

# Numerical analysis and random matrix theory

Tom Trogdon  
[ttrogdon@math.uci.edu](mailto:ttrogdon@math.uci.edu)  
UC Irvine



# Acknowledgements

This is joint work with:

- Percy Deift
- Govind Menon
- Sheehan Olver
- Raj Rao



# Numerical analysis and random matrix theory

- Using techniques from numerical analysis to analyze random matrices

Trotter (1984), Silverstein (1985), Edelman (1989), Dumitriu and Edelman (2002)

- Computing distributions from random matrix theory

Bornemann (2008), Witte, Bornemann and Forrester (2012), Olver and T (2014)

- Generating samples from random matrix distributions

Mezzadri (2006), Edelman and Rao (2005), Menon and Li (2014), Olver, Rao and T (2015)

- Using random matrices to analyze algorithms, statistically

Spielman and Teng (2009), Borgwardt (2012), Smale (1983, 1985), Deift and T (2016, 2017),  
Menon and T (2016), Feldheim, Paquette, and Zeitouni (2014)

- Randomized linear algebra

Liberty, Woolfe, Martinsson, Rokhlin, and Tygert (2007)



# Statistical analysis of algorithms

- Smoothed analysis: Spielman and Teng (2009)
- Simplex algorithm: Borgwardt (2012), Smale (1985)
- Universality: Pfrang, Deift and Menon (2014)



# Universality in numerical computation

w/ Percy Deift, Govind Menon and others



# Universality: A toy example

Consider the numerical solution of the system

$$(I - H)\mathbf{x} = \mathbf{b}, \quad H = H^*, \quad \|H\|_2 < 1.$$

The ensemble. Assume  $H = U \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N) U^*$  where  $(\lambda_j)_{1 \leq j \leq N}$  are iid with

$$\mathbb{P}(\lambda_j \in B) = \int_{B \cap [-1,1]} p(x) dx.$$

The method. A naïve iterative scheme is given by

$$\mathbf{x}_n = H\mathbf{x}_{n-1} + \mathbf{b}, \quad \mathbf{x}_0 = \mathbf{0}.$$



# Universality: A toy example

The halting time. To keep things simple, define  $T = T(H, \epsilon)$  to be the smallest integer  $T$  such that

$$\left\| (I - H)^{-1} - \sum_{k=0}^{T-1} H^k \right\|_2 < \epsilon.$$

The random variable  $T$  is the halting time.

The question. What is the distribution of  $T$ ? Is it universal as  $N \rightarrow \infty$ ?

---



Define

$$\lambda_{\min} = \min \lambda_j, \quad \lambda_{\max} = \max \lambda_j.$$

Then for fixed  $t \geq 0$  as  $N \rightarrow \infty$

$$\mathbb{P}\left(1 - \frac{t}{N} \leq \lambda_j \leq 1\right) = p(1) \frac{t}{N} + o(N^{-1}),$$

$$\mathbb{P}\left(-1 \leq \lambda_j \leq -1 + \frac{t}{N}\right) = p(-1) \frac{t}{N} + o(N^{-1}).$$

Assume  $p(1) \neq 0$ . Then as  $N \rightarrow \infty$

$$\begin{aligned} p(1)N(1 - \lambda_{\max}) &\xrightarrow{\text{dist.}} \exp(1), \\ N(1 + \lambda_{\min}) &\xrightarrow{\text{dist.}} \exp(p(-1)). \end{aligned}$$

Because  $H$  is Hermitian

$$\left\| (I - H)^{-1} - \sum_{k=0}^{n-1} H^k \right\|_2 = \max \left\{ \frac{|\lambda_{\max}|^n}{|1 - \lambda_{\max}|}, \frac{|\lambda_{\min}|^n}{|1 - \lambda_{\min}|} \right\}.$$



For  $\epsilon > 0$ , define  $t = t(H, \epsilon)$  by

$$\max \left\{ \frac{|\lambda_{\max}|^t}{|1 - \lambda_{\max}|}, \frac{|\lambda_{\min}|^t}{|1 - \lambda_{\min}|} \right\} = \epsilon.$$

Then  $T(H, \epsilon) = \max\{0, \lceil t(H, \epsilon) \rceil\}$  and

$$t(H, \epsilon) = \max \left\{ \frac{\log \epsilon^{-1} |1 - \lambda_{\max}|^{-1}}{\log |\lambda_{\max}|^{-1}}, \frac{\log \epsilon^{-1} |1 - \lambda_{\min}|^{-1}}{\log |\lambda_{\min}|^{-1}} \right\}.$$

Define

$$t_{\max} = \frac{\log \epsilon^{-1} |1 - \lambda_{\max}|^{-1}}{\log |\lambda_{\max}|^{-1}}, \quad t_{\min} = \frac{\log \epsilon^{-1} |1 - \lambda_{\min}|^{-1}}{\log |\lambda_{\min}|^{-1}}.$$



# “Universality”

Two estimates are needed as  $N \rightarrow \infty$ :

$$\frac{t_{\max}}{r(1)N \log[N\epsilon^{-1}]} \xrightarrow{\text{dist.}} \frac{1}{\xi}, \quad \xi \sim \exp(1),$$
$$t_{\min} = O(N).$$

Because  $t_{\min}$  is asymptotically smaller than  $t_{\max}$  it does not affect the limit:

$$\frac{T(H, \epsilon)}{r(1)N \log[N\epsilon^{-1}]} \xrightarrow{\text{dist.}} \frac{1}{\xi}, \quad \xi \sim \exp(1).$$



This establishes universality for  $T(H, \epsilon)$  in this simple case. It required three main components:

- A formula to estimate error.
- Universality for key statistics ( $\lambda_{\max}$ ).
- Estimates for the rest ( $\lambda_{\min}$ ).



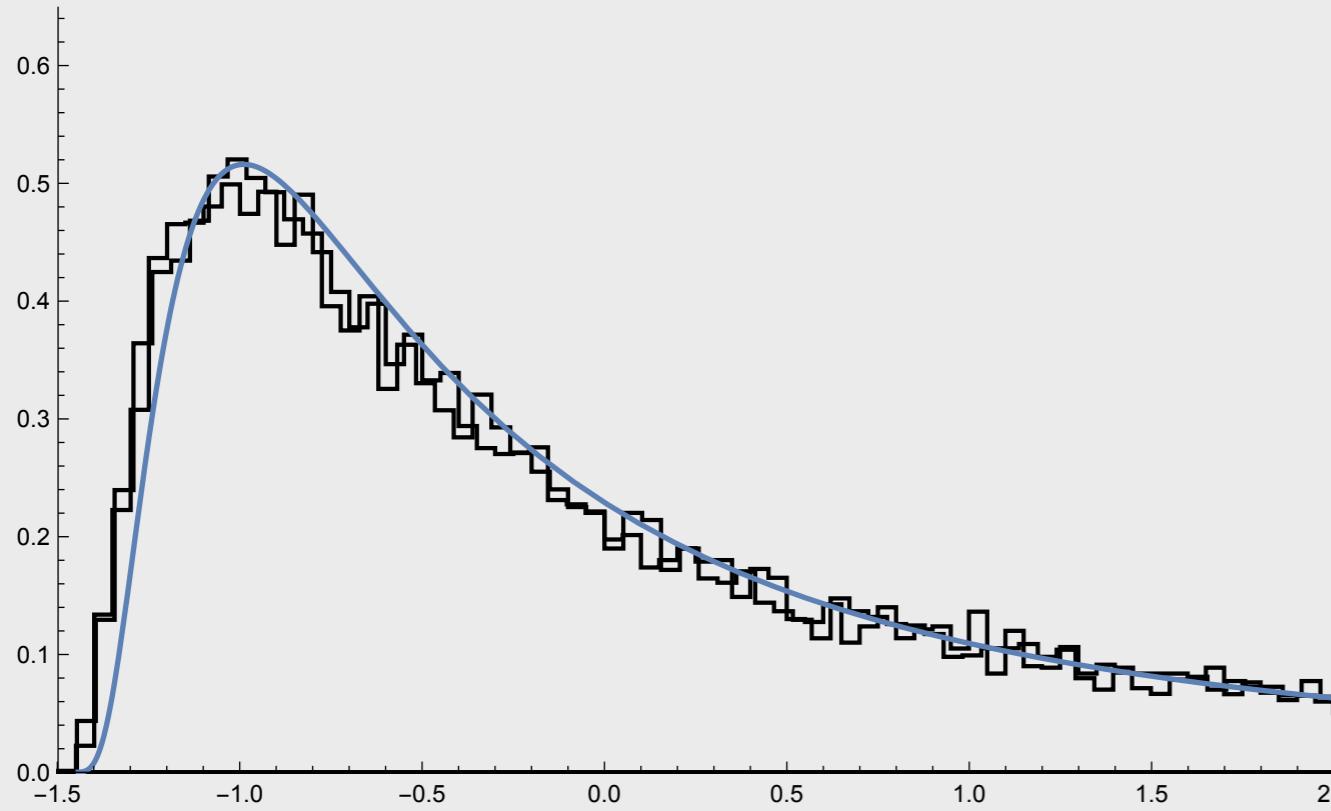
Significant work goes into determining the scaling. If the scaling is not known, but universality is going to be investigated, a normalization must be inferred from the data.

Given a collection of samples of the random variable  $T = T(H, \epsilon)$ , define the empirical fluctuations

---

$$\tau(H, \epsilon) = \frac{T(H, \epsilon) - \langle T \rangle}{\sigma_T},$$

where  $\langle T \rangle$  is the sample median and  $\sigma_T$  is the median absolute deviation.



Two choices for  $p(x)$ :

- $p(x) = 1$
- $p(x) = e^{-x}/(e - e^{-1})$



# The (inverse) power method

The ensembles. A sample covariance matrix (SCM) is an  $N \times N$  real symmetric ( $\beta = 1$ ) or complex Hermitian ( $\beta = 2$ ) matrix

$$H = X^*X/M, \quad X = (X_{ij})_{1 \leq i \leq M, 1 \leq j \leq N} \text{ iid}, \quad M \sim N/d, \quad 0 < d < 1,$$
$$\mathbb{E}X_{ij} = 0, \quad \mathbb{E}|X_{ij}|^2 = 1.$$

For  $\beta = 2$ ,  $\mathbb{E}X_{ij}^2 = 0$  is imposed. We require uniform exponential tails.

The method. The power method is given by ( $\mathbf{x}_0$  specified)

$$\mu_k = \frac{\mathbf{x}_0^* H^{2k-1} \mathbf{x}_0}{\mathbf{x}_0^* H^{2k-2} \mathbf{x}_0} \rightarrow \lambda_N = \lambda_{\max}, \quad k \rightarrow \infty.$$

The halting time. The halting time is given by

$$T(H, \mathbf{x}_0, \epsilon) = \min\{n : |\mu_n - \mu_{n-1}| < \epsilon^2\}.$$

The question. What is the distribution of  $T$ ? Is it universal as  $N \rightarrow \infty$ ?

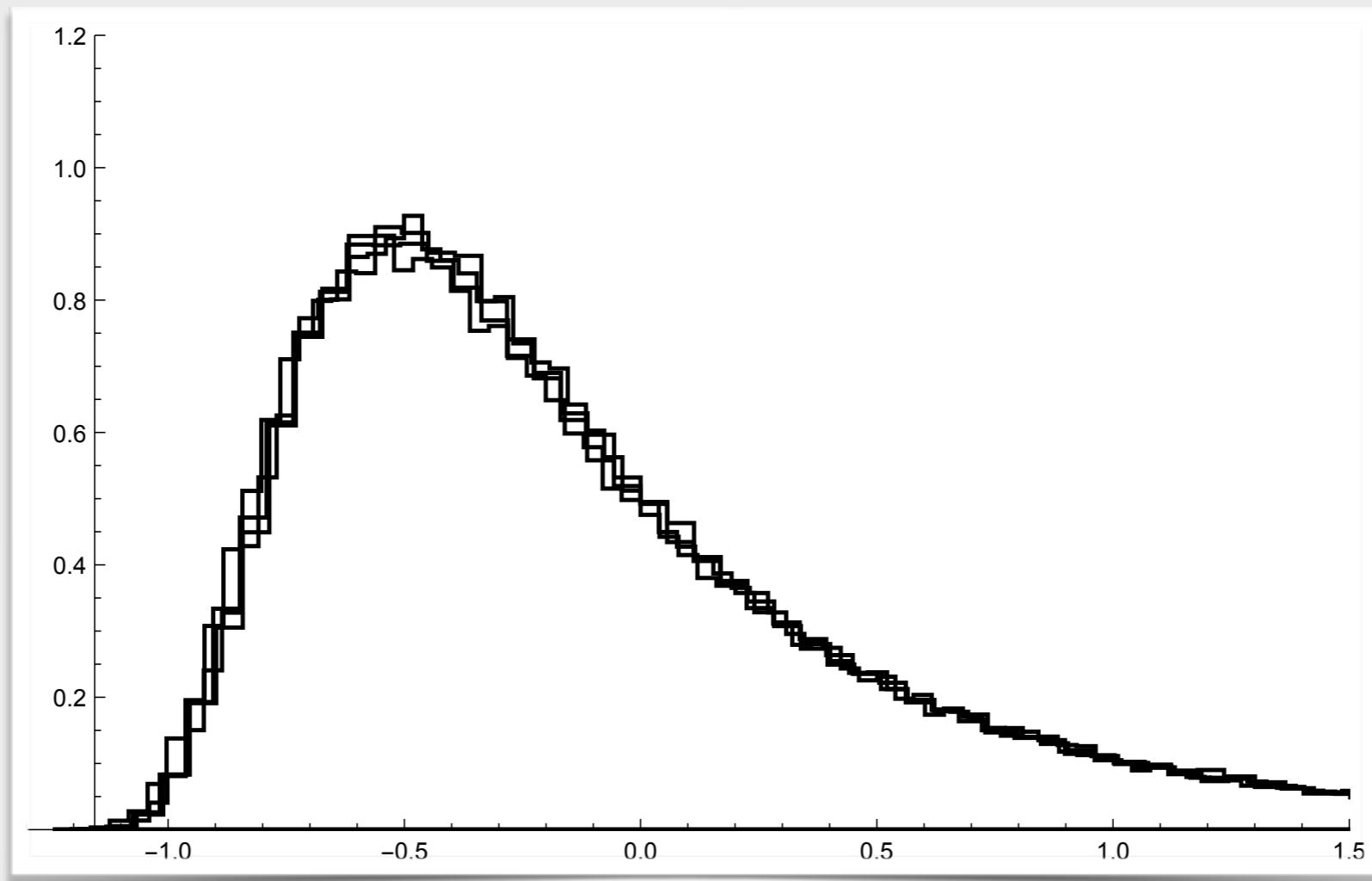


# Observing universality

Using 4 complex SCMs, the fluctuations

$$\tau(H, \epsilon) = \frac{T(H, \epsilon) - \langle T \rangle}{\sigma_T},$$

appear universal.



# A formula to estimate the error

Let  $H = U\Lambda U^*$  and  $U^*\mathbf{x}_0 = [\beta_1, \beta_2, \dots, \beta_N]^T$ ,

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N), \quad 0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N.$$

To estimate the halting time for the power method, we must find the large  $N$  and large  $k$  asymptotics for

$$\mu_k = \frac{\sum_{j=1}^N |\beta_j|^2 \lambda_j^{2k-1}}{\sum_{j=1}^N |\beta_j|^2 \lambda_j^{2k-2}} = \lambda_N \left( \frac{1 + \sum_{j=1}^{N-1} \left| \frac{\beta_j}{\beta_N} \right|^2 \left( \frac{\lambda_j}{\lambda_N} \right)^{2k-1}}{1 + \sum_{j=1}^{N-1} \left| \frac{\beta_j}{\beta_N} \right|^2 \left( \frac{\lambda_j}{\lambda_N} \right)^{2k-2}} \right).$$



# A historical interlude

For eigenvalues:

- The seminal work of Geman (1980) showed that the largest eigenvalue of an SCM converges a.s.
- Silverstein (1985) established that the smallest eigenvalue converges a.s. to  $\lambda_-$  for iid standard normal random variables.
- Johnstone (2001); Johansson(2000); Forrester (1993) gave the first results on the fluctuations of the largest and smallest eigenvalues for (real or complex) standard normal distributions.
- Universality was obtained by Ben Arous and Péché (2005) (see also Tao and Vu (2012)).
- We reference Pillai and Yin (2014) and Bloemendal, Knowles, Yau and Yin (2016) for the most comprehensive results.

For eigenvectors:

- The limits of the eigenvectors have also been considered in various ways, see Silverstein (1986); Bai, Miao and Pan (2007).
- We reference Bloemendal, Knowles, Yau and Yin (2016) for the generality needed to prove our theorems.

We require exponential tails which is stronger than the assumptions in Yin (1986); Geman (1980).



# Universality for key statistics

**Theorem.** For SCMs

$$N^{1/2}(|\beta_N|, |\beta_{N-1}|, |\beta_{N-2}|)$$

converge jointly in distribution to  $(|X_1|, |X_2|, |X_3|)$  where  $\{X_1, X_2, X_3\}$  are iid real ( $\beta = 1$ ) or complex ( $\beta = 2$ ) standard normal random variables. Additionally,

$$N^{2/3} \lambda_+^{-2/3} d^{1/2}(\lambda_+ - \lambda_N, \lambda_+ - \lambda_{N-1}, \lambda_+ - \lambda_{N-2})$$

converge jointly in distribution to random variables  $(\Lambda_{1,\beta}, \Lambda_{2,\beta}, \Lambda_{3,\beta})$  which are the smallest three eigenvalues of the so-called stochastic Airy operator.

This follows from Ramírez, Rider and Virág (2011); Pillai and Yin (2014); Bloemendal, Knowles, Yau and Yin (2016).



# Estimates for the rest

**Theorem (Pillai and Yin (2014)).** For any  $s > 0$

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(|\lambda_n - \gamma_n| \leq N^{-2/3+s} (\max\{n, N-n+1\})^{-1/3} \text{ for all } n\right) = 1.$$

**Theorem (Bloemendaal, Knowles, Yau and Yin (2016)).** For any  $s > 0$

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(|\beta_n| \leq N^{-1/2+s} \text{ for all } n\right) = 1.$$

The first theorem is known as rigidity and it was first shown for Wigner ensembles by Erdős, Yau and Yin (2012) (see also Bourgade, Erdős and Yau (2014)).

The second theorem is known as delocalization.



# Universality for the (inverse) power method

The distribution function  $F_\beta^{\text{gap}}(t)$ , supported on  $t \geq 0$  for  $\beta = 1, 2$  is defined by

$$F_\beta^{\text{gap}}(t) := \lim_{N \rightarrow \infty} \mathbb{P}\left(\frac{1}{2^{-7/6} N^{2/3} \lambda_+^{-2/3} d^{-1/2} (\lambda_N - \lambda_{N-1})} \leq t\right).$$

**Theorem (Deift and T).** Let  $H$  be a real ( $\beta = 1$ ) or complex ( $\beta = 2$ ) SCM and let  $\mathbf{x}_0$  be a random unit vector independent of  $H$ . Assuming  $\epsilon \leq N^{-5/3-\sigma}$

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(\frac{T(H, \mathbf{x}_0, \epsilon)}{2^{-7/6} \lambda_+^{1/3} d^{-1/2} N^{2/3} (\log \epsilon^{-1} - 2/3 \log N)} \leq t\right) = F_\beta^{\text{gap}}(t).$$



# A note on the proof

In the the proof we give the following approximation:

$$\frac{T(H, \mathbf{x}_0, \epsilon)}{N^{2/3}} - \frac{\log \epsilon^{-1} + \log \left(1 - \left| \frac{\lambda_N}{\lambda_{N-1}} \right| \right) + \frac{1}{2} \log 2\lambda_N + \log \left| \frac{\beta_{N-1}}{\beta_N} \right|}{N^{2/3} \log \left| \frac{\lambda_N}{\lambda_{N-1}} \right|} \xrightarrow{\text{prob.}} 0$$

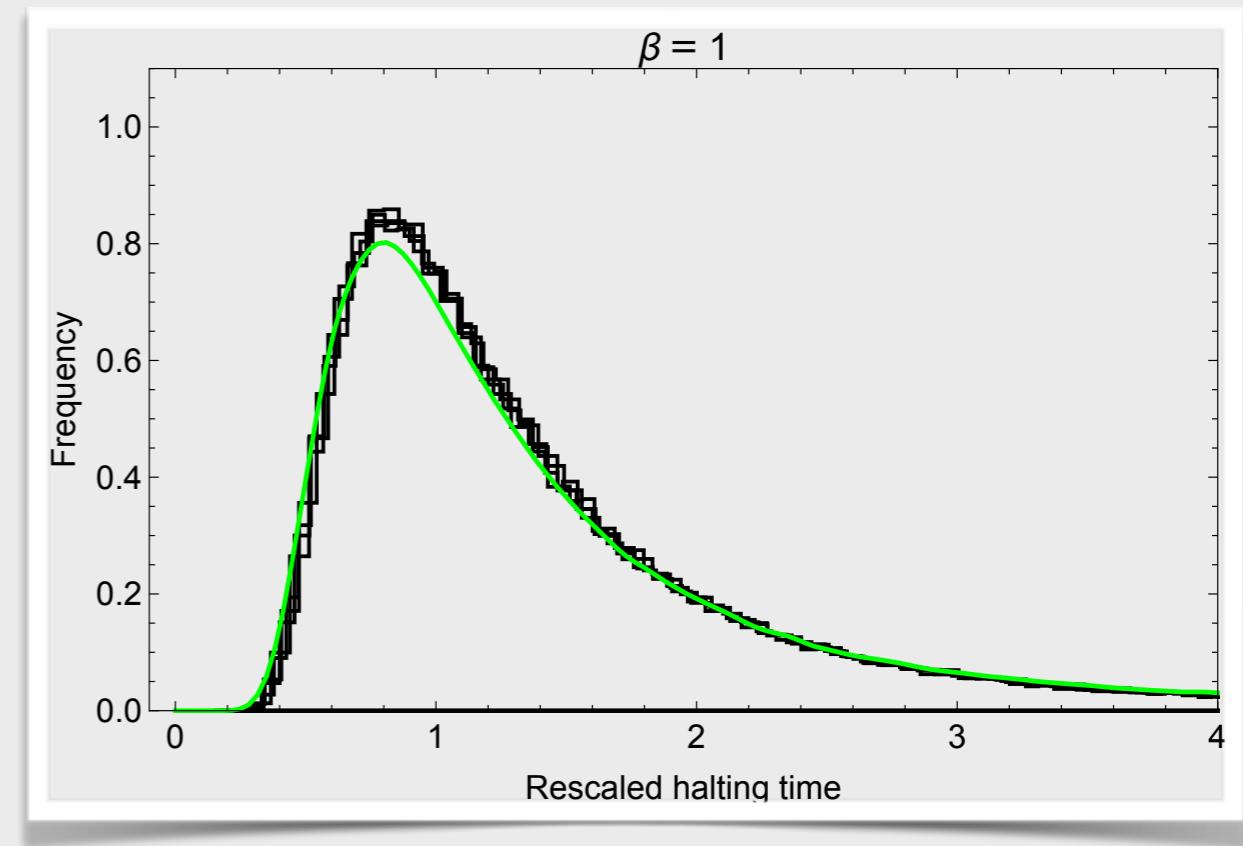
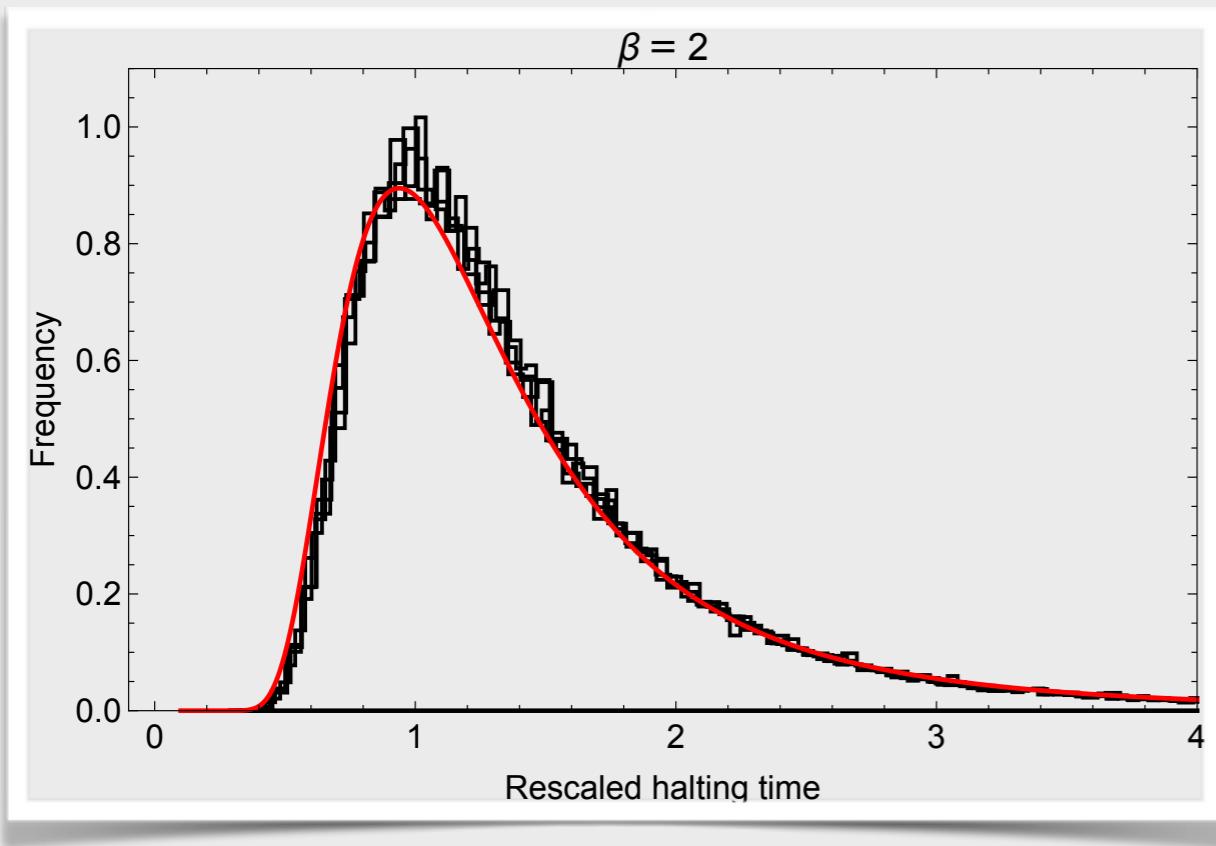
This we obtain:

$$\frac{T(H, \mathbf{x}_0, \epsilon)}{N^{2/3} \log N} - \frac{\log \epsilon^{-1} - \frac{2}{3} \log N}{N^{2/3} \log N \lambda_+^{-1} |\lambda_N - \lambda_{N-1}|} \xrightarrow{\text{prob.}} 0$$

Convergence is at an (at best) logarithmic rate.



# A demonstration



Thanks to Forkmar Bornemann (TUM)

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( \frac{T(H, \mathbf{x}_0, \epsilon)}{2^{-7/6} \lambda_+^{1/3} d^{-1/2} N^{2/3} (\log \epsilon^{-1} - 2/3 \log N + \gamma)} \leq t \right) = F_\beta^{\text{gap}}(t).$$



# Other theorems

Similar techniques give universality theorems for:

- The QR eigenvalue algorithm.
- The Toda algorithm.

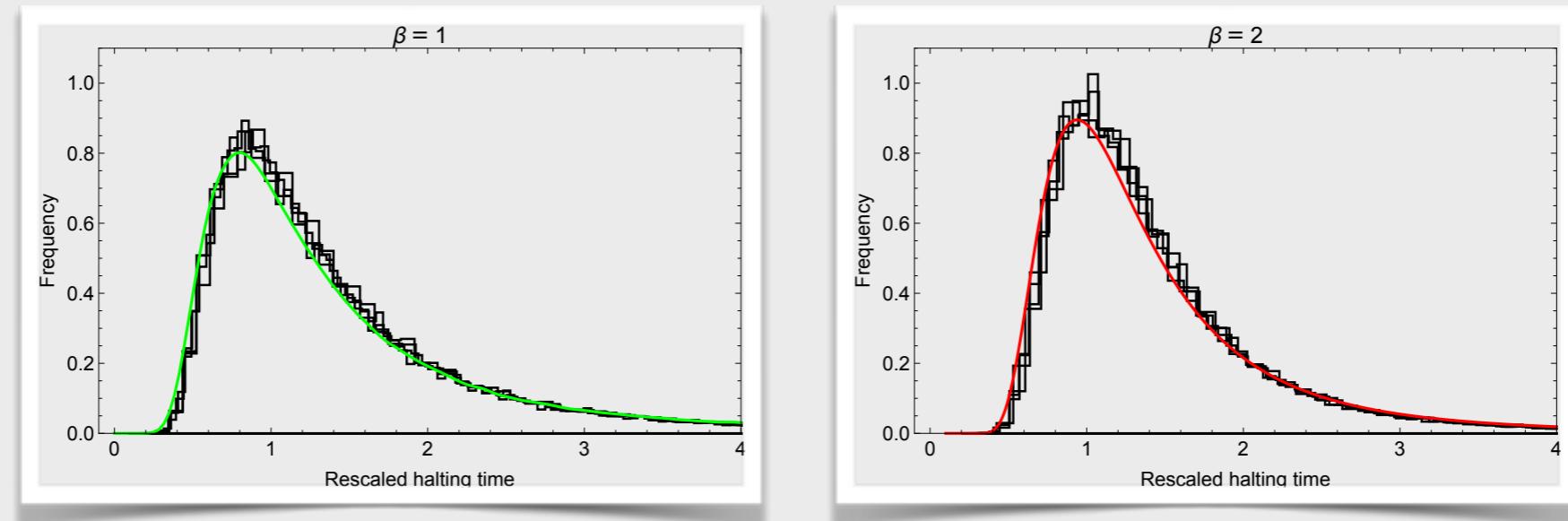
See Deift and T (2016) and Deift and T (2017).



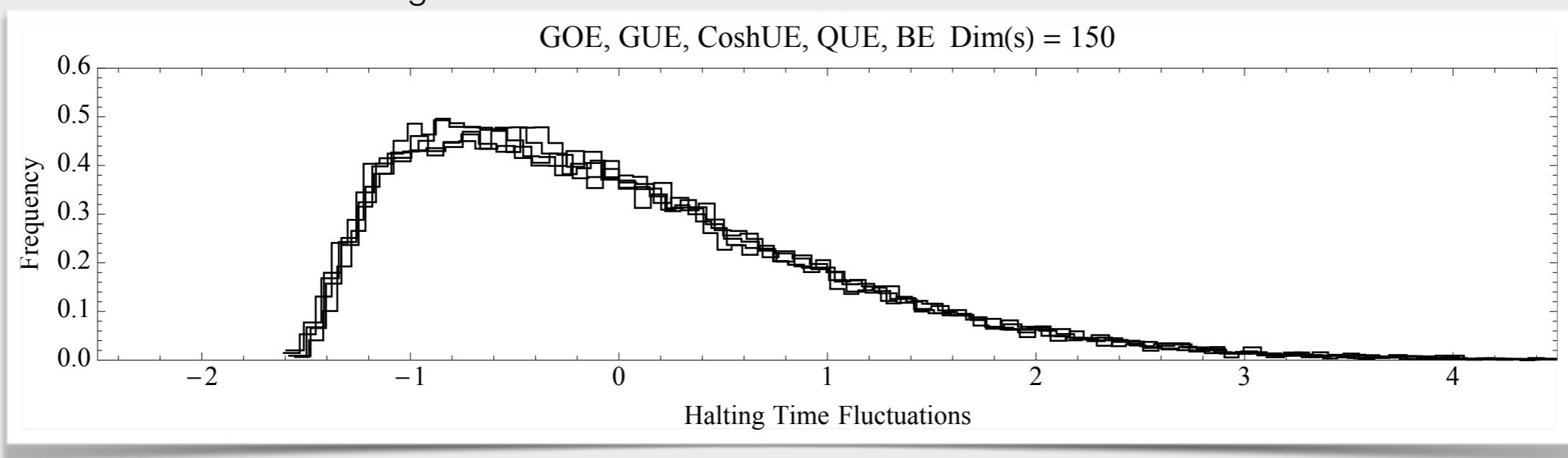
# Other algorithms

w/ Percy Deift, Govind Menon and Sheehan Olver

QR algorithm with  
“simple halting”



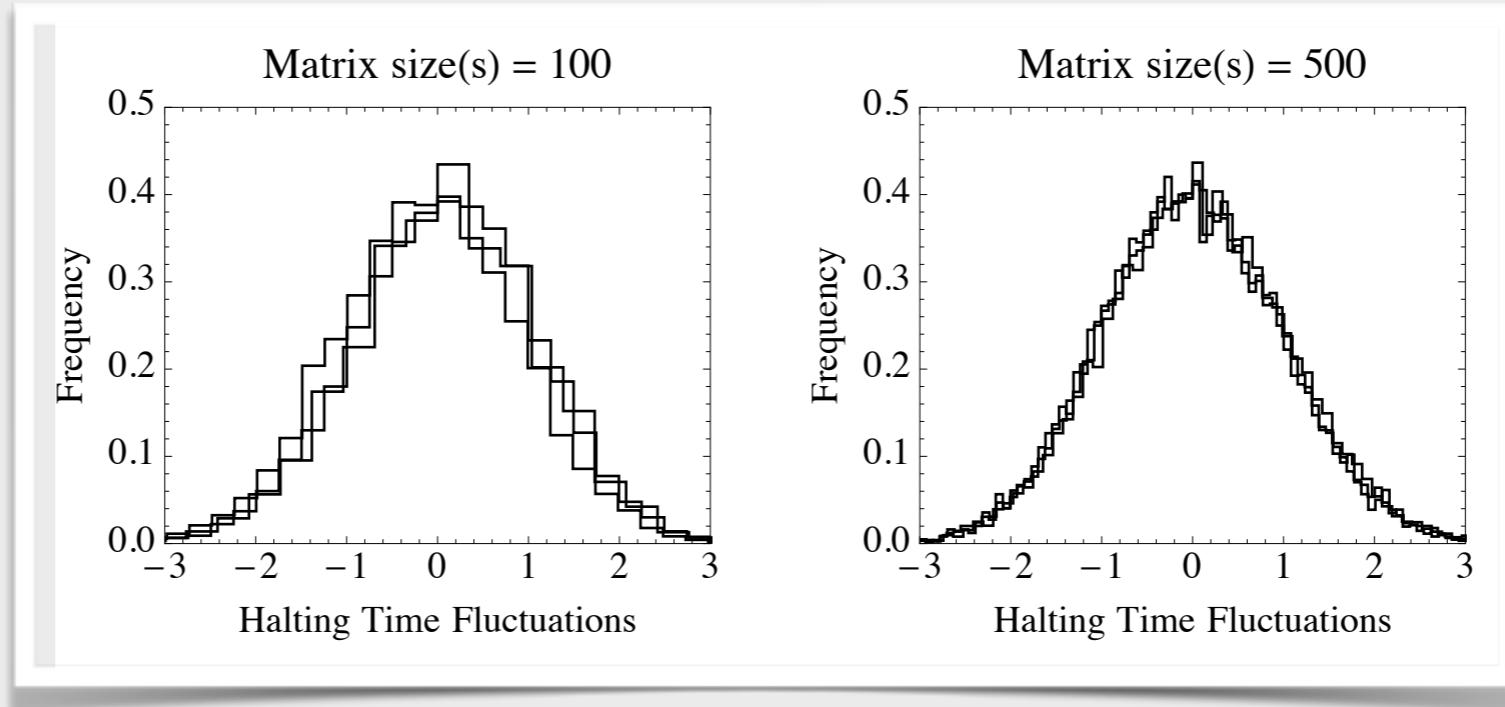
QR algorithm with  
“realistic halting”



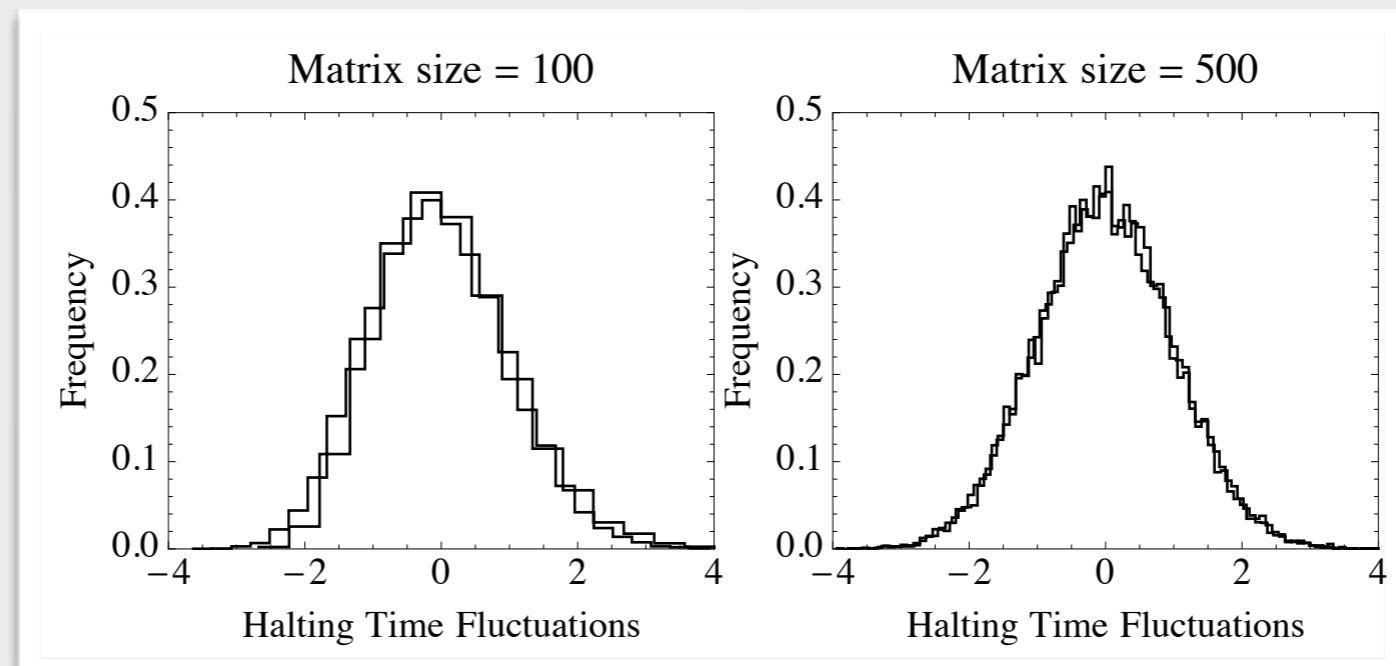
# Other algorithms

w/ Percy Deift, Govind Menon and Sheehan Olver

Conjugate gradient  
algorithm

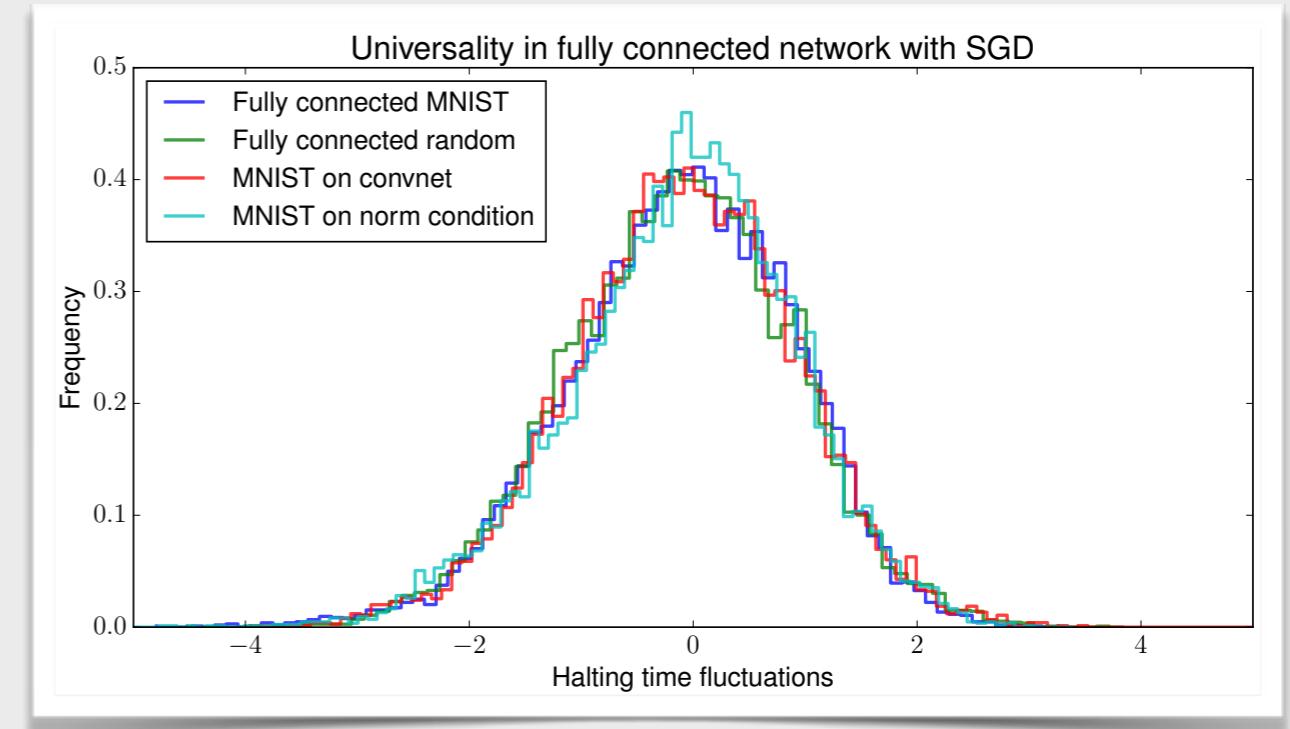
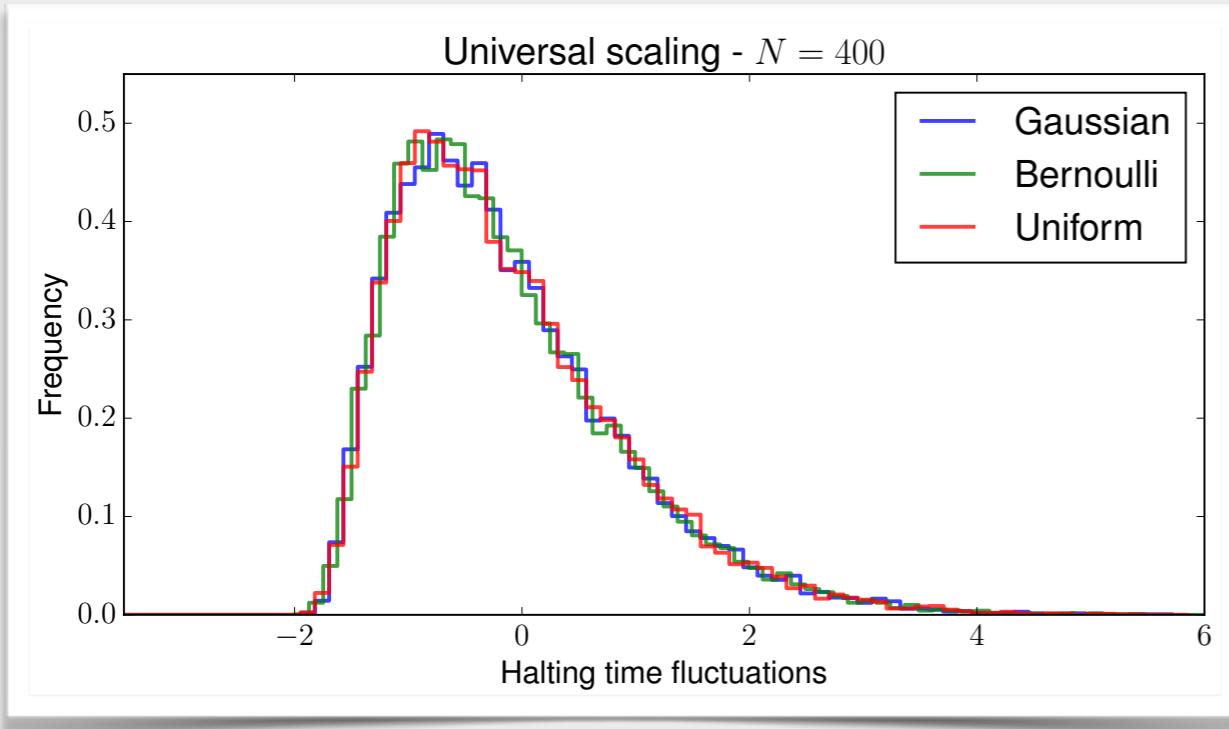


GMRES algorithm

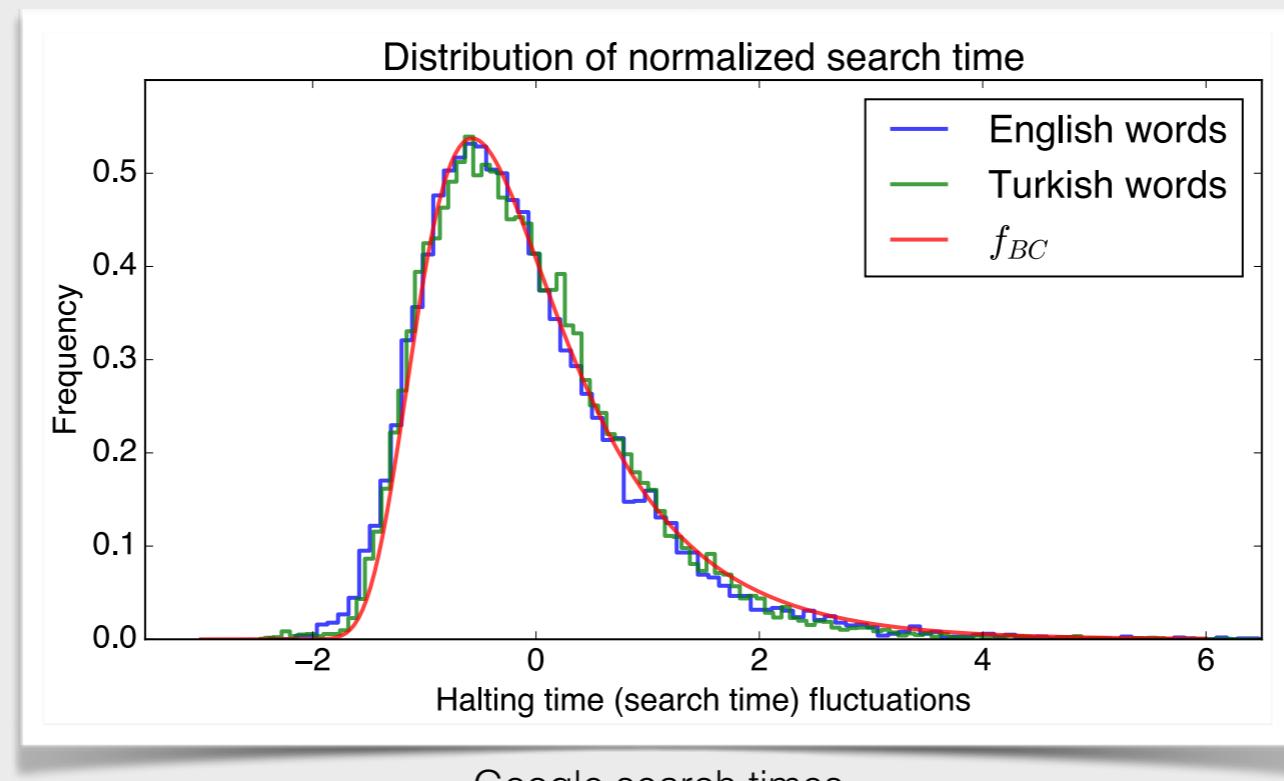


# Other algorithms

w/ Levent Sagun and Yann LeCun



Gradient descent for spin glasses



Google search times



# Smoothed analysis for the conjugate gradient algorithm

w/ Govind Menon



The ensemble. An LUE matrix is an  $N \times N$  complex Hermitian matrix

$$H = X^*X/M, \quad X = (X_{ij})_{1 \leq i \leq M, 1 \leq j \leq N} \text{ iid standard complex normals.}$$

Take  $M \sim N/d$ ,  $d = 1 - N^{-\alpha}$ ,  $1/2 \leq \alpha < 1$ .

The method. The conjugate gradient algorithm is an iterative method to solve  $H\mathbf{x} = \mathbf{b}$ :

$$\mathbf{0} = \mathbf{x}_0 \rightarrow \mathbf{x}_1 \rightarrow \cdots \rightarrow \mathbf{x}_k \rightarrow \cdots \rightarrow \mathbf{x}_N = \mathbf{x}.$$

The halting time. The halting time ( $\mathbf{x}_0 = \mathbf{0}$ ) is given by

$$T(H, \epsilon) = \min\{k : \|H\mathbf{x}_k - \mathbf{b}\|_2 < \epsilon\}.$$

The problem. For a deterministic matrix  $A > 0$ , estimate the moments of  $T(A + \sigma^2 H, \epsilon)$ .



# The scaling

Here we have chosen  $d = 1 - N^{-\alpha}$ .

The rationale for this is the following:

- If  $d > 1$  is fixed, then  $X$  has a fixed aspect ratio (well conditioned, concentrated & bounded condition number).
- If  $d = 1$  then  $X$  has a heavy-tailed condition number (ill conditioned, unbounded & unconcentrated).

The case where  $M = N + \alpha$ ,  $\alpha > 0$ ,  $N \rightarrow \infty$  then  $\alpha \rightarrow \infty$  was studied by Borodin & Forrester (2002).

This describes the transition from the ill-conditioned to the well-conditioned case (from heavy-tailed to sub-exponential distributions).



# The scaling

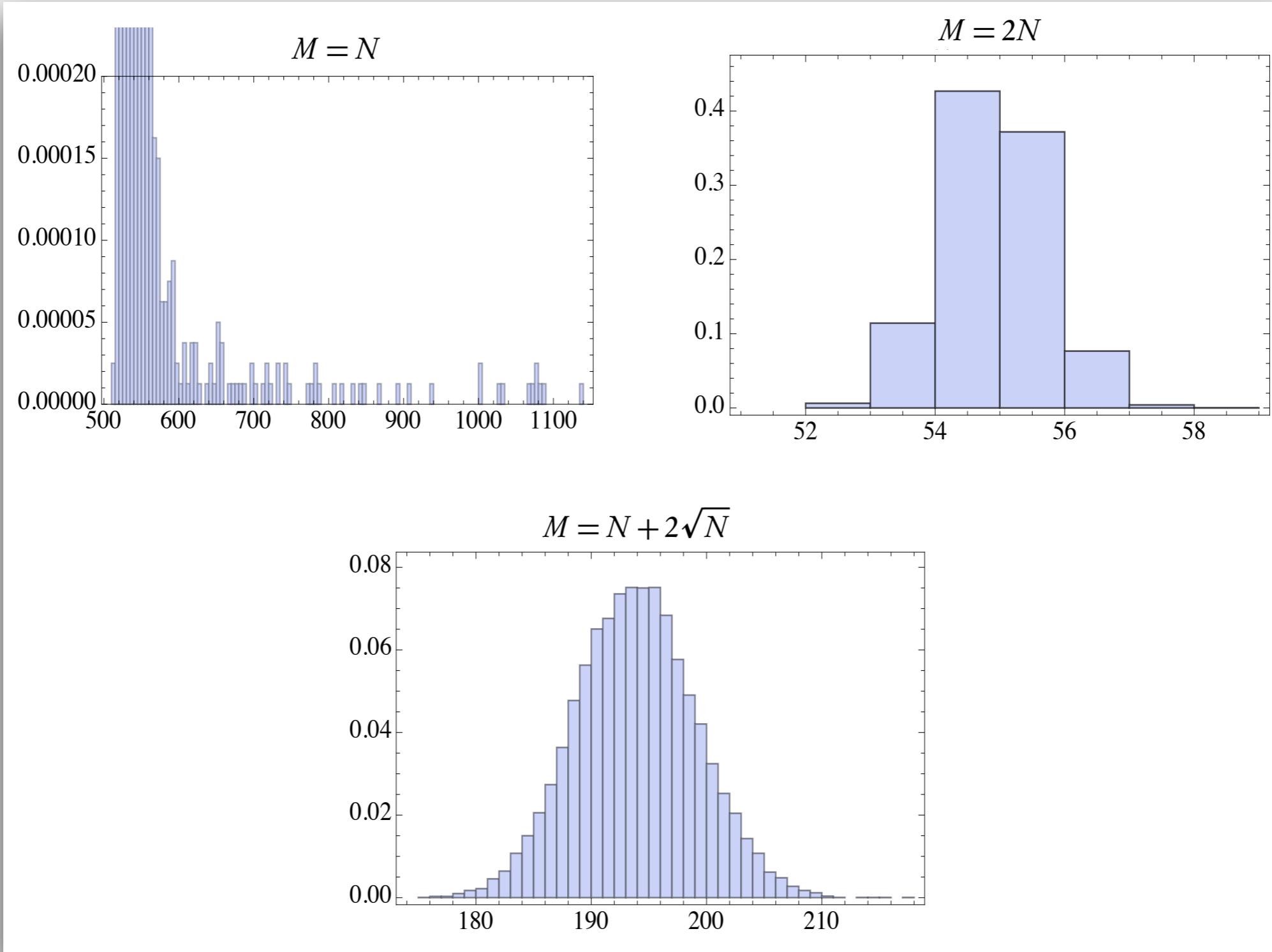
If  $d = 1 - (4c)^{-1/2}N^{-1/2}$  then (see Deift, Menon & T. (2017))

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( \frac{\lambda_{\max}/\lambda_{\min} - 4Nc^{-1}}{4c^{-4/3}N^{2/3}} \leq t \right) = F_2(t) \quad \Leftarrow \quad \text{Tracy-Widom dist.}$$

The condition number grows with  $N$  but the (limit) distribution has sub-exponential tails, giving a finite  $N$  matrix that gives the Borodin–Forrester transition.



# The performance of the conjugate gradient algorithm



# Smoothed analysis for the CG algorithm

**Theorem (Menon and T (2016)).** As  $N \rightarrow \infty$ , for  $1/2 \leq \alpha < 1$ ,

$$\sup_{\|A\| \leq 1, A \geq 0} \mathbb{E}[T^j(A + \sigma^2 H, \epsilon)] = O\left(N^{\alpha j} \left(1 + \frac{1}{\sigma^2}\right)^j \log^j \left[N^\alpha \left(1 + \frac{1}{\sigma^2}\right) \epsilon^{-1}\right]\right).$$

In particular,

$$\sup_{\|A\| \leq 1, A \geq 0} \mathbb{E}[T(A + \sigma^2 H, \epsilon)] = O\left(N^\alpha \left(1 + \frac{1}{\sigma^2}\right) \log \left[N^\alpha \left(1 + \frac{1}{\sigma^2}\right) \epsilon^{-1}\right]\right).$$

and the right-hand side is less than  $N$  (for suff. large  $N$ ).

The proof uses tail estimates on the condition number derived from the Riemann–Hilbert analysis in Deift, Menon, and T (2015).



# Smoothed analysis for the CG algorithm

Let  $\mathbf{x}_k$  be the approximation of  $\mathbf{x}$ ,  $H\mathbf{x} = \mathbf{b}$  found by CG. It is known (see Greenbaum, Pták & Strakoš (1996)) that any non-increasing sequence for

$$\frac{\|\mathbf{x}_k - \mathbf{x}\|_H}{\|\mathbf{x}_0 - \mathbf{x}\|_H}, \quad k = 0, 1, \dots, N-1,$$

is attainable by choosing  $H$  appropriately.

In particular, there exists a pathological case where

$$\frac{\|\mathbf{x}_k - \mathbf{x}\|_H}{\|\mathbf{x}_0 - \mathbf{x}\|_H} = 1, \quad k = 0, 1, \dots, N-1,$$

and  $\mathbf{x}_N = \mathbf{x}$ .

The probability of such a pathological case decays faster than  $N^{-k}$  for any  $k > 0$ .



Detailed estimates from random matrix theory can provide  
a clear mechanism for the probabilistic analysis of  
numerical algorithms...

...and proofs of

universality!



# References

- [Bai et al., 2007] Bai, Z. D., Miao, B. Q., and Pan, G. M. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.*, 35(4):1532–1572.
- [Ben Arous and Péché, 2005] Ben Arous, G. and Péché, S. (2005). Universality of local eigenvalue statistics for some sample covariance matrices. *Commun. Pure Appl. Math.*, 58(10):1316–1357.
- [Bloemendal et al., 2016] Bloemendal, A., Knowles, A., Yau, H.-T., and Yin, J. (2016). On the principal components of sample covariance matrices. *Probab. Theory Relat. Fields*, 164(1-2):459–552.
- [Borgwardt, 1987] Borgwardt, K. H. (1987). *The simplex method: A probabilistic analysis*. Springer-Verlag, Berlin, Heidelberg.
- [Bornemann, 2010] Bornemann, F. (2010). On the numerical evaluation of Fredholm determinants. *Math. Comput.*, 70(4):871–915.
- [Deift and Trogdon, 2016] Deift, P. and Trogdon, T. (2016). Universality for the Toda algorithm to compute the largest eigenvalue of a random matrix.
- [Deift and Trogdon, 2017] Deift, P. and Trogdon, T. (2017). Universality for eigenvalue algorithms on sample covariance matrices.
- [Deift et al., 2016] Deift, P. A., Menon, G., and Trogdon, T. (2016). On the condition number of the critically-scaled Laguerre Unitary Ensemble. *Discret. Contin. Dyn. Syst.*, 36(8):4287–4347.
- [Dumitriu and Edelman, 2002] Dumitriu, I. and Edelman, A. (2002). Matrix models for beta ensembles. *J. Math. Phys.*, 43(11):5830.
- [Edelman, 1988] Edelman, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560.
- [Edelman and Rao, 2005] Edelman, A. and Rao, N. R. (2005). Random matrix theory. *Acta Numer.*, 14:233–297.
- [Forrester, 1993] Forrester, P. J. (1993). The spectrum edge of random matrix ensembles. *Nucl. Phys. B*, 402(3):709–728.
- [Geman, 1980] Geman, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261.
- [Hough et al., 2006] Hough, J. B., Krishnapur, M., Peres, Y., and Virág, B. (2006). Determinantal Processes and Independence. *Probab. Surv.*, 3:206–229.
- [Johansson, 2000] Johansson, K. (2000). Shape Fluctuations and Random Matrices. *Commun. Math. Phys.*, 209(2):437–476.
- [Johnstone, 2001] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Stat.*, 29(2):295–327.
- [Li and Menon, 2013] Li, X. H. and Menon, G. (2013). Numerical solution of Dyson Brownian motion and a sampling scheme for invariant matrix ensembles. *J. Stat. Phys.*, 153(5):801–812.
- [Menon and Trogdon, 2016] Menon, G. and Trogdon, T. (2016). Smoothed analysis for the conjugate gradient algorithm. *Symmetry, Integr. Geom. Methods Appl.*, pages 1–19.
- [Mezzadri, 2006] Mezzadri, F. (2006). How to generate random matrices from the classical compact groups. *Not. AMS*, 54:592–604.
- [Olver et al., 2015] Olver, S., Rao, N. R., and Trogdon, T. (2015). Sampling unitary ensembles. *Random Matrices Theory Appl.*, 4(1):1550002.
- [Olver and Trogdon, 2013] Olver, S. and Trogdon, T. (2013). Numerical solution of Riemann–Hilbert Problems: Random matrix theory and orthogonal polynomials. *Constr. Approx.*, 39(1):101–149.
- [Pfrang et al., 2014] Pfrang, C. W., Deift, P., and Menon, G. (2014). How long does it take to compute the eigenvalues of a random symmetric matrix? *Random matrix theory, Interact. Part. Syst. Integr. Syst. MSRI Publ.*, 65:411–442.
- [Pillai and Yin, 2014] Pillai, N. S. and Yin, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.*, 24(3):935–1001.
- [Scardicchio et al., 2009] Scardicchio, A., Zachary, C. E., and Torquato, S. (2009). Statistical properties of determinantal point processes in high-dimensional Euclidean spaces. *Phys. Rev. E*, 79(4):041108.
- [Silverstein, 1986] Silverstein, J. (1986). Eigenvalues and Eigenvectors of large dimensional sample covariance matrices. *Contemp. Math.*, 50.
- [Silverstein, 1985] Silverstein, J. W. (1985). The Smallest Eigenvalue of a Large Dimensional Wishart Matrix. *Ann. Probab.*, 13(4):1364–1368.
- [Smale, 1983] Smale, S. (1983). On the average number of steps of the simplex method of linear programming. *Math. Program.*, 27(3):241–262.
- [Smale, 1985] Smale, S. (1985). On the efficiency of algorithms of analysis. *Bull. Am. Math. Soc.*, 13(2):87–121.

# Sampling unitary ensembles

w/ Sheehan Olver and Raj Rao



# Sampling unitary ensembles

The problem: Generate a sample from the  $N \times N$  Hermitian random matrix with distribution

$$\frac{1}{Z_N} e^{-N\text{tr}V(H)} dH.$$

If  $V(x)$  is not quadratic (say,  $V(x) = x^4$ ), then the entries of the matrix are not independent.

The density on the eigenvalues is

$$\frac{1}{\hat{Z}_N} \prod_{i < j} |\lambda_j - \lambda_i|^2 e^{-N \sum_i V(\lambda_i)} d\lambda = \frac{1}{\hat{Z}_N} e^{-2NH_N(\lambda)} d\lambda.$$



# Sampling unitary ensembles

One approach: Numerically solve the Dyson BM SDEs:

$$d\lambda = -\nabla H_N(\lambda)dt + \frac{1}{\sqrt{N}}dW.$$

See Menon and Li (2014).

Our approach: Exploit the determinantal structure:

$$\frac{1}{\hat{Z}_N} \prod_{i < j} |\lambda_j - \lambda_i|^2 e^{-N \sum_i V(\lambda_i)} d\lambda = \frac{1}{N!} \det K_N(\lambda_i, \lambda_j) d\lambda,$$

$$K_N(\lambda_i, \lambda_j) = \sum_n p_n(\lambda_i) p_n(\lambda_j) e^{-\frac{1}{2} N V(\lambda_i) - \frac{1}{2} N V(\lambda_j)}.$$



Assume  $\phi_n(x) = p_n(x)e^{-\frac{1}{2}V(x)}$  can be evaluated for  $n = 0, 1, \dots, N - 1$ .

Find the “first” eigenvalue: Use inverse transform sampling to sample the density

---

$$\rho_1(x) := \frac{1}{N} K_N(x, x) \rightarrow \lambda_1.$$

Find and sample the new density:

---

$$\rho_2(x) \propto \det \begin{bmatrix} K_N(x, x) & K_N(x, \lambda_1) \\ K_N(\lambda_1, x) & K_N(\lambda_1, \lambda_1) \end{bmatrix}, \quad \rho_2(x) \rightarrow \lambda_2.$$

Find and sample the new density:

---

$$\rho_3(x) \propto \det \begin{bmatrix} K_N(x, x) & K_N(x, \lambda_1) & K_N(x, \lambda_2) \\ K_N(\lambda_1, x) & K_N(\lambda_1, \lambda_1) & K_N(\lambda_1, \lambda_2) \\ K_N(\lambda_2, x) & K_N(\lambda_2, \lambda_1) & K_N(\lambda_2, \lambda_2) \end{bmatrix}, \quad \rho_3(x) \rightarrow \lambda_3.$$



# A stable algorithm

For  $N$  large, this is not a stable procedure. By implementing the work of Hough, Krishnapur, Peres and Virág (2006) and Scardicchio, Zachary and Torquato (2009), the conditional densities can be computed using stable methods from linear algebra.

Given  $\boldsymbol{\phi}_1(x) := (\phi_n(x))_{0 \leq n \leq N-1}$  and the sample

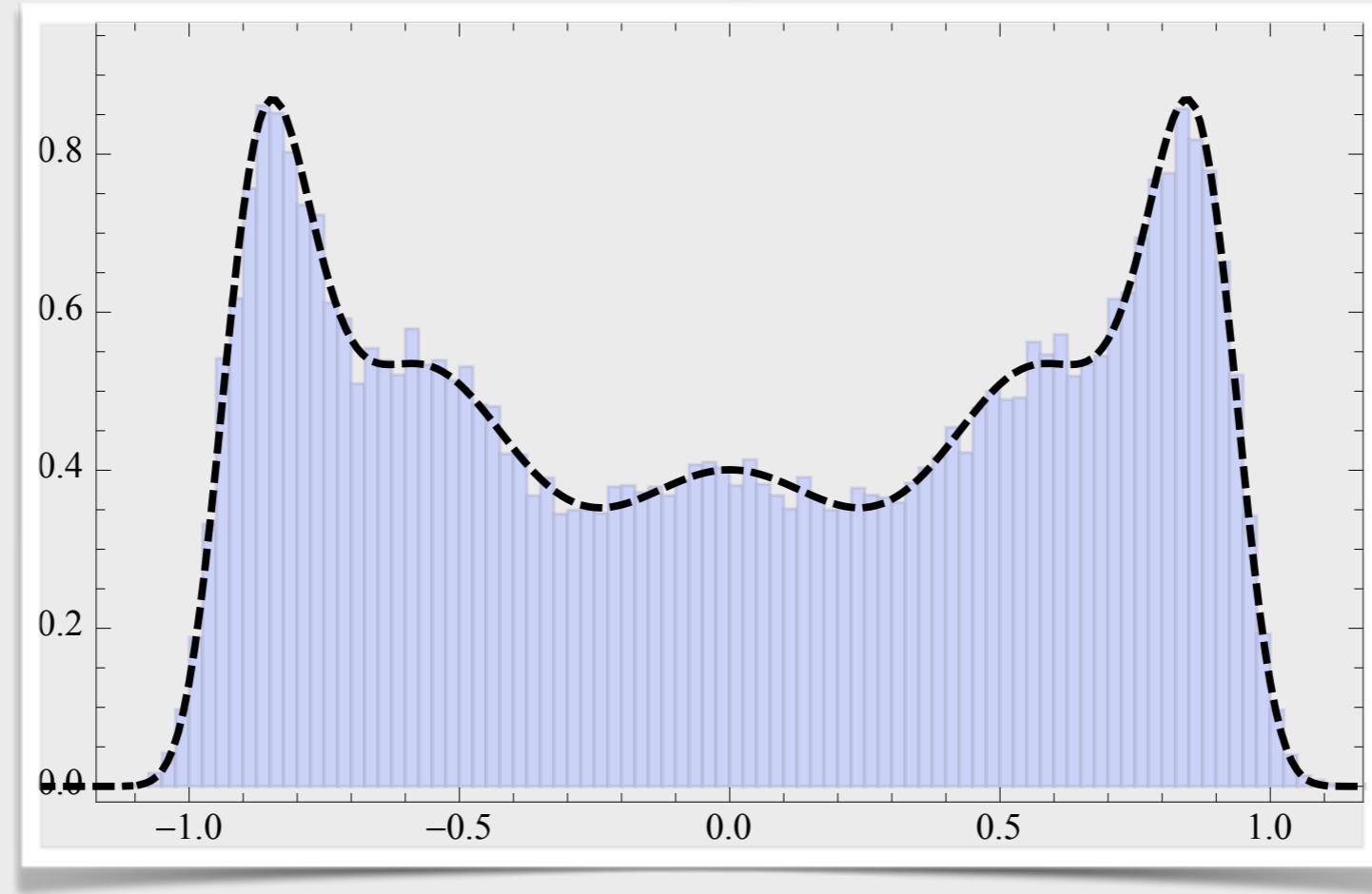
$$\frac{1}{N} K_N(x, x) = \frac{1}{N} \boldsymbol{\phi}_1(x)^T \boldsymbol{\phi}_1(x) \rightarrow \lambda_1,$$

let  $U$  be an  $N \times N - 1$  orthogonal matrix whose columns form an orthonormal basis the orthogonal complement of  $\{\boldsymbol{\phi}_1(\lambda_1)\}$ .

$$\boldsymbol{\phi}_1(\lambda_1)^T U = 0 \quad U^T U = I$$

$$\rho_2(x) = \frac{1}{N-1} \boldsymbol{\phi}_2(x)^T \boldsymbol{\phi}_2(x) \quad \boldsymbol{\phi}_2(x) = U^T \boldsymbol{\phi}_1(x)$$





Sampling with  $V(x) = x^8$  and  $N = 5$   
Note that  $H = U\Lambda U^*$

