What can random matrices tell us about algorithms?

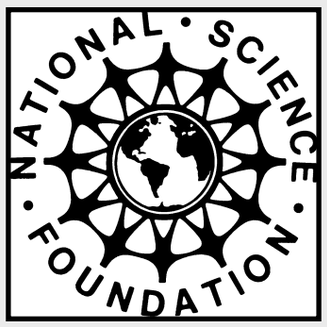
Case studies.

Tom Trogdon
trogdon@uw.edu
University of Washington
Department of Applied Mathematics

Acknowledgements

Joint work with:

- Percy Deift (NYU)
- Xiucai Ding (UC Davis)
- Tyler Chen (NYU)
- Elliot Paquette (McGill)



NSF Funding via:

CAREER: Numerical linear algebra, random matrix theory and applications

Perspectives on Matrix Computations: Theoretical Computer Science Meets Numerical analysis

Perspectives on Matrix Computations: Theoretical Computer Science Meets Numerical analysis

... Meets Mathematical Physics?

Perspectives on Matrix Computations: Theoretical Computer Science Meets Numerical analysis

... Meets Mathematical Physics?

P Deift, T Nanda, and C Tomei. Ordinary differential equations and the symmetric eigenvalue problem. SIAM J. on Numer. Anal., 20:1–22, 1983

Plan

- Lanczos/CG and the power method on sample covariance matrices
- Lanczos and finite-precision effects
- Riemann-Hilbert analysis for CG/Lanczos*

Building blocks of random matrix theory

The real Ginibre Ensemble, $\mathrm{Gin}_{\mathbb{R}}(N,M)$, is the matrix

$$Y = \left(\frac{Y_{ij}}{\sqrt{M}}\right)_{1\le i\le N, 1\le j\le M}, \quad Y_{ij} \text{ iid standard normal random variables.}$$

The complex Ginibre Ensemble, $\mathrm{Gin}_{\mathbb{C}}(N,M)$, is the matrix

$$X = \frac{1}{\sqrt{2}}(Y_1 + \mathrm{i}Y_2), \quad Y_1, Y_2 \text{ independent } \mathrm{Gin}_{\mathbb{R}}(N,M).$$

If $N = M$ we use $\mathrm{Gin}_{\mathbb{F}}(N)$.

The Gaussian Unitary Ensemble, $\mathrm{GUE}(N)$, is the matrix

$$H = \frac{1}{\sqrt{2}}(X + X^*), \quad X \sim \mathrm{Gin}_{\mathbb{C}}(N).$$

Sample covariance matrices (SCMs)

The classical examples of sample covariance matrices are:

The real Wishart Ensemble, $\mathrm{Wishart}_{\mathbb{R}}(N,M)$:

$$W = XX^T, \quad X \sim \mathrm{Gin}_{\mathbb{R}}(N,M).$$

The complex Wishart Ensemble, $\mathrm{Wishart}_{\mathbb{C}}(N,M)$:

$$W = XX^*, \quad X \sim \mathrm{Gin}_{\mathbb{C}}(N,M).$$

Sample covariance matrices (SCMs)

The classical examples of sample covariance matrices are:

The real Wishart Ensemble, $\text{Wishart}_{\mathbb{R}}(N,M)$:

$$W = XX^T, \quad X \sim \text{Gin}_{\mathbb{R}}(N,M).$$

The complex Wishart Ensemble, $\text{Wishart}_{\mathbb{C}}(N,M)$:

$$W = XX^*, \quad X \sim \text{Gin}_{\mathbb{C}}(N,M).$$

But there is, in principle, no reason why the entries of X should be Gaussian, and no reason why the entries in each column need to be independent.

Example: X could be an $N \times M$ matrix of iid random variables, $X_{11} = \pm 1/\sqrt{M}$ with equal probability.

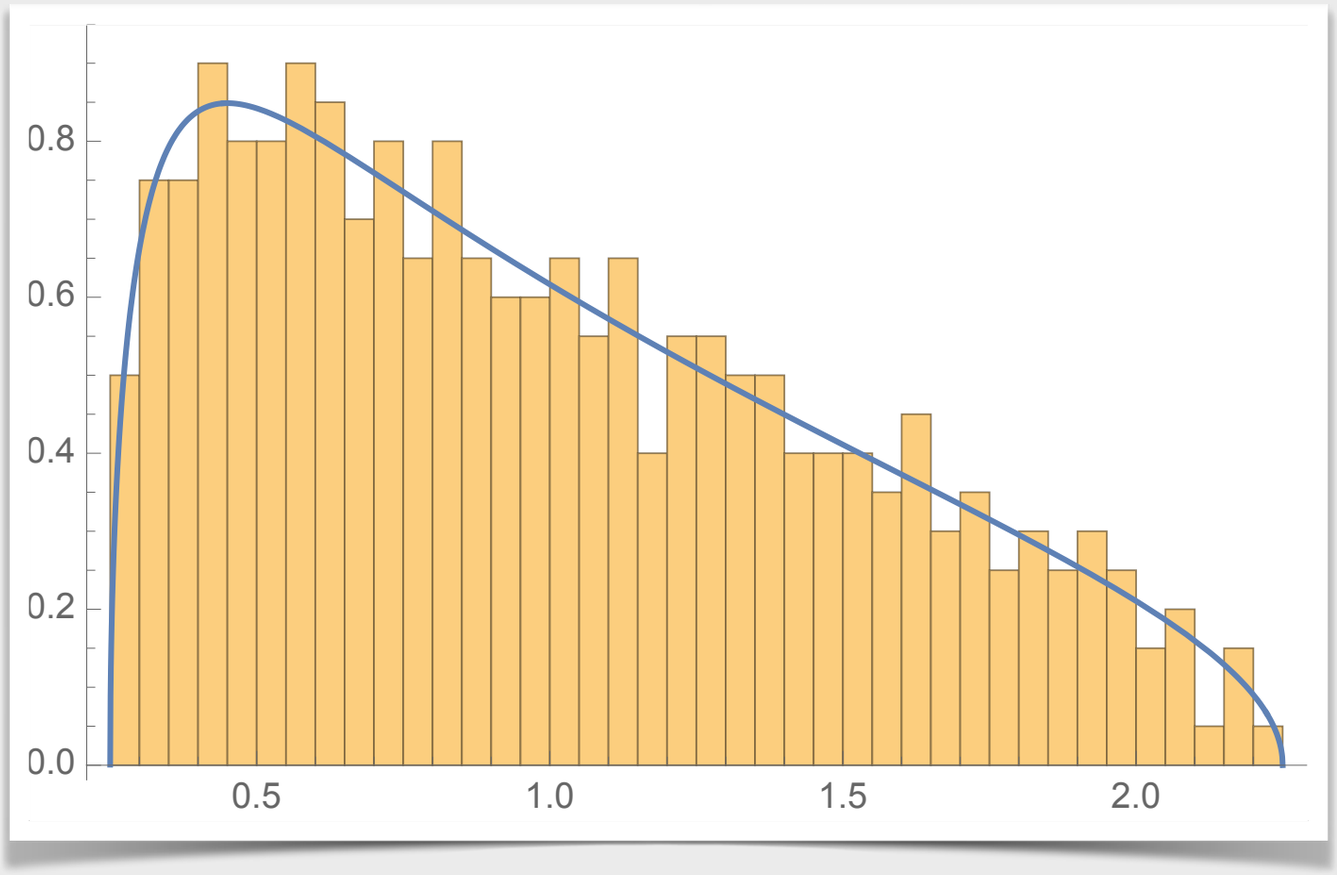
Sample covariance matrices (SCMs)

The classical examples of sample covariance matrices are:

The real Wishart Ensemble, $\mathrm{Wishart}_{\mathbb{R}}(N,M)$:

$$W = XX^T, \quad X \sim \mathrm{Gin}_{\mathbb{R}}(N,M).$$

The complex Wishart Ensemble, $\mathrm{Wishart}_{\mathbb{C}}(N,M)$:



The Marchenko–Pastur distribution gives the macroscopic behavior of the spectrum:

$$p_d(x) = \frac{1}{2\pi d}\frac{\sqrt{[(\lambda_+ - x)(x-\lambda_-)]_+}}{|x|}$$

$$\lambda_\pm = (1 \pm \sqrt{d})^2$$

$$M \sim N/d$$

$$\frac{1}{N}\sum_{j=1}^{N} \delta_{\lambda_j} \rightharpoonup p_d(x)\mathrm{d}x$$

CG and Wishart: An exactly solvable case

The CG algorithm and orthogonal polynomials

The CG algorithm applied to $W\mathbf{x} = \mathbf{b}$ is equivalent to the polynomial minimization problem

$$\min_{P:\,\text{degree } k,\,P(0)=1} \|W^{-1}P(W)\mathbf{b}\|_W^2 = \min_{P:\,\text{degree } k,\,P(0)=1} \int \frac{P(\lambda)^2}{\lambda}\,\mu_{W,\mathbf{b}}(\mathrm{d}\lambda),$$

where

$$\mu_{W,\mathbf{b}} = \sum_{j=1}^{N} |\langle \mathbf{q}_j, \mathbf{b}\rangle|^2 \delta_{\lambda_j},$$

is the so-called eigenvector spectral distribution (VESD).

M Hestenes and E Steifel. Method of Conjugate Gradients for Solving Linear Systems. J. Research Nat. Bur. Standards, 20:409–436, 1952

Z Bai and J W Silverstein. Spectral Analysis of Large Dimensional Random Matrices. Springer Series in Statistics. Springer New York, New York, NY, 2010

The CG algorithm and orthogonal polynomials

The CG algorithm applied to $W\mathbf{x} = \mathbf{b}$ is equivalent to the polynomial minimization problem

$$\min_{P:\,\text{degree } k,\,P(0)=1} \|W^{-1}P(W)\mathbf{b}\|_W^2 = \min_{P:\,\text{degree } k,\,P(0)=1} \int \frac{P(\lambda)^2}{\lambda}\,\mu_{W,\mathbf{b}}(\mathrm{d}\lambda),$$

where

$$\mu_{W,\mathbf{b}} = \sum_{j=1}^{N} |\langle \mathbf{q}_j, \mathbf{b}\rangle|^2 \delta_{\lambda_j},$$

is the so-called eigenvector spectral distribution (VESD).

M Hestenes and E Steifel. Method of Conjugate Gradients for Solving Linear Systems. J. Research Nat. Bur. Standards, 20:409–436, 1952

Z Bai and J W Silverstein. Spectral Analysis of Large Dimensional Random Matrices. Springer Series in Statistics. Springer New York, New York, NY, 2010

A useful (Riemann–Hilbert) tool for orthogonal polynomials:

A S Fokas, A R Its, and A V Kitaev. The isomonodromy approach to matrix models in 2D quantum gravity. Comm. Math. Phys., 147(2):395–430, 1992

P Deift. Orthogonal Polynomials and Random Matrices: a Riemann-Hilbert Approach. Amer. Math. Soc., Providence, RI, 2000

Wishart matrices

Suppose $W \overset{\mathrm{L}}{=} \mathrm{Wishart}(N, M)$.

If $\mathbf{b} = \mathbf{f}_1$ (the first standard basis vector), then Lanczos is mathematically equivalent to Householder tridiagonalization. By the invariance of Wishart, we use $\mathbf{b} = \mathbf{f}_1$.

$$\mathscr{L}_k(\mu_{W,\mathbf{b}}) \overset{\mathrm{L}}{=} \frac{1}{\sqrt{\beta M}} \begin{bmatrix} \chi_{\beta M} & & & & & \\ \chi_{\beta(N-1)} & \chi_{\beta(M-1)} & & & & \\ & \chi_{\beta(N-2)} & \chi_{\beta(M-2)} & & & \\ & & \ddots & \ddots & & \\ & & & \chi_{\beta(N-k+1)} & \chi_{\beta(M-k+1)} \end{bmatrix}$$

J W Silverstein. The Smallest Eigenvalue of a Large Dimensional Wishart Matrix. The Annals of Probability, 13(4):1364–1368, 1985

I Dumitriu and A Edelman. Matrix models for beta ensembles. Journal of Mathematical Physics, 43(11):5830, oct 2002

Wishart matrices

Suppose $W \overset{\mathrm{L}}{=} \mathrm{Wishart}(N, M)$.

If $\mathbf{b} = \mathbf{f}_1$ (the first standard basis vector), then Lanczos is mathematically equivalent to Householder tridiagonalization. By the invariance of Wishart, we use $\mathbf{b} = \mathbf{f}_1$.

$$\mathscr{L}_k(\mu_{W,\mathbf{b}}) \overset{\mathrm{L}}{=} \frac{1}{\sqrt{\beta M}} \begin{bmatrix} \chi_{\beta M} & & & & & \\ \chi_{\beta(N-1)} & \chi_{\beta(M-1)} & & & & \\ & \chi_{\beta(N-2)} & \chi_{\beta(M-2)} & & & \\ & & \ddots & \ddots & & \\ & & & \chi_{\beta(N-k+1)} & \chi_{\beta(M-k+1)} & \end{bmatrix}$$

Cholesky factorization of Lanczos matrix

J W Silverstein. The Smallest Eigenvalue of a Large Dimensional Wishart Matrix. The Annals of Probability, 13(4):1364–1368, 1985

I Dumitriu and A Edelman. Matrix models for beta ensembles. Journal of Mathematical Physics, 43(11):5830, oct 2002

Wishart matrices

Suppose $W \overset{\mathrm{L}}{=} \mathrm{Wishart}(N, M)$.

If $\mathbf{b} = \mathbf{f}_1$ (the first standard basis vector), then Lanczos is mathematically equivalent to Householder tridiagonalization. By the invariance of Wishart, we use $\mathbf{b} = \mathbf{f}_1$.

$$\mathscr{L}_k(\mu_{W,\mathbf{b}}) \overset{\mathrm{L}}{=} \frac{1}{\sqrt{\beta M}} \begin{bmatrix} \chi_{\beta M} & & & & \\ \chi_{\beta(N-1)} & \chi_{\beta(M-1)} & & & \\ & \chi_{\beta(N-2)} & \chi_{\beta(M-2)} & & \\ & & \ddots & \ddots & \\ & & & \chi_{\beta(N-k+1)} & \chi_{\beta(M-k+1)} \end{bmatrix}$$

Cholesky factorization of Lanczos matrix

J W Silverstein. The Smallest Eigenvalue of a Large Dimensional Wishart Matrix. The Annals of Probability, 13(4):1364–1368, 1985

I Dumitriu and A Edelman. Matrix models for beta ensembles. Journal of Mathematical Physics, 43(11):5830, oct 2002

$$\|\mathbf{r}_k(W, \mathbf{b})\|_2 \overset{\mathrm{L}}{=} \prod_{j=0}^{k-1} \frac{\chi_{\beta(N-j-1)}}{\chi_{\beta(M-j)}}$$

$$\|\mathbf{e}_k(W, \mathbf{b})\|_W \overset{\mathrm{L}}{=} \frac{\sqrt{\beta M}}{\chi_{\beta(M-N+1)}} \prod_{j=0}^{k-1} \frac{\chi_{\beta(N-j-1)}}{\chi_{\beta(M-j)}}$$

Wishart matrices

Suppose $W \overset{\mathrm{L}}{=} \mathrm{Wishart}(N, M)$.

If $\mathbf{b} = \mathbf{f}_1$ (the first standard basis vector), then Lanczos is mathematically equivalent to Householder tridiagonalization. By the invariance of Wishart, we use $\mathbf{b} = \mathbf{f}_1$.

$$\mathscr{L}_k(\mu_{W,\mathbf{b}}) \overset{\mathrm{L}}{=} \frac{1}{\sqrt{\beta M}} \begin{bmatrix} \chi_{\beta M} & & & & & \\ \chi_{\beta(N-1)} & \chi_{\beta(M-1)} & & & & \\ & \chi_{\beta(N-2)} & \chi_{\beta(M-2)} & & & \\ & & \ddots & \ddots & & \\ & & & \chi_{\beta(N-k+1)} & \chi_{\beta(M-k+1)} \end{bmatrix}$$

Cholesky factorization of Lanczos matrix

J W Silverstein. The Smallest Eigenvalue of a Large Dimensional Wishart Matrix. The Annals of Probability, 13(4):1364–1368, 1985

I Dumitriu and A Edelman. Matrix models for beta ensembles. Journal of Mathematical Physics, 43(11):5830, oct 2002

$$\|\mathbf{r}_k(W, \mathbf{b})\|_2 \overset{\mathrm{L}}{=} \prod_{j=0}^{k-1} \frac{\chi_{\beta(N-j-1)}}{\chi_{\beta(M-j)}}$$

$$\|\mathbf{e}_k(W, \mathbf{b})\|_W \overset{\mathrm{L}}{=} \frac{\sqrt{\beta M}}{\chi_{\beta(M-N+1)}} \prod_{j=0}^{k-1} \frac{\chi_{\beta(N-j-1)}}{\chi_{\beta(M-j)}}$$

E Paquette and T T. Universality for the Conjugate Gradient and MINRES Algorithms on Sample Covariance Matrices. Communications on Pure and Applied Mathematics, sep 2022

Beyond Wishart: Universal fluctuations

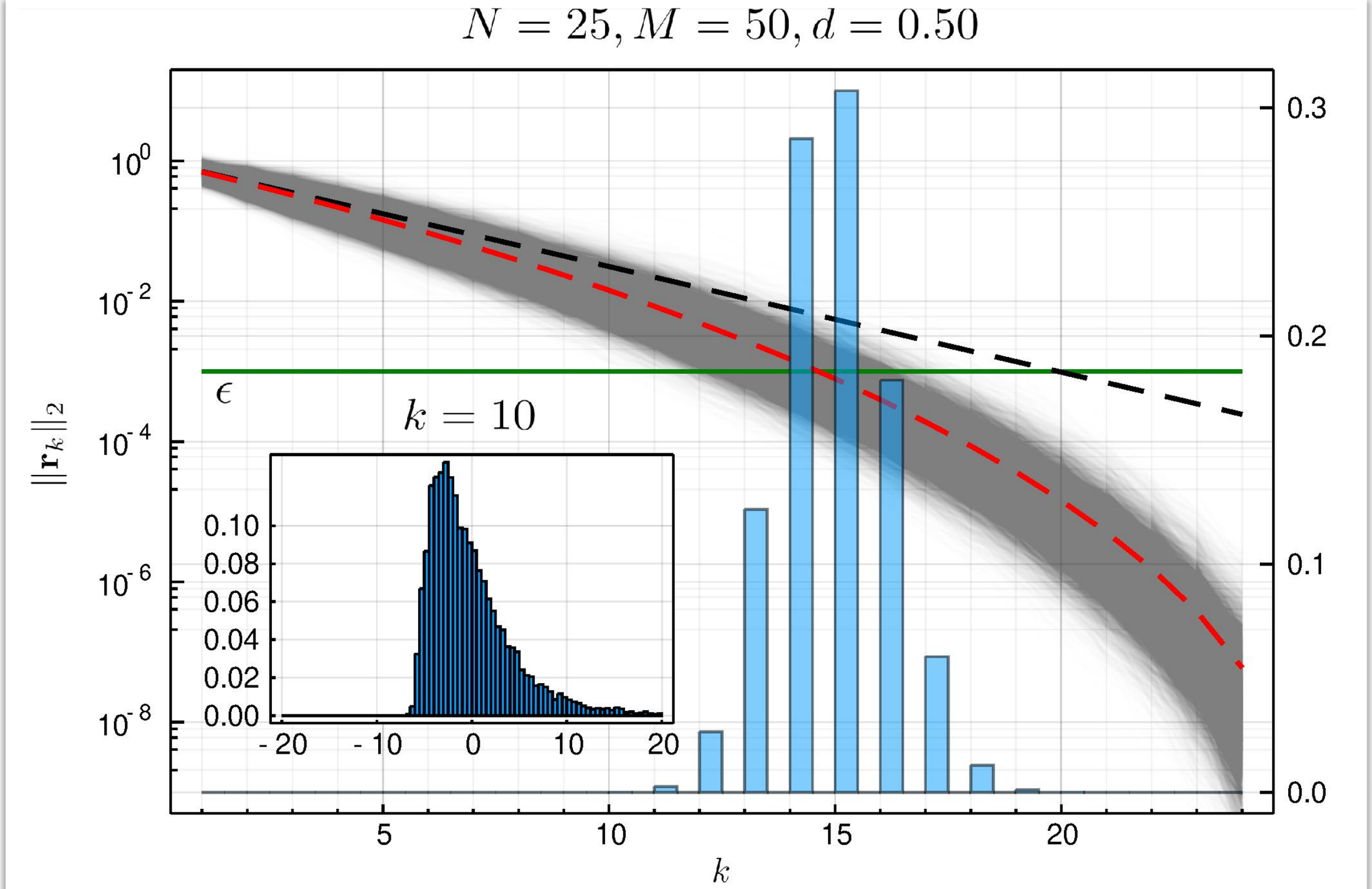
$$\|\mathbf{r}_k(W,\mathbf{b})\|_2^2 \sim \left(\frac{N}{M}\right)^k \left(1+\sqrt{\frac{2k}{\beta M}}\sqrt{1+\frac{N}{M}}\,\mathcal{N}(0,1)\right)$$

We will also consider the halting time:

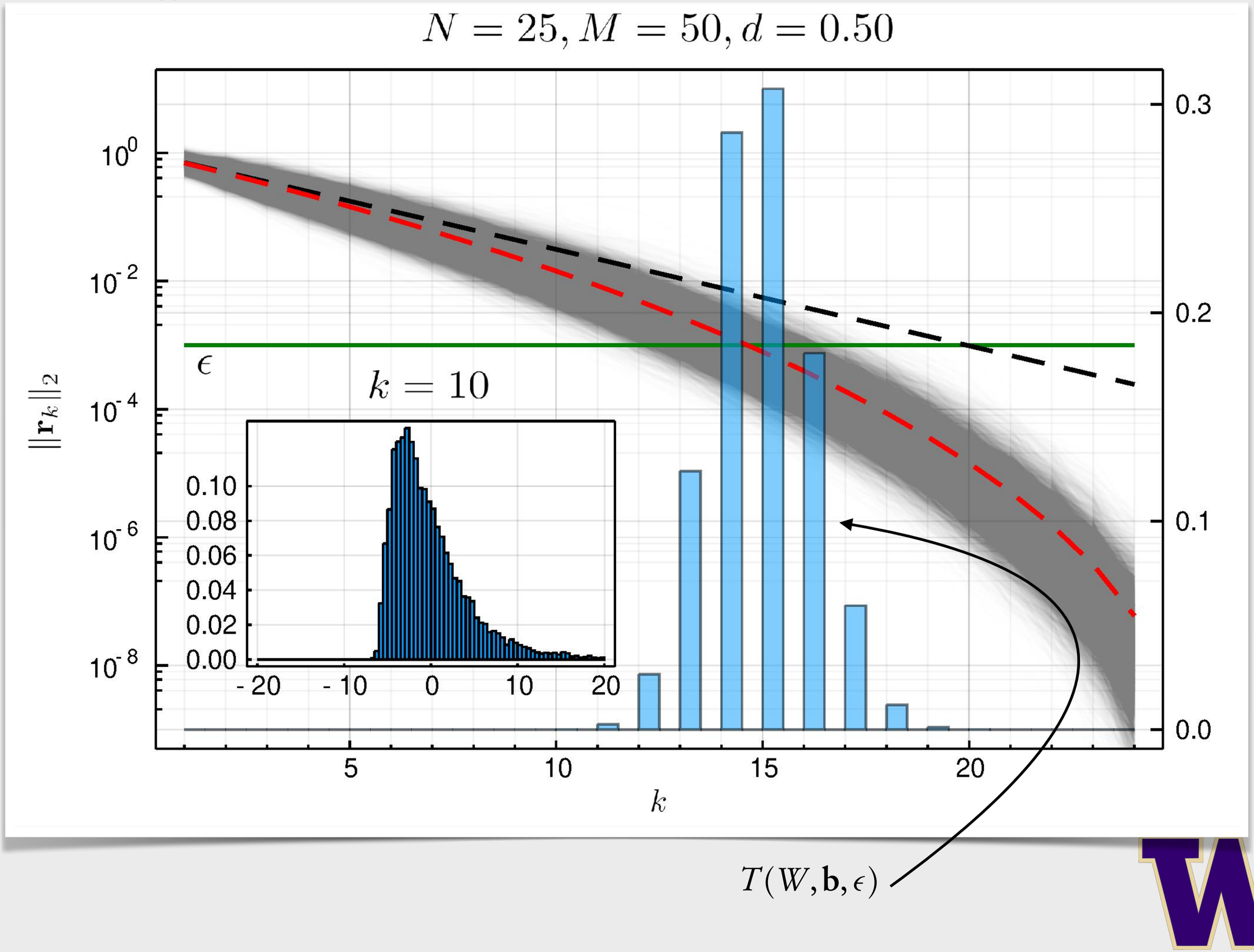
$$T(W,\mathbf{b},\epsilon) := \min\{k : \|\mathbf{r}_k(W,\mathbf{b})\| < \epsilon\}.$$

E Paquette and T T. Universality for the Conjugate Gradient and MINRES Algorithms on Sample Covariance Matrices. Communications on Pure and Applied Mathematics, sep 2022

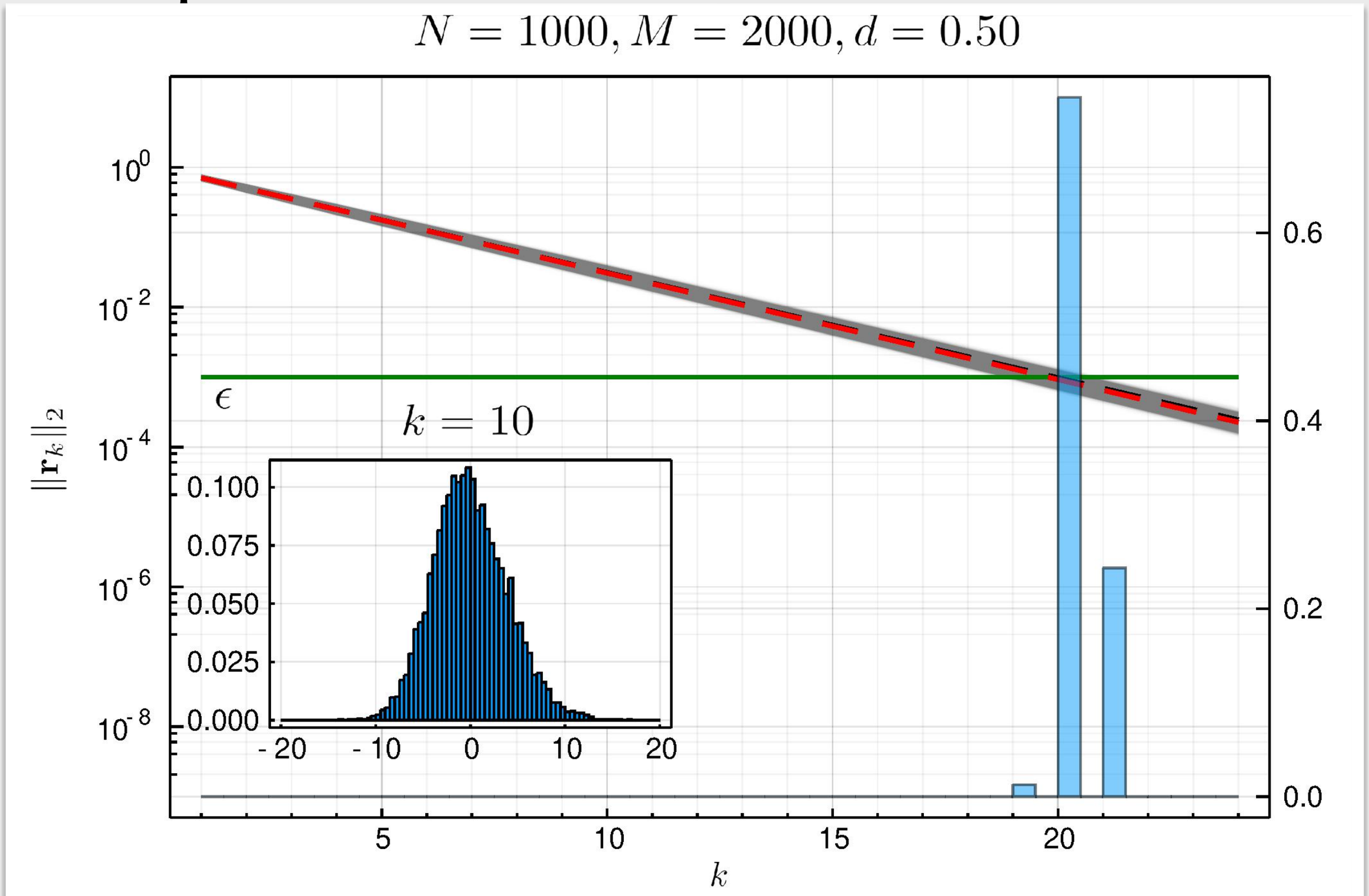
Sample covariance with trivial covariance



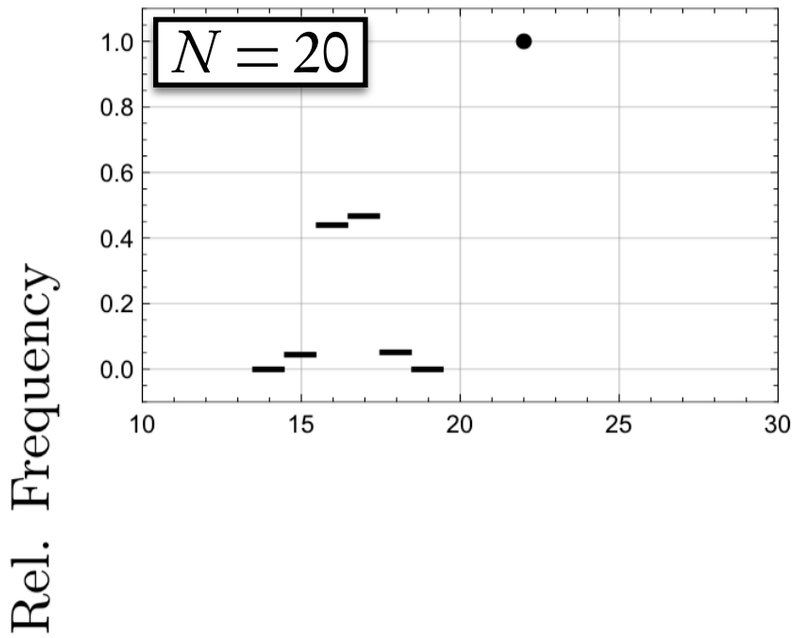
Sample covariance with trivial covariance



Sample covariance with trivial covariance



Concentration occurs if $\dfrac{\log \epsilon^{-1}}{\sqrt{M}} \ll 1$

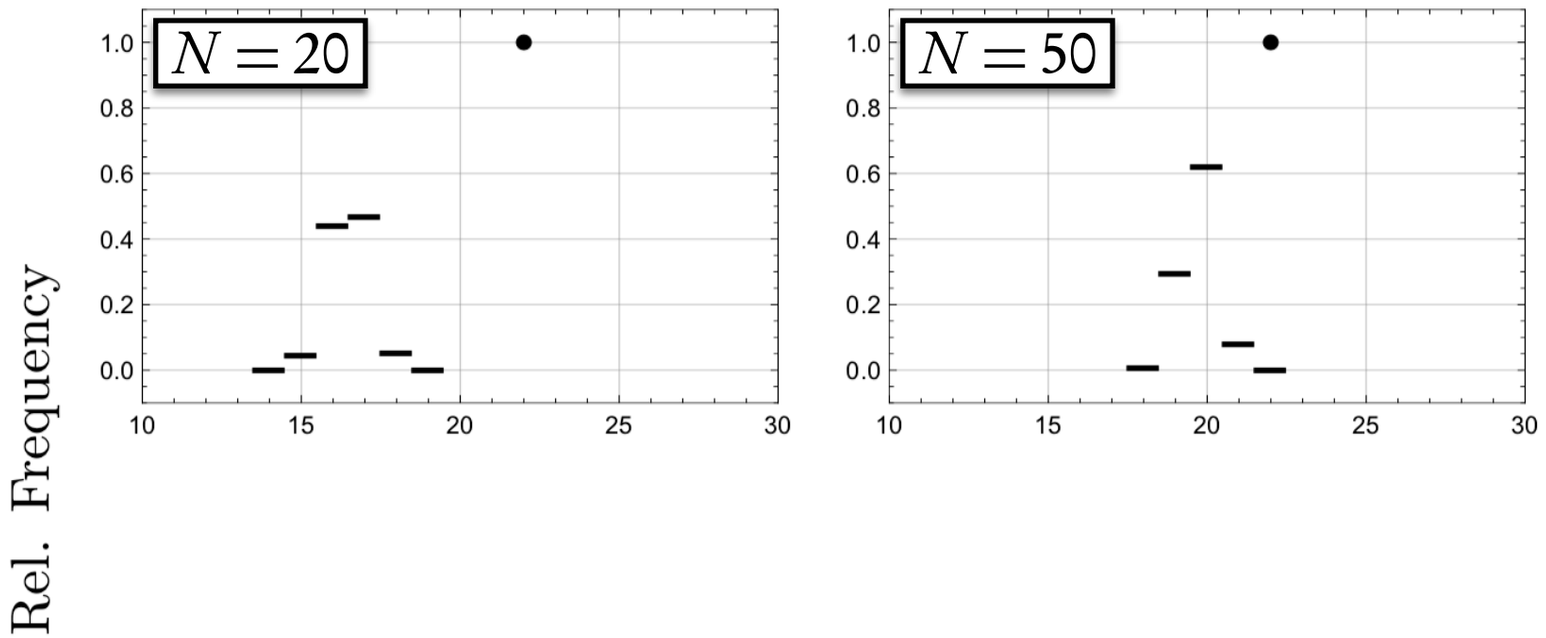


Statistics for $T(W, \mathbf{b}, \epsilon) = \min\{k : \|\mathbf{r}_k\|_2 < \epsilon\}$

$\text{Wishart}_{\mathbb{R}}(N, M)$, $d = 0.2$, $\epsilon \approx 6 \times 10^{-8}$
20,000 samples

$$\forall k \quad \epsilon \neq d^{k/2} \quad \mathbb{P}\left(T(W, \mathbf{b}, \epsilon) \neq \left\lceil \frac{2 \log \epsilon}{\log d} \right\rceil \right) \leq C e^{-cN}$$

$$\epsilon = d^{k/2} \quad \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k) = \frac{1}{2} = \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k + 1)$$

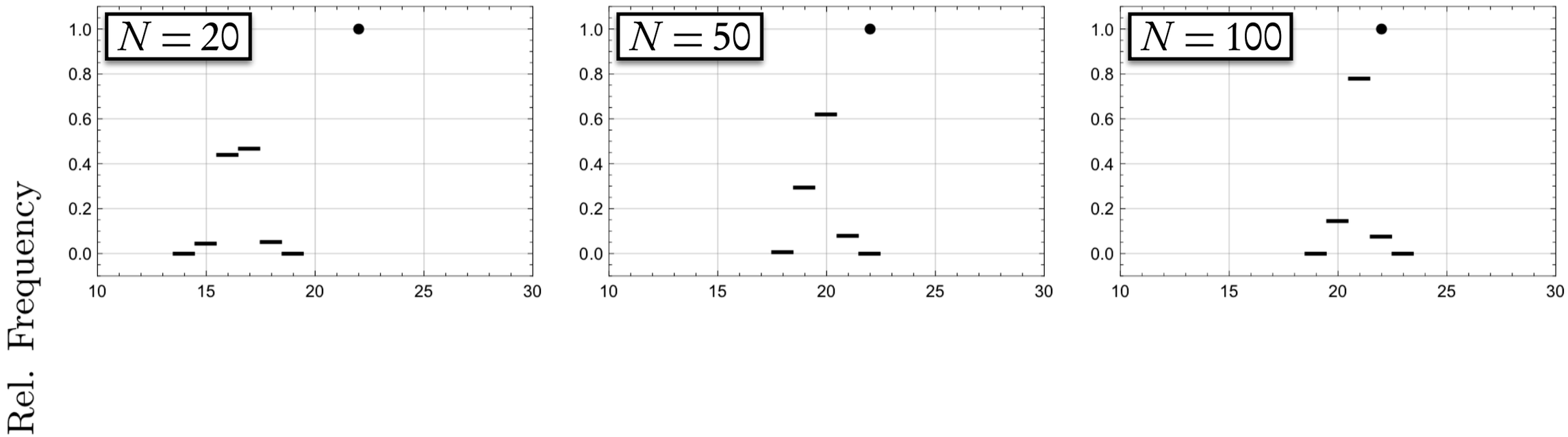


Statistics for $T(W, \mathbf{b}, \epsilon) = \min\{k : ||\mathbf{r}_k||_2 < \epsilon\}$

$\text{Wishart}_{\mathbb{R}}(N, M)$, $d = 0.2$, $\epsilon \approx 6 \times 10^{-8}$
20,000 samples

$$\forall k \quad \epsilon \neq d^{k/2} \quad \mathbb{P}\left(T(W, \mathbf{b}, \epsilon) \neq \left\lceil \frac{2 \log \epsilon}{\log d} \right\rceil \right) \leq C e^{-cN}$$

$$\epsilon = d^{k/2} \quad \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k) = \frac{1}{2} = \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k + 1)$$

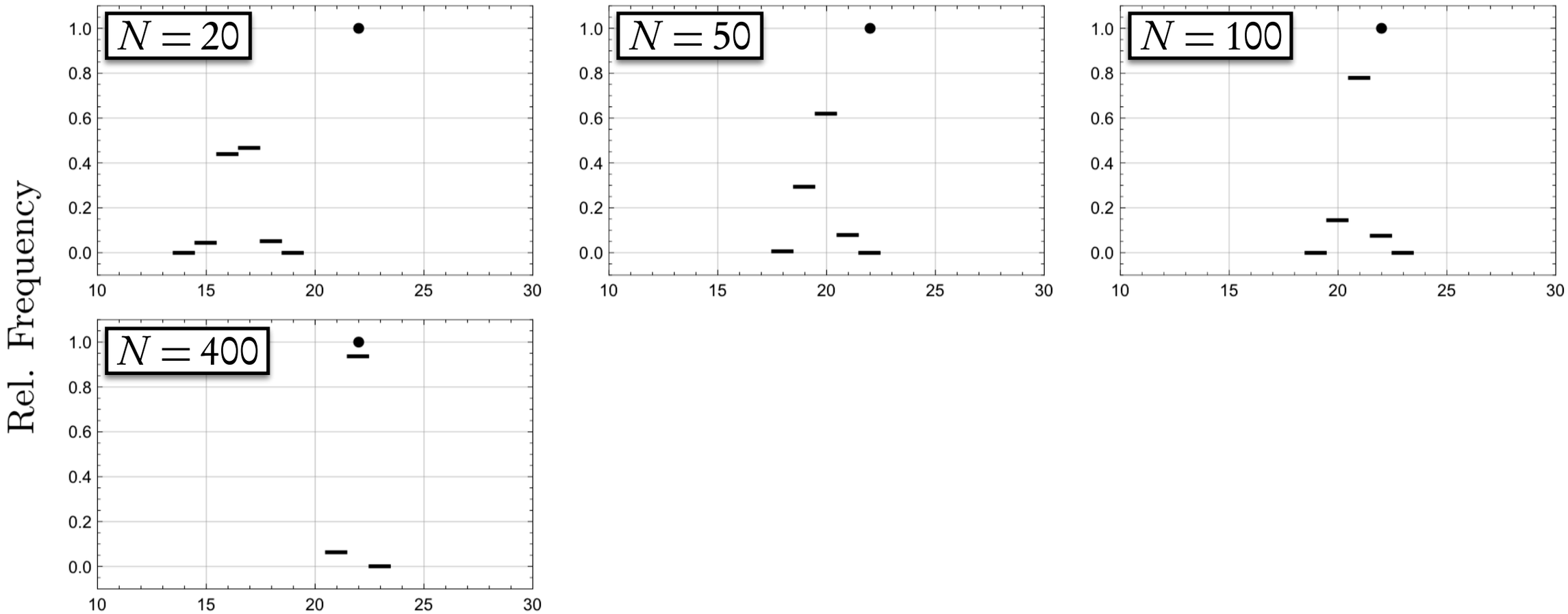


Statistics for $T(W, \mathbf{b}, \epsilon) = \min\{k : ||\mathbf{r}_k||_2 < \epsilon\}$

$\text{Wishart}_{\mathbb{R}}(N, M)$, $d = 0.2$, $\epsilon \approx 6 \times 10^{-8}$
20,000 samples

$$\forall k \quad \epsilon \neq d^{k/2} \quad \mathbb{P}\left(T(W, \mathbf{b}, \epsilon) \neq \left\lceil \frac{2 \log \epsilon}{\log d} \right\rceil \right) \leq C e^{-cN}$$

$$\epsilon = d^{k/2} \quad \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k) = \frac{1}{2} = \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k + 1)$$

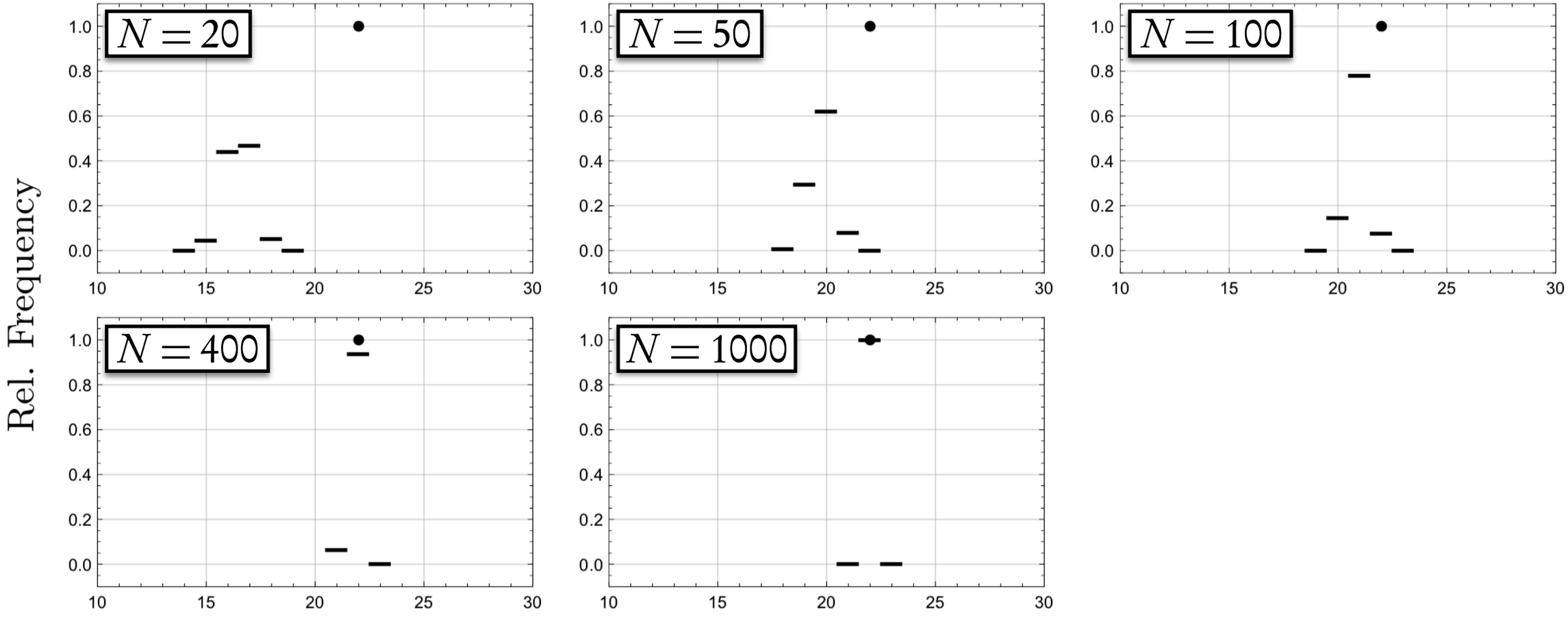


Statistics for $T(W, \mathbf{b}, \epsilon) = \min\{k : ||\mathbf{r}_k||_2 < \epsilon\}$

$\text{Wishart}_{\mathbb{R}}(N, M)$, $d = 0.2$, $\epsilon \approx 6 \times 10^{-8}$
20,000 samples

$$\forall k \quad \epsilon \neq d^{k/2} \quad \mathbb{P}\left(T(W, \mathbf{b}, \epsilon) \neq \left\lceil \frac{2 \log \epsilon}{\log d} \right\rceil\right) \leq C e^{-cN}$$

$$\epsilon = d^{k/2} \quad \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k) = \frac{1}{2} = \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k + 1)$$

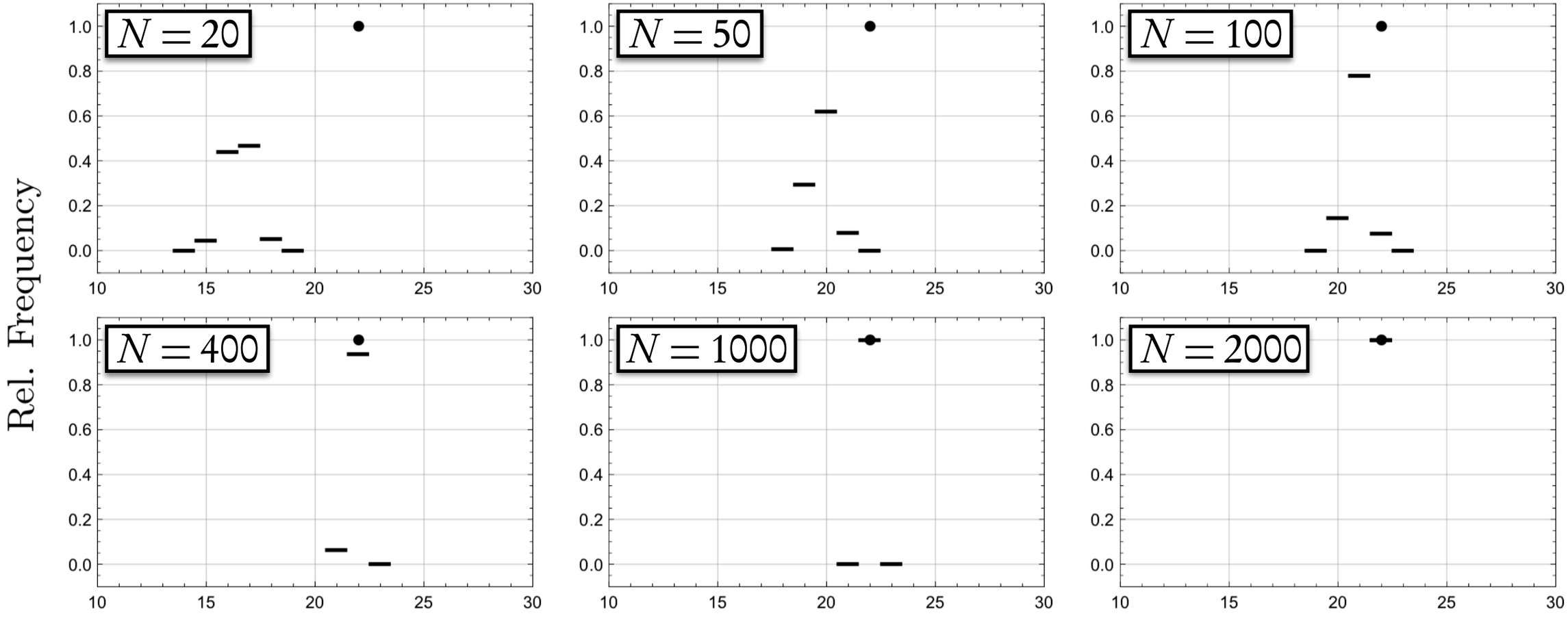


Statistics for $T(W, \mathbf{b}, \epsilon) = \min\{k : \|\mathbf{r}_k\|_2 < \epsilon\}$

$\text{Wishart}_{\mathbb{R}}(N, M)$, $d = 0.2$, $\epsilon \approx 6 \times 10^{-8}$
20,000 samples

$$\forall k \quad \epsilon \neq d^{k/2} \quad \mathbb{P}\left(T(W, \mathbf{b}, \epsilon) \neq \left\lceil \frac{2 \log \epsilon}{\log d} \right\rceil\right) \leq C e^{-cN}$$

$$\epsilon = d^{k/2} \quad \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k) = \frac{1}{2} = \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k + 1)$$



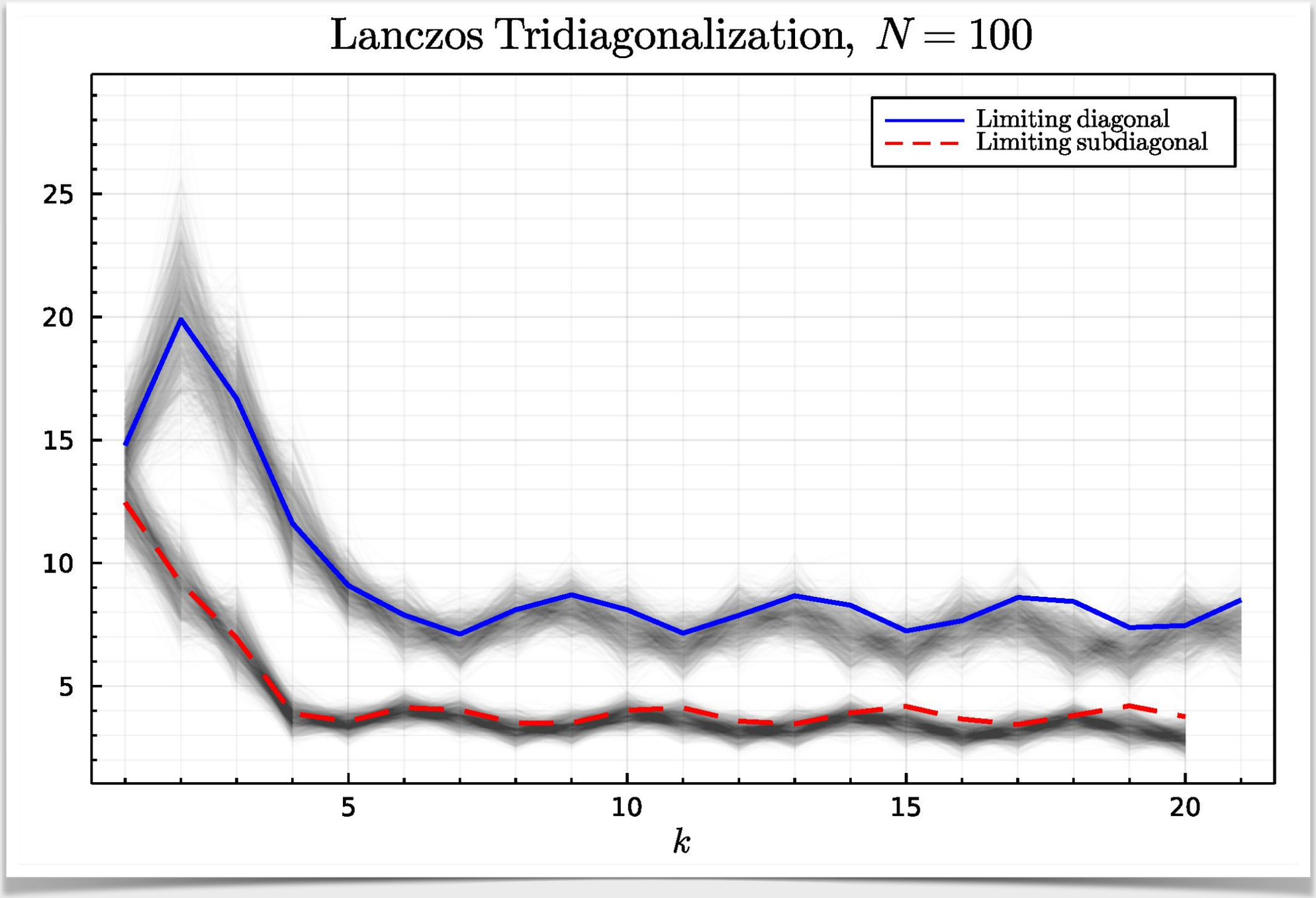
Statistics for $T(W, \mathbf{b}, \epsilon) = \min\{k : \|\mathbf{r}_k\|_2 < \epsilon\}$

$\text{Wishart}_{\mathbb{R}}(N, M)$, $d = 0.2$, $\epsilon \approx 6 \times 10^{-8}$
20,000 samples

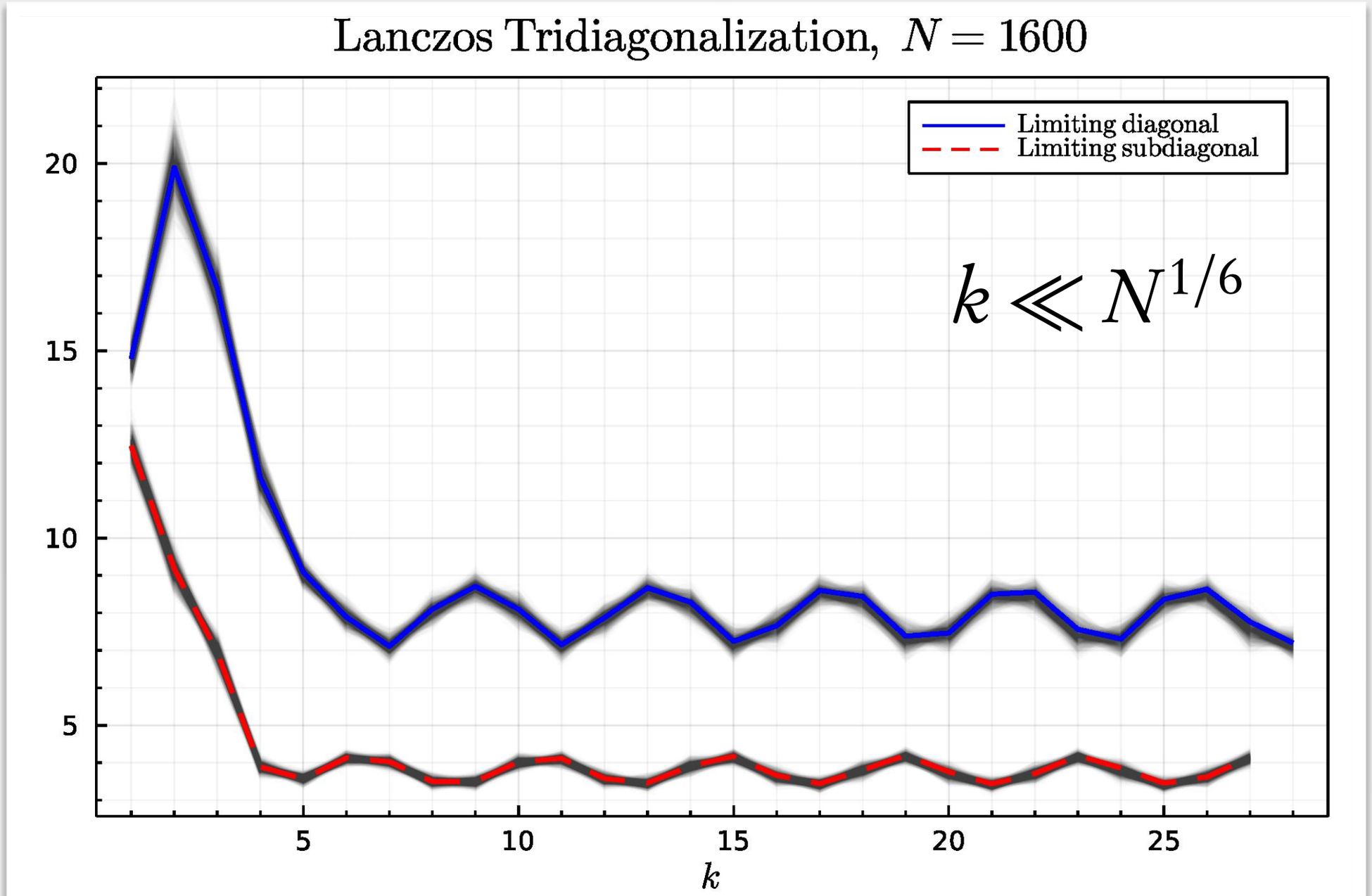
$$\forall k \quad \epsilon \neq d^{k/2} \quad \mathbb{P}\left(T(W, \mathbf{b}, \epsilon) \neq \left\lceil \frac{2 \log \epsilon}{\log d} \right\rceil\right) \leq C e^{-cN}$$

$$\epsilon = d^{k/2} \quad \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k) = \frac{1}{2} = \lim_{N \to \infty} \mathbb{P}(T(W, \mathbf{b}, \epsilon) = k + 1)$$

Spiked covariance with multiple bulk components



Spiked covariance with multiple bulk components



X Ding and T T. The conjugate gradient algorithm on a general class of spiked covariance matrices. Quarterly of Applied Mathematics, 80(1):99–155, nov 2021

X Ding and T T. A Riemann–Hilbert approach to the perturbation theory for orthogonal polynomials: Applications to numerical linear algebra and random matrix theory. arXiv preprint 2112.12354, pages 1–77, 2021

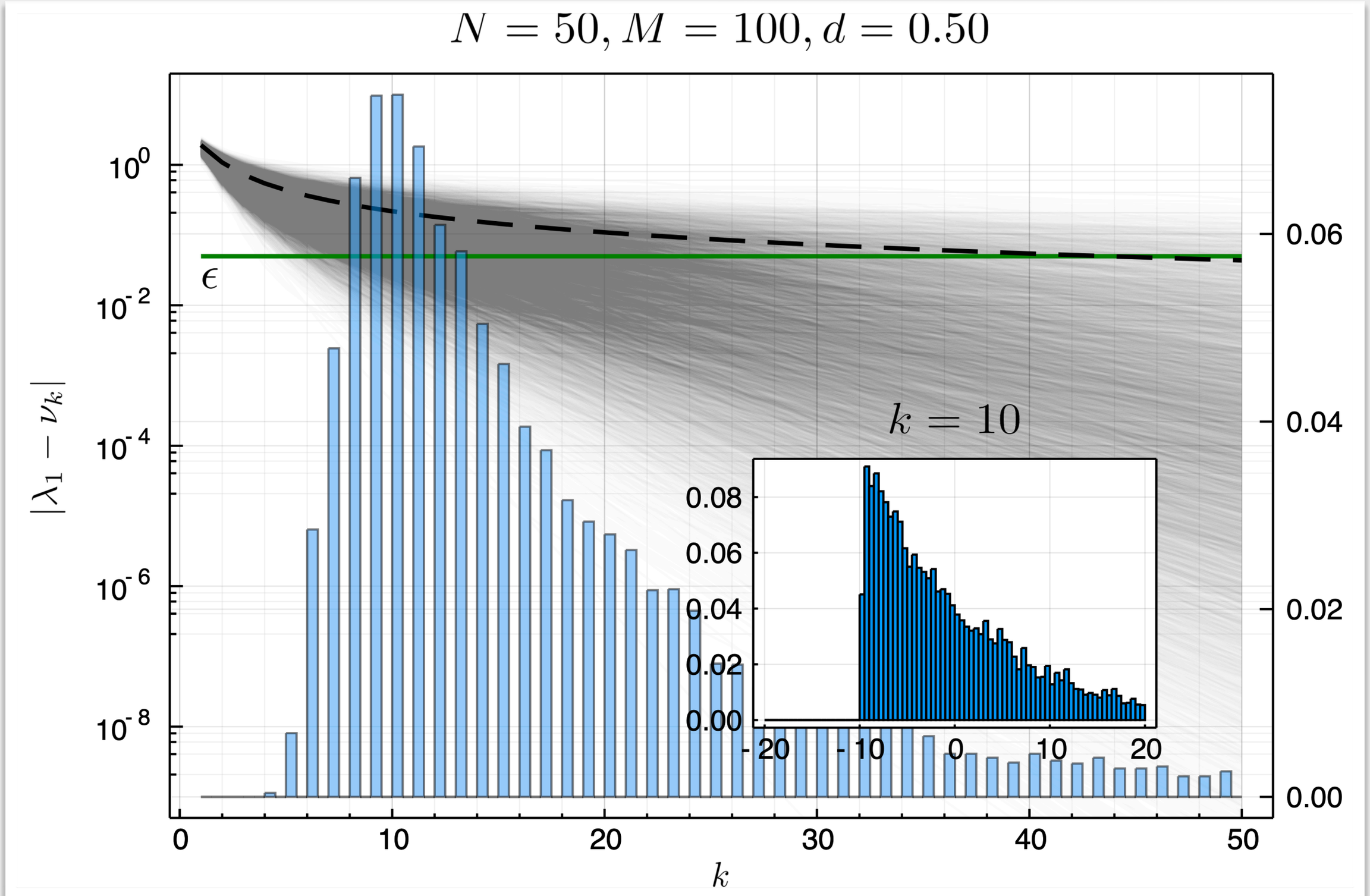
The power method

$$\nu_k(W,\mathbf{b}) = \frac{\int \lambda^{2k-1}\mu_{W,\mathbf{b}}(\mathrm{d}\lambda)}{\int \lambda^{2k-2}\mu_{W,\mathbf{b}}(\mathrm{d}\lambda)}, \quad k = 1,2,\ldots.$$

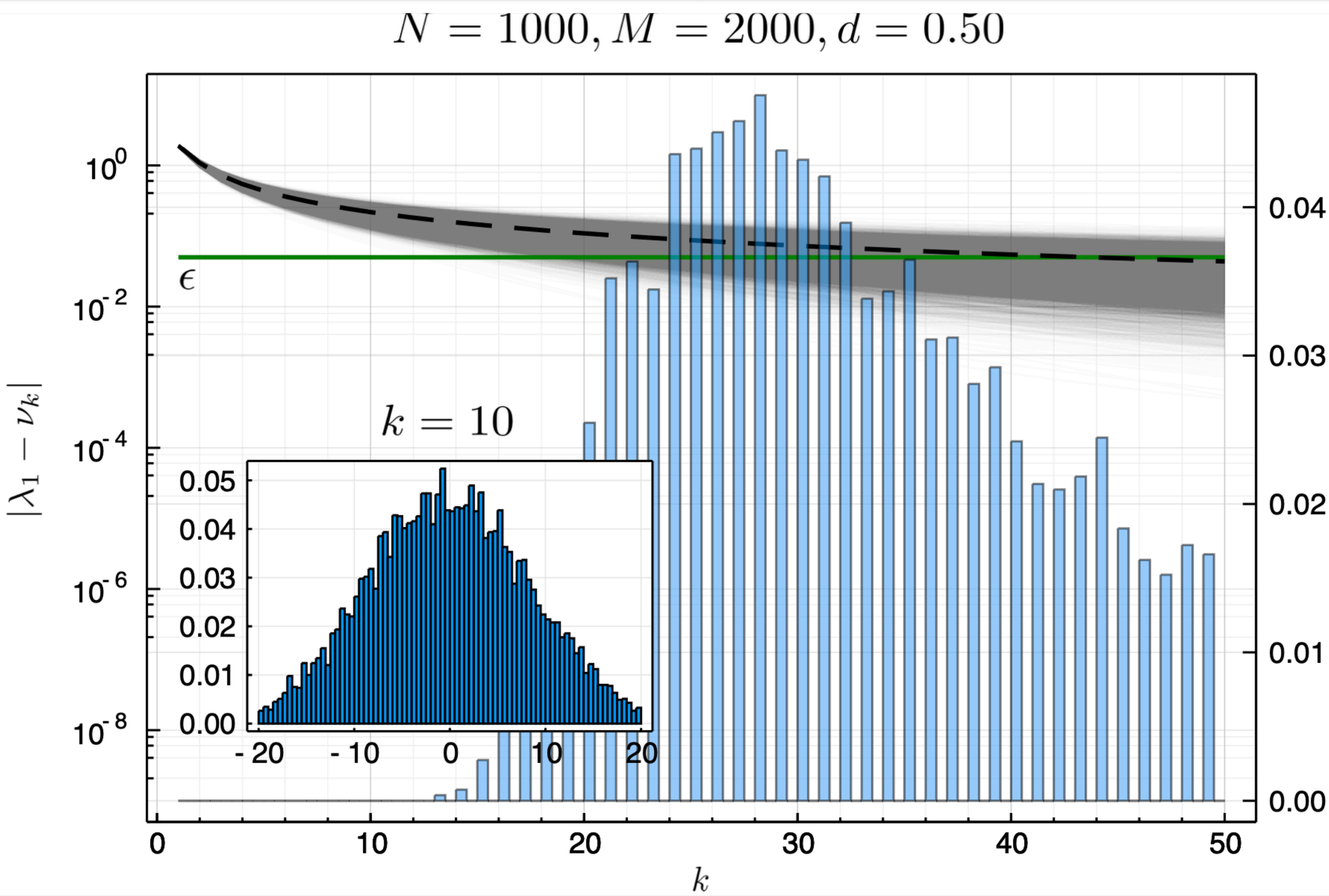
$$\nu_k(W,\mathbf{b}) \xrightarrow{N\to\infty} \frac{\int \lambda^{2k-1}\mu_{\mathrm{MP}}(\mathrm{d}\lambda)}{\int \lambda^{2k-2}\mu_{\mathrm{MP}}(\mathrm{d}\lambda)}, \quad k = 1,2,\ldots.$$

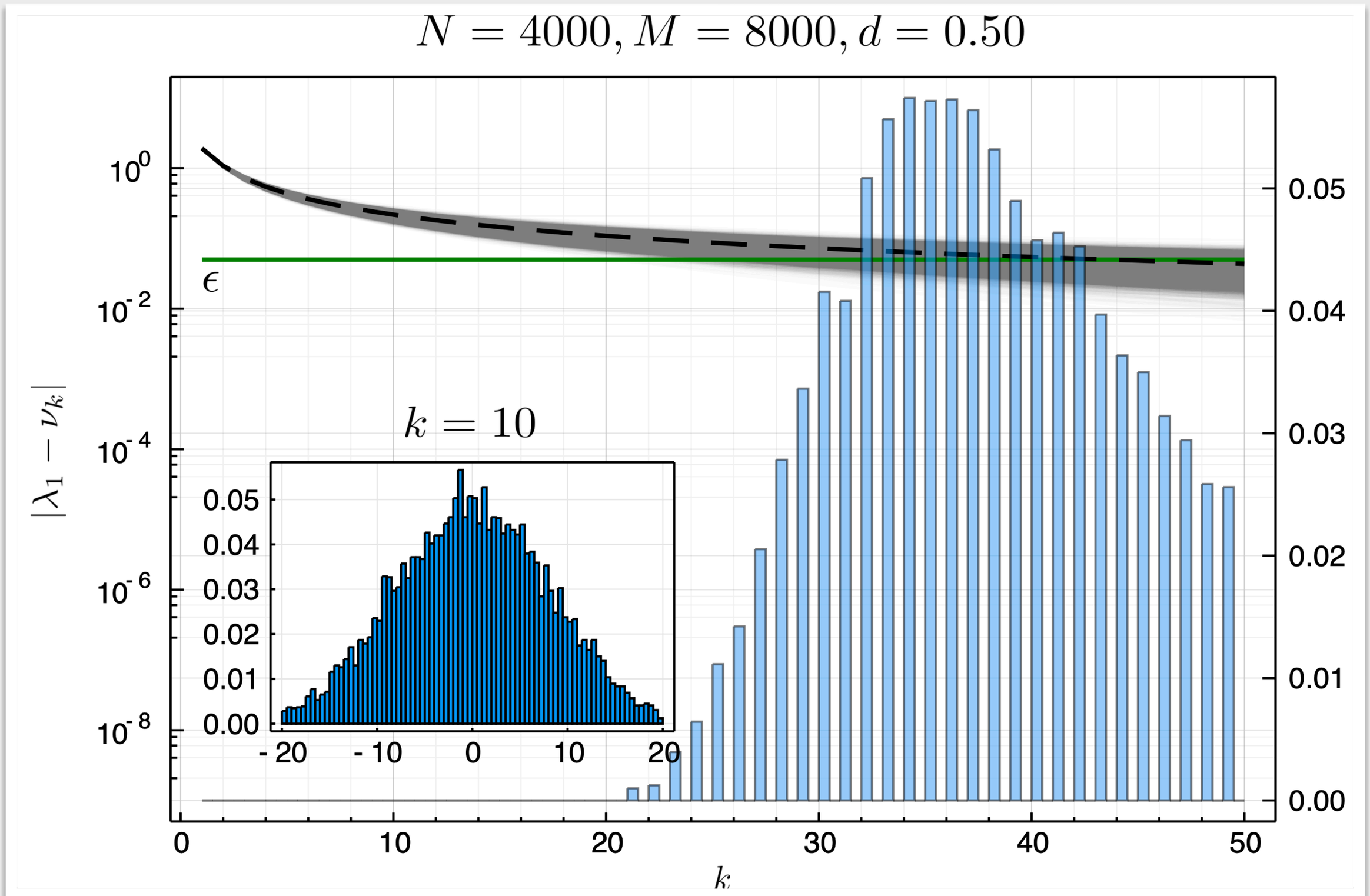
Consider the measure of error

$$E_k(W,\mathbf{b}) = |\lambda_{\max}(W) - \nu_k(W,\mathbf{b})|, \quad k = 1,2,\ldots.$$



J Kuczyński and H Woźniakowski. Estimating the Largest Eigenvalue by the Power and Lanczos Algorithms with a Random Start. SIAM Journal on Matrix Analysis and Applications, 13(4):1094–1122, oct 1992

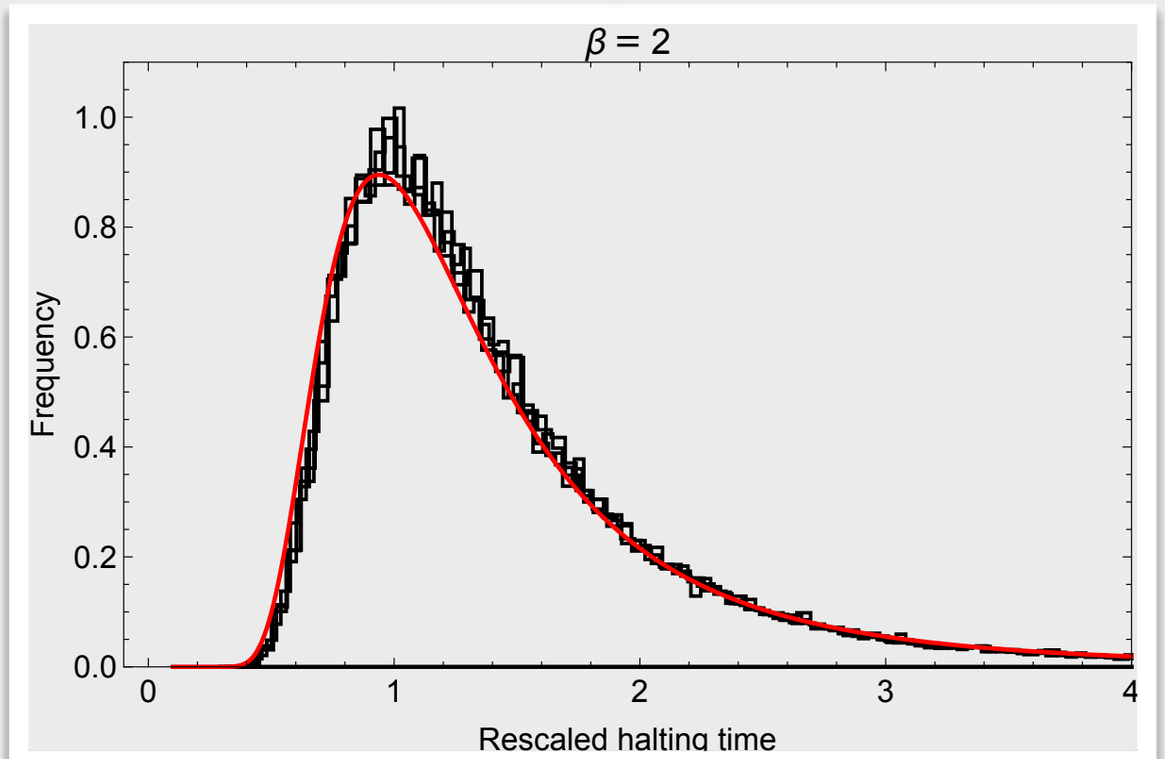

$N = 1000, M = 2000, d = 0.50$
$|\lambda_1 - \nu_k|$
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
ϵ
$k = 10$
0.05
0.04
0.03
0.02
0.01
0.00
- 20
- 10
0
10
20
0.04
0.03
0.02
0.01
0.00
0
10
20
30
40
50
k

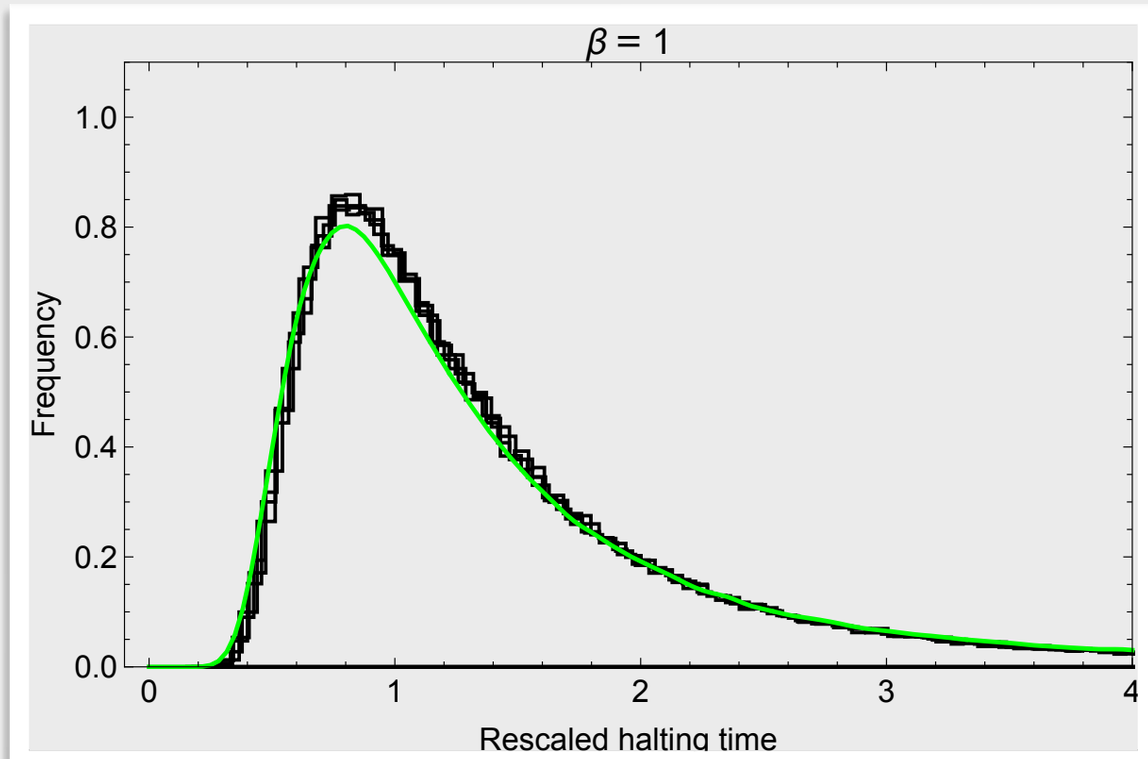


Concentration occurs if $\dfrac{\epsilon^{-1}}{\sqrt{M}} \ll 1$

A better model of the halting time for the power method?

$N = 500$





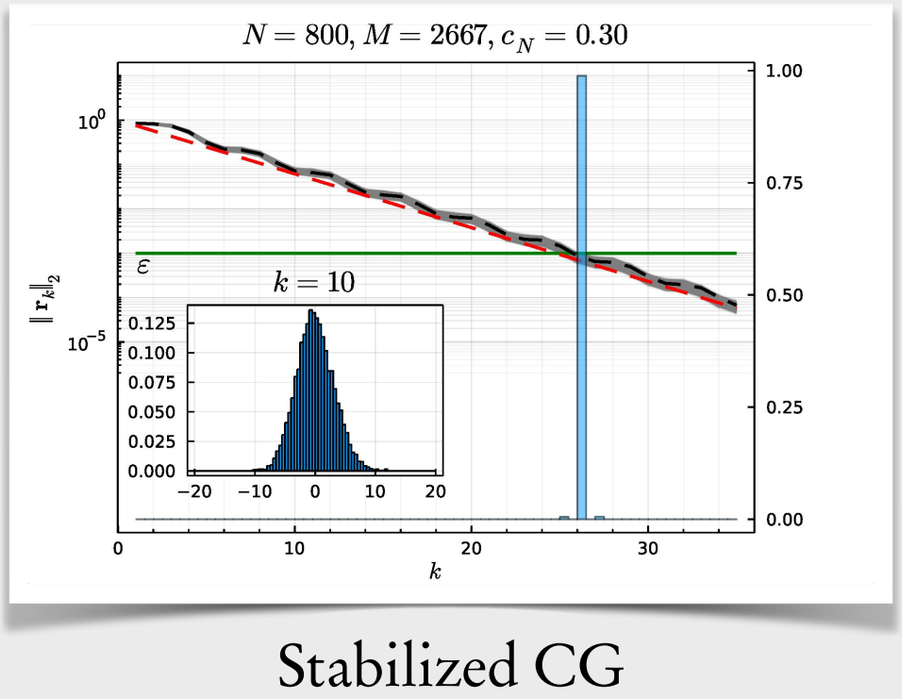
Theorem (Deift and T). Let W be a real ($\beta = 1$) or complex ($\beta = 2$) SCM and let $\mathbf{b}$ be a random unit vector independent of W. Assuming $\epsilon \leq N^{-5/3-\sigma}$, $\sigma > 0$,

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

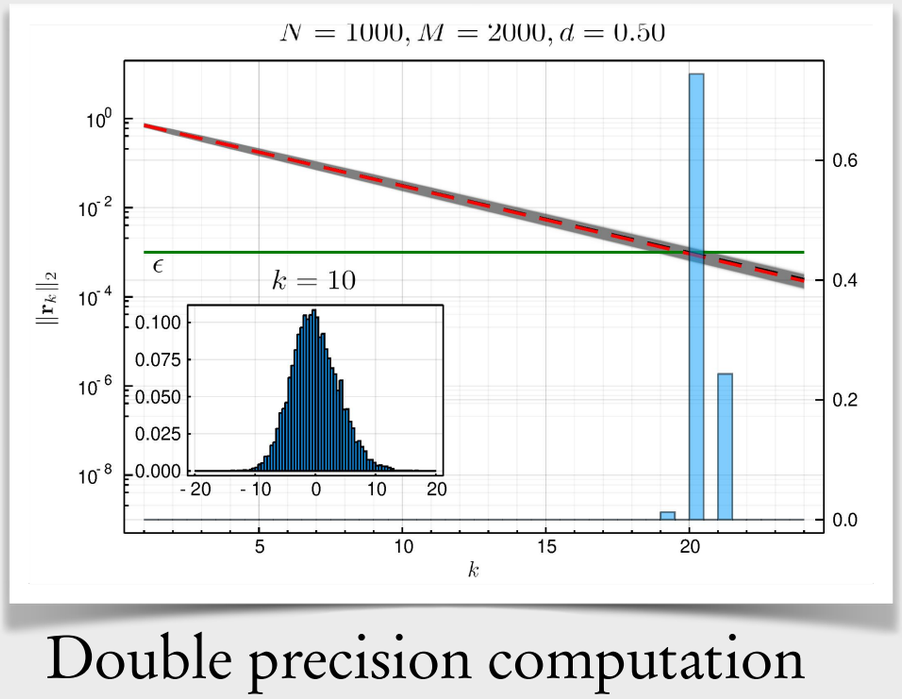
P Deift and T T. Universality for Eigenvalue Algorithms on Sample Covariance Matrices. SIAM J. Numer. Anal., 55(6):2835–2862, jan 2017

The concentration of the output of algorithms applied to random matrices varies.

Lanczos in finite precision on random matrices



Stabilized CG



Double precision computation

We all know that Lanczos is notoriously unstable.

Once a Ritz value effectively converges, Lanczos typically goes unstable.

But if the Ritz values are behaving like zeros of orthogonal polynomials on a single interval, then the fact that they interlace means they never really converge.

So maybe Lanczos is stable?

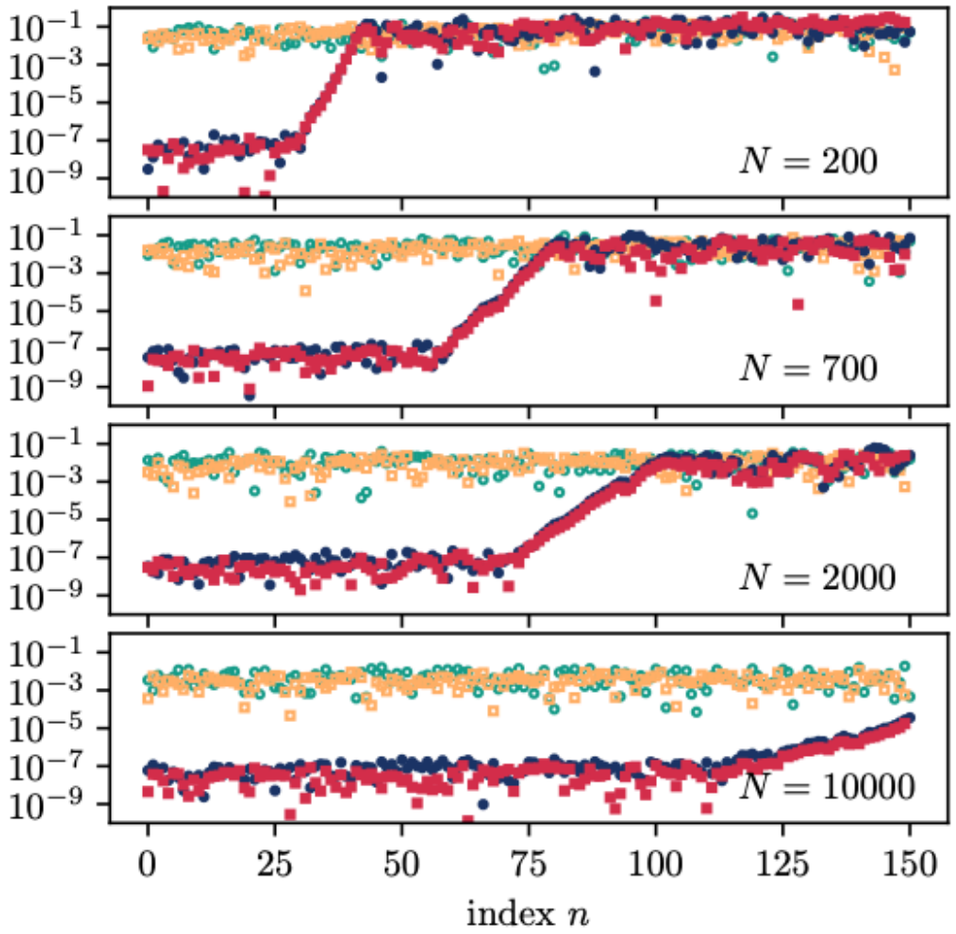
C.C. Paige. Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem. Linear Algebra and its Applications, 34:235–258, dec 1980

Lanczos is forward/backward stable on classical random matrix ensembles

If $\mu_{W,\mathbf{b}} \to \mu$ and μ is regular
in the sense that its orthogonal polynomials
are well-behaved, then Lanczos is
forward/backward stable.

This applies to the Wishart distribution
with overwhelming probability.

Lanczos on a spiked covariance model
immediately demonstrates the instability.



(b) *Forward error of recurrence coefficients* $|\alpha_n - \overline{\alpha}_n|$ (•) *and* $|\beta_n - \overline{\beta}_n|$ (▪) *and distance to limiting values* $|0 - \overline{\alpha}_n|$ (◦) *and* $|1/2 - \overline{\beta}_n|$ (▫).

T Chen and T T. Stability of the Lanczos algorithm on matrices with regular spectral distributions. arXiv preprint 2302.14842, pages 1–35, 2023

Running algorithms on random matrices may highlight important aspects but hide others!

Motivation to understand more and more random matrix models.

Stability of Lanczos

For an admissible symmetric matrix $A \in \mathbb{R}^{N\times N}$ and a unit vector $\mathbf{b} \in \mathbb{R}^N$ let $T_k(A,\mathbf{b})$ and $\overline{T}_k(A,\mathbf{b})$ be the $k \times k$ Lanczos matrices at iteration k.

There exists $\alpha > 0$ be such that if $\epsilon_{\text{mach}} = O(k^{-\alpha})$ then there exists $\gamma > \beta > 0, \delta > 0$ such that:

1. $T_k(A,\mathbf{b}_*) = \overline{T}_k(A,\mathbf{b})$ where $||\mathbf{b}_* - \mathbf{b}|| = O(k^{-\gamma})$, and
2. $||T_k(A,\mathbf{b}) - \overline{T}_k(A,\mathbf{b})|| = O(k^{-\beta} + N^{-\delta})$.

Some details

For CG, we have, approximately,

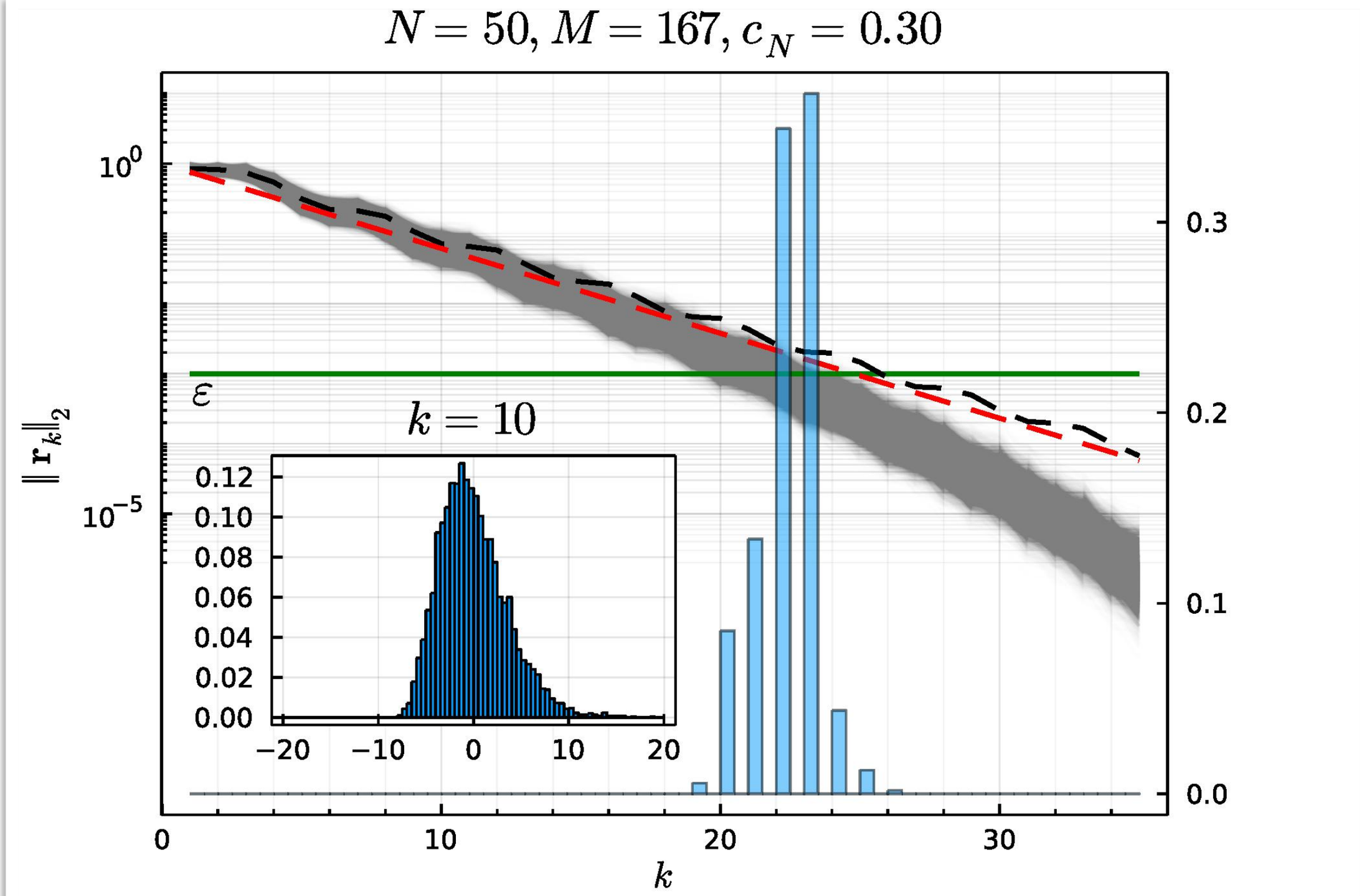
$$\|\mathbf{e}_k(W,\mathbf{b})\|_W \sim c\left|\int p_k(\lambda;\mu_{W,\mathbf{b}})\frac{\mu_{W,\mathbf{b}}(\mathrm{d}\lambda)}{\lambda}\right| = \underbrace{c\left|\int p_k(\lambda;\mu_{\mathrm{MP}})\frac{\mu_{\mathrm{MP}}(\mathrm{d}\lambda)}{\lambda}\right|}_{\frac{d^{k/2}}{\sqrt{1-d}}}\left(1+O\left(\sqrt{\frac{k}{M}}\right)\right),$$

where $p_k(x;\mu)$ is $L^2(\mu)$ normalized. It turns out that perturbations of this quantity, preserve its exponential scale because modified moments are well-behaved.

For the power method

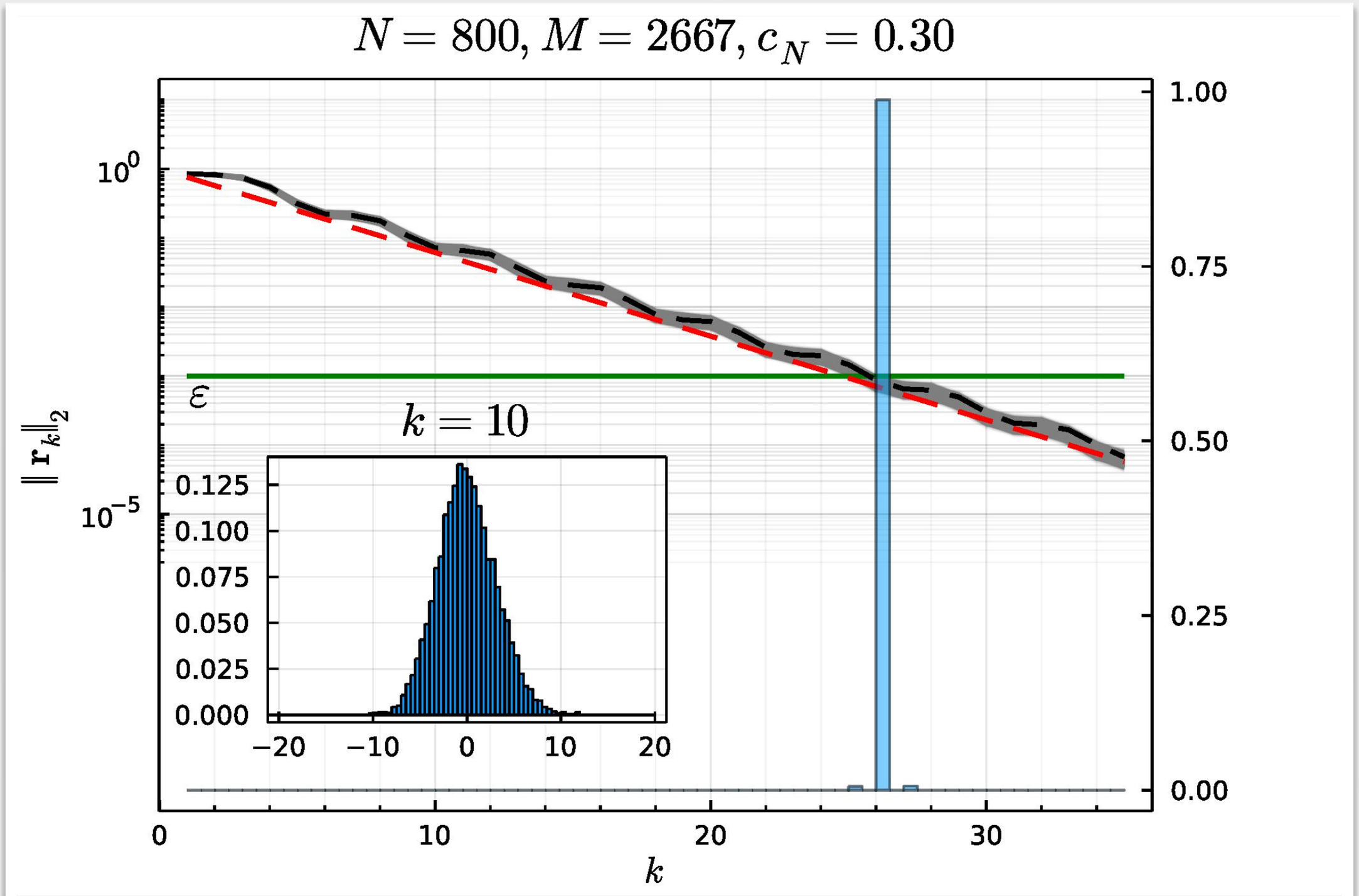
$$E_k(W,\mathbf{b}) = \left(\underbrace{(1+\sqrt{d})^2 - \frac{\int \lambda^{2k-1}\mu_{\mathrm{MP}}(\mathrm{d}\lambda)}{\int \lambda^{2k-2}\mu_{\mathrm{MP}}(\mathrm{d}\lambda)}}_{O(k^{-1})} + O\left(k^{-2/3}\sqrt{\frac{\log k}{M}}\right)\right).$$

Spiked covariance with multiple bulk components



X Ding and T T. A Riemann–Hilbert approach to the perturbation theory for orthogonal polynomials: Applications to numerical linear algebra and random matrix theory. arXiv preprint 2112.12354, pages 1–77, 2021

Spiked covariance with multiple bulk components



X Ding and T T. A Riemann–Hilbert approach to the perturbation theory for orthogonal polynomials: Applications to numerical linear algebra and random matrix theory. arXiv preprint 2112.12354, pages 1–77, 2021

Wishart matrices

Suppose $W \overset{\mathrm{L}}{=} \mathrm{Wishart}(N, M)$.

If $\mathbf{b} = \mathbf{f}_1$ (the first standard basis vector), the Lanczos is mathematically equivalent to Householder tridiagonalization. By the invariance of Wishart, we use $\mathbf{b} = \mathbf{f}_1$.

$$\mathscr{L}_k(\mu_{W,\mathbf{b}}) \overset{\mathrm{L}}{=} \frac{1}{\sqrt{\beta M}} \begin{bmatrix} \chi_{\beta M} & & & & & \\ \chi_{\beta(N-1)} & \chi_{\beta(M-1)} & & & & \\ & \chi_{\beta(N-2)} & \chi_{\beta(M-2)} & & & \\ & & \ddots & \ddots & & \\ & & & \chi_{\beta(N-k+1)} & \chi_{\beta(M-k+1)} \end{bmatrix}$$

Cholesky factorization of Lanczos matrix

J W Silverstein. The Smallest Eigenvalue of a Large Dimensional Wishart Matrix. The Annals of Probability, 13(4):1364–1368, 1985

I Dumitriu and A Edelman. Matrix models for beta ensembles. Journal of Mathematical Physics, 43(11):5830, oct 2002

$$\|\mathbf{r}_k(W, \mathbf{b})\|_2 \overset{\mathrm{L}}{=} \prod_{j=0}^{k-1} \frac{\chi_{\beta(N-j-1)}}{\chi_{\beta(M-j)}}$$

$$\|\mathbf{e}_k(W, \mathbf{b})\|_W \overset{\mathrm{L}}{=} \frac{\sqrt{\beta M}}{\chi_{M-N+1}} \prod_{j=0}^{k-1} \frac{\chi_{\beta(N-j-1)}}{\chi_{\beta(M-j)}}$$

The CLT

For general sample covariance matrices, we identify Gaussian processes $\mathscr{G}^{(1)} = (\mathscr{G}_j^{(1)})_{j\geq 1}$, $\mathscr{G}^{(2)} = (\mathscr{G}_j^{(2)})_{j\geq 1}$ so that for $\mathfrak{d} = N/M \to d \in (0,1]$

$$\sqrt{\frac{\beta M}{2}}\left(\|\mathbf{e}_k(W,\mathbf{b})\|_W^2 - \frac{\mathfrak{d}^k}{1-\mathfrak{d}}\right)_{k\geq 1} \xrightarrow{\text{dist.}} \mathscr{G}^{(1)}, \quad d \neq 1,$$

$$\sqrt{\frac{\beta M}{2}}\left(\|\mathbf{r}_k(W,\mathbf{b})\|_2^2 - \mathfrak{d}^k\right)_{k\geq 1} \xrightarrow{\text{dist.}} \mathscr{G}^{(2)},$$

in the sense of convergence of finite-dimensional marginals.

In particular, this implies

$$\sqrt{\frac{\beta M}{2}}\left(\|\mathbf{r}_k(W,\mathbf{b})\|_2^2 - \mathfrak{d}^k\right) \xrightarrow{\text{dist.}} \mathcal{N}(0,\sigma_{k,\mathbf{r}}^2), \quad \sigma_{k,\mathbf{r}}^2 = k d^{2k}\left(1+\frac{1}{d}\right).$$

Importantly, we have universality in the sense that this holds for a wide class of sample covariance matrices W (with trivial covariance) that goes far beyond the Gaussian case of Wishart (LUE & LOE).