Accurate Eigenvalues of Random Matrices

Plamen Koev

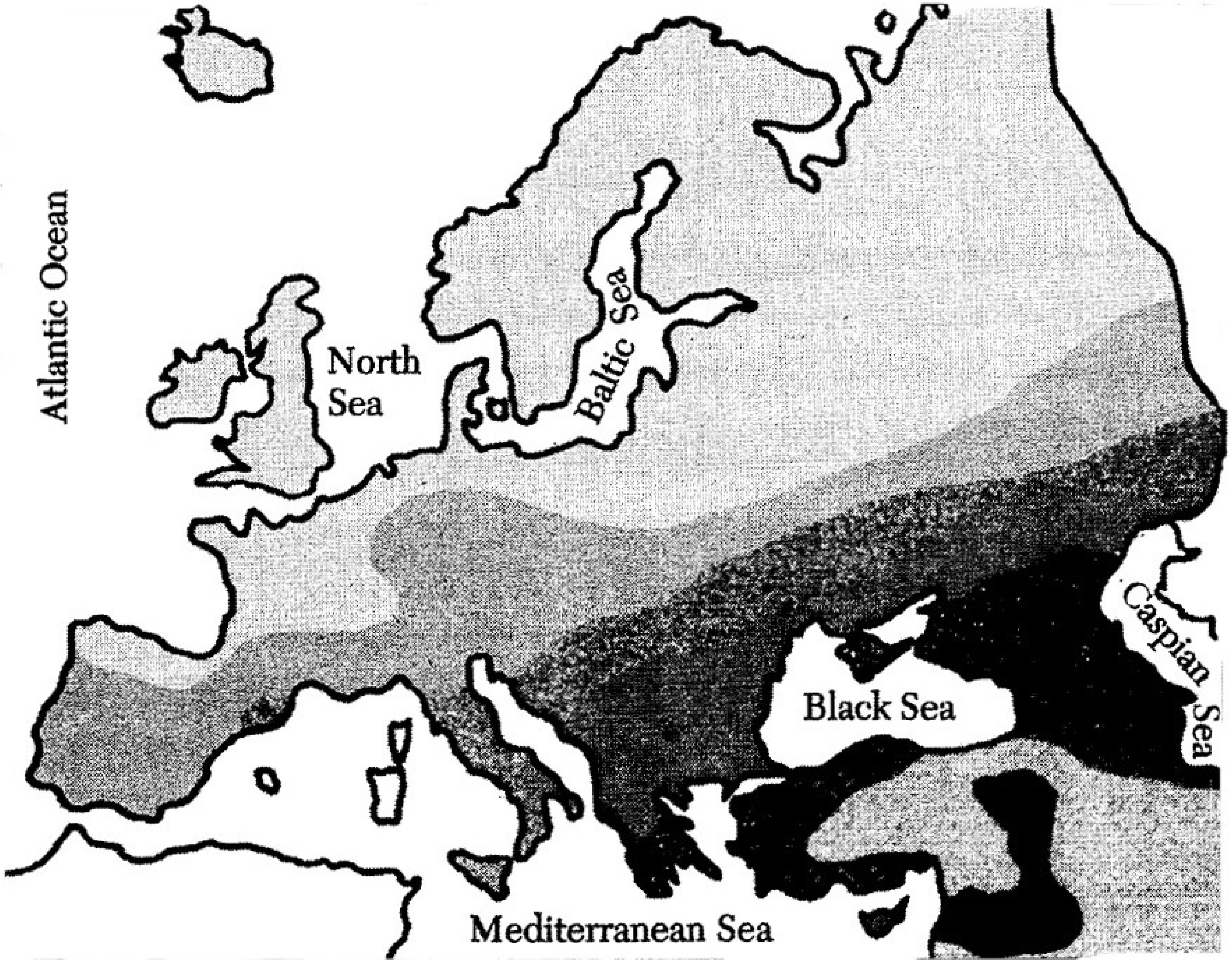
Department of Mathematics, North Carolina State University

Joint with: Cy Chan, Jim Demmel, Vesselin Drensky, Alan Edelman, Iain Johnstone, Raymond Kan

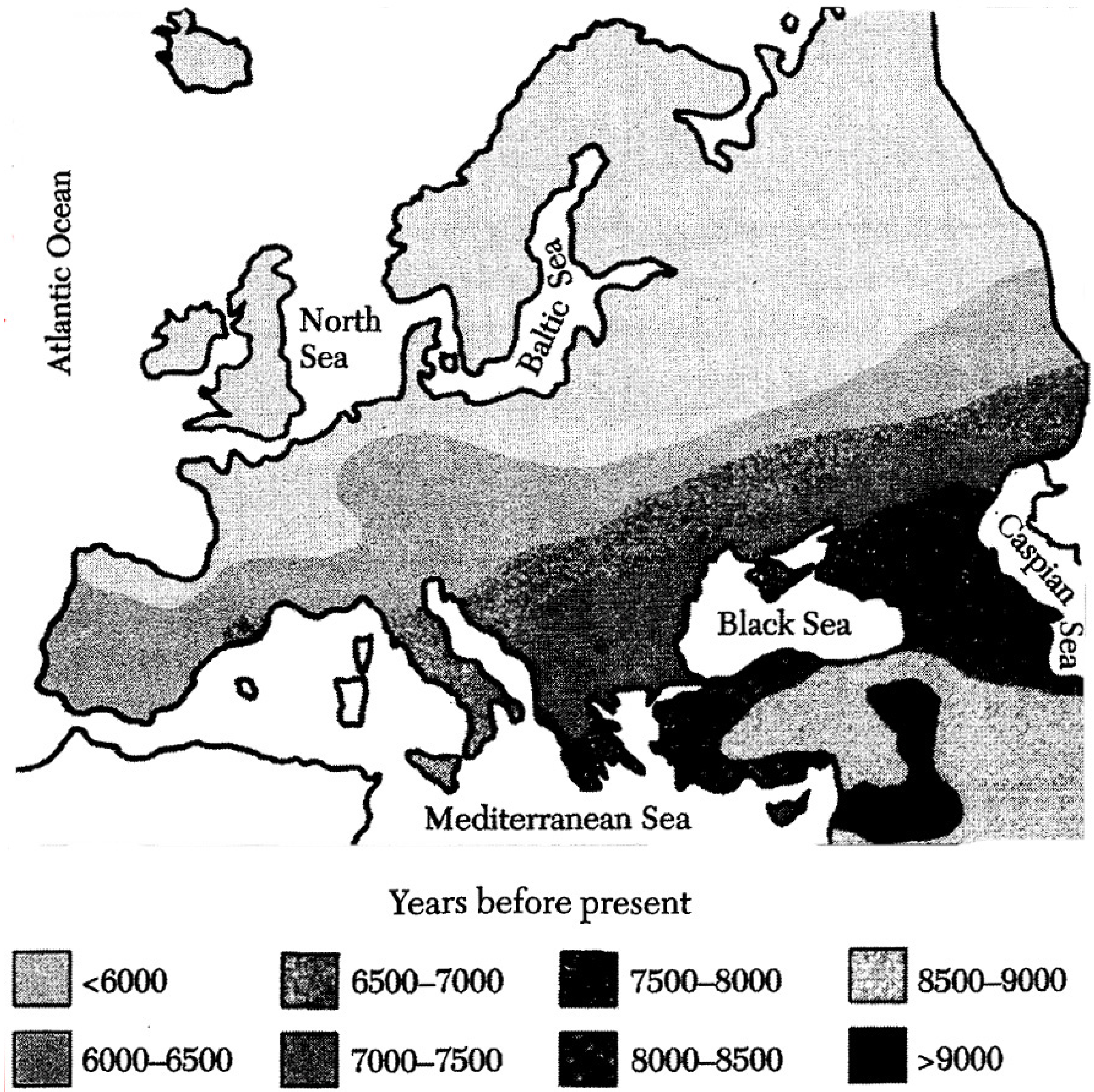
Supported by NSF

International Workshop on Accurate Solutions of Eigenvalue Problems, Dubrovnik, Croatia, June 8–12, 2008

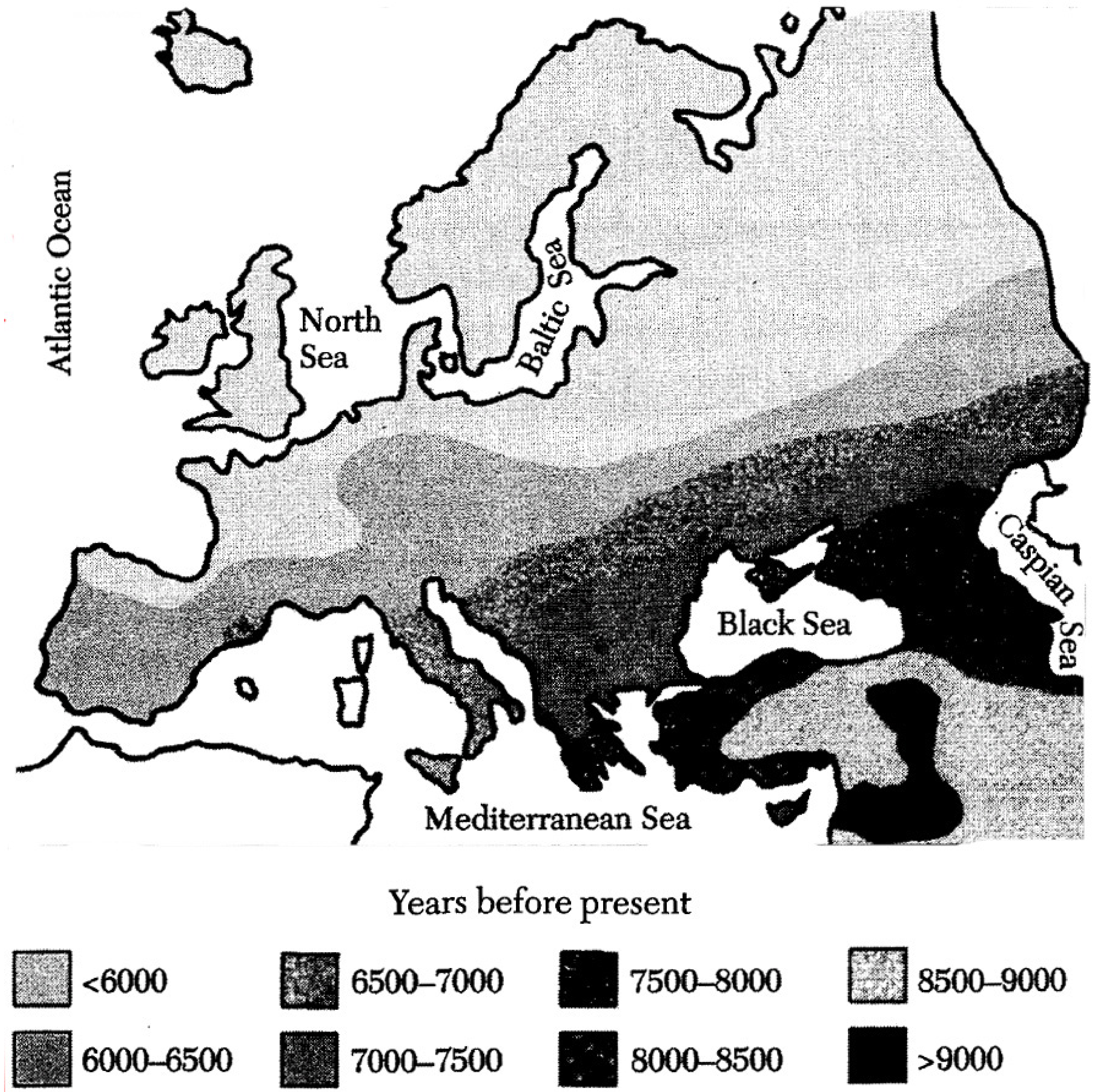
Motivating Example (Cavalli–Sforza '00)



Spread of Agriculture in Europe



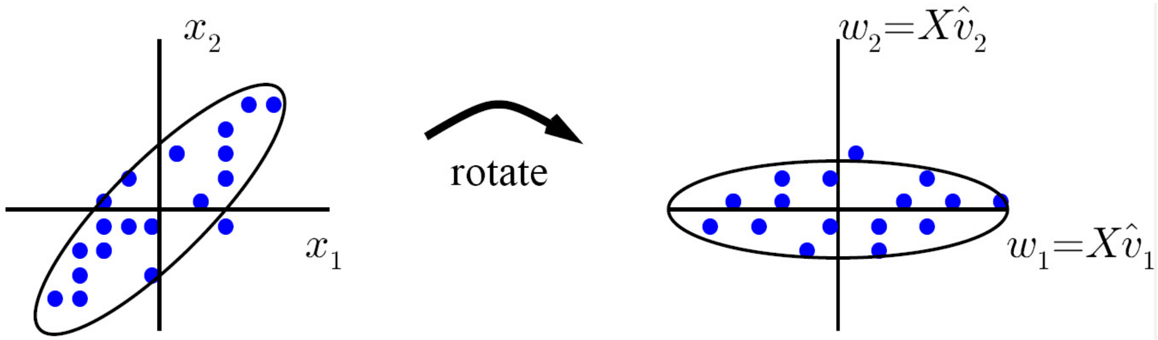
Transfer of Technology or Spread of Farmers?



Methodology = PCA (read: SVD)

- **Data: DNA sequences: 95 genes at 400 locations in Europe:**

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,400} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,400} \\ & \vdots & & \\ x_{95,1} & x_{95,2} & \cdots & x_{95,400} \end{bmatrix}$$

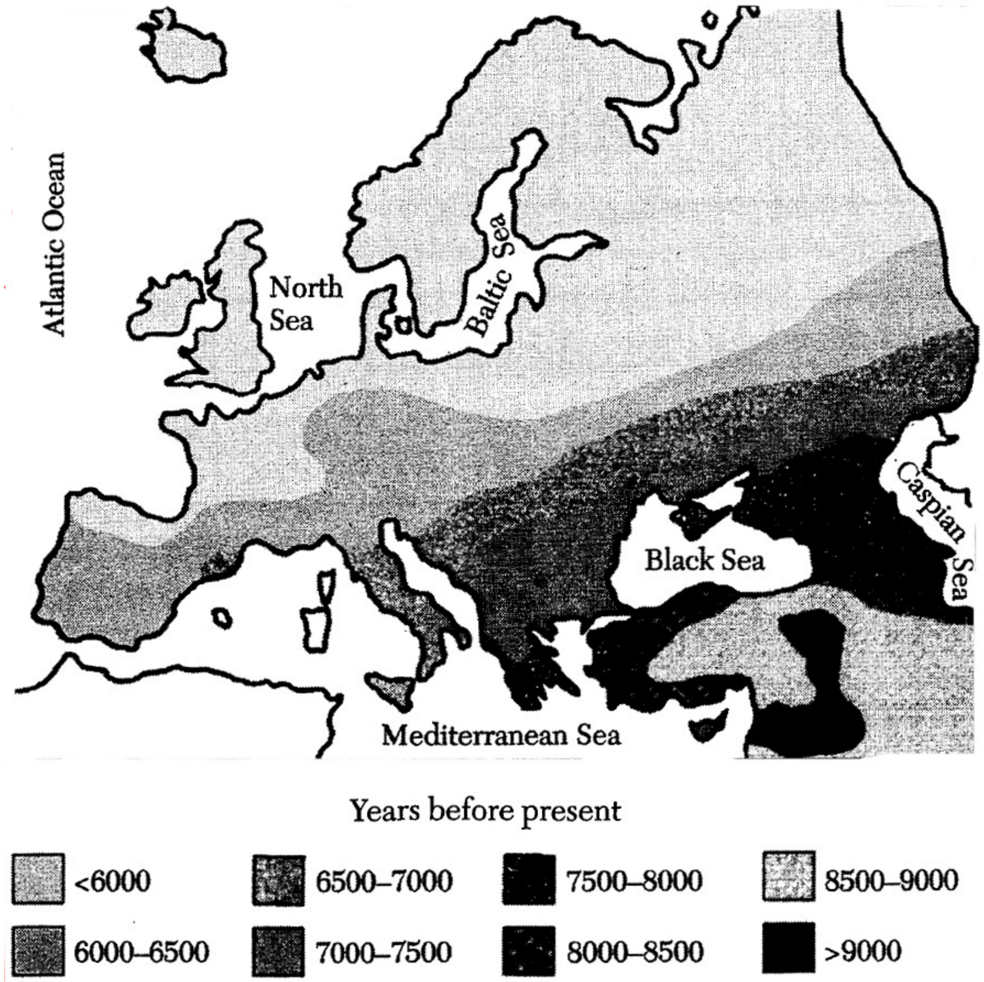


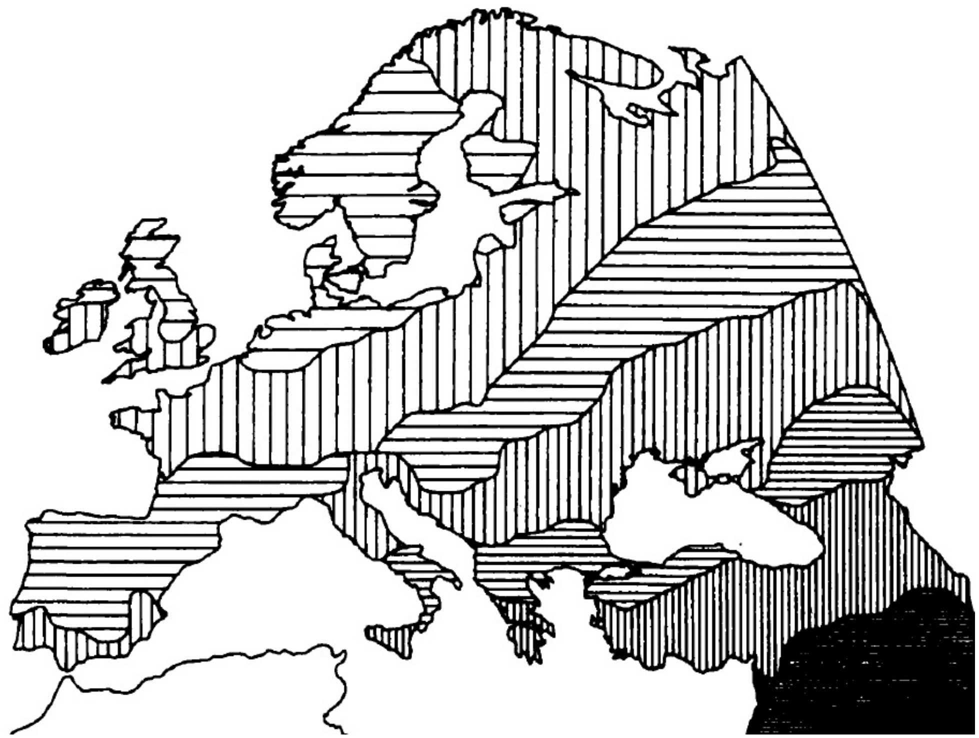
- **Principal Components = singular vectors = directions of maximal variance (usually first 3 PC explain $> 50\%$ of total variance)**

- **Reveals structure, if it exists, but does not detect it or explain what it is**

- **Next: Contour plot of first principal component (singular vector)**

Data Supports Thesis that Farmers Spread





Why Eigenvalues of Random Matrices

- **Conventional model:**

$$X \in \mathbb{C}^{m\times n}, \quad X \sim \mathcal{N}_m(0,\Sigma), \quad n\text{-variate Gaussian},$$

  i.e., x_{ij}—normal random variables and $E(X^*X) = \Sigma$.

- **Key question:** existence and nature of interdependence between $X(:,i)$'s
  i.e., $\Sigma =?\ \Sigma = I?\ \Sigma = \Sigma_0?$ ...

- $A \equiv X^*X$ is called $n \times n$ Wishart with m DOF and covariance Σ

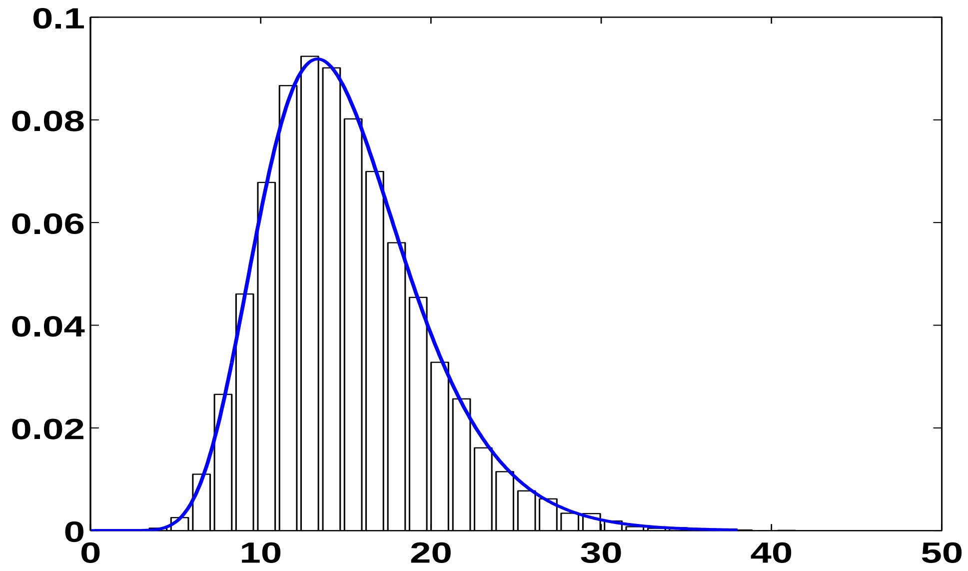
- Cast as tests on $\lambda_{\max}(A)$, a "test statistic"

- 5%, 1% benchmarks

- Thus (distributions of) eigenvalues of Wishart critical

Computational Aspects of Eigenvalues of Random Matrices

- **Theory: 1960s: easy**
- **Surprise: Explicit formulas, but ...**
- **Algorithms: hard; only very recent**
- **Matrix size relatively small $=$ population size**
- **Example:** $\lambda_{\max}$ **of** 4×4 **Wishart with 7 DOF,** $\Sigma = I$



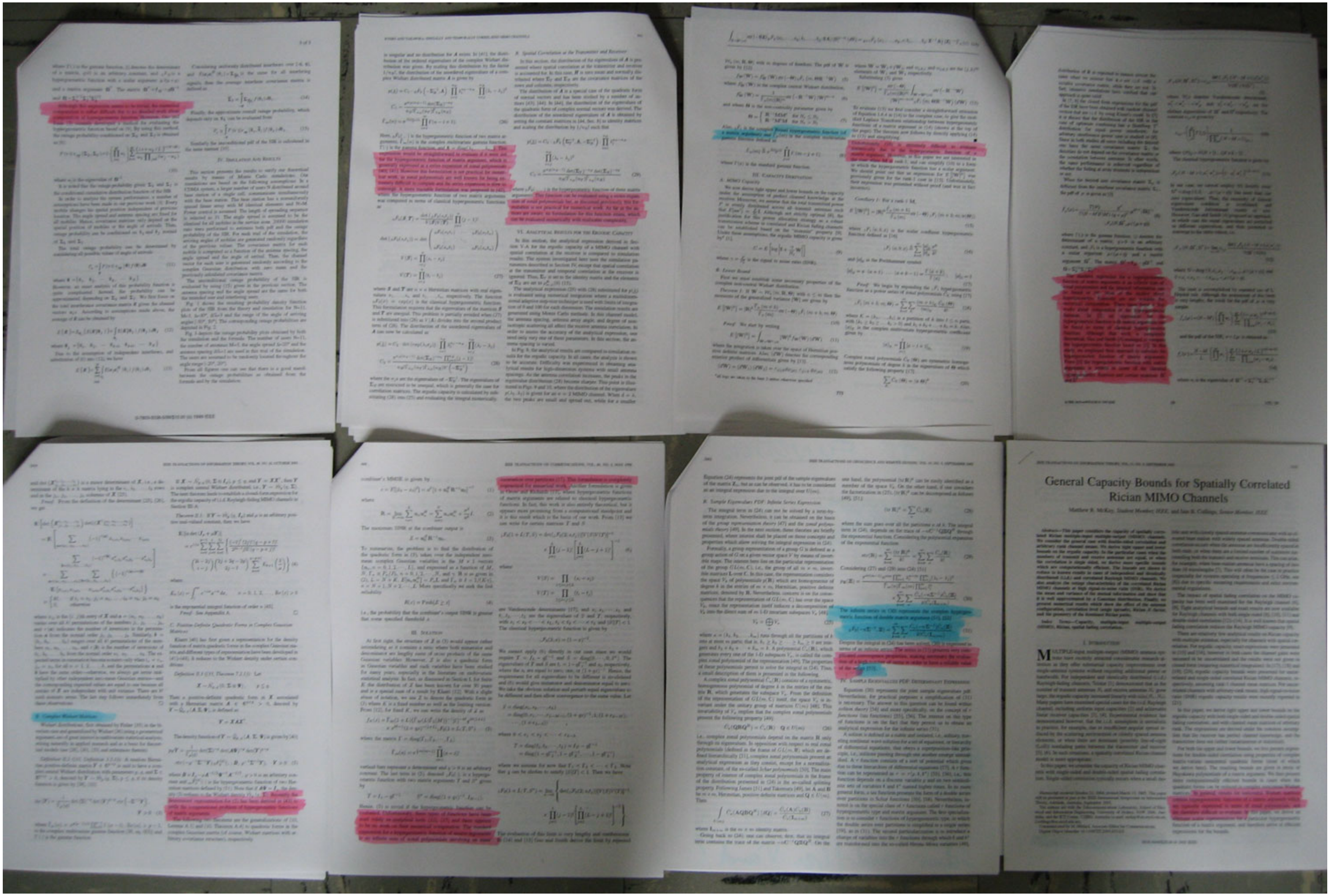
Exact vs Empirical with 20,000 replications

- **Formulas intrinsically accurate (sums of positives)**

Computing Eigenvalues of Wisharts is Really Hard!

- **Sample complaints:**
  - G.J. Byers and F. Takawira. Spatially and temporally correlated MIMO channels: modeling and capacity analysis. IEEE Transactions on Vehicular Technology, 53:634–643, May 2004.
  - Pinyuen Chen, G.J. Genello, and M.C. Wicks. Estimating the number of signals in presence of colored noise. In Radar Conference 2004. Proceedings of the IEEE, pages 432–437, 26–29 April 2004.
  - H. Gao, P.J. Smith, and M.V. Clark. Theoretical reliability of MMSE linear diversity combining in Rayleigh-fading additive interference channels. IEEE Transactions on Communications, 46:666–672, May 1998.
  - A.J. Grant. Performance analysis of transmit beamforming. IEEE Transactions on Communications, 53:738–744, April 2005.
  - M. Kang and M.-S. Alouini. Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems. IEEE Journal on Selected Areas in Communications, 21(3):418–431, 4 2003.
  - C. Lopez-Martinez, E. Pottier, and S.R. Cloude. Statistical assessment of eigenvector-based target decomposition theorems in radar polarimetry. IEEE Transactions on Geoscience and Remote Sensing, 43:2058–2074, 2005.
  - M.R. McKay and I.B. Collings. Capacity bounds for correlated rician MIMO channels. In 2005 IEEE International Conference on Communications. ICC 2005., volume 2, pages 772–776, 16-20 May 2005.
  - M.R. McKay and I.B. Collings. General capacity bounds for spatially correlated Rician MIMO channels. IEEE Transactions on Information Theory, 51:3121–3145, September 2005.
  - A. Ozyildirim and Y. Tanik. Outage probability analysis of a CDMA system with antenna arrays in a correlated Rayleigh environment. In IEEE Military Communications Conference Proceedings, 1999. MILCOM 1999, volume 2, pages 939–943, 31 Oct.–3 Nov. 1999.
  - A. Ozyildirim and Y. Tanik. SIR statistics in antenna arrays in the presence of correlated Rayleigh fading. In IEEE VTS 50th Vehicular Technology Conference, 1999. VTC 1999 - Fall, volume 1, pages 67–71, 19-22 September 1999.
  - Hyundong Shin and Jae Hong Lee. Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole. IEEE Transactions on Information Theory, 49:2636–2647, October 2003.
  - V. Smidl and A. Quinn. Fast variational PCA for functional analysis of dynamic image sequences. In Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003., volume 1, pages 555–560, 18-20 September 2003.

Computing Eigenvalues of Wisharts is Really Hard!



Computing Eigenvalues of Wisharts is Really Hard!

quadratic forms can be reduced to complex noncentral Wishart matrices. In general, results for noncentral Wishart matrices contain hypergeometric functions of a matrix argument which are typically expressed in terms of zonal polynomials, and are therefore difficult to evaluate. In this paper, we derive an alternate scalar representation for a particular hypergeometric

Unfortunately (10) is extremely difficult to evaluate numerically due to the hypergeometric function of a matrix argument. However, in this paper we are interested in the case when M is

Hypergeometric Fn of Matrix Argument, Previous Best Algorithm

Electronic Transactions on Numerical Analysis.
Volume 11, pp. 121-130, 2000.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

APPROXIMATION OF HYPERGEOMETRIC FUNCTIONS WITH MATRICIAL ARGUMENT THROUGH THEIR DEVELOPMENT IN SERIES OF ZONAL POLYNOMIALS*

R. GUTIÉRREZ†, J. RODRIGUEZ‡, AND A. J. SÁEZ§

Hypergeometric Fn of Matrix Argument, Previous Best Algorithm

Electronic Transactions on Numerical Analysis.
Volume 11, pp. 121-130, 2000.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

APPROXIMATION OF HYPERGEOMETRIC FUNCTIONS WITH MATRICIAL ARGUMENT THROUGH THEIR DEVELOPMENT IN SERIES OF ZONAL POLYNOMIALS*

R. GUTIÉRREZ†, J. RODRIGUEZ‡, AND A. J. SÁEZ§

more time would be needed. (We spent about 8 days to obtain the 627 zonal polynomials of degree 20 with a 350 MHz Pentium II processor.)

Hypergeometric Fn of Matrix Argument, Previous Best Algorithm

Electronic Transactions on Numerical Analysis.
Volume 11, pp. 121-130, 2000.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

APPROXIMATION OF HYPERGEOMETRIC FUNCTIONS WITH MATRICIAL ARGUMENT THROUGH THEIR DEVELOPMENT IN SERIES OF ZONAL POLYNOMIALS*

R. GUTIÉRREZ†, J. RODRIGUEZ‡, AND A. J. SÁEZ§

more time would be needed. (We spent about 8 days to obtain the 627 zonal polynomials of degree 20 with a 350 MHz Pentium II processor.)

- **New result (2.33 GHz Pentium):**

```
>> tic; mhg(20,2,[],[],[1:20]/210), toc
ans =
   2.71828182845905
elapsed_time =
   0.03100000000000
```

Deriving the Distribution of $\lambda_{\max}(X^T X)$

- **Distribution of $\lambda_{\max}$ (complex case):**

Variable	Density
$x \sim \mathcal{N}(0, \sigma)$	$f(x) = Ce^{-\frac{\lvert x\rvert^2}{\sigma}}$
$X = (X_1, \ldots, X_n) \sim \mathcal{N}_m(0, \Sigma)$	$f(X) = C\lvert\Sigma\rvert^{-1} e^{-X^*\Sigma^{-1}X}$
$A = X^*X \sim \mathcal{W}_m(n, \Sigma)$	$f(A) = C\lvert\Sigma\rvert^{-m}\lvert A\rvert^{m-n} e^{-\mathrm{tr}(\Sigma^{-1}A)}$

$$P(\lambda_{\max} < x) = P(A < xI) = \int_{0<A<xI} f(A)\,dA$$

Distribution of $\lambda_{\max}$ of Wishart

$$
\begin{aligned}
P(\lambda_{\max} < x) &= P(A < xI) \\
&= \int_{0<A<xI} C|\Sigma|^{-m}|A|^{m-n}e^{-\mathrm{tr}(\Sigma^{-1}A)}dA \\
(A \to Q^*\Lambda Q) &= \int_{\Lambda\in[0,x]^n} C\prod_{i<j}|\lambda_i - \lambda_j|^2 \prod_{i=1}^{n} \lambda_i^{m-n}e^{-\lambda_i}d\Lambda
\end{aligned}
$$

Selberg Integral, 1940s

$$
= \sum_{k=0}^{\infty}\sum_{\kappa\vdash k} c_\kappa \cdot s_\kappa(x_1, \ldots, x_n)
$$

- s_κ = Schur functions, orthogonal basis of $\Pi(x_1, \ldots, x_n)$
- Indexed by partitions κ:

  Leading terms: x_1, x_1^2, x_1x_2, x_1^3, $x_1^2x_2$, $x_1x_2x_3$, ...

Distribution of $\lambda_{\max}$ of Wishart

$$
\begin{aligned}
P(\lambda_{\max} < x) &= P(A < xI) \\
&= \int_{0<A<xI} C |\Sigma|^{-m} |A|^{m-n} e^{-\mathrm{tr}(\Sigma^{-1}A)} dA \\
(A \to Q^* \Lambda Q) &= \int_{\Lambda \in [0,x]^n} C \prod_{i<j} |\lambda_i - \lambda_j|^2 \prod_{i=1}^{n} \lambda_i^{m-n} e^{-\lambda_i} d\Lambda
\end{aligned}
$$

Selberg Integral, 1940s

$$
= \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} c_\kappa \cdot s_\kappa(x_1, \ldots, x_n)
$$

$$
\sim {}_1F_1(m; n+m; x_1, \ldots, x_n)
$$

Hypergeometric Function of Matrix Argument

- **Need to truncate for $k \leq M$ for some M**
- **Even a single s_κ is exponential if computed naively**

Cost of Computing Schur Polynomials Naively

Degree	Partition κ	s_κ	Number of terms
1	(1)	$x_1 + \cdots + x_n$	$\mathcal{O}(n)$
2	(2)	$\sum_{i \le j} x_i x_j$	$\mathcal{O}(n^2)$
2	$(1,1)$	$\sum_{i < j} x_i x_j$	$\mathcal{O}(n^2)$
3	$(1,1,1)$	$\sum_{i<j<k} x_i x_j x_k$	$\mathcal{O}(n^3)$
$\lvert\kappa\rvert$	κ	$\sum x_1^{\kappa_1} \cdots x_n^{\kappa_n}$	$\mathcal{O}\left(n^{\lvert\kappa\rvert}\right)$

- **New result: $\mathcal{O}(n)$ each**, as long as one wants them all (and we do!)

Computing the Schur Polynomial "Cleverly"

- s_κ of higher degree "contain" s_λ of lower degree. Redundancy.
- E.g.:

$$\begin{aligned} s_{(1,1)}(x_1, \ldots, x_n) &= \sum_{i<j} x_i x_j \\ &= x_1x_2 + (x_1 + x_2)x_3 + \cdots + (x_1 + \cdots + x_{n-1})x_n \end{aligned}$$

- In general:

$$s_\kappa(x_1, x_2, \ldots, x_n) = \sum_{\lambda<\kappa} s_\lambda(x_1, x_2, \ldots, x_{n-1}) \cdot x_n^{|\kappa|-|\lambda|}$$

- Connection with representation theory:
  - s_κ are the irreducible characters of $\mathbf{GL}_n(\mathbb{C})$
  - the characters of $\mathbf{GL}_{n-1}(\mathbb{C})$ induce those of $\mathbf{GL}_n(\mathbb{C})$
- Result long known (Macdonald), but missed for 40 years

Computing the Schur Polynomial "Cleverly"

- **Example:** $s_{(1,1)}(x_1, \ldots, x_n)$

$$= \sum_{i<j} x_i x_j \qquad (\ \sim n^2 \ \text{ operations })$$

$$= \underbrace{x_1}_{s_1} x_2 + (\underbrace{x_1 + x_2}_{s_2}) x_3 + (\underbrace{x_1 + x_2 + x_3}_{s_3}) x_4 + \cdots + (\underbrace{x_1 + \cdots + x_{n-1}}_{s_{n-1}}) x_n$$

- **New cost:** $3n - 2$ **instead of** n^2
- **Generalizes to all** κ**:**
  Cost of $s_\kappa(x_1, \ldots, x_n)$ **goes down from** $\mathcal{O}(n^{|\kappa|})$ **to** $\mathcal{O}(N_\kappa n)$
- **It gets better: We can get rid of** $N_\kappa \equiv \{\#\lambda | \lambda < \kappa\}$

Analogy with the FFT

- **Idea: (DFT)$_{ij}$—characters of $\mathbb{Z}/n\mathbb{Z}$ $\longleftrightarrow$ s_λ—characters of GL$_n(\mathbb{C})$**
- **Write our main identity**

$$s_\kappa(x_1, x_2, \ldots, x_n) = \sum_{\lambda < \kappa} s_\lambda(x_1, x_2, \ldots, x_{n-1}) \cdot x_n^{|\kappa|-|\lambda|}$$

in matrix form: $\mathcal{C}_n = \mathcal{C}_{n-1} \cdot Y_n(x_n)$, where

$$Y_2(x) = \left[\begin{array}{cccc|cccc|cccc|cccc} 1 & x & x^2 & x^3 & & & & & & & & & & & & \\ & 1 & x & x^2 & x & x^2 & x^3 & x^4 & & & & & & & & \\ & & 1 & x & & x & x^2 & x^3 & x^2 & x^3 & x^4 & x^5 & & & & \\ & & & 1 & & & x & x^2 & & x & x^2 & x^3 & x^3 & x^4 & x^5 & x^6 \end{array}\right]$$

$$\begin{bmatrix} 1 & x & x^2 & x^3 \\ & 1 & x & x^2 \\ & & 1 & x \\ & & & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -x & & \\ & 1 & -x & \\ & & 1 & -x \\ & & & 1 \end{bmatrix}$$

- **Matrix-vector multiplication by Y_n costs $\mathcal{O}(n)$ per s_λ instead of $\mathcal{O}(nN_\kappa)$**

Our New Fast Algorithm

- **Let A be the lower shift matrix $a_{i+1,i} = 1$; $B = A^T$**
- **Structure of Y_n :**

$$\begin{aligned}
U_n(x_n) &= I_{(N+1)^{n-1}} + x_n(A \otimes B_{n-1}) + \cdots + x_n^N(A^N \otimes B_{n-1}^N) \\
&= \left(I_{(N+1)^{n-1}} - x_n(A \otimes B_{n-1})\right)^{-1}, \\
C_n(x_n) &= U_n(x_n)K_{n-1}(x_n), \\
K_n(x_n) &= I_{N+1} \otimes C_n(x_n), \\
B_n &= B_{n-1} \otimes I_{N+1} = B \otimes I_{(N+1)^{n-2}}, \\
Q_n(x_n) &= \left(I_{(N+1)^{n-1}} \,|\, x_n B_n \,|\, \ldots \,|\, x_n^N B_n^N\right) \\
Y_n &= Q_n(x_n)K_n(x_n)
\end{aligned}$$

- **New algorithm:**

for i=n:-1:1

  for all λ such that $|\lambda| \leq M$ in reverse lexicographic order

    $s_\lambda = s_\lambda + s_{\lambda^{(i)}} x_n$

(where $\lambda^{(i)} \equiv (\lambda_1, \ldots, \lambda_i - 1, \ldots, \lambda_n)$)

- **Final cost: $\mathcal{O}(n)$ per each s_λ, optimal**

Open Problems: Many; Here are Two

- Generalize the FFT idea to **real, $\alpha = 2$** case
- How does one multiply quickly by the matrix:

$$A = \begin{bmatrix} 1 & 1 & \frac{\alpha+1}{2} & \frac{(\alpha+1)(2\alpha+1)}{6} \\ & 1 & 1 & \frac{\alpha+1}{2} \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -2 & & \\ & 1 & -2 & \\ & & 1 & -2 \\ & & & 1 \end{bmatrix}^{-1/2}$$

Open problem 2: Are the decimal digits of π random?

- **Test 1: Histogram: Passed**
- **Test 2: Longest increasing subsequence $\leq s$**

$$\int_L e^{\sum_{i=1}^n x_i^2} \prod_{i<j} (x_i - x_j)^2 dx_1 \cdots dx_{n-1}$$

  - $L = \{\max_{1 \leq i \leq n} x_i \leq s, x_1 + \cdots + x_n = 0\}$
  - **same as $\lambda_{\max}$ of Gaussian matrix with trace 0**

Conclusions

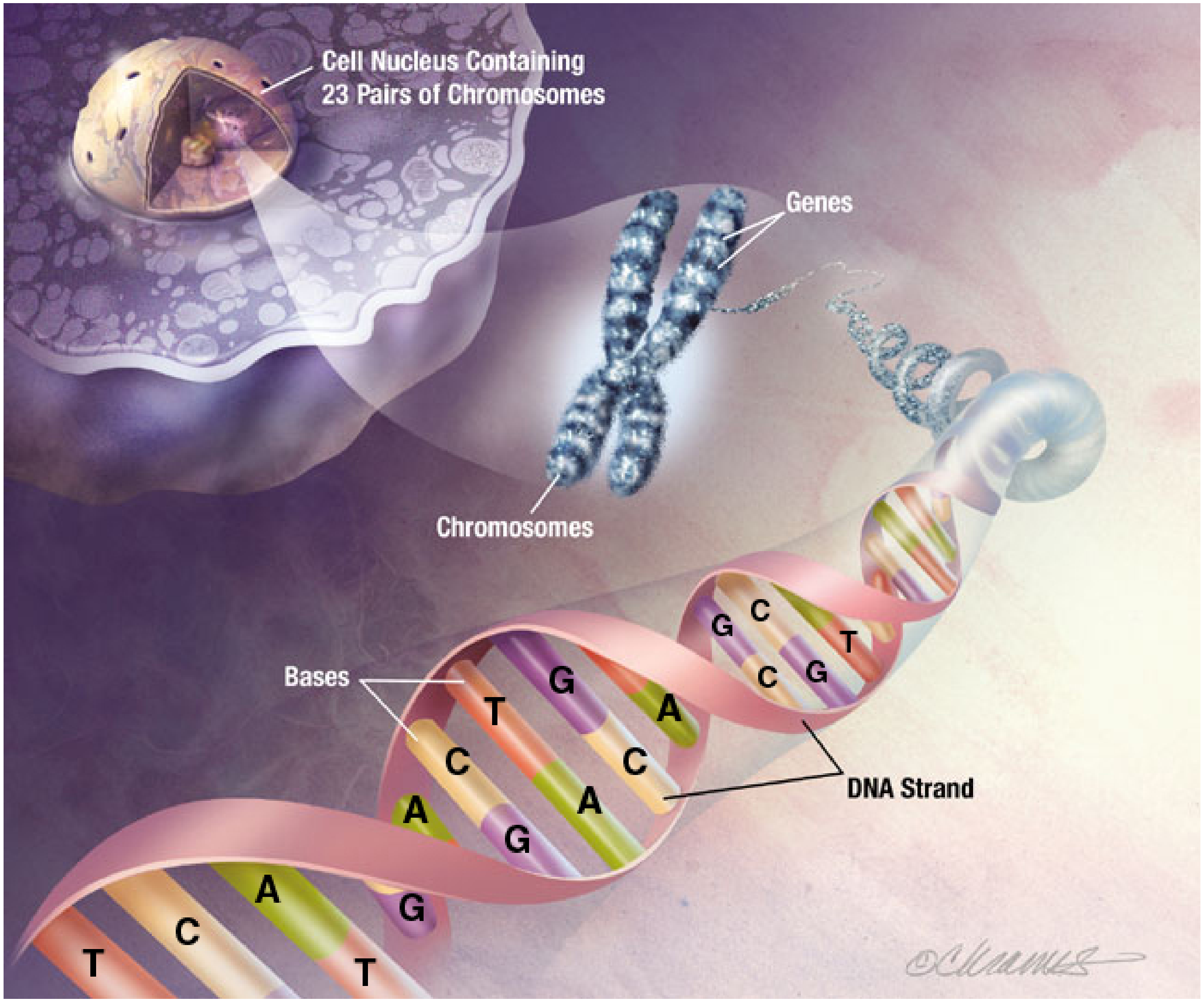
- **Eigenvalues of random matrices matter, even for small matrices**
- **Papers and software at:** `http://www4.ncsu.edu/~pskoev`
- **Impact on important applications:**
  - **3D target classification**
  - **Genomics**
  - **Wireless communications**

Future work

- **New algorithms based on saddle point approximations**
- **Automatic convergence detection**
- **FFT generalization to zonal polynomials**
- **Tracy–Widom finite inference**

Example: Genomics

- **Population Classification with Nick Patterson, Broad Institute**



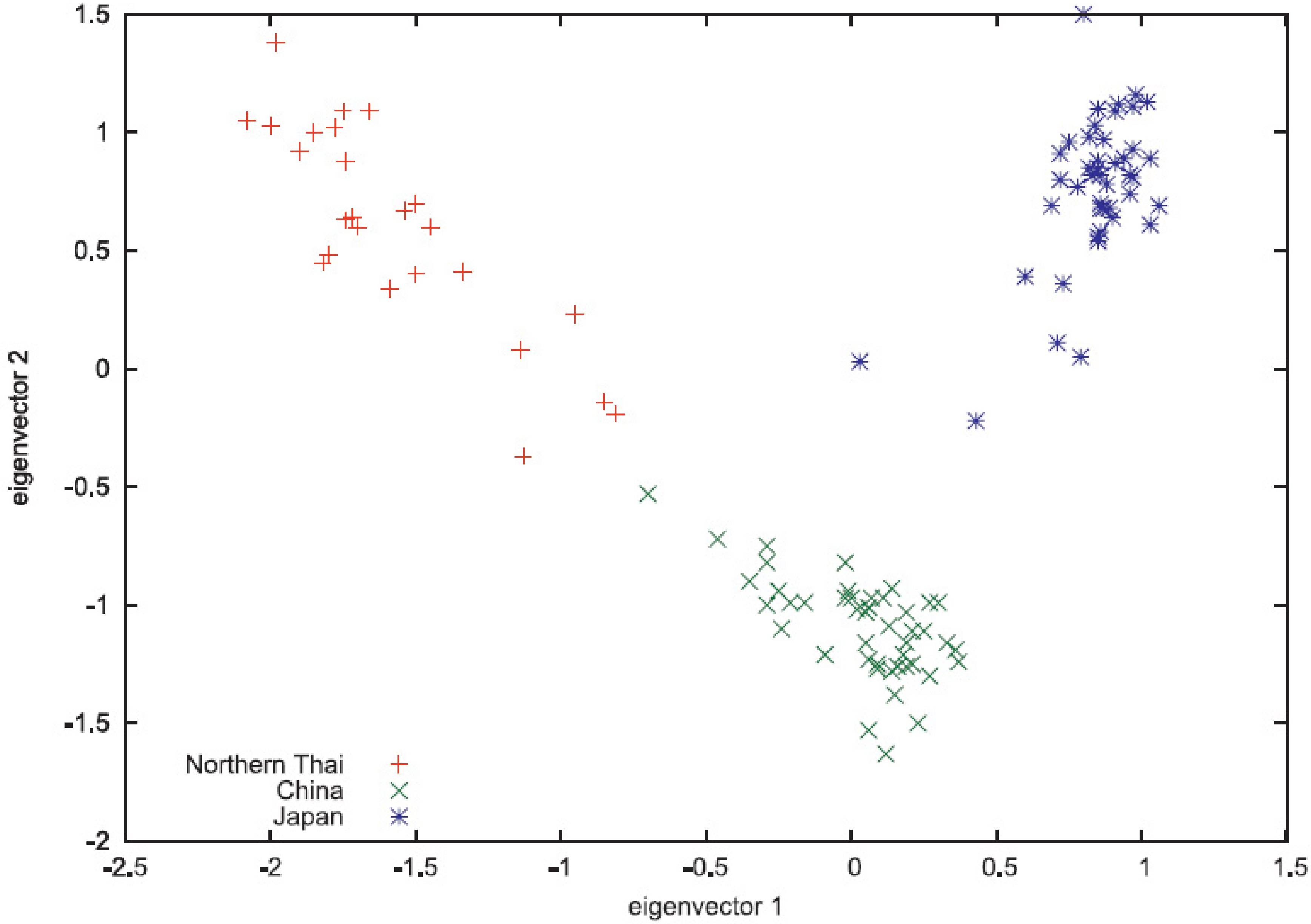
Example: Genomics

- Given: DNA of m people
- Question: Is there structure in the observed population?
- Equivalently: DNA independent samples from same distribution?

$$\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n) = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{1} \\ \mathbf{0} & \mathbf{3} & \mathbf{2} & \ldots & \mathbf{0} \\ \vdots & \vdots & & & \\ \mathbf{1} & \mathbf{0} & \mathbf{3} & \ldots & \mathbf{2} \end{bmatrix}$$

- Recenter to make the mean in each column 0
- If no population structure
  - all columns of A have same multivariate distribution
  - $\lambda_{\max}(XX^T)$ has same distribution as $\lambda_{\max}$(Wishart)
- Critical: We need the distribution of $\lambda_{\max}$(Wishart)!

Population Classification



3D Target Recognition (with Mike Jeffris, MITRE Corp.)

Blazer

HMMWV

M1 A1 Abrams

Leopard

T62

Challenger

?

Old: 2D Target Recognition

Views:  . . .

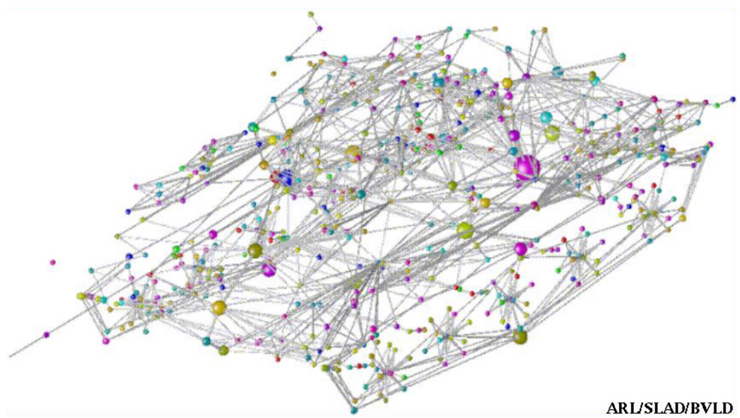
× **Sizes:**  . . .

× **Types:**  . . .

Inefficient

Enter 3D

Synthetic Aperture Radar



- **Works in fog, smoke, cloud cover; returns 3D images**
- **Tank** $= n \times 3$ **matrix**

3D Target Recognition: The Math Problem

- **Database:** $X_1, X_2, \ldots, X_m$ **(**$n \times 3$**)**
- **Observe: Tank** i **(**X_i**) + errors (**$E \sim N(0, \sigma^2 I_3 \otimes I_n)$**), rotated**

$$X = Q \cdot (X_i + E)$$

- **The covariance matrix** $S \equiv X^T X$ **becomes the tank's signature**
- S **is a non-central** 3×3 **Wishart**
- **Inverse problem:** $i = ?$
- **Hypothesis testing—based joint eigenvalue density of** S**:**

$$\log L(i|X) = \mathrm{tr}\left(-\frac{1}{2}\Sigma^{-1}S - \frac{1}{2}\Omega\right) + \log\left({}_0F_1\left(\frac{1}{2}m; \frac{1}{4}\Omega\Sigma^{-1}S\right)\right)$$

- **Requires the computation of** ${}_0F_1$**!**