How to Generate Random Matrices from the Classical Compact Groups

Francesco Mezzadri

Since Random Matrix Theory (RMT) was introduced by Wishart [17] in 1928, it has found applications in a variety of areas of physics, pure and applied mathematics, probability, statistics, and engineering. A few examples—far from being exhaustive—include: analytic number theory, combinatorics, graph theory, multivariate statistics, nuclear physics, quantum chaos, quantum information, statistical mechanics, structural dynamics, and wireless telecommunications. The reasons for the ever growing success of RMT are mainly two. Firstly, in the limit of large matrix dimension the statistical correlations of the spectra of a family, or ensemble, of matrices are independent of the probability distribution that defines the ensemble, but depend only on the invariant properties of such a distribution. As a consequence random matrices turn out to be very accurate models for a large number of mathematical and physical problems. Secondly, RMT techniques allow analytical computations to an extent that is often impossible to achieve in the contexts that they are modelling. This predictive ability of RMT is particularly powerful whenever in the original problem there are no natural parameters to average over.

Although the advantage of using RMT lies in the possibility of computing explicit mathematical and physical quantities analytically, it is sometimes necessary to resort to numerical simulations. The purpose of this article is twofold. Firstly, we provide the reader with a simple method for generating random matrices from the classical compact groups that most mathematicians—not necessarily familiar with computer programming —should be able to implement in a code of only a few lines. This is achieved in the section "A Correct and Efficient Algorithm". Secondly, we discuss in detail the main ideas, which turn out be fairly general and quite interesting, behind this algorithm.

Francesco Mezzadri is a Lecturer in Applied Mathematics at the University of Bristol, United Kingdom. His email address is f.mezzadri@bristol.ac.uk.

An $N \times N$ unitary matrix $U = (u_{jk})$ is defined by the relation $U^*U = UU^* = I$, which in terms of the matrix elements reads

$$\sum_{k=1}^{N} u_{jk}^* u_{kl} = \sum_{k=1}^{N} \overline{u}_{kj} u_{kl} = \delta_{jl} \quad \text{and} \quad \sum_{k=1}^{N} u_{jk} u_{kl}^* = \sum_{k=1}^{N} \overline{u}_{jk} u_{lk} = \delta_{jl}, \tag{1}$$

where $U^* = (u_{jk}^*)$ is the conjugate transpose of U, i.e., $u_{jk}^* = \overline{u}_{kj}$. In this article we will use the symbol $\overline{\ }$ to denote complex conjugation, in order to distinguish it from *, which is reserved for the conjugate transpose of a matrix. The constraints (1) simply state that the columns (rows) of a unitary matrix form an orthonormal basis in $\mathbb{C}^N$. The set $\mathrm{U}(N)$ of unitary matrices forms a compact Lie group whose real dimension is N^2; it is then made into a probability space by assigning as a distribution the unique measure invariant under group multiplication, known as *Haar measure.* Such a probability space is often referred to as Circular Unitary Ensemble (CUE).

Usually the correct ensemble to model a particular situation depends on the symmetries of the problem. Ensembles of unitary matrices are constructed in two steps: we first identify a subset $\mathsf{U} \subset \mathrm{U}(N)$ by imposing further restrictions on U; then we assign to U a probability measure with the appropriate invariant properties. As well as $\mathrm{U}(N)$, we will discuss how to generate random matrices from the orthogonal $\mathrm{O}(N)$ and unitary symplectic $\mathrm{USp}(2N)$ groups with probability distributions

given by the respective unique invariant measures. We shall also consider the two remaining *Dyson circular ensembles* [2], namely the Circular Orthogonal Ensemble (COE) and Circular Symplectic Ensemble (CSE). Other symmetric spaces appear in the applications [18], but we will not concern ourselves with them.

Writing an algorithm to generate random unitary matrices that is both correct and numerically stable presents some pitfalls. The reason is that the conditions (1) imply that the matrix elements are not independent and thus are statistically correlated. The main ideas discussed in this article are centered around the QR decomposition and go back to Wedderburn [16], Heiberger [5] (corrected by Tanner and Thisted [15]), Stewart [14], and Diaconis and Shahshahani [1]. However, the technical literature may be difficult to access for a reader without a background in numerical analysis or statistics, while the implementation of such techniques is elementary. Another method discussed in the literature involves an explicit representation of the matrix elements of U in terms of N^2 independent parameters (Euler angles) [19], but it does not seem to be equally efficient or convenient.

Some Examples and Motivations

Before discussing how to generate random matrices it is helpful to give a few examples that show how they appear in the applications.

In quantum mechanics all the information about an isolated physical system at a given time t_0 is contained in a state vector ψ_0 belonging to a Hilbert space $\mathcal{H}$—in general infinite dimensional. The time evolution of ψ_0, i.e., its dynamics, is determined by a unitary operator U. In other words, at a time $t > t_0$, ψ_0 has evolved into

$$\psi = U\psi_0. \tag{2}$$

The fact that U is unitary guarantees that $\|\psi\| = \|\psi_0\| = 1$, which is required by the probabilistic interpretation of quantum mechanics.

If the dynamics is complicated—as in heavy nuclei or in quantum systems whose classical limits are characterized by chaotic dynamics—writing down an explicit expression for U may be hopeless. Therefore, we can attempt to replace U by a random operator and check if the predictions that we obtain are consistent with the empirical observations. It is also reasonable to simplify the problem even further and replace U by a random unitary matrix of finite, but large, dimension. Then the main question is: What are the matrix space and the probability distribution that best model our system?

In physics the symmetries of a problem are often known a priori, even if the details of the dynamics remain obscure. Now, suppose that our system is invariant under time reversal but does not have any other symmetry. From general considerations (see Mehta [9] p. 36) we know that U is always conjugate by a unitary transformation to a symmetric matrix. Therefore, we can always choose U so that

$$U = U^t, \tag{3}$$

where U^t denotes the transpose of U. Since there are no other symmetries in the problem, this is the only constraint that we can impose. Therefore, the appropriate matrices that model this physical system should be symmetric. Let us denote by O the set of unitary symmetric matrices. If $U \in \mathsf{O}$ it can be proved (see Meta [9] p. 499) that it admits the representation

$$U = WW^t, \quad W \in \mathrm{U}(N). \tag{4}$$

This factorization is not unique. Let $\mathrm{O}(N)$ be the group of real matrices O such that $OO^t = O^tO = I$ and set $W' = WO$. By definition we have

$$U = W'W'^t = WOO^tW^t = WW^t. \tag{5}$$

This statement is true also in the opposite direction: if $WW^t = W'W'^t$ there exists an $O \in \mathrm{O}(N)$ such that $W' = WO$. Therefore, O is isomorphic to the left coset space of $\mathrm{O}(N)$ in $\mathrm{U}(N)$, i.e.,

$$\mathsf{O} \cong \mathrm{U}(N)/\mathrm{O}(N). \tag{6}$$

Since a measure space with total mass equal to one is a probability space, in what follows we shall use the two terminologies interchangeably. An ensemble of random matrices is defined by a matrix space and a probability measure on it. We have found the former; we are left to identify the latter. Haar measure, which will be discussed in detail in the section "Haar Measure and Invariance", provides a natural probability distribution on $\mathrm{U}(N)$; "natural" in the sense that it equally weighs different regions of $\mathrm{U}(N)$, thus it behaves like a uniform distribution. From the factorization (4) the probability distribution on $\mathrm{U}(N)$ induces a measure on O. As a consequence, if W is Haar distributed the resulting measure on O will be uniform too. In the section "A Group Theoretical Interpretation", we shall see that such a measure is the unique probability distribution induced by Haar measure on O. Therefore, it provides a natural choice to model a time reversal invariant quantum system. The space O together with this measure is the COE ensemble.

If a quantum system does not have any symmetry, then there are no restrictions on $\mathrm{U}(N)$, and the natural choice of probability distribution is Haar measure. This is the CUE ensemble. If the system is invariant under time reversal and has a half-integer spin, then the appropriate ensemble is the CSE. The matrix space of the CSE is the subset $\mathsf{S} \subset \mathrm{U}(2N)$ whose elements admit the representation

$$U = -WJW^tJ, \quad W \in \mathrm{U}(2N), \tag{7}$$

where

$$J = \begin{pmatrix} 0 & I_N \\ -I_N & 0 \end{pmatrix}. \tag{8}$$

From the factorization (7) the probability distribution on $\mathrm{U}(2N)$ induces a measure on S. As previously, such a measure is fixed by assigning Haar measure to $\mathrm{U}(2N)$.

The set S is isomorphic to a coset space too. The unitary symplectic group $\mathrm{USp}(2N)$ is the subgroup of $\mathrm{U}(2N)$ whose elements obey the relation

$$SJS^t = J. \tag{9}$$

Therefore, the matrix U in equation (7) does not change if we replace W with $W' = WS$, where $S \in \mathrm{USp}(2N)$. Similarly, if W and W' are such that

$$U = -WJW^tJ = -W'JW'^tJ, \quad W, W' \in \mathrm{U}(2N), \tag{10}$$

then $W'W^{-1} \in \mathrm{USp}(2N)$. Therefore,

$$\mathsf{S} \cong \mathrm{U}(2N)/\mathrm{USp}(2N). \tag{11}$$

The probability distribution of the CSE is the unique invariant measure induced on the coset space (11) by Haar measure on $\mathrm{U}(2N)$.

From equations (4) and (7) all we need to generate random matrices in the CUE, COE, and CSE ensembles is an algorithm whose output is Haar distributed unitary matrices. The rest of this article will concentrate on generating random matrices from all three classical compact groups $\mathrm{U}(N)$, $\mathrm{O}(N)$, and $\mathrm{USp}(2N)$ with probability distributions given by the respective Haar measures. These groups are not only functional to constructing matrices in the COE and CSE, but are also important ensembles in their own right. Indeed, the work of Montgomery [11], Odlyzko [12], Katz and Sarnak [6], Keating and Snaith [7, 8], and Rubinstein [13] has shown beyond doubt that the local statistical properties of the Riemann zeta function and other L-functions can be modelled by the characteristic polynomials of Haar distributed random matrices. Over the last few years the predictive power of this approach has brought about impressive progress in analytic number theory that could not have been achieved with traditional techniques. (See [10] for a collection of review articles on the subject.)

Haar Measure and Invariance

Since the algorithm we shall discuss is essentially based on the invariant properties of Haar measure, in this section we introduce the main concepts that are needed to understand how it works. We nevertheless begin with another ensemble: *the Ginibre ensemble*. Besides being a simpler illustration of the ideas we need, generating a matrix in the Ginibre ensemble is the first step toward producing a random unitary matrix.

The space of matrices for the Ginibre ensemble is $\mathrm{GL}(N, \mathbb{C})$, the set of all the invertible $N \times N$ complex matrices $Z = (z_{jk})$; the matrix elements are independent identically distributed (i.i.d.) standard normal complex random variables. In other words, the probability density function (p.d.f.) of z_{jk} is

$$p(z_{jk}) = \frac{1}{\pi} e^{-|z_{jk}|^2}. \tag{12}$$

By definition the matrix entries are statistically independent, therefore the joint probability density function (j.p.d.f.) for the matrix elements is

$$\begin{aligned} P(Z) &= \frac{1}{\pi^{N^2}} \prod_{j,k=1}^{N} e^{-|z_{jk}|^2} \\ &= \frac{1}{\pi^{N^2}} \exp\left(- \sum_{j,k=1}^{N} |z_{jk}|^2 \right) \\ &= \frac{1}{\pi^{N^2}} \exp\left(- \operatorname{Tr} Z^* Z \right). \end{aligned} \tag{13}$$

Since $P(Z)$ is a probability density, it is normalized to one, i.e.,

$$\int_{\mathbb{C}^{N^2}} P(Z)\, dZ = 1, \tag{14}$$

where $dZ = \prod_{j,k=1}^{N} dx_{jk} dy_{jk}$ and $z_{jk} = x_{jk} + iy_{jk}$. The j.p.d.f. $P(Z)$ contains all the statistical information on the Ginibre ensemble.

Since $\mathbb{C}^{N \times N} \cong \mathbb{C}^{N^2}$, we will use the two notations according to what is more appropriate for the context. Thus, we can write

$$d\mu_{\mathrm{G}}(Z) = P(Z)\, dZ \tag{15}$$

and think of $d\mu_{\mathrm{G}}$ as an infinitesimal volume or measure in $\mathbb{C}^{N^2}$. If $f : \mathbb{C}^{N \times N} \longrightarrow \mathbb{C}^{N \times N}$, we say that $d\mu_{\mathrm{G}}$ is invariant under f if

$$d\mu_{\mathrm{G}}(f(Z)) = d\mu_{\mathrm{G}}(Z). \tag{16}$$

Lemma 1. *The measure of the Ginibre ensemble is invariant under left and right multiplication of Z by arbitrary unitary matrices, i.e.,*

$$d\mu_{\mathrm{G}}(UZ) = d\mu_{\mathrm{G}}(ZV) = d\mu_{\mathrm{G}}(Z), \quad U, V \in \mathrm{U}(N). \tag{17}$$

Proof. First we need to show that $P(UZ) = P(Z)$; then we must prove that the Jacobian of the map

$$Z \mapsto UZ \tag{18}$$

(seen as a transformation in $\mathbb{C}^{N^2}$) is one. Since by definition $U^*U = I$, we have

$$\begin{aligned} P(UZ) &= \frac{1}{\pi^{N^2}} \exp\left(- \operatorname{Tr} Z^* U^* U Z \right) \\ &= \frac{1}{\pi^{N^2}} \exp\left(- \operatorname{Tr} Z^* Z \right) = P(Z). \end{aligned} \tag{19}$$

Now, the map (18) is isomorphic to

$$X = \underbrace{U \oplus \cdots \oplus U}_{N\text{times}}. \tag{20}$$

It follows immediately that X is a $N^2 \times N^2$ unitary matrix, therefore $|\det X| = 1$. The proof of right invariance is identical. □

Because the elements of a unitary matrix are not independent, writing an explicit formula for the infinitesimal volume element of $\mathrm{U}(N)$ is more complicated than for the Ginibre ensemble. An $N \times N$ unitary matrix contains $2N^2$ real numbers and the constraints (1) form a system of N^2 real equations. Therefore, $\mathrm{U}(N)$ is isomorphic to a N^2-dimensional manifold embedded in $\mathbb{R}^{2N^2}$. Such a manifold is compact and has a natural group structure that comes from matrix multiplication. Thus, an infinitesimal volume element on $\mathrm{U}(N)$ will have the form

$$d\mu(U) = m(\alpha_1, \ldots, \alpha_{N^2}) d\alpha_1 \cdots d\alpha_{N^2}, \tag{21}$$

where $\alpha_1, \ldots, \alpha_{N^2}$ are local coordinates on $\mathrm{U}(N)$. Every compact Lie group has a unique (up to an arbitrary constant) left and right invariant measure, known as *Haar measure.* In other words, if we denote Haar measure on $\mathrm{U}(N)$ by $d\mu_{\mathrm{H}}(U)$, we have

$$d\mu_{\mathrm{H}}(VU) = d\mu_{\mathrm{H}}(UW) = d\mu_{\mathrm{H}}(U), \quad V, W \in \mathrm{U}(N). \tag{22}$$

Although an explicit expression for Haar measure on $\mathrm{U}(N)$ in terms of local coordinates can be written down (see Życzkowski and Kus [19] for a formula), we will see that in order to generate matrices distributed with Haar measure we only need to know that it is invariant and unique.

Haar measure normalized to one is a natural choice for a probability measure on a compact group because, being invariant under group multiplication, any region of $\mathrm{U}(N)$ carries the same weight in a group average. It is the analogue of the uniform density on a finite interval. In order to understand this point consider the simplest example: $\mathrm{U}(1)$. It is the set $\{e^{i\theta}\}$ of the complex numbers with modulo one, therefore it has the topology of the unit circle $\mathbb{S}^1$. Since in this case matrix multiplication is simply addition mod 2π, $\mathrm{U}(1)$ is isomorphic to the group of translations on $\mathbb{S}^1$. A probability density function that equally weighs any part of the unit circle is the constant density $\rho(\theta) = 1/(2\pi)$. This is the standard Lebesgue measure, which is invariant under translations. Therefore, it is the unique Haar measure on $\mathrm{U}(1)$.

Note that it is not possible to define an "unbiased" measure on a non-compact manifold. For example, we can provide a finite interval with a constant p.d.f. $\rho(x)$, but not the whole real line $\mathbb{R}$, since the integral $\int_{-\infty}^{\infty} \rho(x)dx$ would diverge.

The QR Decomposition and a Numerical Experiment

By definition the columns of an $N \times N$ unitary matrix are orthonormal vectors in $\mathbb{C}^N$. Thus, if we take an arbitrary complex $N \times N$ matrix Z of full rank and apply Gram-Schmidt orthonormalization to its columns, the resulting matrix Q is unitary. It turns out that if the entries of Z are i.i.d. standard complex normal random variables, i.e., if Z belongs to the Ginibre ensemble, then Q is distributed with Haar measure (see Eaton [3], p. 234, for a proof). Unfortunately, the implementation of this algorithm is numerically unstable. However, we may observe that

$$Z = QR, \tag{23}$$

where R is upper-triangular and invertible. In other words, the Gram-Schmidt algorithm realizes the *QR decomposition.* This factorization is widely used in numerical analysis to solve linear least squares problems and as the first step of a particular eigenvalue algorithm. Indeed, every linear algebra package has a routine that implements it. In most cases, however, the algorithm adopted is not the Gram-Schmidt orthonormalization but uses the *Householder reflections*, which are numerically stable.

Because of this simple observation, at first one might be tempted to produce a matrix in the Ginibre ensemble and then to use a black box QR decomposition routine. Writing such a code is straightforward. For example, if we choose the SciPy library in Python, we may implement the following function:

```
from scipy import *
def wrong_distribution(n):
"""A Random matrix with the wrong
distribution"""
z = (randn(n,n) + 1j*randn(n,n))/
sqrt(2.0)
q,r = linalg.qr(z)
return q
```

Unfortunately, as Edelman and Rao observed [4], the output is not distributed with Haar measure. It is instructive to give an explicit example of this phenomenon.

A unitary matrix can always be diagonalized in $\mathrm{U}(N)$. Therefore, its eigenvalues $\{e^{i\theta_1}, \ldots, e^{i\theta_N}\}$ lie on the unit circle. A classical calculation in RMT (see Mehta [9] pp. 203–205) consists of computing the statistical correlations among the arguments θ_j. The simplest correlation function to determine is the density of the eigenvalues $\rho(\theta)$, or—as sometimes it is called—the one-point correlation. Since Haar measure is the analogue of a uniform distribution, each set of eigenvalues must have the same weight, therefore the normalized eigenvalue density is

$$\rho(\theta) = \frac{1}{2\pi}. \tag{24}$$

It is important to point out that equation (24) does not mean that the eigenvalues are statistically uncorrelated.

Testing (24) numerically is very simple. We generated 10,000 random unitary matrices using `wrong_distribution(n)`. The density of the eigenvalues of such matrices is clearly not constant (Figure 1(a)). Figure 1(b) shows the histogram of the spacing distribution, which deviates from the theoretical prediction too. This statistic is often plotted because it encodes the knowledge of all the spectral correlations and is easy to determine empirically. For unitary matrices it is defined as follows. Take the arguments of the eigenvalues and order them in ascending order:

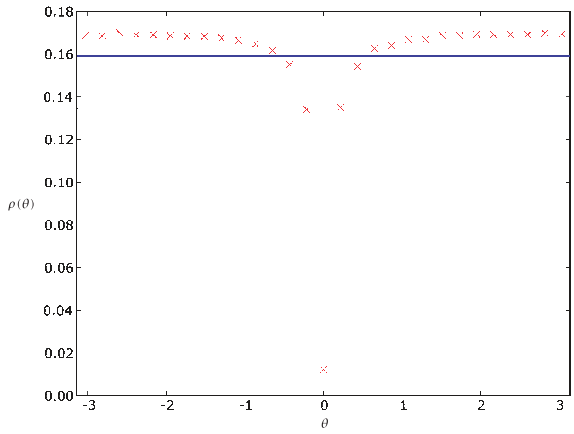
$$\theta_1 \le \theta_2 \le \ldots \le \theta_N. \tag{25}$$

The normalized distances, or spacings, between consecutive eigenvalues are

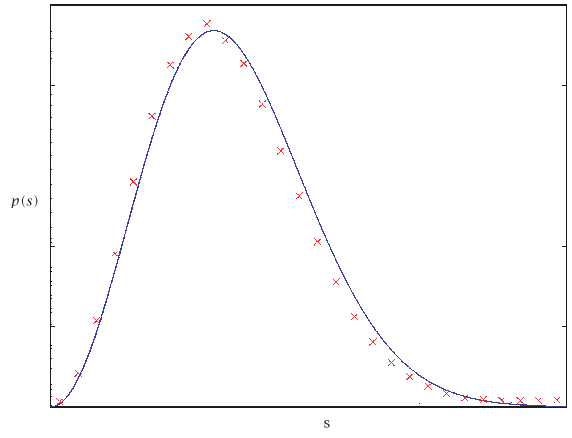
$$s_j = \frac{N}{2\pi}(\theta_{j+1} - \theta_j), \quad j = 1, \ldots, N. \tag{26}$$

The spacing distribution $p(s)$ is the probability density of s. (For a discussion on the spacing distribution see Mehta [9] p. 118.)

It is worth emphasizing that the QR decomposition is a standard routine. The most commonly known mathematical software packages like Matlab, Mathematica, Maple, and SciPy for Python essentially use a combination of algorithms found in LAPACK routines. Changing software would not alter the outcome of this numerical experiment.



(a) Eigenvalue density



(b) Spacing distribution

Figure 1. Empirical histograms of the density of the eigenvalues and of the spacing distributions compared with the theoretical curves for the CUE. The data are computed from the eigenvalues of ten thousand 50 x 50 random unitary matrices obtained from the routine wrong_distribution(n).

A Correct and Efficient Algorithm

What is wrong with standard QR factorization routines? Where do they differ from the Gram-Schmidt orthonormalization? Why is the probability distribution of the output matrix not Haar measure?

The main problem is that QR decomposition is not unique. Indeed, let $Z \in \mathrm{GL}(N, \mathbb{C})$ and suppose that $Z = QR$, where Q is unitary and R is invertible and upper-triangular. If

$$\Lambda = \begin{pmatrix} e^{i\theta_1} & & \\ & \ddots & \\ & & e^{i\theta_N} \end{pmatrix} = \mathrm{diag}\left(e^{i\theta_1}, \ldots, e^{i\theta_N}\right), \tag{27}$$

then

$$Q' = Q\Lambda \quad \text{and} \quad R' = \Lambda^{-1}R \tag{28}$$

are still unitary and upper-triangular, respectively. Furthermore,

$$Z = QR = Q'R'. \tag{29}$$

Therefore, the QR decomposition defines a multi-valued map

$$\mathrm{QR} : \mathrm{GL}(N, \mathbb{C}) \longrightarrow \mathrm{U}(N) \times \mathrm{T}(N), \tag{30}$$

where $\mathrm{T}(N)$ denotes the group of invertible upper-triangular matrices.

In order to make the mapping (30) single-valued, we need to specify the algorithm that achieves the factorization. In most applications such a choice is dictated only by the performance and stability of the code. For our purposes, however, the subset of $\mathrm{U}(N) \times \mathrm{T}(N)$, in which the output of the QR decomposition is chosen, is fundamental and we need to pay particular attention to it. It is convenient from a mathematical point of view to introduce a variation of the mapping (30), which is not only single-valued but also one-to-one. In this way we will not have to refer all the time to a specific algorithm. Indeed, the idea is that we should be able to alter the output of a QR decomposition routine without even knowing the algorithm implemented.

We first need

Lemma 2. *Equation* (29) *implies* (28)*, where* $\Lambda \in \Lambda(N)$ *and* $\Lambda(N)$ *denotes the group of all unitary diagonal matrices* (27)*.*

Proof. Equation (29) can be rearranged as

$$Q^{-1}Q' = RR'^{-1}. \tag{31}$$

Since $\mathrm{U}(N)$ and $\mathrm{T}(N)$ are groups, both sides of equation (31) must belong to $\mathrm{U}(N) \cap \mathrm{T}(N)$. By definition the inverse of a unitary matrix U is its conjugate transpose and the inverse of an upper-triangular matrix is upper-triangular. Therefore, if a matrix is both unitary and upper-triangular it must be diagonal, i.e., $\Lambda(N) = \mathrm{U}(N) \cap \mathrm{T}(N)$. □

This lemma suggests that, more naturally, instead of the QR factorization (30) we should consider a one-to-one map

$$\overline{\mathrm{QR}} : \mathrm{GL}(N, \mathbb{C}) \longrightarrow \mathrm{U}(N) \times \Gamma(N), \tag{32}$$

where $\Gamma(N) = \mathrm{T}(N)/\Lambda(N)$ is the right coset space of $\Lambda(N)$ in $\mathrm{T}(N)$. We construct (32) as follows: we first define it on a class of representatives of $\Gamma(N)$ using the QR factorization; then we extend it to the whole $\Gamma(N)$. However, since the QR decomposition is not unique, there is a certain degree of arbitrariness in this definition. We need to find a map under which the measure of the Ginibre ensemble induces Haar measure on $\mathrm{U}(N)$. The main tool to achieve this goal is the invariance under group multiplication of Haar measure and its uniqueness. Thus, our choice of the decomposition (32) must be such that if

$$Z \mapsto (Q, \gamma) \quad \text{then} \quad UZ \mapsto (UQ, \gamma) \tag{33}$$

with the *same* $\gamma \in \Gamma(N)$ and for *any* $U \in \mathrm{U}(N)$. This property implies that left multiplication of Z by a unitary matrix reduces, after the decomposition, to the left action of $\mathrm{U}(N)$ into itself. But lemma 1 states that

$$d\mu_{\mathrm{G}}(UZ) = d\mu_{\mathrm{G}}(Z) \tag{34}$$

for any $U \in \mathrm{U}(N)$. As a consequence, if the map (32) satisfies (33) the induced measure on $\mathrm{U}(N)$ will be invariant under left multiplication too and therefore must be Haar measure.

How do we construct the map (32)? A class of representatives of $\Gamma(N)$ can be chosen by fixing the arguments of the elements of the main diagonal of $R \in \mathrm{T}(N)$. Let us impose that such elements all be real and strictly positive. Using (28) we can uniquely factorize any $Z \in \mathrm{GL}(N, \mathbb{C})$ so that the main diagonal of R has this property. It follows that if $Z = QR$, then

$$UZ = UQR, \quad U \in \mathrm{U}(N). \tag{35}$$

This QR decomposition of UZ is unique within the chosen class of representatives of $\Gamma(N)$. Therefore, the resulting map (32) obeys (33). Finally, we arrive at

Theorem 1. *Suppose that the map* (32) *satisfies the hypothesis* (33)*. Then it decomposes the measure* (15) *of the Ginibre ensemble as*

$$d\mu_{\mathrm{G}}(Z) = d\mu_{\mathrm{H}}(Q) \times d\mu_{\Gamma(N)}(\gamma). \tag{36}$$

Proof. We have

$$d\mu_{\mathrm{G}}(UZ) = d\mu_{\mathrm{G}}(Z) \quad \text{by lemma 1} \tag{37a}$$

$$= d\mu(UQ, \gamma) = d\mu(Q, \gamma) \quad \text{by equation (33)} \tag{37b}$$

$$= d\mu_{\mathrm{H}}(Q) \times d\mu_{\Gamma(N)}(\gamma) \quad \text{by the uniqueness of Haar measure.} \tag{37c}$$

□

The choice of the class of representatives that we made coincides exactly with outcome of the Gram-Schmidt orthonormalization. The output of standard QR decomposition routines are such that if $Z \mapsto (Q, R)$ then $UZ \mapsto (Q', R')$ with $Q' \neq UQ$ and $R' \neq R$. Therefore, the corresponding map (32) does not obey (33) and theorem 1 does not hold.

We can now give a recipe to create a random unitary matrix with distribution given by Haar measure.

(1) Take an $N \times N$ complex matrix Z whose entries are complex standard normal random variables.
(2) Feed Z into *any* QR decomposition routine. Let (Q, R), where $Z = QR$, be the output.
(3) Create the following diagonal matrix

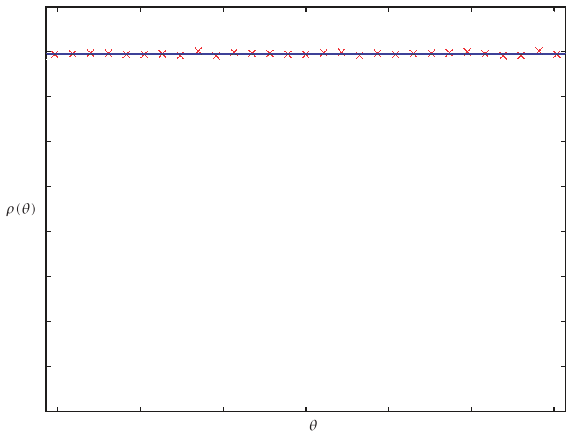
$$\Lambda = \begin{pmatrix} \frac{r_{11}}{|r_{11}|} & & \\ & \ddots & \\ & & \frac{r_{NN}}{|r_{NN}|} \end{pmatrix}, \tag{38}$$

where the r_{jj}s are the diagonal elements of R.
(4) The diagonal elements of $R' = \Lambda^{-1}R$ are *always* real and strictly positive, therefore the matrix $Q' = Q\Lambda$ is distributed with Haar measure.

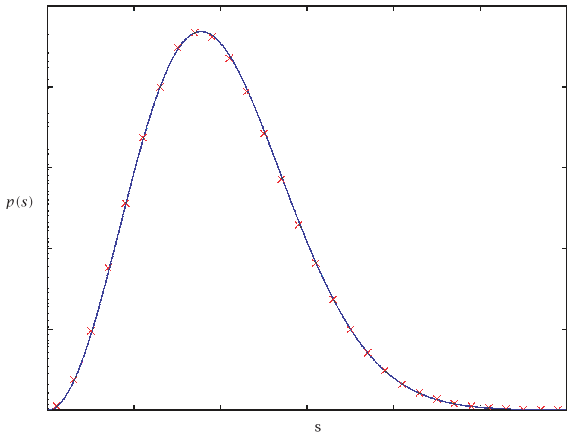
The corresponding Python function is:

```
from scipy import *
def haar_measure(n):
"""A Random matrix distributed with
Haar measure"""
z = (randn(n,n) + 1j*randn(n,n))/
sqrt(2.0)
q,r = linalg.qr(z)
d = diagonal(r)
ph = d/absolute(d)
q = multiply(q,ph,q)
return q
```

If we repeat the numerical experiment discussed in the section "The QR Decomposition and a Numerical Experiment", using this routine, we obtain the histograms in Figure 2, which are consistent with the theoretical predictions.



(a) Eigenvalue density



(b) Spacing distribution

Figure 2. Empirical histograms of the density of the eigenvalues and of the spacing distributions compared with the theoretical curves for the CUE. The data are computed from the eigenvalues of ten thousand 50x 50 random unitary matrices output of the function `haar_measure(n)`.

The Unitary Symplectic Group $\mathrm{USp}(2N)$

Up to now we have only considered $\mathrm{U}(N)$. The discussion for $\mathrm{O}(N)$ is identical, except that the input matrix of the QR decomposition routine must be real. Unfortunately, however, for $\mathrm{USp}(2N)$ there are no black box routines that we can use, and we must put more effort into writing an algorithm.

The algebra of unitary symplectic matrices can be rephrased in terms of Hamilton's quaternions; it is convenient for our purposes to use this formalism. A quaternion $q \in \mathbb{H}$ is a linear combination

$$q = a \cdot 1 + b i_1 + c i_2 + d i_3, \quad a, b, c, d \in \mathbb{R}, \tag{39}$$

where 1 is the identity and i_1, i_2, i_3 are the quaternion units; they obey the algebra

$$i_1^2 = i_2^2 = i_3^2 = i_1 i_2 i_3 = -1. \tag{40}$$

We can also define the conjugate of q,

$$\overline{q} = a \cdot 1 - b i_1 - c i_2 - d i_3, \tag{41}$$

as well the norm

$$\|q\|^2 = q\,\overline{q} = \overline{q}\,q = a^2 + b^2 + c^2 + d^2. \tag{42}$$

When $c = d = 0$, $\mathbb{H}$ reduces to $\mathbb{C}$ and $\overline{q}$ is simply the complex conjugate of q.

In analogy with $\mathbb{R}^N$ and $\mathbb{C}^N$—provided we are careful with the fact that multiplication in $\mathbb{H}$ is not commutative—we can study the space $\mathbb{H}^N$. Elements in $\mathbb{H}^N$ are N-tuples $\mathbf{q} = (q_1, \ldots, q_N)$. The bilinear map

$$\langle \mathbf{p}, \mathbf{q} \rangle = \sum_{j=1}^{N} \overline{p}_j q_j, \quad \mathbf{p}, \mathbf{q} \in \mathbb{H}^N, \tag{43}$$

is the analogue of the usual Hermitian inner product in $\mathbb{C}^N$, and the norm of a quaternion vector is simply

$$\|\mathbf{q}\|^2 = \langle \mathbf{q}, \mathbf{q} \rangle = \sum_{j=1}^{N} \|q_j\|^2. \tag{44}$$

Similarly, $\mathrm{GL}(N, \mathbb{H})$ is the group of all the $N \times N$ invertible matrices with quaternion elements.

The quaternion units admit a representation in terms of the 2×2 matrices

(45a)

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad e_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad e_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$

where

$$1 \mapsto I_2, \quad i_1 \mapsto e_1, \quad i_2 \mapsto e_2 \quad \text{and} \quad i_3 \mapsto e_3. \tag{45b}$$

Thus, $q = a \cdot 1 + b i_1 + c i_2 + d i_3$ is mapped into the complex matrix

$$A = \begin{pmatrix} z & w \\ -\overline{w} & \overline{z} \end{pmatrix} \tag{46a}$$

where $z = a + ib$ and $w = c + id$. In addition

$$\overline{q} \mapsto A^* = \begin{pmatrix} \overline{z} & -w \\ \overline{w} & z \end{pmatrix}. \tag{46b}$$

Equations (46) generalize to an arbitrary $N \times N$ quaternion matrix $\mathcal{Q}$, which can be represented in terms of a $2N \times 2N$ complex matrix Q using the decomposition

$$\mathcal{Q} \mapsto Q = Q_0 \otimes I_2 + Q_1 \otimes e_1 + Q_2 \otimes e_2 + Q_3 \otimes e_3, \tag{47}$$

where Q_0, Q_1, Q_2, and Q_3 are arbitrary $N \times N$ *real* matrices. Proceeding in the same fashion, if $\mathcal{Q} \in \mathrm{GL}(N, \mathbb{H})$ we define its conjugate transpose $\mathcal{Q}^* = (q^*_{jk})$ by setting $q^*_{jk} = \overline{q}_{kj}$.

The symplectic group $\mathrm{Sp}(N)$ is the subset of $\mathrm{GL}(N, \mathbb{H})$ whose matrices satisfy the identity $S^* S =$

$SS^* = \mathcal{I}$. Because of the analogy between $\mathrm{U}(N)$ and $\mathrm{Sp}(N)$, the latter is sometimes called the *hyperunitary group* and is denoted by $\mathrm{U}(N,\mathbb{H})$. The usefulness of the quaternion algebra lies in

Theorem 2. *The groups* $\mathrm{Sp}(N)$ *and* $\mathrm{USp}(2N)$ *are isomorphic, i.e.*

$$\mathrm{USp}(2N) \cong \mathrm{Sp}(N). \tag{48}$$

Proof. It is convenient to replace the skew-symmetric matrix J in the definition (9) with

$$\Omega = \begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & 0 & 1 \\ & & & -1 & 0 \end{pmatrix} = I \otimes e_2. \tag{49}$$

This substitution is equivalent to a permutation of the rows and columns of J, therefore it is simply a conjugation by a unitary matrix.

We first prove that if $S \in \mathrm{Sp}(N)$, then its complex representation S belongs to $\mathrm{USp}(2N)$. By equation (47) S^* is mapped to

$$S_0^t \otimes I_2 - S_1^t \otimes e_1 - S_2^t \otimes e_2 - S_3^t \otimes e_3 = -\Omega\, S^t\, \Omega, \tag{50}$$

which follows from the identities

$$(A\otimes B)^t = A^t \otimes B^t \quad \text{and} \quad (A\otimes B)(C\otimes D) = AC \otimes BD, \tag{51}$$

and from the algebra (40) of the quaternion units. As a consequence,

$$SS^* \mapsto -S\,\Omega\,S^t\,\Omega = I. \tag{52}$$

Therefore, the matrix S is symplectic. Combining equations (46b) and (50) gives

$$-\Omega\, S^t\, \Omega = S^*, \tag{53}$$

thus $S \in \mathrm{USp}(2N)$.

We now need to show that if $S \in \mathrm{USp}(2N)$ then it is the representation of a matrix $S \in \mathrm{Sp}(N)$. This statement follows if we prove that S admits a decomposition of the form (47), where S_0, S_1, S_2, and S_3 must be real $N \times N$ matrices. If this is true, then the argument of the first part of the proof can simply be reversed.

Let us allow the coefficients a, b, c, and d in the definition (39) to be complex numbers. The definitions of conjugate quaternion (41) and conjugate transpose of a quaternion matrix, however, remain the same. The matrices (45a) form a basis in $\mathbb{C}^{2\times 2}$. Therefore, any 2×2 complex matrix can be represented as a linear combination of I_2, e_1, e_2, and e_3. Thus, any matrix $Q \in \mathbb{C}^{2N\times 2N}$ admits a decomposition of the form (47), but now the matrices Q_0, Q_1, Q_2, and Q_3 are allowed to be complex. In other words, Q is always the representation of a quaternion matrix $\mathcal{Q}$, but in general the quaternion units have complex coefficients. The important fact that we need to pay attention to is that

$$\mathcal{Q}^* \mapsto Q^*, \tag{54}$$

if and only if the coefficients of the quaternion units are real numbers. This is a straightforward consequence of the representation (46a).

Let $S \in \mathrm{USp}(2N)$ be the complex representation of the quaternion matrix S, but assume that S^* is not mapped into S^*. It is still true, however, that

$$S^* \mapsto -\Omega\, S^t\, \Omega, \tag{55}$$

because equation (50) is only a consequence of matrix manipulations. But since S is unitary symplectic $S^* = -\Omega\, S^t\, \Omega$, which is a contradiction. □

The algebra of $\mathrm{Sp}(N)$ is the generalization to Hamilton's quaternions of the algebra of $\mathrm{U}(N)$. Therefore, it is not surprising that the discussion in the section "A Correct and Efficient Algorithm" is not affected by replacing $\mathrm{GL}(N,\mathbb{C})$ and $\mathrm{U}(N)$ with $\mathrm{GL}(N,\mathbb{H})$ and $\mathrm{Sp}(N)$ respectively. Thus, since $\mathrm{Sp}(N)$ and $\mathrm{USp}(2N)$ are isomorphic, $\mathrm{USp}(2N)$ and $\mathrm{Sp}(N)$ have the same Haar measure $d\mu_{\mathrm{H}}$. In particular, we can introduce the quaternion Ginibre ensemble, which is the set $\mathrm{GL}(N,\mathbb{H})$ equipped with the probability density

$$P(\mathcal{Z}) = \frac{1}{\pi^{2N^2}} \exp\left(-\operatorname{Tr} \mathcal{Z}^*\mathcal{Z}\right) = \frac{1}{\pi^{2N^2}} \exp\left(-\sum_{j,k=1}^{N} ||z_{jk}||^2\right). \tag{56}$$

Quaternion matrices can be factorized by the QR decomposition too: for any $\mathcal{Z} \in \mathrm{GL}(N,\mathbb{H})$ we can always write

$$\mathcal{Z} = \mathcal{Q}\mathcal{R}, \tag{57}$$

where $\mathcal{Q} \in \mathrm{Sp}(N)$ and $\mathcal{R}$ is an invertible and upper-triangular quaternion matrix. Now, let

$$\Lambda(N,\mathbb{H}) = \Big\{\Lambda \in \mathrm{T}(N,\mathbb{H}) \,\Big|\, \Lambda = \operatorname{diag}(q_1,\ldots,q_N),\ ||q_j|| = 1,\quad j = 1,\ldots,N\Big\}, \tag{58}$$

where $\mathrm{T}(N,\mathbb{H})$ is the group of invertible upper-triangular quaternion matrices. Furthermore, let $\Gamma(N,\mathbb{H}) = \mathrm{T}(N,\mathbb{H})/\Lambda(N,\mathbb{H})$ be the right coset space of $\Lambda(N,\mathbb{H})$ in $\mathrm{T}(N,\mathbb{H})$. We have the following

Theorem 3. *There exists a one-to-one map*

$$\mathcal{Q}\mathcal{R} : \mathrm{GL}(N,\mathbb{H}) \longrightarrow \mathrm{Sp}(N) \times \Gamma(N,\mathbb{H}) \tag{59}$$

such that

$$\mathcal{Z} \mapsto (\mathcal{Q},\gamma) \quad \textit{and} \quad \mathcal{U}\mathcal{Z} \mapsto (\mathcal{U}\mathcal{Q},\gamma), \tag{60}$$

where $\gamma = \Lambda(N,\mathbb{H})\mathcal{R}$. *Furthermore, it factorizes the measure* $d\mu_{\mathrm{G}}$ *of the Ginibre ensemble as*

$$d\mu_{\mathrm{G}}(\mathcal{Z}) = d\mu_{\mathrm{H}}(\mathcal{Q}) \times d\mu_{\Gamma(N,\mathbb{H})}(\gamma). \tag{61}$$

We leave proving these generalizations as an exercise for the reader.

Householder Reflections

Theorem 3 provides us with the theoretical tools to generate a random matrix in $\mathrm{USp}(2N)$. However, when we implement these results in computer code, we need to devise an algorithm whose output satisfies the condition (60). The first one that comes to one's mind is Gram-Schmidt orthonormalization. But given that black box routines for quaternion matrices do not exist on the market and that we are forced to write the complete code ourselves, we may as well choose one that is numerically stable and that as it turns out, requires the same effort. The most common algorithm that achieves the QR decomposition uses the *Householder reflections.* For the sake of clarity, we will discuss this method for $\mathrm{O}(N)$; the generalizations to $\mathrm{U}(N)$ and $\mathrm{Sp}(N)$ are straightforward.

Given an arbitrary vector $\mathbf{v} \in \mathbb{R}^m$, the main idea of the Householder reflections is to construct a simple orthogonal transformation H_m (dependent on $\mathbf{v}$) such that

$$H_m \mathbf{v} = \|\mathbf{v}\| \, \mathbf{e_1}, \tag{62}$$

where $\mathbf{e_1} = (1, 0, \ldots, 0) \in \mathbb{R}^m$. For any real matrix $X = (x_{jk})$, H_N is determined by replacing $\mathbf{v}$ in equation (62) with the first column of X. The product $H_N X$ will have the structure

$$H_N X = \begin{pmatrix} r_{11} & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \cdots & * \end{pmatrix}, \tag{63}$$

where

$$r_{11} = \|\mathbf{v}\| = \sqrt{\textstyle\sum_{j=1}^{N} x_{j1}^2}. \tag{64}$$

Then, define the matrix

$$\widetilde{H}_{N-1} = \begin{pmatrix} 1 & 0 \\ 0 & \boxed{H_{N-1}} \end{pmatrix}, \tag{65}$$

where

$$H_{N-1}(\mathbf{v}')\mathbf{v}' = \|\mathbf{v}'\| \, \mathbf{e}_1. \tag{66}$$

In this case $\mathbf{v}'$ is the $(N-1)$-dimensional vector obtained by dropping the first element of the second column of the matrix (63). We proceed in this fashion until the matrix

$$R = \widetilde{H}_1 \widetilde{H}_2 \cdots \widetilde{H}_{N-1} H_N X \tag{67}$$

is upper-triangular with diagonal entries $r_{11}, \ldots, r_{NN}$. The product

$$Q = H_N^t \widetilde{H}_{N-1}^t \cdots \widetilde{H}_2^t \widetilde{H}_1^t \tag{68}$$

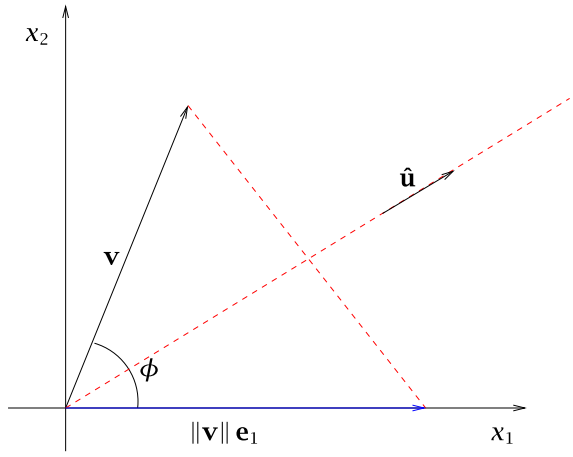


Figure 3. Householder reflection in $\mathbb{R}^2$.

is by construction an orthogonal matrix. In equations (67) and (68) $\widetilde{H}_m$ denotes the block matrix

$$\widetilde{H}_m = \begin{pmatrix} I_{N-m} & \\ & H_m \end{pmatrix}, \tag{69}$$

where H_m is defined in equation (62).

The matrix H_m is constructed using elementary geometry. Consider a vector in the plane, $\mathbf{v} = (x_1, x_2)$, and assume, for simplicity, that $x_1 > 0$. Furthermore, let $\hat{\mathbf{u}}$ denote the unit vector along the interior bisector of the angle ϕ that $\mathbf{v}$ makes with the x_1-axis, i.e.,

$$\hat{\mathbf{u}} = \frac{\mathbf{v} + \|\mathbf{v}\| \, \mathbf{e}_1}{\|\mathbf{v} + \|\mathbf{v}\| \, \mathbf{e}_1\|}, \tag{70}$$

where $\mathbf{e}_1$ is the unit vector along the x_1-axis. The reflection of $\mathbf{v}$ along the direction of $\hat{\mathbf{u}}$ is $\|\mathbf{v}\| \, \mathbf{e}_1$ (see Figure 3). Reflections are distance-preserving linear transformations, therefore their representations in an orthonormal basis are orthogonal matrices. In this simple example it can be constructed from elementary linear algebra:

$$H_2(\mathbf{v}) = -I + 2\hat{\mathbf{u}}\hat{\mathbf{u}}^t. \tag{71}$$

Finally, we obtain

$$H_2(\mathbf{v})\mathbf{v} = \|\mathbf{v}\| \, \mathbf{e}_1. \tag{72}$$

It is worth noting that H_2 depends only on the direction of $\mathbf{v}$ and not on its modulus. Thus we can rewrite equation (72) as

$$H_2(\hat{\mathbf{v}})\hat{\mathbf{v}} = \mathbf{e}_1, \tag{73}$$

where $\hat{\mathbf{v}} = \mathbf{v}/\|\mathbf{v}\|$.

The generalization to arbitrary dimensions is straightforward. For any vector $\mathbf{v} \in \mathbb{R}^m$, the Householder reflection is defined as

$$H_m(\hat{\mathbf{v}}) = \mp \left(I - 2\hat{\mathbf{u}}\hat{\mathbf{u}}^t\right), \tag{74}$$

where

$$\hat{\mathbf{u}} = \frac{\mathbf{v} \pm \|\mathbf{v}\| \, \mathbf{e}_1}{\|\mathbf{v} \pm \|\mathbf{v}\| \, \mathbf{e}_1\|}. \tag{75}$$

Furthermore, we have

$$H_m(\hat{\mathbf{v}})\hat{\mathbf{v}} = \mathbf{e}_1. \tag{76}$$

How do we choose the sign in the right-hand side of equation (75)? From a mathematical point of view such a choice is irrelevant: in both cases $H_m(\hat{\mathbf{v}})$ maps $\mathbf{v}$ into a vector whose only component different from zero is the first one. However, numerically it can be important. The square of the denominator in (75) is

$$\left|\left|\mathbf{v} \pm \|\mathbf{v}\|\,\mathbf{e}_1\right|\right|^2 = 2\,\|\mathbf{v}\|\,(\|\mathbf{v}\| \pm x_1)\,, \tag{77}$$

where x_1 is the first component of $\mathbf{v}$. If x_1 is comparable in magnitude to $\|\mathbf{v}\|$ and negative (positive) and we choose the plus (minus) sign, then the term

$$\|\mathbf{v}\| \pm x_1, \tag{78}$$

could be very small and cancellations with significant round-off errors may occur. Therefore, the Householder transformation to be implemented in computer code should be

$$H_m(\hat{\mathbf{v}}) = -\operatorname{sgn}(x_1)\,(I - 2\hat{\mathbf{u}}\hat{\mathbf{u}}^t)\,, \tag{79}$$

where

$$\hat{\mathbf{u}} = \frac{\hat{\mathbf{v}} + \operatorname{sgn}(x_1)\mathbf{e}_1}{\left|\left|\hat{\mathbf{v}} + \operatorname{sgn}(x_1)\mathbf{e}_1\right|\right|}. \tag{80}$$

The additional factor of $\operatorname{sgn}(x_1)$ in the right-hand side of equation (79) assures that there is no ambiguity in the sign of the right-hand side of equation (76). In turn, it guarantees that all the diagonal elements of the upper-triangular matrix R are positive. This is not the definition of Householder reflection used in standard QR decomposition routines. Usually,

$$H'_m(\hat{\mathbf{v}}) = I - 2\hat{\mathbf{u}}\hat{\mathbf{u}}^t, \tag{81}$$

with the same $\hat{\mathbf{u}}$ as in (75). Therefore,

$$H'_m(\hat{\mathbf{v}})\hat{\mathbf{v}} = \mp\mathbf{e}_1. \tag{82}$$

As a consequence, the signs of the diagonal elements of R are random. This is the reason why the output of black box QR decomposition routines must be modified in order to obtain orthogonal matrices with the correct distribution.

Besides being numerically stable, this algorithm has another advantage with respect to Gram-Schmidt orthonormalization. In most applications of numerical analysis Q need not be computed explicitly, only $Q\mathbf{w}$ does, where $\mathbf{w}$ is a specific vector. Generating all the Householder reflections is an $O(N^2)$ process and computing $H_N\mathbf{w}$ requires $O(N)$ operations—it just evaluates the scalar product $(\hat{\mathbf{u}}, \mathbf{w})$. Successively multiplying $H_N, \ldots, H_1$ into $\mathbf{w}$ is an $O(N^2)$ process. Therefore, it takes in total $O(N^2)$ operations to compute $Q\mathbf{w}$. Instead, Gram-Schmidt orthonormalization is an $O(N^3)$ process. However, if Q is explicitly needed, computing the product (68) requires $O(N^3)$ operations too.

The generalizations to $\mathrm{U}(N)$ and $\mathrm{Sp}(N)$ are straightforward. The only differences are in the definitions of the Householder reflections. A suitable choice for $\mathrm{U}(N)$ is

$$H_m(\hat{\mathbf{v}}) = -e^{-i\theta}\,(I - 2\hat{\mathbf{u}}\hat{\mathbf{u}}^*)\,. \tag{83}$$

The unit vector $\hat{\mathbf{u}}$ is

$$\hat{\mathbf{u}} = \frac{\hat{\mathbf{v}} + e^{i\theta}\mathbf{e}_1}{\|\hat{\mathbf{v}} + e^{i\theta}\mathbf{e}_1\|}, \tag{84}$$

where $\mathbf{v} = (x_1, \ldots, x_m) \in \mathbb{C}^m$ and $x_1 = e^{i\theta}\,|x_1|$. The matrix $H_m(\hat{\mathbf{v}})$ is unitary and

$$H_m(\hat{\mathbf{v}})\hat{\mathbf{v}} = \mathbf{e}_1. \tag{85}$$

Note that the introduction of $e^{i\theta}$ in equations (83) and (84) takes into account both the potential cancellations and the correct values of the arguments of the diagonal elements of the upper-triangular matrix R: equation (85) implies that all the r_{jj}s are real and strictly positive.

For $\mathrm{Sp}(N)$ we have

$$H_m(\hat{\mathbf{v}}) = -\overline{q}\,(I - 2\hat{\mathbf{u}}\hat{\mathbf{u}}^*)\,, \tag{86}$$

with

$$\hat{\mathbf{u}} = \frac{\hat{\mathbf{v}} + q\mathbf{e}_1}{\|\hat{\mathbf{v}} + q\mathbf{e}_1\|}, \tag{87}$$

where $\mathbf{v} = (x_1, \ldots, x_m) \in \mathbb{H}^m$ and $x_1 = q\,\|x_1\|$. Also in this case

$$H_m(\hat{\mathbf{v}})\hat{\mathbf{v}} = \mathbf{e}_1. \tag{88}$$

A Group Theoretical Interpretation

We now know how to generate random matrices from any of the classical compact groups $\mathrm{U}(N)$, $\mathrm{O}(N)$, and $\mathrm{Sp}(N)$. In order to achieve this goal, we have used little more than linear algebra. However simple and convenient this approach is (after all linear algebra plays a major role in writing most numerical algorithms), it hides a natural group theoretical structure behind the Householder reflections, which was uncovered by Diaconis and Shahshahani [1]. Indeed, generating a random matrix as a product of Householder reflections is only one example of a more general method that can be applied to any finite or compact Lie group. Our purpose in this section is to give a flavor of this perspective. For the sake of clarity, as before, we will discuss the orthogonal group $\mathrm{O}(N)$; the treatment of $\mathrm{U}(N)$ and $\mathrm{Sp}(N)$ is, once again, almost identical.

The need for a more general and elegant approach arises also if one observes that there is one feature of the QR decomposition that may not be entirely satisfactory to a pure mathematician: Why in order to generate a random point on an $N(N-1)/2$-dimensional manifold—$\mathrm{O}(N)$ in this case—do we need to generate N^2 random numbers? It does not look like the most efficient option, even if it is a luxury that can be easily afforded on today's computers.

We will first show how the key ideas that we want to describe apply to finite groups, as in this setting they are more transparent. Suppose that

we need to generate a random element g in a finite group Γ_N. In this context, if Γ_N has p elements, uniform distribution simply means that the probability of extracting any $g \in \Gamma_N$ is $1/p$. In addition, we assume that there exists a chain of subgroups of Γ_N:

$$\Gamma_1 \subset \Gamma_2 \subset \cdots \subset \Gamma_N. \tag{89}$$

In practical situations it is often easier to generate a random element $\tilde{g}$ in a smaller subgroup, say $\Gamma_{m-1} \in \Gamma_N$, than in Γ_N itself; we may also know how to take a random representative g_m in the left coset $C_m = \Gamma_m/\Gamma_{m-1}$. Now, write the decomposition

$$\Gamma_m \cong C_m \times \Gamma_{m-1}. \tag{90}$$

Once we have chosen a set of representatives of C_m, an element $g \in \Gamma_m$ is uniquely factorized as $g = g_m \tilde{g}$, where $g_m \in C_m$. If both g_m and $\tilde{g}$ are uniformly distributed in C_m and Γ_{m-1} respectively, then g is uniformly distributed in Γ_m.

We can apply this algorithm iteratively starting from Γ_1 and eventually generate a random element in Γ_N. In other words, we are given the decomposition

$$\Gamma_N \cong C_N \times \cdots \times C_2 \times \Gamma_1. \tag{91}$$

An element $g \in \Gamma_N$ has a unique representation as a product

$$g = g_N \cdots g_1, \tag{92}$$

where g_m is a representative in C_m. If the g_ms are uniformly distributed in C_m so is g in Γ_N. This is known as the *subgroup algorithm* [1].

This technique applies to random permutations of N letters. The chains of subgroups is

$$\{\mathrm{Id}\} \subset S_2 \subset \cdots \subset S_N, \tag{93}$$

where S_m is the m-th symmetric group. Other examples include generating random positions of Rubik's cube and random elements in $\mathrm{GL}(N, \mathbb{F}_p)$, where $\mathbb{F}_p$ is a finite field with p elements.

For $\mathrm{O}(N)$ the decompositions (91) and (92) are hidden behind the factorization (68) in terms of Householder reflections. Indeed, the subgroup algorithm for $\mathrm{O}(N)$ is contained in

Theorem 4. *Let $\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_N$ be uniformly distributed on $\mathbb{S}^0, \ldots, \mathbb{S}^{N-1}$ respectively, where*

$$\mathbb{S}^{m-1} = \left\{ \hat{\mathbf{v}}_m = (x_1, \ldots, x_m) \in \mathbb{R}^m \,\middle|\, \textstyle\sum_{j=1}^m x_j^2 = 1 \right\} \tag{94}$$

is the unit sphere in $\mathbb{R}^m$. Furthermore, let $H_m(\hat{\mathbf{v}})$ be the m-th Householder reflection defined in equation (79)*. The product*

$$O = H_N(\hat{\mathbf{v}}_N) H_{N-1}(\hat{\mathbf{v}}_{N-1}) \cdots H_2(\hat{\mathbf{v}}_2) H_1(\hat{\mathbf{v}}_1) \tag{95}$$

is a random orthogonal matrix with distribution given by Haar measure on $\mathrm{O}(N)$.

Proof. Suppose we construct $O \in \mathrm{O}(N)$ distributed with Haar measure by factorizing a matrix X in the Ginibre ensemble as described in the section "Householder Reflections". The random matrix O is the product of Householder reflections (68) and each factor $H_m(\hat{\mathbf{v}}_m)$ is a function of the unit vector $\hat{\mathbf{v}}_m \in \mathbb{S}^{m-1}$ only. We need to show that such $\hat{\mathbf{v}}_m$s are independent and uniformly distributed in $\mathbb{S}^{m-1}$ for $m = 1, \ldots, N$.

At each step in the construction of the upper-triangular matrix (67), the matrix multiplied by the m-th Householder reflection, i.e.,

$$X_m = H_m(\hat{\mathbf{v}}_m) \cdots H_{N-1}(\hat{\mathbf{v}}_{N-1}) H_N(\hat{\mathbf{v}}_N) X, \tag{96}$$

is still in the Ginibre ensemble. All its elements are, therefore, i.i.d. normal random variables. This is a consequence of the invariance

$$d\mu_{\mathrm{G}}(OX) = d\mu_{\mathrm{G}}(X), \quad O \in \mathrm{O}(N), \tag{97}$$

of the measure of the Ginibre ensemble. Now, $\hat{\mathbf{v}}_m = (x_1, \ldots, x_m)$ is constructed by taking the m-th dimensional vector $\mathbf{v}_m$ obtained by dropping the first $N-m$ elements of the $(N-m+1)$-th column of X_m. The components of $\mathbf{v}_m$ are i.i.d. normal random variables. It follows that the p.d.f. of $\mathbf{v}_m$ is

$$P(\mathbf{v}_m) = \frac{1}{\pi^{m/2}} \prod_{j=1}^m \exp\left(-x_j^2\right) \tag{98}$$

$$= \frac{1}{\pi^{m/2}} \exp\left(- \sum_{j=1}^m x_j^2 \right) = \frac{1}{\pi^{m/2}} \exp\left(-\|\mathbf{v}_m\|^2\right).$$

Since $P(\mathbf{v}_m)$ depends only on the length of $\mathbf{v}_m$, and not on any angular variable, the unit vector $\hat{\mathbf{v}}_m = \mathbf{v}_m / \|\mathbf{v}_m\|$ is uniformly distributed in $\mathbb{S}^{m-1}$ and is statistically independent of $\hat{\mathbf{v}}_k$ for $k \neq m$. □

Theorem 4 is more transparent than relying on the QR decomposition, which seems only a clever technical trick. If nothing else, the counting of the number of degrees of freedom matches. In fact, the dimension of the unit sphere $\mathbb{S}^{m-1}$ is $m-1$. Thus, the total number of independent real parameters is

$$\sum_{m=1}^N (m-1) = \frac{N(N-1)}{2}. \tag{99}$$

Why is theorem 4 the subgroup algorithm for $\mathrm{O}(N)$? As we shall see in theorem 5, the factorization (95) is unique—provided that we restrict to the definition (79) of the Householder reflections. This means that

$$\mathrm{O}(N) \cong \mathbb{S}^{N-1} \times \cdots \times \mathbb{S}^1 \times \mathrm{O}(1), \tag{100}$$

where

$$\mathrm{O}(1) \cong \mathbb{S}^0 = \{-1, 1\}. \tag{101}$$

If we proceed by induction, we obtain

$$\mathrm{O}(N) = \mathbb{S}^{N-1} \times \mathrm{O}(N-1). \tag{102}$$

Therefore, a matrix $O \in \mathrm{O}(N)$ admits a unique representation as

$$O = H_N(\hat{\mathbf{v}}_N)\Omega, \tag{103}$$

where

$$\Omega = \begin{pmatrix} 1 & 0 \\ 0 & \boxed{\widetilde{O}} \end{pmatrix} \tag{104}$$

and $\widetilde{O} \in \mathrm{O}(N-1)$. A consequence of theorem 4 is that if $\hat{\mathbf{v}}_N$ is uniformly distributed in $\mathbb{S}^{N-1}$ and $\widetilde{O}$ is distributed with Haar measure on $\mathrm{O}(N-1)$, then O is Haar distributed too. The final link with the subgroup algorithm is given by

Theorem 5. *The left coset space of* $\mathrm{O}(N-1)$ *in* $\mathrm{O}(N)$ *is isomorphic to* $\mathbb{S}^{N-1}$*, i.e.,*

$$\mathrm{O}(N)/\mathrm{O}(N-1) \cong \mathbb{S}^{N-1}. \tag{105}$$

A complete class of representatives is provided by the map [1] $H_N : \mathbb{S}^{N-1} \longrightarrow \mathrm{O}(N)$*,*

$$H_N(\hat{\mathbf{v}}) = \begin{cases} -\operatorname{sgn}(x_1)\,(I - 2\hat{\mathbf{u}}\hat{\mathbf{u}}^t) & \text{if } \hat{\mathbf{v}} \neq \mathbf{e}_1, \\ I_N & \text{if } \hat{\mathbf{v}} = \mathbf{e}_1, \end{cases} \tag{106}$$

where

$$\hat{\mathbf{u}} = \frac{\hat{\mathbf{v}} + \operatorname{sgn}(x_1)\mathbf{e}_1}{||\hat{\mathbf{v}} + \operatorname{sgn}(x_1)\mathbf{e}_1||} \tag{107}$$

and x_1 *is the first component of* $\hat{\mathbf{v}}$*.*

Proof. The group of $N \times N$ matrices Ω defined in equation (104) is isomorphic to $\mathrm{O}(N-1)$. Since

$$\Omega\mathbf{e}_1 = \mathbf{e}_1, \tag{108}$$

$\mathrm{O}(N-1)$ can be identified with the subgroup of $\mathrm{O}(N)$ that leave $\mathbf{e}_1$ invariant, i.e.,

$$\mathrm{O}(N-1) = \{O \in \mathrm{O}(N) \mid O\mathbf{e}_1 = \mathbf{e}_1\}. \tag{109}$$

Now, if two matrices O and O' belong to the same coset, then

$$O\mathbf{e}_1 = O'\mathbf{e}_1 = \hat{\mathbf{v}} \tag{110}$$

and vice versa. In other words, cosets are specified by where $\mathbf{e}_1$ is mapped. Furthermore, since $\|O\mathbf{e}_1\| = 1$, we see that they can be identified with the points in the unit sphere. Finally, the map (106) is one-to-one and is such that

$$H_N(\hat{\mathbf{v}})\mathbf{e}_1 = \hat{\mathbf{v}}. \tag{111}$$

Therefore, H_N spans a complete class of representatives. □

Incidentally, theorem 4 implies

Corollary 1. *Let* $d\mu_{\mathrm{O}(N)}$ *and* $d\mu_{\mathrm{O}(N-1)}$ *be the Haar measures on* $\mathrm{O}(N)$ *and* $\mathrm{O}(N-1)$ *respectively. Then*

$$d\mu_{\mathrm{O}(N)} = d\mu_{\mathbb{S}^{N-1}} \times d\mu_{\mathrm{O}(N-1)}, \tag{112}$$

where $d\mu_{\mathbb{S}^{N-1}}$ *is the uniform measure on* $\mathbb{S}^{N-1}$*.*

What is the meaning of $d\mu_{\mathbb{S}^{N-1}}$? Given that we are dealing with uniform random variables, it is quite natural that we end up with the uniform measure. In this case, however, it has a precise group theoretical interpretation. Left multiplication of the right-hand side of equation (103) by $O' \in \mathrm{O}(N)$ induces a map on the coset space:

$$O'H_N(\hat{\mathbf{v}})\Omega = H_N(\hat{\mathbf{v}}')\Omega'\Omega = H_N(\hat{\mathbf{v}}')\Omega''. \tag{113}$$

Since the decomposition (103) is unique the transformation $\hat{\mathbf{v}} \mapsto \hat{\mathbf{v}}'$ is well defined. This map can be easily determined. A coset is specified by where $\mathbf{e}_1$ is mapped, therefore

$$O'H_N(\hat{\mathbf{v}})\mathbf{e}_1 = O'\hat{\mathbf{v}} = \hat{\mathbf{v}}' = H_N(\hat{\mathbf{v}}')\mathbf{e}_1. \tag{114}$$

If $\hat{\mathbf{v}}$ is uniformly distributed on the unit circle so is $\hat{\mathbf{v}}' = O\hat{\mathbf{v}}$. Thus, $d\mu_{\mathbb{S}^{N-1}}$ is the unique measure on the coset space $\mathrm{O}(N)/\mathrm{O}(N-1)$ invariant under the left action of $\mathrm{O}(N)$. Its uniqueness follows from that of Haar measure and from the factorization (112).

Corollary 1 is a particular case of a theorem that holds under general hypotheses for topological compact groups. Indeed, let Γ be such a group, Ξ a closed subgroup and $C = \Gamma/\Xi$. Furthermore, let $d\mu_\Gamma$, $d\mu_C$ and $d\mu_X$ be the respective invariant measures, then

$$d\mu_\Gamma = d\mu_\Xi \times d\mu_C. \tag{115}$$

Acknowledgements

This article stems from a lecture that I gave at the summer school on Number Theory and Random Matrix Theory held at the University of Rochester in June 2006. I would like to the thank the organizers David Farmer, Steve Gonek, and Chris Hughes for inviting me. I am also particularly grateful to Brian Conrey, David Farmer, and Chris Hughes for the encouragement to write up the content of this lecture.

References

[1] P. Diaconis and M. Shahshahani, The subgroup algorithm for generating uniform random variables, *Prob. Eng. Inf. Sc.* **1** (1987), 15–32.

[2] F. M. Dyson, The threefold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics, *J. Math. Phys.* **3** (1962), 1199–1215.

[3] M. L. Eaton, *Multivariate Statistics: A Vector Space Approach*, Wiley and Sons, New York, NY, 1983.

[4] A. Edelman and N. R. Rao, Random matrix theory, *Acta Num.* **14** (2005), 233–297.

[5] R. M. Heiberger, Algorithm AS127. Generation of random orthogonal matrices, *App. Stat.* **27** (1978), 199–206.

[1] *The Householder reflections defined in equation* (79) *are not continuous at* $\mathbf{e}_1$*. Indeed, it can be proved that there is no continuous choice of coset representatives. In the section "Householder Reflections", this distinction was superfluous: if* $\mathbf{v}$ *is randomly generated, the probability that* $\mathbf{v} = \alpha\mathbf{e}_1$ *is zero.*

[6] N. M. KATZ and P. SARNAK, *Random matrices, Frobenius eigenvalues, and monodromy*, Amer. Math. Soc. Colloquium Publications, 45, Amer. Math. Soc., Providence, RI, 1999.
[7] J. P. KEATING and N. C. SNAITH, Random matrix theory and $\zeta(1/2+it)$, *Commun. Math. Phys.* **214** (2000), 57–89.
[8] ———, Random matrix theory and L-functions at $s = 1/2$, *Commun. Math. Phys.* **214** (2000), 91–110.
[9] M. L. MEHTA, *Random matrices*, Elsevier, San Diego, CA, 2004.
[10] *Recent perspectives in random matrix theory and number theory*, LMS Lecture Note Series, 322, (F. Mezzadri and N. C. Snaith, eds.), Cambridge University Press, Cambridge, 2005.
[11] H. L. MONTGOMERY, The pair correlation of zeros of the zeta function, *Analytic Number Theory: Proc. Symp. Pure Math. (St. Louis, MO, 1972)*, vol. 24, Amer. Math. Soc., Providence, RI, 1973, pp. 181–93.
[12] A. M. ODLYZKO, *The 10^{20}-th zero of the Riemann zeta function and 70 million of its neighbors,* 1989, `http://www.dtc.umn.edu/~odlyzko/unpublished/index.html`.
[13] M. RUBINSTEIN, Low-lying zeros of L-functions and random matrix theory, *Duke Math. J.* **109** (2001), 147–181.
[14] G. W. STEWART, The efficient generation of random orthogonal matrices with an application to condition estimators, *SIAM J. Num. Anal.* **17** (1980), 403–409.
[15] M. A. TANNER and R A. THISTED, A remark on AS127. Generation of random orthogonal matrices, *App. Stat.* **31** (1982), 190–192.
[16] R. W. M. WEDDERBURN, *Generating random rotations*, Research report, Rothamsted Experimental Station (1975).
[17] J. WISHART, The generalised product moment distribution in samples from a normal multivariate population, *Biometrika* **20A** (1928), 32–52.
[18] M. R. ZIRNBAUER, Riemannian symmetric superspaces and their origin in random-matrix theory, *J. Math. Phys.* **37** (1996), 4986–5018.
[19] K. ŻYCZKOWSKI and M. KUS, Random unitary matrices, *J. Phys. A: Math. Gen.* **27** (1994), 4235–4245.