane, which leaves, of course, $\mathrm{d}A$ and $\operatorname{tr} A^T A$ invariant, we see that in terms of the rotated columns $|\det \tilde{B}| = \sqrt{\frac{1+u_0^2}{1+\xi^T\xi}} |\det B|$, and thus, finally, we obtain that

$$P(z; u) = \frac{\sqrt{1 + u_0^2}}{(1 + \xi^T \xi)^{\frac{m+1}{2}}} \int \mathrm{d}A \, f(\operatorname{tr} A^T A) |\det B| \tag{4.26}$$

which coincides with (4.11), due to (4.12) and (4.13). □

Remark 4.2. The fact that the integral $C = \int \mathrm{d}A |\det B| f(\operatorname{tr} A^T A)$ is convergent and independent of the function $f(u)$ can be proved by decomposing A into its singular values [5, 6], essentially in a manner similar to our proof of lemma (2.1), but with slight modifications in (2.13) and (2.14) due to the fact that A is a rectangular matrix rather than a square matrix. We shall not get into these technicalities here, which the reader may find in [5, 6]. Note, however, that C may be determined from the normalization of $P(z; u)$

$$\int \mathrm{d}^m z \, P(z; u) = 1 = C\beta \int \frac{\mathrm{d}^m z}{(\beta^2 + z^T z)^{\frac{m+1}{2}}} = \frac{C\pi^{\frac{m+1}{2}}}{\Gamma\left(\frac{m+1}{2}\right)}.$$

Thus,

$$C = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\pi^{\frac{m+1}{2}}} = \frac{2}{S_{m+1}}. \tag{4.27}$$

Remark 4.3. We note that for $n = 0$ (i.e., when $A = B$) and $u = 0$, (4.3) degenerates into $Bz = b$, which is precisely the case $n = 1$ (and $z \equiv Z$) in the conditions for theorem (2.1), which we analysed in example (1). Thus, (4.11), evaluated at $n = 1$ and $u = 0$, must coincide with (2.30), as one can easily check it does.

Since theorem (4.2) states the explicit form (4.11) of $P(z; u)$, we can now use it to derive, e.g., the probability density of the distribution of a single component z_i and that of the ratio of two different components, mentioned in Girko's theorem (4.1).

Corollary 4.2. *The m components z_i are identically distributed, with the probability density of any one of the components $z_i = \zeta$ given by (4.7) of corollary (4.1).*

Proof. That the z_i are identically distributed is an immediate consequence of the rotational invariance of $P(z; u)$ in (4.11). The proof is completed by performing the necessary integrals

$$\begin{aligned} p(\zeta; \beta) &= \int \mathrm{d}z_2 \cdots \mathrm{d}z_m \, P(z; u)_{|z_1=\zeta} = C\beta \int \frac{\mathrm{d}z_2 \cdots \mathrm{d}z_m}{\left(\beta^2 + \zeta^2 + \sum_{i=2}^m z_i^2\right)^{\frac{m+1}{2}}} \\ &= \frac{C\pi^{\frac{m-1}{2}}}{\Gamma\left(\frac{m+1}{2}\right)} \frac{\beta}{\beta^2 + \zeta^2}. \end{aligned} \tag{4.28}$$

Thus, from (4.27) we obtain the desired result that $p(\zeta; \beta) = \frac{\beta}{\pi(\beta^2+\zeta^2)}$. This result, should have been anticipated, since the universal formula (4.11) holds, in particular, for the Gaussian distribution (4.8). □

Finally, we have

Corollary 4.3. *The ratios z_i/z_j $(i \neq j; i, j = 1, \ldots m)$ have the density $P(r) = p(r; 1)$.*

Proof. The ratio z_i/z_j is dimensionless, and thus its distribution cannot depend on the width β, which is the only dimensionful quantity in (4.11). The proof amounts to performing the necessary integrals, e.g., for the random variable z_1/z_2

$$\begin{aligned} P(r) &= \int \mathrm{d}^m z \, P(z; u)\delta\left(r - \frac{z_1}{z_2}\right) = C\beta \int \frac{|z_2| \, \mathrm{d}z_2 \cdots \mathrm{d}z_m}{\left[\beta^2 + (r^2+1)z_2^2 + \sum_{i=3}^m z_i^2\right]^{\frac{m+1}{2}}} \\ &= \frac{2CS_{m-2}}{r^2+1} \frac{\Gamma\left(\frac{m-2}{2}\right)\Gamma\left(\frac{3}{2}\right)}{2\Gamma\left(\frac{m+1}{2}\right)} = \frac{1}{\pi(r^2+1)} = p(r; 1). \end{aligned} \tag{4.29}$$

□

Acknowledgments

I thank Asa Ben-Hur, Shmuel Fishman and Hava Siegelmann for a stimulating collaboration on complexity in analogue computation, which inspired me to consider the problems studied in this paper. I also thank Ofer Zeitouni for making some comments. Shortly after the first version of this work appeared, its connection with the theory of matrix variate distributions was pointed out to me by Peter Forrester. I am indebted to him for enlighting me on that connection and for suggesting [3, 4]. This research was supported in part by the Israeli Science Foundation.

References

[1] Ben-Hur A, Feinberg J, Fishman S and Siegelmann H T 2003 Probabilistic analysis of a differential equation for linear programming *J. Complexity* **19** 474 (*Preprint* CC/0110056)

Ben-Hur A, Feinberg J, Fishman S and Siegelmann H T 2004 Probabilistic analysis of the phase space flow for linear programming *Phys. Lett.* A **323** 204–9

Ben-Hur A 2000 Computation: a dynamical system approach *PhD Thesis* Technion, Haifa

[2] Brockett R W 1991 Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems *Linear Algebra Appl.* **146** 79–91

Faybusovich L 1991 Dynamical systems which solve optimization problems with linear constraints *IMA J. Math. Control Inform.* **8** 135–49

Helmke U and Moore J B 1994 *Optimization and Dynamical Systems* (London: Springer)

[3] Dickey J M 1967 Matricvariate generalizations of the multivariate *t* distribution and the inverted multivariate *t* distribution *Ann. Math. Stat.* **38** 511

[4] Gupta A K and Nagar D K 2000 *Matrix Variate Distributions* (Boca Raton, FL: Chapman and Hall/CRC) Chapter 4

[5] Hua L K 1963 *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains (Translations of Mathematical Monographs* vol 6*)* (Providence, RI: American Mathematical Society)

[6] Some useful references from the physics literature are Cicuta G M, Molinari L, Montaldi E and Riva F 1987 *J. Math. Phys.* **28** 1716

Anderson A, Myers R C and Periwal V 1991 *Phys. Lett.* B **254** 89

Anderson A, Myers R C and Periwal V 1991 *Nucl. Phys.* B **360** 463 (Section 3)

Feinberg J and Zee A 1997 *J. Stat. Mech.* **87** 473–504

[7] Mehta M L 1991 *Random Matrices* 2nd edn (San Diego, CA: Academic) Chapter 17

[8] Girko V L 1974 On the distribution of solutions of systems of linear equations with random coefficients *Theory Probab. Math. Stat.* **2** 41