Universality in numerical computation.
A case study: Eigenvalue computation

About 8 or 9 years ago, Christian Pfrang,
([PDM], published 2014)
Gravind Menon and P.D. initiated a statistical study
of the performance of various standard algorithms A to
calculate the eigenvalues of random real symmetric matrices
H. Let $\Sigma_N$ denote the set of real $N \times N$ symmetric
matrices. Associated with each algorithm $A$, there is, in
the discrete case such as QR, a map $\phi = \phi_A: \Sigma_N \rightarrow \Sigma_N$
with the properties

- (isospectral) $\text{spec}(\phi_A(X)) = \text{spec}(H)$
- (convergence) The iterates $X_{k+1} = \phi_A(X_k)$, $k > 0$,
  $X_0 = H$ given, converge to a diagonal matrix $X_\infty$.
  $X_k \to X_\infty$ as $k \to \infty$. 
and in the continuum case, such as Toda, there is a flow
\[ t \mapsto X(t) \in \Sigma_N \]
with the properties

* (isospectral) \( \text{spec} (X(t)) \) is constant

* (convergence) the flow \( X(t), t > 0, X(0) = H \) given, converges to a diagonal matrix \( X_\infty \), \( X(t) \to X_\infty \) as \( t \to \infty \).

In both cases, necessarily, the (diagonal) entries of \( X_\infty \) are the eigenvalues of the given matrix \( H \).

Given \( \varepsilon > 0 \), it follows, in the discrete case, that
for some \( m \), the off-diagonal entries of \( X_m \) are \( O(\varepsilon) \) and hence the diagonal entries of \( X_m \) give the eigenvalues of \( X_\infty = H \) to \( O(\varepsilon) \). The situation is similar for continuous algorithms \( t \mapsto X(t) \). Rather than running the algorithm until all the off-diagonal entries are \( O(\varepsilon) \), it is customary to run the algorithm with deflations as follows,
For an $N \times N$ matrix $Y$ in block form

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}$$

with $Y_{11}$ of size $k \times k$ and $Y_{22}$ of size $N-k \times N-k$ for some $k \in \{1, 2, \ldots, N-1\}$, the process of projecting $Y_{11}$ along $Y_{12}$ is called deflation. For a given $\varepsilon$, algorithm $A$ and matrix $H \in \Sigma_N$, define the $k$-deflation time $T^{(k)}(H) = \min_{X^k} T^{(k)}(H)$, $1 \leq k \leq N-1$, to be the smallest value of $m$ such that $X^m$, the $m$th iterate of algorithm $A$ with $X_0 = H$, has block form

$$X^m = \begin{bmatrix} X^{(k)}_{11} & X^{(k)}_{12} \\ X^{(k)}_{21} & X^{(k)}_{22} \end{bmatrix}$$

with $X^{(k)}_{11}$ of size $k \times k$ and $X^{(k)}_{22}$ of size $N-k \times N-k$ and $\|X^{(k)}_{12}\| = \|X^{(k)}_{21}\| \leq \varepsilon$. The deflation time $T(H)$ is then defined as

$$T(H) = T_{e,A}(H) = \min_{1 \leq k \leq N-1} T^{(k)}(H).$$
If \( k = 1, \ldots, N-1 \) is such that \( T(H) = T_{e,i}(H) \), it follows that the eigenvalues of \( H = X_0 \) are given by the eigenvalues of the block-diagonal matrix \( \text{diag}(X_{11}^{(k)}, X_{22}^{(k)}) \) to \( O(\varepsilon) \). After running the algorithm to time \( T_{e,i}(H) \), the algorithm restarts by applying the basic algorithm \( A \) separately to the smaller matrices \( X_{11}^{(k)} \) and \( X_{22}^{(k)} \) until the next deflation time, and so on. There are again similar considerations for continuous algorithms.

As the algorithm proceeds, the number of matrices after each deflation doubles. This is counterbalanced by the fact that the matrices are smaller and smaller in size, and the calculations are clearly parallelizable. Allowing for parallel computation, the number of deflations to compute all the eigenvalues of a given matrix \( H \) to a given accuracy \( \varepsilon \), will vary from
\[ O(\log N) \rightarrow O(N), \]

In the work of Dorrance et al. The authors considered the deflation time \( T = T_e, a \) for \( N \times N \) matrices chosen from a given ensemble \( \mathcal{E} \). Henceforth in these lectures we suppress the dependence on \( e, N, A \), and \( \mathcal{E} \), and simply write \( T \) with these variables understood. For a given algorithm \( A \) and ensemble \( \mathcal{E} \), the authors computed \( T(H) \) for 5,000 - 15,000 samples of matrices \( H \) chosen from \( \mathcal{E} \), and recorded the normalized deflation time

\[ \overline{T(H)} = \frac{T(H) - \langle T \rangle}{\sigma} \]

where \( \langle T \rangle \) and \( \sigma^2 = \langle (T - \langle T \rangle)^2 \rangle \) are the sample average and sample variance of \( T(H) \), respectively. Surprisingly, the authors found that for the given algorithm \( A \), and \( e \) and \( N \) in a suitable scaling range with \( N \to \infty \), the histogram was
of $\bar{T}$ was universal, independent of the ensemble $\mathcal{E}$. In other words, the fluctuations in the deflation time $\bar{T}$, suitably scaled, were universal, independent of $\mathcal{E}$. The following figure displays, in a slightly different form, some of the numerical results from P特朗g et al. Figure 1 (a) displays data for the QR algorithm, which is discrete, and Figure 1 (b) displays data for the Toda algorithm, which is continuous.

![Figure 1](image_url)

**Figure 1**: Universality for $\bar{T}$ when (a) $A$ is the QR eigenvalue algorithm and when (b) $A$ is the Toda algorithm. Panel (a) displays the overlay of two histograms for $\bar{T}$ in the case of QR, one for each of the two ensembles $\mathcal{E} = \text{BE}$, consisting of iid mean-zero Bernoulli random variables (see Definition A.1) and $\mathcal{E} = \text{GOE}$, consisting of iid mean-zero normal random variables. Here $\epsilon = 10^{-10}$ and $N = 100$. Panel (b) displays the overlay of two histograms for $\bar{T}$ in the case of the Toda algorithm, and again $\mathcal{E} = \text{BE}$ or GOE. And here $\epsilon = 10^{-8}$ and $N = 100$. 
Subsequently, Govind Menon, Sheehan Olver, Tom Trogdon and I raised the question of whether the universality results of Phang et al. were limited to eigenvalue algorithms for real symmetric matrices, or whether they were present more generally in numerical computation. And indeed, they found similar universality results for a wide variety of numerical algorithms, including:

(a) other algorithms such as the QR algorithm with shifts, the Jacobi eigenvalue algorithm, and also algorithms applied to complex Hermitian ensembles

(b) the conjugate gradient and GMRES algorithms to solve linear system $Hx = b$ with $H$ and $b$ random

(c) an iterative algorithm to solve the Dirichlet problem $\Delta u = 0$ on a random star-shaped region $\Omega \subset \mathbb{R}^2$ with random boundary data $f$ on $\partial \Omega$, and
(d) a genetic algorithm to compute the equilibrium measure for orthogonal polynomials on the line.

In [DMHT] the authors also discussed similar universality results obtained by Balasubramanian and Correll in a series of experiments with live participants recording.

e) decision making times for a specific task.

Whereas (a) and (b) concern finite dimensional problems, (c) shows that universality is also present in problems that are genuinely infinite dimensional. And whereas (a), (b), and (c) concern, in effect, deterministic dynamical systems acting on random initial data, problem (d) shows that universality is also present in genuinely stochastic algorithms.

The demonstration of universality in problems (a)–(d) raised the following issue: Given the common view of neuroscientists...
That the brain is just a big computer with hardware and software, one should be able to find evidence of universality in some neural computations. It is this issue that the authors in DMOTJ refer to the work of Bakhtin and Correll.

All of the above results are numerical. In order to establish universality as a bona fide phenomenon in numerical analysis, and not just an artifact, suggested, however strongly, by certain computations as above, D. and Trojedon [DT 2017(1)] sought out and proved universality for an algorithm of interest. The algorithm they analyzed was the Toda eigenvalue algorithm to compute the largest eigenvalue of a random real symmetric (alternatively, Hermitian) matrix. The goal of these lectures is to describe this work in detail. In
Subsequent work D. and Tropp also proved universality for other eigenvalue algorithms, including QR on sample covariance matrices. We will also describe this work, but in less detail.

We now describe the Toda and QR algorithms. Here we focus on the case where the matrices $H$ are real symmetric, i.e., $H \in \Sigma_n$. The Hermitian case will be considered in later lectures.

The Toda equations have the form

$$\frac{dX}{dt} = [X, B(X)] = X B(X) - B(X) X, \quad X(0) = H$$

where $B(X) = X_+ - X_-^T = -B(X)^T$, where $X_-$ is the strictly lower triangular part of $X$. These equations have a unique local solution, $X(t)$, $0 \leq t < t^*$ for some $0 < t^* \leq \infty$. 
Now for $0 \leq t < t^*$

\[
\frac{d}{dt} \begin{bmatrix} x^2 \\ x \end{bmatrix} = \frac{d}{dt} x^2 + x \frac{d}{dt} x \\
= (x B - B x) x + x (x B - B x) \\
= [x^2, B]
\]

so

\[
\frac{d}{dt} x^2 = [x^2, B] \\
0 \leq t < t^*
\]

Also as $B(x) = -B x^T$, we have

\[
\frac{d}{dt} x^T = [x^T, B] = B^T x^T - x^T B^T \\
= x^T B - B x^T \\
= [x^T, B(x)] \\
0 \leq t < t^*
\]

But for $X(0) = H \in \Sigma_N$

\[
X^T(0) = H^T = H
\]

Thus as $X^T(t)$ and $X(t)$ solve the same ode, and $X^T(0) = X(0)$, we must have for $0 \leq t < t^*$

\[
(1.1) \quad X(t) = x(t)^T
\]

is the Toda equations preserve $\Sigma_N$. From (1.1), as the trace of
a commutator is always zero, we have
\[
\frac{\partial}{\partial t} X^k = 0
\]
(12.1) Thus \( \frac{\partial}{\partial t} X(t) = \frac{\partial}{\partial t} U^2 \).

But as \( X = X^T \), this implies
\[
\sum_{i,j=1}^N X_{ij}(t) = \sum_{i,j=1}^N U_{ij}.
\]
(12.2)

In particular
\[
| X_{ij}(t) | \leq \sqrt{tr} \; H < \infty \quad \forall i, j, \quad 0 \leq t \leq t^*.
\]
(12.3)

which gives an a priori bound on the entries of \( X(t) \).

It follows by standard ode techniques that in fact
\[ t^* = \infty, \quad \text{i.e.} \quad (10.1) \quad \text{has a unique global solution} \quad X(t) = X(t)T. \]

By induction, one sees as in (11.1) that
\[
\frac{\partial}{\partial t} X^k = [X^k, B(x)]', \quad k = 1, 2, \ldots
\]
and so
\[
\frac{\partial}{\partial t} X^k(t) = X^k(0),
\]
(12.4)

which implies that the eigenvalues of \( X(t) \) are constants of the motion.

Furthermore,
\[
t \rightarrow X(t) \quad \text{is isospectral}
\]
(12.5)
But more is true.

Let $Q(t)$, $t > 0$, $Q(0) = I$ be the solution of the equation

$$\frac{dQ}{dt} = QBX(t), \quad Q(0) = I$$

where $X(t)$ solves (10.1). As (13.1) is linear, the solution of (13.1) is unique and global.

Then

$$\frac{d}{dt} \dot{Q} \dot{Q}^T = \dot{Q} \dot{Q}^T + \dot{Q} \dot{Q}^T$$

$$-QB \dot{Q}^T + Q(-B) \dot{Q}^T = 0$$

and no

$$Q(t)Q(t)^T = \text{const} = I$$

is

(13.2) $Q(t)$ is orthogonal for all $t > 0$

Now set

$$\ddot{X}(t) = Q(t) H Q(t)$$

Then

$$\frac{d}{dt} \ddot{X} = \dot{Q}^T H Q + \dot{Q}^T H \dot{Q} = -B \dot{Q}^T H Q + \dot{Q}^T H \dot{Q} \dot{X}$$

$$= \left[ \dddot{X}, B(X) \right]$$
so that $\tilde{X}(t)$ solves the same equation as $X(t)$, and $\tilde{X}(0) = H = X(0)$ and so $\tilde{X}(t) = X(t)$.

Thus

$$X(t) = Q(t)^T H Q(t)$$

So we see that not only is $\psi^{-1}X(t)$ iso-spectral as in (12.5), but in fact

$$X(t) \text{ is orthogonally equivalent to } X(0) = H, \quad t \geq 0.$$

In a later lecture we will use an argument of Moser [Mos] to show that indeed

$$X(t) \to X_0 \quad \text{as } t \to \infty$$

where

$$X_0 = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

As noted earlier, the $\lambda_i$'s are the eigenvalues of $X(0) = H$.

We will also show that generically, the flow (10.13) is

$$\text{sorting } \lambda_1 > \lambda_2 > \cdots > \lambda_n.$$
As concluded in Deift-Nanda-Tomei (DNT 1983), the Toda flow gives rise to an eigenvalue algorithm, in order to compute the eigenvalues of a given matrix \( A \in \mathbb{R}^{n \times n} \), solve (10.1) for \( X(t) \) with \( X(0) = H \). Then \( \alpha \mapsto \alpha \)

\[
X(t) \rightarrow X_\alpha = \text{diag} (\lambda_1, \ldots, \lambda_n)
\]

where \( \lambda_i \) are the eigenvalues of \( H \). This is the Toda algorithm.

We will also show that (10.11) is a completely integrable Hamiltonian flow which can be integrated explicitly.

The history of the Toda system, or Toda lattice, is as follows. The lattice was introduced by H. Toda in 1967 and describes the motion of \( N \) particles \( x_i, i=1,2,\ldots,N \), on the line under the Hamiltonian

\[
H_{\text{Toda}}(x_i,y_i) = \frac{1}{2} \sum_{i=1}^{N} y_i^2 + \frac{1}{2} \sum_{i=1}^{N-1} e^{x_i-x_{i+1}}.
\]
In 1974 Flaschka (and independently Manakov) showed that Hamilton's equations

\[ x_i = \frac{\partial H_{\text{oda}}}{\partial y_i}, \quad y_i = -\frac{\partial H_{\text{oda}}}{\partial x_i} \]

can be written in so-called Lax pair form (10.17) when \( X = X(t) \) is tridiagonal and

\[
(16.1) \quad \begin{cases} 
 X_{i+1} = X_i, 
 X_{i+1} = \frac{1}{2} e^{\frac{1}{2} (x_i - x_{i-1})}, 
 i = 1, \ldots, N - 1 
\end{cases}
\]

and \( B(X) \) is the (tridiagonal) skew-symmetric matrix

\[
B(X) = X - X^T \] as above. Not only, as noted by Flaschka, is the flow \( t \to X(t) \) isospectral, so that the eigenvalues \( \lambda_i(t) = \lambda_i(t = 0), i = 1, \ldots, N \) are constants of motion for the Toda flow, but the \( \lambda_i \)'s are independent and Poisson commute in the underlying symplectic structure,
The Toda lattice is integrable in the sense of Liouville. In later work, Moser showed how to solve the Toda lattice explicitly and he also showed how to evaluate the long-time behavior of the system. The Toda system (10.1) is the natural extension of the original tridiagonal Toda lattice to full $N \times N$ matrices. Although it is not a priori clear, this extended system is also Hamiltonian and completely integrable in the sense of Liouville. We will derive and explain all the above properties of the Toda system in later chapters.

We note here that it is a basic mantra in the modern theory of integrable systems, that if a dynamical
system can be written in lax-pair form

$$\frac{dS}{dt} = [S, U]$$

for some \( U = U(s) \), then the flow \( t \to S(t) \) is isopectral so that the system has (at least) \( N \) integrals of the motion, where \( N = \dim S \). This follows in general, by the argument due to Lax—following (13.1),

The QR algorithm works in the following way.

Let \( X_0 \) be an invertible matrix in \( \Sigma_N \). Then \( X_0 \) has a QR factorization

(18.1) \[ X_0 = Q_0 R_0 \]

where \( Q_0 \) is orthogonal, \( Q_0 Q_0^T = Q_0^T Q_0 = I \), and \( R_0 \) is upper triangular with \( (R_0)_{ii} > 0 \), \( i = 1, \ldots, N \). The factorization is unique.
Set

\[ X_1 = R_0 Q_0 \]

Substituting \( R_0 = Q_0^T X_0 \) from (18.1) we see that

\[ X_1 = Q_0^T X_0 Q_0 \]

Denote the mapping

\[ X \rightarrow Q R \rightarrow R Q = X' \]

By \( \Phi_{ar} : \omega \rightarrow X' = \Phi_{ar}(X) \). From (19.2) we see that

\[ \Phi_{ar} : \Sigma_N \rightarrow \Sigma_N \text{ is isospectral} \]

Now \( X_1 \) has its own \( Q R \) factorization,

\[ X_1 = Q_1 R_1 \]

Set

\[ X_k = R_k, Q_k \]

\[ = Q_k^T X_k Q_k \]

etc.

In this way we obtain an isospectral sequence

\[ X_0, X_1, \ldots, X_k, \ldots \]

of matrices. Generically \( X_k \) converges as \( k \to \infty \) to a
diagonal matrix $X_0 = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Again, necessarily, $X_0$'s are the eigenvalues of $X_0$. This construction is at the heart of the so-called QR algorithm that plays an outsize role in numerical analysis and occupies a prime position in software (see LINPACK) for eigenvalue computation. It turns out that the QR algorithm is also Hamiltonian and completely integrable in the following way. That is a

(20.1) Stroboscopic Reverser (SOMT 1985)

For $X \in \Sigma_N$, let

$$H_{\text{QR}}(X) = \text{tr}(X \log X - X)$$

Then $H_{\text{QR}}$ generates a Hamiltonian flow

$$t \mapsto X(t), \quad X(0) = H \text{ given}$$

with the property that

(20.2) $X(k) = X_{\text{QR}}$, \quad k = 0, 1, 2, \ldots$
where $X_n$ are the QR iterates generated by $X(0) = X_0 = H$. Moreover the flow $t \mapsto X_{an}(t)$ is completely integrable and commutes with the Toda flow generated by (10.1). We will describe, and explain, all the above properties of the QR algorithm in later chapters.

The goal of these lectures is to prove the result of D and Trofimov [DT 2017 (1)] on universality for the Toda eigenvalue algorithm to compute the largest eigenvalue of a matrix $H \in \Sigma_N$.

The result is the following: Let $X(t)$, $t \geq 0$, solve (10.1) with $X(0) = H \in \Sigma_N$. Set

$$E(t) = \sum_{n=2}^{N} (X_{1n}(t))^2$$

so that $E(t) = 0$ implies $X_{11}(t)$ is an eigenvalue of $H$. Thus
with $E(t)$ as in (21.1), the halting time (or the 1-deflation time) for the Toda algorithm is given by

$$T^{(1)}(t) = \inf \{ t \geq 0 : E(t) \leq \varepsilon^2 \}$$

Note that by the min-max principle, if $E(t) < \varepsilon^2$, \[ |Y_{11}(t) - \lambda_1| < \varepsilon \] for some eigenvalue $\lambda_1$ of $X(0) = H$.

As the Toda algorithm is generically sorting in the sense that as $X(t) \to X_\infty = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$, $t \to \infty$, the eigenvalues appear ordered, $\lambda_1 > \lambda_2 > \cdots > \lambda_N$. It follows that for generic $X(0) = H$, $\lambda_1$ above is $\lambda_1$. The top eigenvalue is $\lambda_1$.

Order the eigenvalues of a matrix $H \in \Sigma_N$, $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$.

For invariant and Wigner random matrix ensembles (see later lectures for appropriate definitions) there is a constant $c_v > 0$, which depends on the ensemble, such that the following limit exists.
\[ (23.1) \quad F_{\text{gap}}(t) = \lim_{t \to 0} \text{Prob} \left( \frac{1}{c_{V}^{3/2} c_{V}^{-1/3} N^{2/3} (A_{1} \to 2)} \leq t \right), \quad t > 0 \]

The basic result in [OT 2017 (1)] is the following:

\[ (23.2) \quad \text{Theorem} \]

Let \( 0 < \sigma < 1 \) be fixed and let \((\varepsilon, N)\) be in the scaling region

\[ (23.3) \quad \frac{\log \varepsilon^{-1}}{\log N} \geq \frac{\sigma + 5}{3} \]

Then if \( H \) is distributed according to any invariant or Wigner ensemble we have for \( t > 0 \)

\[ (23.4) \quad \lim_{N \to \infty} \text{Prob} \left( \frac{T^{(1)}}{c_{V}^{3/2} c_{V}^{-2/3} N^{2/3} (\log \varepsilon^{-1})^{3/3} (\log N)^{2/3}} \leq t \right) = F_{\text{gap}}(t) \]

Thus the halting time for the Toda algorithm to compute the top eigenvalue of a random matrix ensemble is universal and behaves statistically like the inverse of
The gap $\lambda_1 - \lambda_2$ of the two top eigenvalues of a random matrix. We will prove this result, and more, including the case where $H$ is Hermitian, in the lectures that follow. As noted above, we will also describe related results for the QR algorithm and other eigenvalue algorithms, but in less detail.

Note finally that the scaling regime (23.3) in which random matrix behavior is guaranteed to appear, includes a common arena for numerical computing. Indeed, for $\varepsilon = 10^{-16}$ and $N < 10^9$, we have

$$\frac{\log \varepsilon}{\log N} > \frac{16}{q} > \frac{5}{3}$$
The outline for these lectures is as follows:

- Introduction to Hamiltonian mechanics and integrable systems
- Toda lattice and its properties and its generalizations
- Properties of random matrix ensembles
- Proof of universality for Toda algorithm to compute the top eigenvalue of a random matrix.