Universality in numerical computations with random data

Percy A. Deift[a,1], Govind Menon[b], Sheehan Olver[c], and Thomas Trogdon[a]

[a]Courant Institute, New York University, New York, NY 10012; [b]Division of Applied Mathematics, Brown University, Providence, RI 02912; and [c]School of Mathematics and Statistics, The University of Sydney, Sydney, NSW 2006, Australia

Contributed by Percy A. Deift, July 16, 2014 (sent for review June 13, 2014)

The authors present evidence for universality in numerical computations with random data. Given a (possibly stochastic) numerical algorithm with random input data, the time (or number of iterations) to convergence (within a given tolerance) is a random variable, called the halting time. Two-component universality is observed for the fluctuations of the halting time—i.e., the histogram for the halting times, centered by the sample average and scaled by the sample variance, collapses to a universal curve, independent of the input data distribution, as the dimension increases. Thus, up to two components—the sample average and the sample variance—the statistics for the halting time are universally prescribed. The case studies include six standard numerical algorithms as well as a model of neural computation and decision-making. A link to relevant software is provided for readers who would like to do computations of their own.

random matrix theory | decision times | numerical analysis

In earlier work (2), P.A.D. and G.M. (together with C. Pfrang) considered the problem of computing the eigenvalues of a real $n \times n$ random symmetric matrix $M = (M_{ij})$. The authors considered matrices chosen from different ensembles E using a variety of different algorithms A (2). Let S_n denote the space of real $n \times n$ symmetric matrices. Standard eigenvalue algorithms involve iterations of isospectral maps $\varphi = \varphi_A$: $S_n \to S_n$, $\mathrm{spec}(\varphi_A(M)) = \mathrm{spec}(M)$ for $M \in S_n$. If $M \in S_n$ is given, one considers the sequence of matrices $M_{k+1} = \varphi(M_k)$, $k \geq 0$, with $M_0 = M$. Clearly, $\mathrm{spec}(M_{k+1}) = \mathrm{spec}(M_k) = \ldots = \mathrm{spec}(M)$, and under appropriate conditions $M_k = \varphi_A^{(k)}(M)$ converges to a diagonal matrix, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Necessarily, the λ_i's are the desired eigenvalues of M.

In ref. 2, the authors discovered the following phenomenon: For a given accuracy ϵ, a given matrix size n (ϵ small, n large, in an appropriate scaling range), and a given algorithm A, the fluctuations in the time to compute the eigenvalues to accuracy ϵ with the given algorithm A, were universal, independent of the choice of ensemble E. More precisely, Pfrang et al. (2) considered fluctuations in the deflation time T (the notion of deflation time is generalized to the notion of halting time in subsequent calculations). Recall that if an $n \times n$ matrix has block form

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

where M_{11} is $k \times k$ and M_{22} is $(n-k) \times (n-k)$ for some $1 \leq k \leq n-1$, then one says that the block diagonal matrix $\hat{M} = \mathrm{diag}(M_{11}, M_{22})$ is obtained from M by deflation. If $\|M_{12}\| = \|M_{21}\| \leq \epsilon$, then the eigenvalues $\{\lambda_i\}$ of M differ from the eigenvalues $\{\hat{\lambda}_i\}$ of $\hat{M}$ by $\mathcal{O}(\epsilon)$. Let $T = T_{\epsilon,n,A,E}(M)$ be the time (= number of steps = number of iterations of φ_A) it takes to deflate a random matrix M, chosen from an ensemble E, to order ϵ, using algorithm A, i.e., T is the smallest time such that for some k, $1 \leq k \leq n-1$, $\left\| \left(\varphi_A^{(T)}(M)\right)_{12} \right\| = \left\| \left(\varphi_A^{(T)}(M)\right)_{21} \right\| \leq \epsilon$.

As explained in ref. 2, T is a useful measure of the time required to compute the eigenvalues of M: generically, at worst, $\mathcal{O}(n)$ deflations are needed to compute the eigenvalues of M, and at best, $\mathcal{O}(\log n)$. The fluctuations $\tau_{\epsilon,n,A,E}(M)$ of T are defined by

$$\tau_{\epsilon,n,A,E}(M) = \frac{T_{\epsilon,n,A,E}(M) - \langle T_{\epsilon,n,A,E} \rangle}{\sigma_{\epsilon,n,A,E}}, \qquad [1]$$

where $\langle T_{\epsilon,n,A,E} \rangle$ is the sample average of $T_{\epsilon,n,A,E}(M)$ taken over matrices M from E, and $\sigma^2_{\epsilon,n,A,E}$ is the sample variance. For a given E, a typical sample size in ref. 2 was of order 5,000–10,000 matrices M, and the output of the calculations in ref. 2 was recorded in the form of a histogram for $\tau_{\epsilon,n,A,E}$.

Most of the calculations in ref. 2 concerned three eigenvalue algorithms: the QR algorithm, the QR algorithm with shifts (the version of QR used in practice), and the Toda algorithm. The QR algorithm is based on the factorization of a(n invertible) matrix M as $M = QR$, where Q is orthogonal and R is upper triangular with $R_{ii} > 0$. Given $M \in S_n$, with $M = QR$, $M' = \varphi_A(M) = \varphi_{\mathrm{QR}}(M) \equiv RQ$. Clearly, $M' = Q^T M Q \in S_n$ and $\mathrm{spec}(M') = \mathrm{spec}(M)$. Practical implementation of the QR algorithm requires the use of a shift, i.e., the QR algorithm with shifts (3). As shown in ref. 2, shifting does not affect universality. The Toda algorithm involves the solution $M(t)$ of the Toda equation $\frac{dM}{dt} = [B(M), M] = B(M)M - M\,B(M)$, where $B(M) = M_+ - M_+^T$, M_+ is the upper triangular part of M, and $M(t=0) = M$. For all $t > 0$, $\mathrm{spec}(M(t)) = \mathrm{spec}(M)$, and as $t \to \infty$, we again have $M(t) \to \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, where $\{\lambda_i\}$ are the eigenvalues of M. For the convenience of the reader, in Fig. 1 we reproduce, in particular, histograms for $\tau_{\epsilon,n,A,E}$, from ref. 2 for the QR algorithm (A = QR) with two different ensembles and varying values of n and ϵ.

From Fig. 1, we see that eigenvalue computation with the QR algorithm exhibits two-component universality, i.e., the fluctuations $\tau_{\epsilon,n,A,E}$ obey a universal law for all ensembles E under consideration. The same is true for all three algorithms considered in ref. 2. The laws are different, however, for different algorithms A.

In this paper, the work in ref. 2 has been extended in various ways as follows. All matrix ensembles are described in *Definition of Matrix Ensembles*.

The Jacobi Algorithm

In the first set of computations, the authors consider the eigenvalue problem for random matrices $M \in S_n$ using the Jacobi algorithm (4): for $M \in S_n$, choose $i < j$ such that $|M_{ij}| \geq \max_{1 \leq i' < j' \leq n} |M_{i'j'}|$, and

Significance

Universal fluctuations are shown to exist when well-known and widely used numerical algorithms are applied with random data. Similar universal behavior is shown in stochastic algorithms and also in an algorithm that models neural computation. The question of whether universality is present in all, or nearly all, computation is raised.

Author contributions: P.A.D., G.M., S.O., and T.T. designed research, performed research, analyzed data; and wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: deift@cims.nyu.edu.

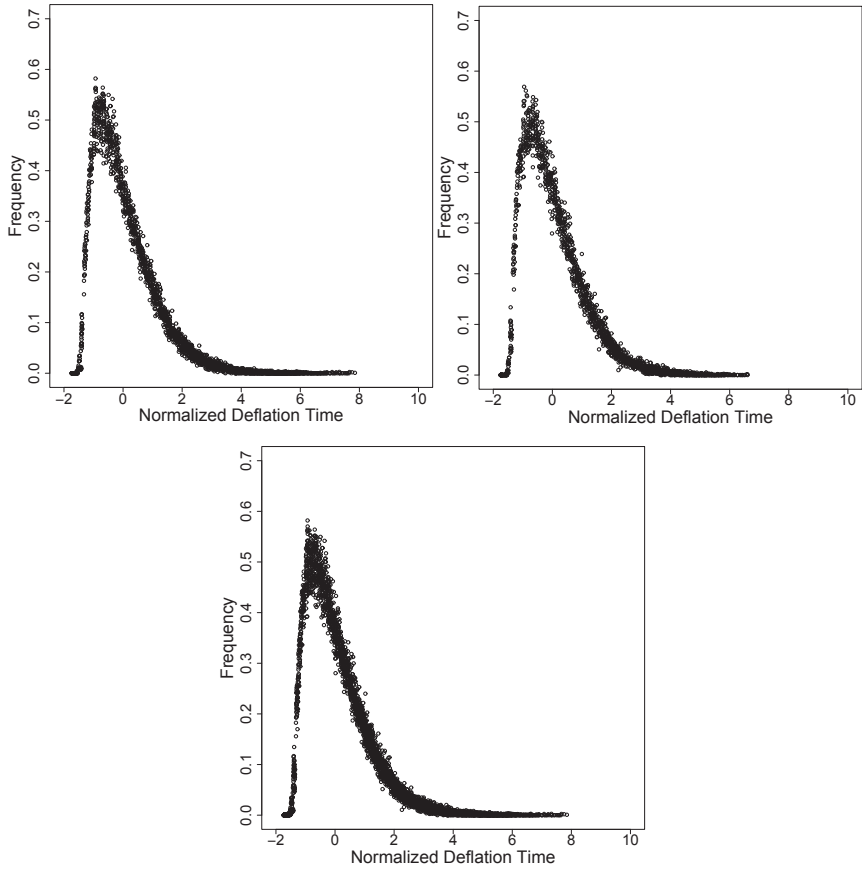

Fig. 1. The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when A = QR. Overlaid histograms demonstrate the collapse of the histogram of $\tau_{\epsilon,n,A,E}$ to a single curve. See *Definition of Matrix Ensembles* for the definitions of our choices for E. (*Upper Left*) E = GOE, and 40 histograms for $\tau_{\epsilon,\epsilon,A,E}$, are plotted one on top of the other for $\epsilon = 10^{-k}$, k = 2, 4, 6, 8, and n = 10, 30, . . .,190. The histograms are created with ∼10,000 samples. *Upper Right* contains same information as *Upper Left*, but for E = *BE*. (*Lower*) All 40 + 40 histograms are overlaid, and universality is evident: The data appears to follow a universal law for the fluctuations. Figure reprinted with permission from ref. 1.

let $G^{(ij)} \equiv G^{(ij)}(\theta)$ be the corresponding Givens rotation matrix: $G^{(ij)}_{i'j'} = \delta_{i'j'}$, for $i', j' \neq i, j$, and

$$\begin{bmatrix} G^{(ij)}_{ii} & G^{(ij)}_{ij} \\ G^{(ij)}_{ji} & G^{(ij)}_{jj} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}, \left(G^{(ij)}\right)^T G^{(ij)} = I.$$

Here, $\theta = \theta(M)$ is chosen so that $((G^{(ij)})^T M G^{(ij)})_{ij} = 0$ and then $\varphi_{\text{Jacobi}}(M) \equiv (G^{(ij)})^T M G^{(ij)}$. Clearly, $M' = \varphi_{\text{Jacobi}}(M) \in S_n$ and $\text{spec}(M') = \text{spec}(M)$ and again (4), $M_k = \varphi^{(k)}_{\text{Jacobi}}(M) \to \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$. The Jacobi algorithm has a very different character from QR/Toda-type algorithms, which are intimately connected to completely integrable Hamiltonian systems (ref. 5 and references therein).† Deflation, which is a useful measure for eigenvalue computation times for QR/Toda-type algorithms, is not useful for the Jacobi algorithm. In place of $T_{\epsilon,n,A,E}$, we record the halting time $k_{\epsilon,n,A,E}$: the number of iterations it takes for the Jacobi algorithm to reduce the Frobenius norm of the off-diagonal elements to be less than a given ϵ.‡ Histograms are produced for an appropriate analog of $\tau_{\epsilon,n,A,E}$:

$$\tau_{\epsilon,n,A,E}(M) = \frac{k_{\epsilon,n,A,E}(M) - \langle k_{\epsilon,n,A,E} \rangle}{\sigma_{\epsilon,n,A,E}}. \quad [2]$$

Computations for A = Jacobi are given in Fig. 2. Again, two-component universality is evident.

Ensembles with Dependent Entries

In all of the above cases, including the calculations for the Jacobi algorithm, the matrices M are real and the entries M_{ij} are independent, subject only to the symmetry requirement $M_{ij} = M_{ji}$. In the second set of computations in this paper, the authors consider $n \times n$ Hermitian matrices $M = M^*$ taken from various unitary ensembles (7) with probability distributions proportional to $e^{-n\text{tr}V(M)}dM$, where V: $\mathbb{R} \to \mathbb{R}$ grows sufficiently rapidly as $|x| \to \infty$, and dM is Lebesgue measure on the algebraically independent entries $M_{ij} = \text{Re } M_{ij} + \sqrt{-1}\ \text{Im } M_{ij}$ of M. Unless $V(x)$ is proportional to x^2, the entries of M for such ensembles are dependent, and it is a nontrivial matter to sample the matrices. A novel technique for sampling such unitary ensembles was introduced

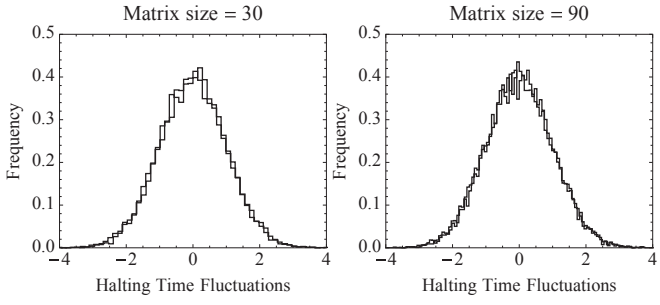

Fig. 2. The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when A = Jacobi, E = GOE, BE and $\epsilon = \sqrt{n}\ 10^{-10}$. (*Left*) Two histograms, one on top of the other, for GOE and BE, when n = 30. (*Right*) Same information for n = 90. All histograms are produced with 16,000 samples. We see two-component universality emerge for n sufficiently large: the histograms follow a universal (independent of E) law.

†The Jacobi algorithm is well-suited to parallel computation, and also has other advantages over QR in the context of modern, large-scale computation (5).

‡This criterion is sufficient to conclude that one element on the diagonal of the transformed matrix is within $\epsilon n^{-1/2}$ of an exact eigenvalue of the original matrix.

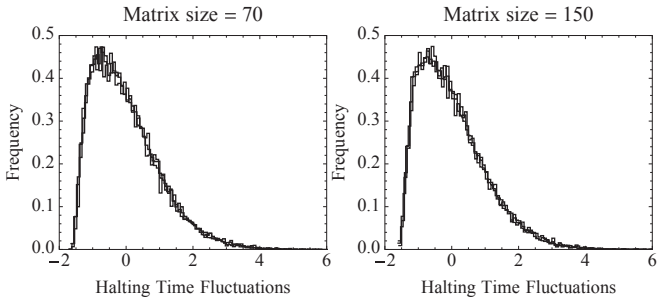

Fig. 3. The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when A = QR, E = QUE, COSH, GUE and $\epsilon = 10^{-10}$. Here we are using deflation time (or halting time), as in ref. 1. (*Left*) Three histograms, one each for GUE, COSH, and QUE, when $n = 70$. (*Right*) Same information for $n = 150$. All histograms are produced with 16,000 samples. Two-component universality emerges for n sufficiently large: the histograms follow a universal (independent of E) law, which is surprising because COSH and QUE have eigenvalue distributions that differ significantly from GUE in that they do not follow the so-called "semicircle law." In addition, these histograms appear to collapse to the same curve in Fig. 1, which is a further surprise, given the well-known fact that orthogonal and unitary ensembles give rise to different (eigenvalue) universality classes.

recently (8) by S.O. and T.T., together with N. R. Rao, taking advantage of the representation of the eigenvalues of M as a determinantal point process whose kernel is given in terms of orthogonal polynomials (9). Using this sampling technique, the authors of the present paper have considered the QR algorithm for various unitary ensembles.[§] Histograms for the halting (= deflation) time fluctuations $\tau_{\epsilon,n,A,E}$, $A = QR$, are given in Fig. 3 and, again, two-component universality is evident.

The Conjugate Gradient Algorithm

In a third set of computations in this paper, the authors start to address the question of whether two-component universality is just a feature of eigenvalue computation, or is present more generally in numerical computation. In particular, the authors consider the solution of the linear system of equations $Wx = b$, where W is real and positive definite, using the conjugate gradient (CG) method. The method is iterative (see ref. 10 and also Remark 1 below), and at iteration k of the algorithm, an approximate solution x_k of $Wx = b$ is found and the residual $r_k = Wx_k - b$ is computed. For any given $\epsilon > 0$, the method is halted when $\|r_k\|_2 < \epsilon$, and the halting time $k_\epsilon(W, b)$ recorded.[¶] The authors consider $n \times n$ matrices W chosen from two different positive definite ensembles E (see *Definition of Matrix Ensembles*) and vectors $b = (b_j)$ chosen independently with iid entries $\{b_j\}$. Given ϵ (small) and n (large), and $(W, b) \in E$, the authors record the halting time $k_{\epsilon,n,A,E}$, A = CG, and compute the fluctuations $\tau_{\epsilon,n,A,E}(W, b)$. The histograms for $\tau_{\epsilon,n,A,E}$ are given in Fig. 4 and, again, two-component universality is evident.

The Generalized Minimal Residual Algorithm

In a fourth set of computations, the authors again consider the solution of $Wx = b$, but here W has the form $I + X$, and $X \equiv X_n$ is a random, real nonsymmetric matrix and $b = (b_j)$ is independent with uniform iid entries $\{b_j\}$. Because $W = I + X$ is (almost surely) no longer positive definite, the CG algorithm breaks down, and the authors solve $(I + X)x = b$ using the generalized minimal residual (GMRES) algorithm (11). Again, the algorithm is iterative, and at iteration k of the algorithm an approximate solution x_k of $(I + X)x = b$ is found and the residual $r_k = (I + X)x_k - b$ is computed. As before, for any given $\epsilon > 0$, the method is halted when $\|r_k\|_2 < \epsilon$ and $k_{\epsilon,n,A,E}(X, b)$ is recorded. As in the conjugate gradient problem, the authors compute the histograms for the fluctuations of the halting time $\tau_{\epsilon,n,A,E}$ (2) for two ensembles E, where now A = GMRES. The results are given in Fig. 5, where again two-component universality is evident.

Remark 1. *The computations in the CG and GMRES sections are particularly revealing for the following reason. Both the CG and GMRES algorithms proceed by generating approximations x_n to the solution in progressively larger subspaces V_k of $\mathbb{R}^n$, $x_k \in V_k$, $\dim V_k = k$ (almost surely). These algorithms terminate in at most n steps, in the absence of rounding errors. If the matrix W in the case of CG, or $I + X$ in the case of GMRES, is too close to the identity, then the algorithm will converge in $\mathcal{O}(1)$ steps, essentially independent of n. However, if W or $I + X$ is too far from the identity, the algorithm will converge only after n steps (GMRES) or be dominated by rounding errors (CG). Thus, in both cases there are no meaningful statistics. What the calculations these sections reveal is that if the ensembles for CG and GMRES are such that the matrices W and $I + X$, respectively, are typically not too close to and not too far from the identity, then the algorithms exhibit significant statistical fluctuations, and two-component universality is immediately evident. (For further discussion, see legends for Figs. 4* and *5). Analogous considerations apply below.*

Discretization of a Random PDE

In a fifth set of computations, the authors raise the issue of whether two-component universality is just a feature of finite-dimensional computation, or is also present in problems that are intrinsically infinite dimensional. In particular, is the universality present in numerical computations for PDEs? As a case study, the authors consider the numerical solution of the Dirichlet problem $\Delta u = 0$ in a star-shaped region $\Omega \subset \mathbb{R}^2$ with $u = f$ on $\partial\Omega$. The boundary is described by a periodic function of the angle θ, $r = r(\theta)$, and similarly $f = f(\theta)$, $0 \le \theta \le 2\pi$. Two ensembles, Bernoulli–Dirichlet ensemble (BDE) and uniform Dirichlet ensemble (UDE; as described in *Definition of Matrix Ensembles*),

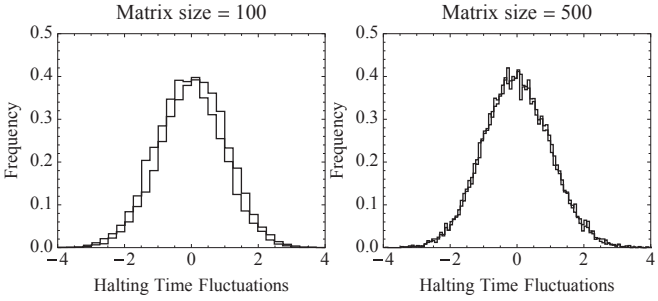

Fig. 4. The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when A = CG and E = cLOE, cPBE with $\epsilon = 10^{-10}$. (*Left*) Two histograms, one for cLOE and cPBE, when $n = 100$. (*Right*) Same information for $n = 500$. All histograms are produced with 16,000 samples. Two-component universality is evident for n sufficiently large: the histograms follow a universal (independent of E) law. The critical scaling (see *Definition of Matrix Ensembles*) has significant impact on the distribution of the condition number and forces $\langle \tau_{\epsilon,N,A,E} \rangle \sim n\alpha$, $\alpha < 1$. If the scaling $m = 2n$ is chosen in the ensemble E, then the CG method converges too quickly and the halting time tends to take only 10–15 different values for each value of m. No interesting limiting statistics are present. Conversely, if $m = n$, the CG method converges slowly ($\langle k_{\epsilon,m,A,E} \rangle \gg m$) and rounding errors dominate the computation. Experiments do not indicate two-component universality if $m = 2n$ or $m = n$. The scaling $m = n + 2\sqrt{n}$ identifies a critical scaling region. Within this scaling region, we see two-component universality emerge for n sufficiently large: the histograms follow a universal (independent of E) law.

[§]Here, $M = QR$ where Q is unitary and, again, R is upper triangular with $R_{ii} > 0$.

[¶]The notation $\|\cdot\|_2$ is used to denote the standard ℓ^2 norm on n-dimensional Euclidean space.

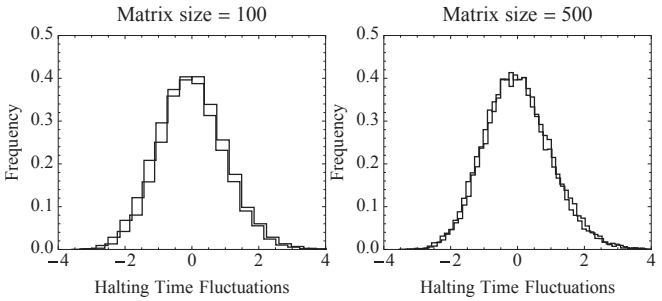

Fig. 5. The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when A = GMRES, E = cSGE, cSBE, and $\epsilon = 10^{-8}$. (*Left*) Two histograms, one for cSGE and one for cSBE, when $n = 100$. (*Right*) Same information for $n = 500$. All histograms are produced with 16,000 samples. The critically scaled ensembles cSBE and cSGE are of the form $I + X_n$ with $\|X_n\| \sim 2$. If the matrix is too close to the identity, the halting time will take almost constant values, i.e., $k_{\epsilon,n,A,E} = 8$, independent of n. If the matrix is too far from the identity, the fact that it is unstructured makes GMRES perform poorly and the algorithm typically completes in n steps, the maximum possible number of iterations (see Remark 1). With the proper scaling of X, we see two-component universality emerge for n sufficiently large: the histograms follow a universal (independent of E) law.

are derived from a discretization of the problem with specific choices for r, defined by a random Fourier series. The boundary condition f is chosen randomly by letting $\left\{f\left(\frac{2\pi j}{n}\right)\right\}_{j=0}^{n-1}$ be iid uniform on $[-1, 1]$. Histograms for the halting time $\tau_{\epsilon,n,A,E}$ from these computations are given in Fig. 6 and again, two-component universality is evident. What is surprising, and quite remarkable, about these computations is that the histograms for $\tau_{\epsilon,500,A,E}$ in this case are the *same* as the histograms for $\tau_{\epsilon,500,A,E}$ in Fig. 5 (see Fig. 6 for the overlaid histograms). In other words, UDE and BDE are structured with random components, whereas scaled shifted Ginibre ensemble (cSGE) and critically scaled shifted Bernoulli ensemble (cSBE) have no structure, yet they produce the same statistics (modulo two components).

A Genetic Algorithm

In all of the computations discussed so far, the randomness in the computations resides in the initial data.[‖] In the sixth set of computations, the authors consider an algorithm which is intrinsically stochastic. They consider a genetic algorithm to compute Fekete points (12, p. 142). Such points $P^* = (P_1^*, P_2^*, \ldots, P_N^*) \in \mathbb{R}^N$ are the global minimizers of the objective function

$$H(P) = \frac{2}{N(N-1)} \sum_{1 \le i \ne j \le N} \log|P_i - P_j|^{-1} + \frac{1}{N} \sum_{i=1}^{N} V(P_i)$$

for real-valued functions $V = V(x)$, which grow sufficiently rapidly as $|x| \to \infty$. It is well-known (12) that as $N \to \infty$, the counting measures $\delta_{P^*} = \frac{1}{N}\sum_{i=1}^{N} \delta_{P_i^*}$ converge to the so-called "equilibrium measure" μ_V, which plays a key role in the asymptotic theory of the orthogonal polynomials generated by measure $e^{-NV(x)}dx$ on $\mathbb{R}$. Genetic algorithms involve two basic components, mutation and cross-over. The authors implement the genetic algorithm in the following way.

The algorithm: Fix a distribution $\mathfrak{D}$ on $\mathbb{R}$. Draw an initial population $\mathcal{P}_0 = \mathcal{P} = \{P_i\}_{i=1}^n$ consisting of $n = 100$ vectors in $\mathbb{R}^N$, N large, with elements that are iid uniform on $[-4, 4]$. The random map $F_{\mathfrak{D}}(\mathcal{P}) : (\mathbb{R}^N)^n \to (\mathbb{R}^N)^n$ is defined by one of the following two procedures.

Mutation. Pick one individual $P \in \mathcal{P}$ at random (uniformly), and then pick two integers n_1, n_2 from {1, 2, …, N} at random (uniformly and independent). Three new individuals are created.

– $\tilde{P}_1$ — draw n_1 iid numbers $\{x_1, \ldots, x_{n_1}\}$ from $\mathfrak{D}$ and perturb the first n_1 elements of P: $(\tilde{P}_1)_i = (P)_i + x_i$, $i = 1, \ldots, n_1$, and $(\tilde{P}_1)_i = (P)_i$ for $i > n_1$.

– $\tilde{P}_2$ — draw $N - n_2$ iid numbers $\{y_{n_2+1}, \ldots, y_N\}$ from $\mathfrak{D}$ and perturb the last $N - n_2$ elements of P: $(\tilde{P}_2)_i = (P)_i + y_i$, $i = n_2 + 1, \ldots, N$, and $(\tilde{P}_2)_i = (P)_i$ for $i \le n_2$.

– $\tilde{P}_3$ — draw $|n_1 - n_2|$ iid numbers $\{z_1, \ldots, z_{|n_1-n_2|}\}$ from $\mathfrak{D}$ and perturb elements $n_1^* = 1 + \min(n_1, n_2)$ through $n_2^* = \max(n_1, n_2)$: $(\tilde{P}_3)_i = (P)_i + z_{i-n_1^*+1}$, $i = n_1^*, \ldots, n_2^*$, and $(\tilde{P}_3)_i = (P)_i$ for $i \notin \{n_1^*, \ldots, n_2^*\}$.

Cross-Over. Pick two individuals P, Q from $\mathcal{P}$ at random (independent and uniformly), and then pick two numbers n_1, n_2 from {1, 2, …, N} (independent and uniformly). Two new individuals are created.

– $\tilde{P}_4$ — Replace the n_1th element of P with the n_2th element of Q and perturb it (additively) with a sample of $\mathfrak{D}$.

– $\tilde{P}_5$ — Replace the n_1th element of Q with the n_2th element of P and perturb it (additively) with a sample of $\mathfrak{D}$.

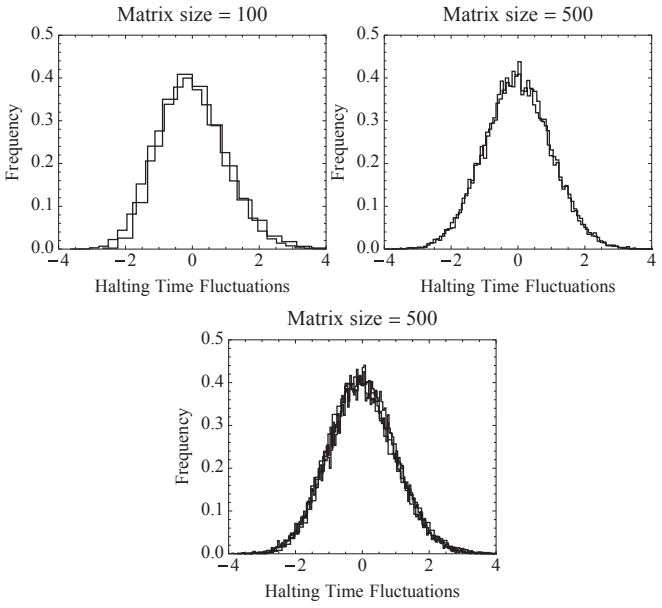

Fig. 6. The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when A = GMRES, E = UDE, BDE, and $\epsilon = 10^{-8}$. (*Upper Left*) Two histograms, one for UDE and one for BDE, when $n = 100$. (*Upper Right*) Same information for $n = 500$. (*Lower*) Four histograms, two taken from Fig. 5 (E = cSGE, cSBE) and two from *Upper Right* (E = UDE, BDE). All histograms are produced with 16,000 samples. It is interesting to note two properties. First, as we observe from our computations, BDE and UDE are of the form $I + X_n$, where X_n has a norm that grows proportional to some fractional power of n. Though this type of growth in the unstructured case of Fig. 5 would cause GMRES to take its maximum possible number of iterations, i.e., $k = n$, nevertheless, nontrivial statistics emerge. In light of Remark 1, we conjecture that structure is necessary for GMRES to perform well when the perturbation of the identity has an unbounded spectral radius in the large n limit. The second and most important feature is that two-component universality for matrices of the form $I + X_n$ persists as the computations are moved from structured randomness (UDE and BDE) to unstructured randomness (cSBE and cSGE): the histograms follow a universal (independent of E) law.

[‖]Aside from round-off errors, see Fig. 4 legend.

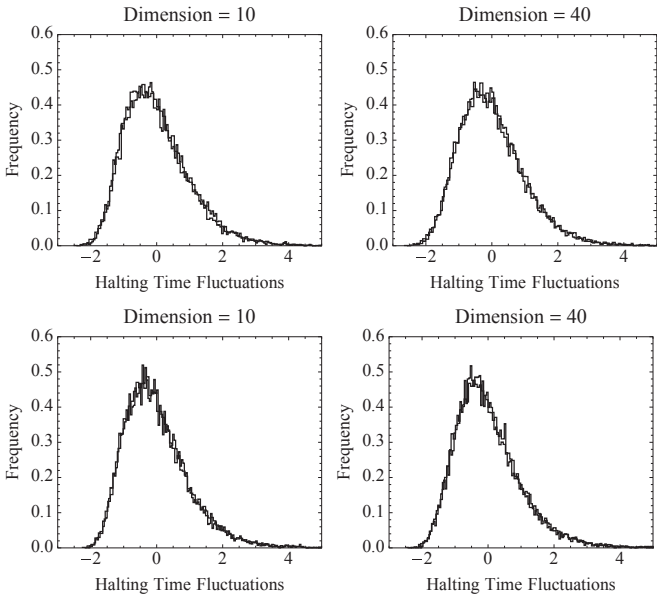

Fig. 7. The observation of two-component universality for $\tau_{\epsilon,N,A,E}$ when A = genetic, $\epsilon = 10^{-2}$, and $E \simeq \mathfrak{D}$, where $\mathfrak{D}$ is chosen to be either uniform on $[-1/(10N), 1/(10N)]$ or taking values $\pm 1/(10N)$ with equal probability. The top row is created with the choice $V(x) = x^2$ and the bottom row with $V(x) = x^4 - 3x^2$. (*Left*) Each of the plots display two histograms, one for each choice of $\mathfrak{D}$ when $n = 10$. (*Right*) Same information for $n = 40$. All histograms are produced with 16,000 samples. It is evident that the histograms collapse onto a universal curve, one for each V.

At each step, the application of either cross-over or mutation is chosen with equal probability. The new individuals are appended to $\mathcal{P}$ and $\mathcal{P} \mapsto \mathcal{P}' = F_{\mathfrak{D}}(\mathcal{P}) \in (\mathbb{R}^N)^n$ is constructed by choosing the 100 P_i's in $\tilde{\mathcal{P}}$ that yield the smallest values of $H(P)$.** The algorithm produces a sequence of populations $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_k, \ldots$ in $(\mathbb{R}^N)^n$, $\mathcal{P}_{k+1} = F_{\mathfrak{D}}(\mathcal{P})$, $n = 100$, and halts, with halting time recorded, for a given ϵ, when $\min_{P \in \mathcal{P}_k} H(P) - \inf_{P \in \mathbb{R}^N} H(P) < \epsilon$.

The histograms for the fluctuations $\tau_{\epsilon,N,A,E}$, with A = genetic are given in Fig. 7, for two choices of V, $V(x) = x^2$ and $V(x) = x^4 - 3x^2$, and different choices of $E \simeq \mathfrak{D}$. For $V(x) = x^2$, $\inf_{P \in \mathbb{R}^N} H(P)$ is known explicitly, and for $V(x) = x^4 - 3x^2$, $\inf_{P \in \mathbb{R}^N} H(P)$ is approximated by a long run of the genetic algorithm. As before, two-component universality is evident.

Curie–Weiss Model

In the seventh and final set of computations, the authors pick up on a common notion in neuroscience that the human brain is a computer with software and hardware; if this is indeed so, then one may speculate that two-component universality should certainly be present in some cognitive actions. Indeed, such a phenomenon is in evidence in the recent experiments of Bakhtin and Correll (13). In ref. 13, data from experiments with 45 human participants was analyzed. The participants are shown 200 pairs of images. The images in each pair consist of nine black disks of variable size. The disks in the images within each pair have approximately the same area so that there is no a priori bias. The participants are then asked to decide which of the two images covers a larger (black) area, and the time T required to make a decision is recorded. For each participant, the decision times for the 200 pairs are collected and the fluctuation histogram is tabulated.†† The experimental results are in good agreement with a dynamical Curie–Weiss model frequently used in describing decision processes (14), and because each of the 45 participants operates, presumably, in his or her own stochastic neural environment, this is a remarkable demonstration of two-component universality in cognitive action.

At its essence, the Curie–Weiss model is Glauber dynamics on the hypercube $\{-1, 1\}^N$ with a microscopic approximation of a drift-diffusion process. Consider N variables $\{X_i(t)\}_{i=1}^N$, $X_i(t) \in \{-1, 1\}$. The state of the system at time t is $X(t) = (X_1(t), X_2(t), \ldots, X_N(t))$. The transition probabilities are given through the expressions

$$\mathbb{P}(X_i(t + \Delta t) \neq X_i(t) | X(t) = x) = c_i(x)\Delta t + o(\Delta t),$$

where $c_i(x)$ is the spin flip intensity. The observable considered is $M(X(t)) = \frac{1}{N}\sum_{i=1}^N X_i(t) \in [-1, 1]$, and the initial state of the system is chosen so that $M(X(0)) = 0$, a state with no a priori bias, as in the case of the experimental setup. The halting (or decision) time for this model is $k = \inf\{t: |M(X(t))| \geq \epsilon\}$, the time at which the system makes a decision. Here, $\epsilon \in (0, 1)$ may not be small.

This model is simulated by first sampling an exponential random variable with mean $\lambda(t) = (\sum_i c_i(X(t)))^{-1}$ to find the time Δt at which the system changes state. Sampling the random variable Y, $\mathbb{P}(Y = i) = c_i(X(t))\lambda(t)$, $i = 1, 2, \ldots, N$ produces an integer j, determining which spin flipped. Define $X_i(t + s) \equiv X_i(t)$ if $s \in [0, \Delta t)$ for $i = 1, 2, \ldots, N$ and $X_i(t + \Delta t) \equiv X_i(t)$, $X_j(t + \Delta t) \equiv -X_j(t)$ for $i \neq j$. This procedure is repeated with t replaced by $t + \Delta t$ to evolve the system.

Central to the application of the model is the assumption on the statistics of the spin flip intensity $c_i(x)$. If one changes the basic statistics of the c_i's, will the limiting histograms for the fluctuations of k be affected as N becomes large? In response to this question, the authors consider the following choices for $E \simeq c_i(x)$ ($\beta = 1.3$): $c_i(x) = o_i(x) = e^{-\beta x_i M(x)}$ (the case studied in ref. 13, $c_i(x) = u_i(x) = e^{-\beta x_i (M(x) - M^3(x)/5)}$, or $c_i(x) = v_i(x) = e^{-\beta x_i (M(x) + M^8(x))}$. The resulting histograms for the fluctuations $\tau_{\epsilon,N,A,E}$ of T are given in Fig. 8. Once again, two-component universality is evident. Thus, the universality in the decision process models mirrors the universality observed among the 45 participants in the experiment of Bakhtin and Correll (13).

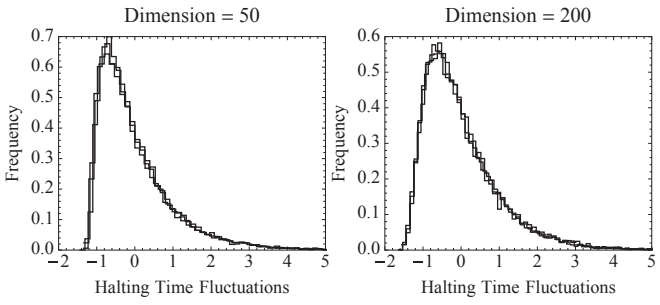

Fig. 8. The observation of two-component universality for $\tau_{\epsilon,N,A,E}$ when A = Curie–Weiss, $E \simeq o_i$, u_i, v_i, $\epsilon = 0.5$, and $\beta = 1.3$. (*Left*) Three histograms, one for each choice of E when $n = 50$. (*Right*) Same information for $n = 200$. All histograms are produced with 16,000 samples. The histogram for $E = o_i$ corresponds to the case studied in refs. 12 and 13. It is clear from these computations that the fluctuations collapse on to the universal curve for $E = o_i$. Thus, reasonable changes in the spin-flip intensity do not appear to change the limiting histogram, which indicates why the specific choice made in ref. 12 of $E = o_i$ is perhaps enough to capture the behavior of many individuals.

**After mutation we have $\tilde{\mathcal{P}} = \mathcal{P} \cup \{\tilde{P}_1, \tilde{P}_2, \tilde{P}_3\}$, and after cross-over, $\tilde{\mathcal{P}} = \mathcal{P} \cup \{\tilde{P}_4, \tilde{P}_5\}$.

††In ref. 12 the authors do not display the histogram for the fluctuations directly, but such information is easily inferred from their figures (see figure 6 in ref. 12).

Conclusions

Two distinct themes are combined in this work: (*i*) the notion of universality in random matrix theory and statistical physics and (*ii*) the use of random ensembles in scientific computing. The origin of both these ideas dates to the 1950s in the work of Wigner (7, 15) and von Neumann and Goldstine (16). There has been considerable progress in the rigorous understanding of universality in random matrix theory (refs. 17 and 18 and references therein). In contrast, the performance of numerical algorithms on random ensembles is less understood, although results in this area include probabilistic bounds for condition numbers and halting times for numerical algorithms (19–21).

The work presented here reveals empirical evidence for two-component universality in several numerical algorithms. The results of ref. 2 and the Jacobi, QR, CG, and GMRES sections reveal universal fluctuations of halting times for iterative algorithms in numerical linear algebra on random matrix ensembles with both dependent and independent entries. In each instance, the process of numerical computation on a random matrix may be viewed as the evolution of a random ensemble by a deterministic dynamical system. In a similar light, the genetic algorithm and *Curie–Weiss Model* may be seen as stochastic dynamical systems with that in *Curie–Weiss Model* having a close connection with neural computation. In all these examples, the empirical observations presented here suggest new universal phenomena in nonequilibrium statistical mechanics. The results from solving the Dirichlet problem reveal that numerical computations with a structured ensemble with some random components may have the same statistics (modulo two-components) as an unstructured ensemble, which brings to mind the situation in the 1950s when Wigner introduced random matrices as a model for scattering resonances of neutrons off heavy nuclei: the neutron–nucleus system has a well-defined and structured Hamiltonian, but nevertheless the resonances for neutron scattering are well-described statistically by the eigenvalues of an (unstructured) random matrix.

Materials and Methods

All algorithms discussed here are implemented in Mathematica. A package is available for download (21) that contains all relevant data and the code to generate this data. The package supports parallel evaluation for most algorithms and runs easily on personal computers.

Definition of Matrix Ensembles

Gaussian Ensembles. The Gaussian orthogonal ensemble is given by $(X+X^T)/\sqrt{4n}$, where X is an $n \times n$ matrix of standard iid Gaussian variables. The Gaussian unitary ensemble is given by $(X+X^*)/\sqrt{8n}$, where X is an $n \times n$ matrix of standard iid complex Gaussian variables.

Bernoulli Ensemble. The Bernoulli ensemble is given by an $n \times n$ matrix X consisting of iid random variables that take the values $\pm 1/\sqrt{n}$ with equal probability subject only to the constraint $X^T = X$.

Positive Definite Ensembles. The cLOE is given by $W = XX^T/m$, where X is an $n \times m$ matrix with standard iid Gaussian entries. The cPBE is given by $W = XX^T/m$, where X is an $n \times m$ matrix consisting of iid Bernoulli variables taking the values ± 1 with equal probability. In both cases, the critical scaling refers to the choice $m = n + 2\lfloor\sqrt{n}\rfloor$.

Shifted Ensembles. The cSBE is given by $I + X/\sqrt{n}$, where X is an $n \times n$ matrix consisting of iid Bernoulli variables taking the values ± 1 with equal probability. The cSGE is given by $I + X/\sqrt{n}$, where X is an $n \times n$ matrix of standard iid Gaussian variables. With this scaling, $\mathbb{P}(|\|X/\sqrt{n}\| - 2| > \epsilon) \to 0$ as $n \to \infty$ (22).

Unitary Ensembles. The Quartic unitary ensemble (QUE) is a complex, unitary ensemble with probability distribution proportional to $e^{-n\mathrm{tr}M^4}dM$. The Cosh unitary ensemble (COSH) has its distribution proportional to $e^{-\mathrm{tr}\,\cosh\,M}dM$.

Dirichlet Ensembles. We consider the numerical solution of the equation $\Delta u = 0$ in Ω and $u = f$ on $\partial\Omega$. Here, we let Ω be the star-shaped region interior to the curve $(x, y) = (r(\theta)\cos(\theta), r(\theta)\sin(\theta))$, where $r(\theta)$ for $0 \le \theta < 2\pi$ is given by $r(\theta) = 1 + \sum_{j=1}^{m}(X_j \cos(j\theta) + Y_j \sin(j\theta))$, and X_j and Y_j are iid random variables on $[-1/(2m), 1/(2m)]$. The boundary integral equation

$$\pi u(P) - \int_{\partial\Omega} u(P)\frac{\partial}{\partial n_Q}\log|P-Q|dS_Q = -f(P), \quad P \in \partial\Omega$$

is solved by discretizing in θ with n points and applying the trapezoidal rule with $n = 2m$ (23). For the BDE, X_m and Y_m are Bernoulli variables taking values $\pm 1/(2m)$ with equal probability. For the UDE, X_m and Y_m are uniform variables on $[-1/(2m), 1/(2m)]$.

ACKNOWLEDGMENTS. Support for this work was provided in part by National Science Foundation Grants NSF-DMS-130318 (to T.T.), NSF-DMS-1300965 (to P.A.D.), and NSF-DMS-07-48482 (to G.M.) and the Australian Research Council through the Discovery Early Career Research Award (to S.O.).

1. Pfrang C, Deift P, Menon G (2014) How long does it take to compute the eigenvalues of a random symmetric matrix? *Random Matrices* (MSRI Publications, Cambridge Univ Press, Cambridge, UK), Vol 65, in press.
2. Parlett BN (1998) *The Symmetric Eigenvalue Problem* (Soc for Industrial and Applied Mathematics, Philadelphia).
3. Golub GH, Loan CFV (2013) *Matrix Computations* (Johns Hopkins Univ Press, Baltimore).
4. Deift P, Nanda T, Tomei C (1983) Ordinary differential equations and the symmetric eigenvalue problem. *SIAM J Numer Anal* 20(1):1–22.
5. Demmel J, Veselić K (1992) Jacobi's method is more accurate than QR. *SIAM J Matrix Anal Appl* 13(4):1204–1245.
6. Mehta ML (2004) *Random Matrices* (Academic, New York).
7. Olver S, Nadakuditi RR, Trogdon T (2014) Sampling unitary ensembles. arXiv: 1404.0071.
8. Li XH, Menon G (2013) Numerical solution of Dyson Brownian motion and a sampling scheme for invariant matrix ensembles. *J Stat Phys* 153(5):801–812.
9. Saad Y (2003) *Iterative Methods for Sparse Linear Systems* (Soc for Industrial and Applied Mathematics, Philadelphia).
10. Saad Y, Schultz MH (1986) GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J Sci Stat Comput* 7(3):856–869.
11. Saff EB, Totik V (1997) *Logarithmic Potentials with External Fields* (Springer, New York).
12. Bakhtin Y, Correll J (2012) A neural computation model for decision-making times. *J Math Psychol* 56(5):333–340.
13. Bakhtin Y (2011) Decision making times in mean-field dynamic Ising model. Ann. *Henri Poincaré* 13(5):1291–1303.
14. Wigner EP (1951) On the statistical distribution of the widths and spacings of nuclear resonance levels. *Math Proc Camb Philos Soc* 47(4):790–798.
15. Goldstine HH, von Neumann J (1951) Numerical inverting of matrices of high order, part II. *Proc Amer Math Soc* 53(11):188–202.
16. Deift P, Gioev D (2009) *Random Matrix Theory: Invariant Ensembles and Universality*. Courant Lecture Notes in Mathematics (Amer Math Soc, Providence, RI), Vol 18.
17. Erdős L, Yau H (2012) Universality of local spectral statistics of random matrices. *Bull Am Math Soc* 49(3):377–414.
18. Demmel J (1988) The probability that a numerical analysis problem is difficult. *Math Comput* 50(182):449–480.
19. Edelman A (1988) Eigenvalues and condition numbers of random matrices. *SIAM J Matrix Anal Appl* 9(4):543–560.
20. Smale S (1985) On the efficiency of algorithms of analysis. *Bull Am Math Soc* 13(2): 87–121.
21. Trogdon T (2014) Numerical universality. Available at https://bitbucket.org/trogdon/numericaluniversality.
22. Geman S (1980) A limit theorem for the norm of random matrices. *Ann Probab* 8(2): 252–261.
23. Atkinson K, Han W (2009) *Theoretical Numerical Analysis* (Springer, New York).