New FFT-like Algorithm for Eigenvalues of Random Matrices

Cy Chan, Vesselin Drensky, Alan Edelman, and Plamen Koev

Department of Mathematics
Massachusetts Institute of Technology

New York University, March 2, 2007

Outline

- **What constitutes a random matrix?**
- **Why are the eigenvalues so important?**
  - **Genomics**
  - **3D Target recognition**
- **Computing the eigenvalue distributions**

What is a Random Matrix?

$B \in \mathbb{R}^{m \times n}, \quad b_{ij} \sim \mathcal{N}(0,1)$ **MATLAB notation:** `B=randn(m,n)`

$$A \equiv B^T \cdot B$$

"$n \times n$ central Wishart matrix with m DOF"

- **With covariance matrix Σ (positive semi-definite)**

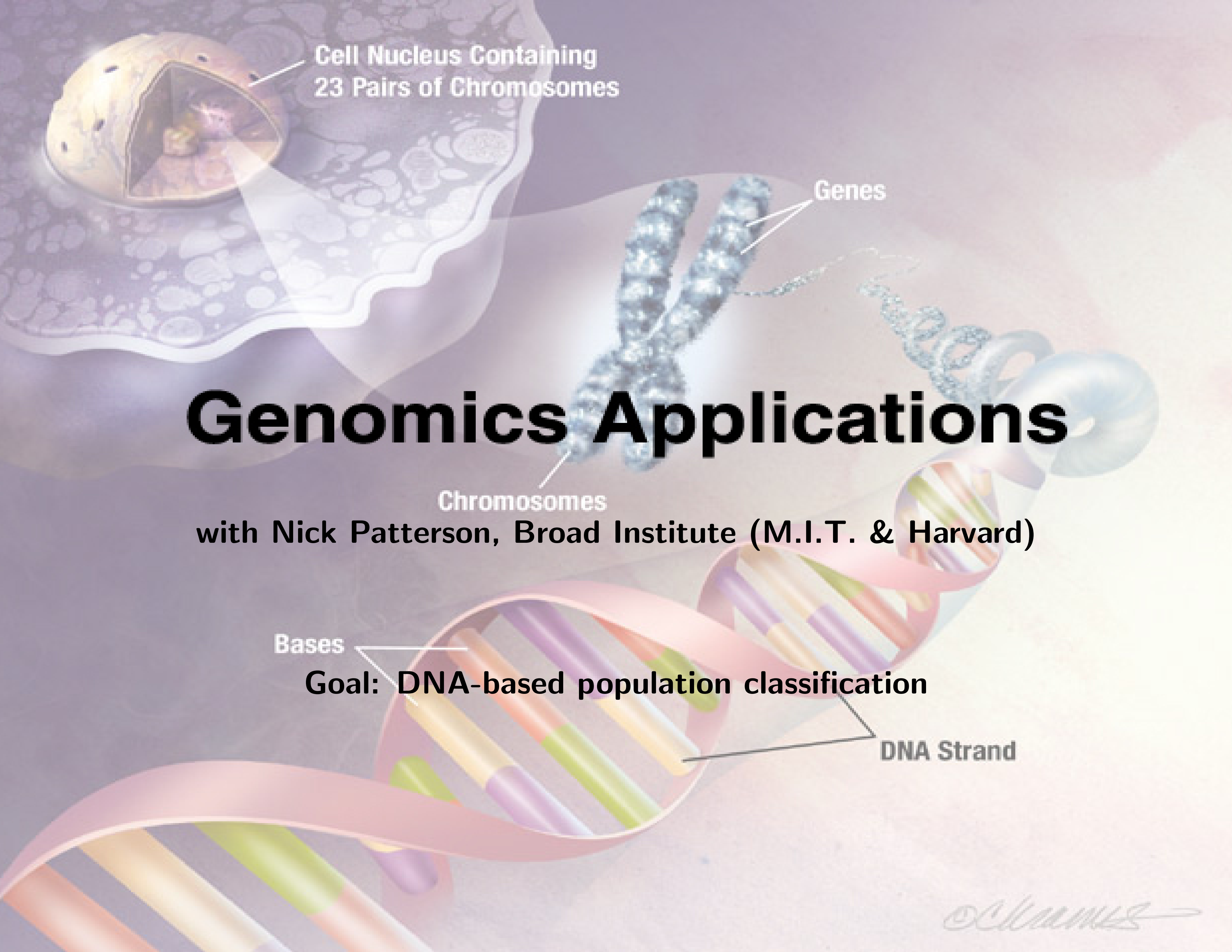
$$A \equiv \Sigma^{1/2} \cdot B^T \cdot B \cdot \Sigma^{1/2}$$

- **Note:** $W_m(1, I) \sim \chi^2_m$

- **(Generalized) eigenvalues critical in multivariate statistical analysis**
- **What if the entries are not normal?**

(Distributions of) Eigenvalues of Random Matrices

- Critical in multivariate statistical analysis
  - Extracting meaningful information from high-dimensional data
  - Clustering, classification
  - Inverse problems / parameter estimation
  - PCA, Canonical correlation analysis, MANOVA
  - Discriminant analysis
  - Multivariate hypothesis testing
- Theory mature since the 1960s
- New result: Fast algorithm for hypergeometric function of matrix argument
  (Solved 40-year-old computational problem)
- New applications:
  - Wireless communications
  - Biostatistics (high dimensional clustering/classification)
  - Genomics (population classification)
  - Automatic 3D Target classification/recognition



Genomics Applications

with Nick Patterson, Broad Institute (M.I.T. & Harvard)

Goal: DNA-based population classification

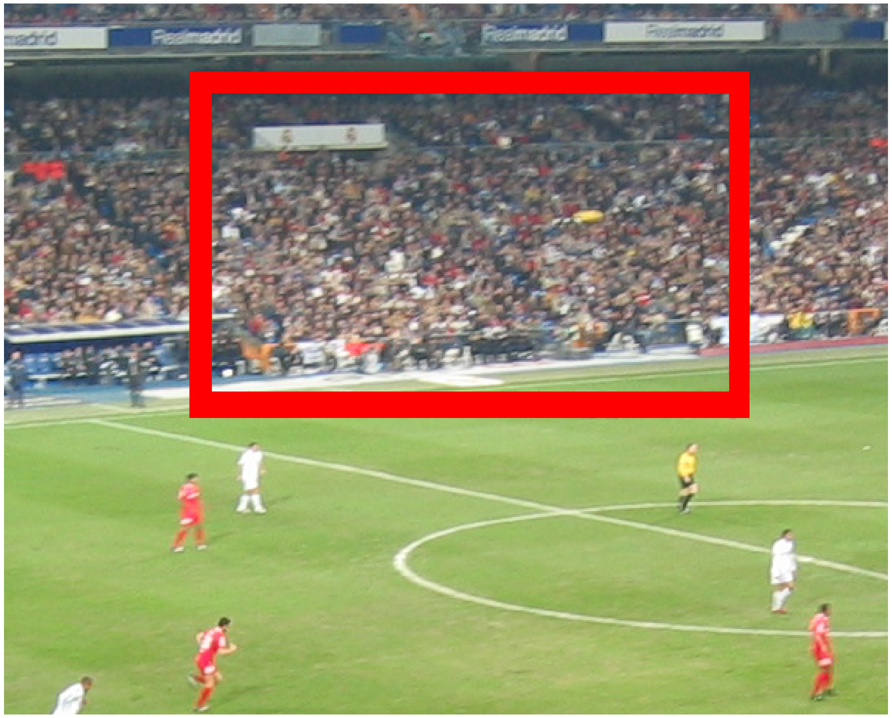
Question of interest: Population classification

Chelsea, London

Real Madrid

Question of interest: Population classification

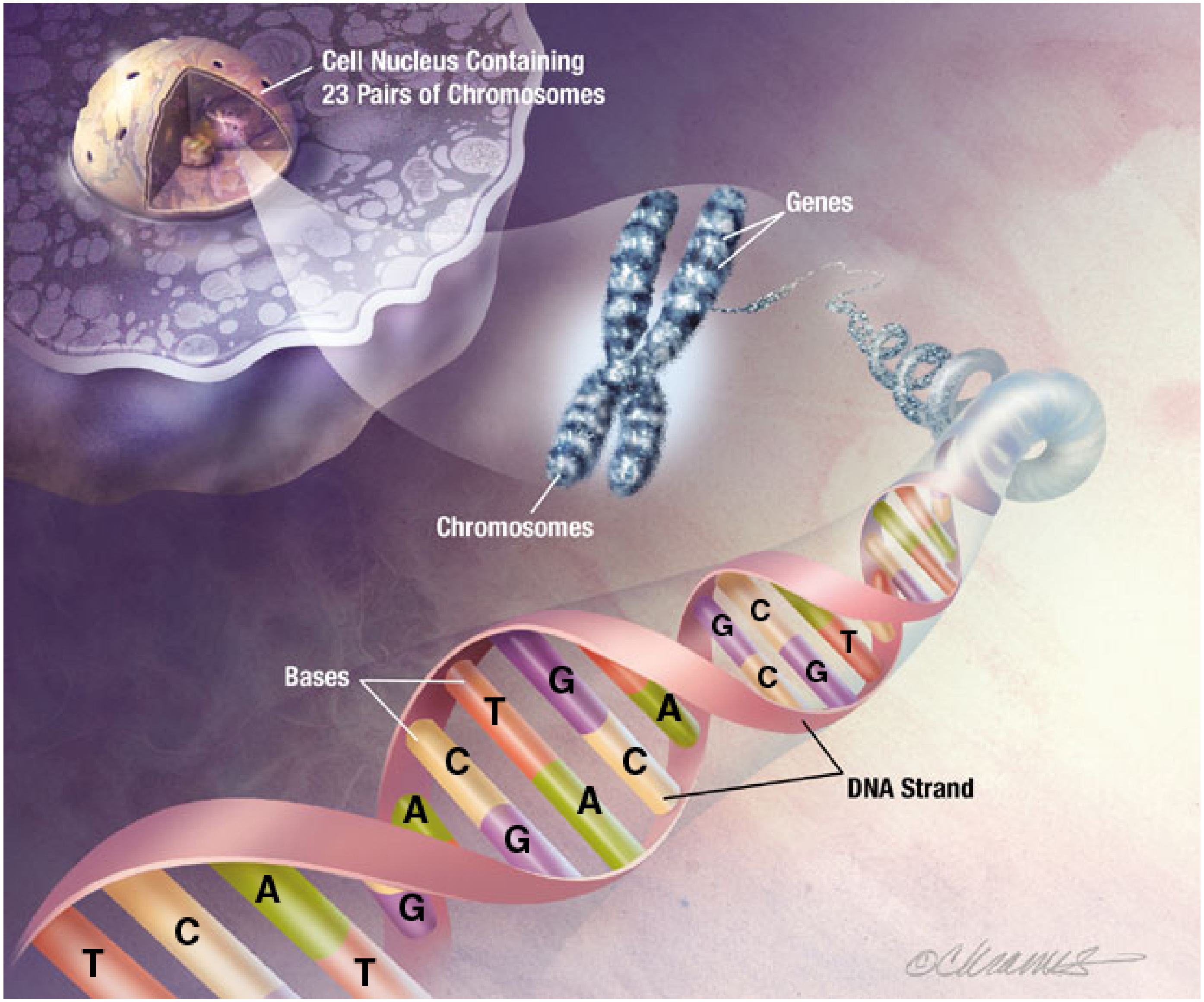
Chelsea, London



Real Madrid

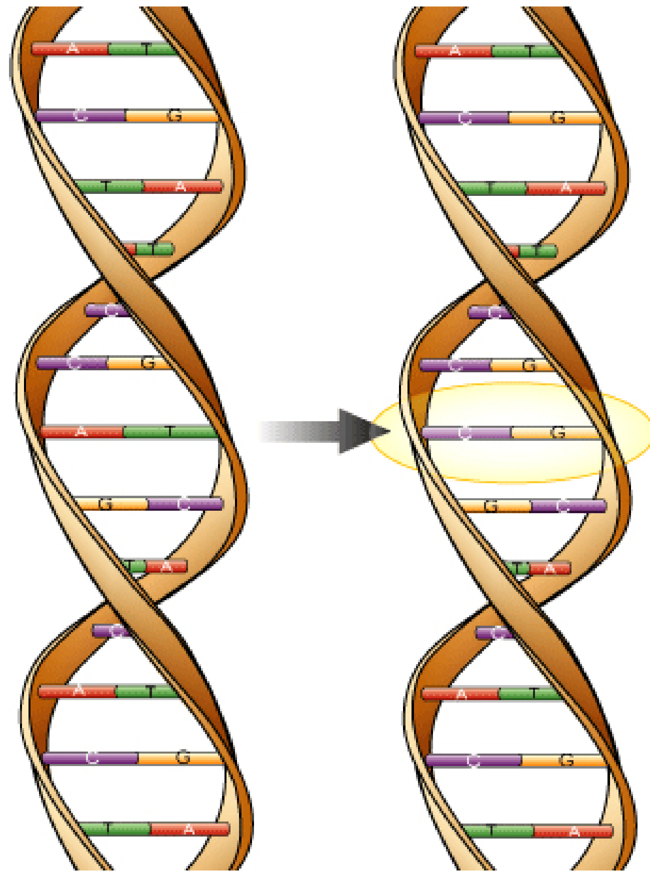
Separate the fans using only DNA information

Genomics 101



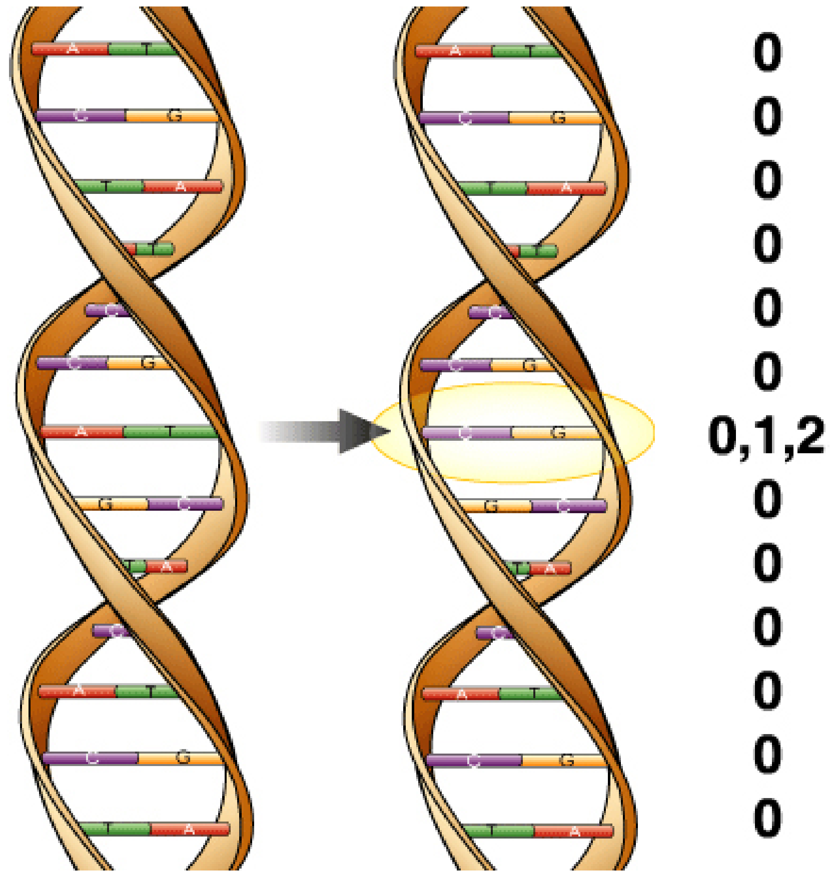
DNA = word in a 4-letter alphabet $\{A, C, G, T\}$

Single Nucleotide Polymorphism (SNP)



Happens during replication, once in 10,000,000 bases

Single Nucleotide Polymorphism (SNP)



Happens during replication, once in 10,000,000 bases

Where random matrices come in ...

 $= \underbrace{[\,0\ 0\ 0\ 1\ 0\ 0\ 3\ 0\ \ldots\ 2\ 3\ 0\,]}_{\approx 500{,}000}$

 $= \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & \ldots & 2 & 0 & 0 \\ 0 & 3 & 2 & 0 & 0 & 0 & 0 & 0 & \ldots & 3 & 1 & 0 \\ \vdots & \vdots & & & & & & & & \vdots & & \\ 1 & 0 & 3 & 0 & 2 & 0 & 1 & 2 & \ldots & 3 & 0 & 1 \end{bmatrix} \equiv A$

- Recenter to make the mean in each row 0
- If no population structure
  - all rows of A have the same multivariate distribution
  - $\lambda_{\max}(AA^T)$ has the same distribution as $\lambda_{\max}(\text{Wishart})$
- Otherwise—PCA to reveal structure
- Critical: We need the distribution of $\lambda_{\max}(\text{Wishart})$!

3D Target Recognition

(with Michael Jeffris, MITRE Corporation)

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

Blazer

HMMWV

M1 A1 Abrams

Leopard

T62

Challenger

?

Old: 2D Target Recognition

Views:

. . .

× **Sizes:**

. . .

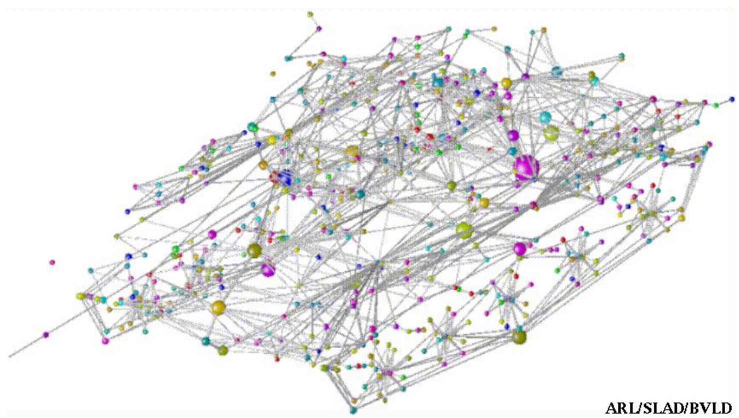
× **Types:**

. . .

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 ${}_0F_1$—
  the Hypergeometric Function of a Matrix Argument

Eigenvalue distributions $= {}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; X)$

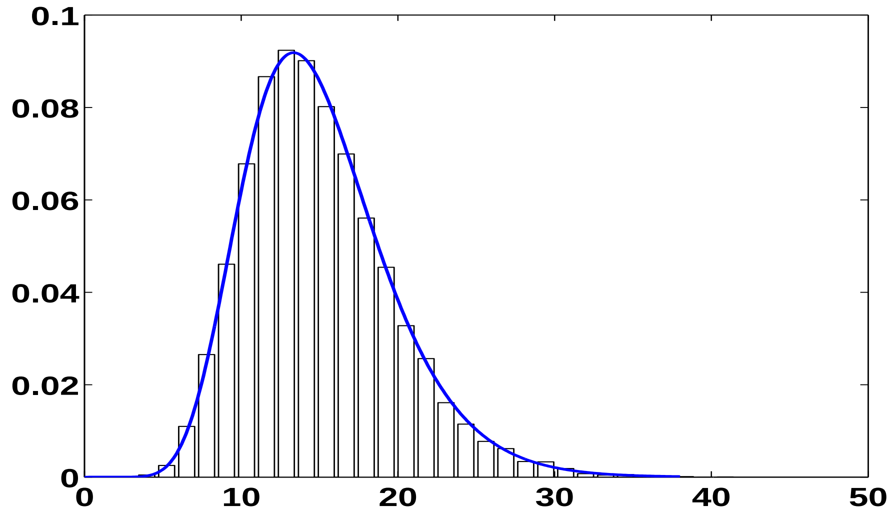
- **Good news: (Joint) Eigenvalue densities/distributions are known explicitly**
- **Joint density of the eigenvalues of $S \sim W_m(n, \Sigma, \Omega)$ is proportional to**

$$e^{\mathrm{tr}\left(-\frac{1}{2}\Sigma^{-1}S - \frac{1}{2}\Omega\right)} \cdot {}_0F_1\left(\frac{1}{2}m; \frac{1}{4}\Omega\Sigma^{-1}S\right)$$

- **Similarly for $\lambda_{\max}(W_m(n, \Sigma))$ (James, 1964):**

$$P(\lambda_{\max} < x) \sim x^{m/2} \cdot {}_1F_1\left(\frac{m}{2}; \frac{n+m+1}{2}; -\frac{1}{2}x\Sigma^{-1}\right)$$

- **E.g., $\lambda_{\max}(W_7(4, I))$**



- **Bad news: It's the hypergeometric function of a matrix argument**

The Hypergeometric Function of a Matrix Argument

- **Univariate:**

$$ {}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;x) \equiv \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{k!(b_1)_k \cdots (b_q)_k} \cdot x^k $$

  where $(a)_k = a(a+1)\cdots(a+k-1)$.

- **of a Matrix Argument X (with eigenvalues $x_1,\ldots,x_n$):**

$$ {}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;X) \equiv \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} \frac{(a_1)_\kappa \cdots (a_p)_\kappa}{k!(b_1)_\kappa \cdots (b_q)_\kappa} \cdot C_\kappa^\alpha(X), $$

  where

  – $(a)_\kappa \equiv \prod_{(i,j)\in\kappa} (a-(i-1)/\alpha + j - 1)$ — Pochhammer symbol

  – $C_\kappa^\alpha(X) = C_\kappa^\alpha(x_1,\ldots,x_n)$ — Schur/zonal/Jack polynomial

- Scalar-valued symmetric function in $x_1,\ldots,x_n$

- Multivariate degrees indexed by partitions κ

$$ \underbrace{(1)}_{\kappa\vdash 1}; \quad \underbrace{(2),(1,1)}_{\kappa\vdash 2}; \quad \underbrace{(3),(2,1),(1,1,1)}_{\kappa\vdash 3}; \quad \ldots $$

- **Notoriously difficult to compute and/or approximate!**

Computing ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; X)$ is really hard!

- **Reported for example in about 50% of papers in the IEEE Journals on wireless communication, signal processing, etc.**

  (search for "hypergeometric function" and "matrix argument" at `ieeexplore.ieee.org`)

  – 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.

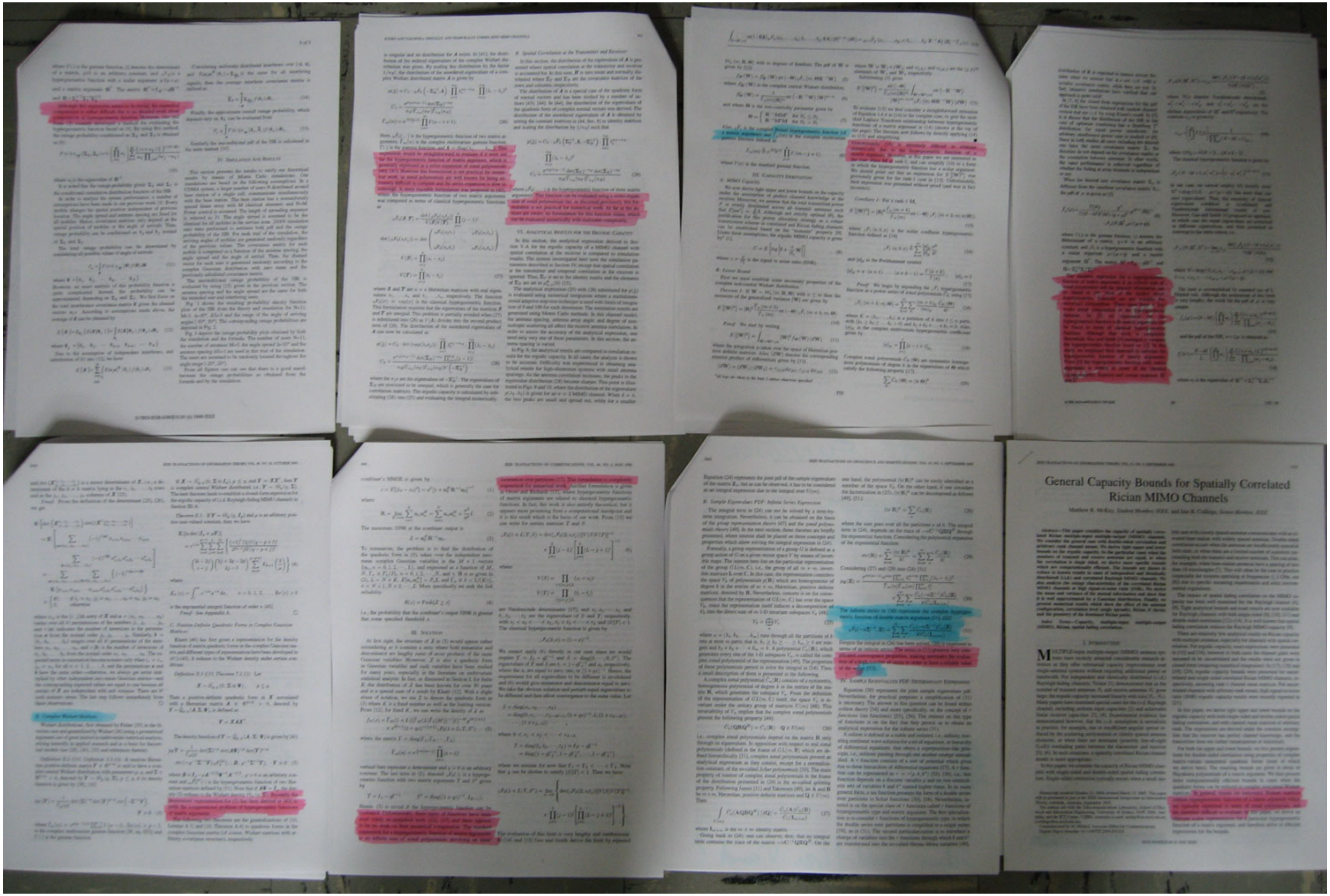
  – A. Ozyildirim and Y. Tanik. Outage probability analysis of a CDMA system with antenna arrays in a correlated Rayleigh environment. In IEEE Military Communications Conference Proceedings, 1999. MILCOM 1999, volume 2, pages 939–943, 31 Oct.–3 Nov. 1999.

  – 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 ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; X)$ is really hard!



Computing ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; X)$ 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

Previous best algorithm for ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; X)$

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§

Previous best algorithm for ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; X)$

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.)

Previous best algorithm for ${}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;X)$

Electronic Transactions on Numerical Analysis.
Volume 11, pp. 121-130, 2000.
Copyright © 2000, Kent State University.
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
```

Outline for today's talk

- **The computation of accurate eigenvalues of structured matrices**
  - **An application in Electrical Impedance Tomography**
  - **New result: accurate eigenvalue algorithm for totally positive matrices**
- **Computing eigenvalues of random matrices**
  - **Application in Automatic Target Recognition**
  - **New fast algorithm for $_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; X)$**

The Hypergeometric Function of a Matrix Argument

$$ {}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; X) \equiv \sum_{k=0}^{M} \sum_{\kappa \vdash k} \frac{(a_1)_\kappa \ldots (a_p)_\kappa}{k!(b_1)_\kappa \ldots (b_q)_\kappa} \cdot C_\kappa^\alpha(X) $$

- $C_\kappa^\alpha(X) = C_\kappa^\alpha(x_1, \ldots, x_n)$ **symmetric polynomial in** x_i
- **Number of terms in each** $C_\kappa^\alpha(X)$: $\mathcal{O}(n^{|\kappa|})$
- **New result: Computing** $C_\kappa^\alpha(X)$ **in** $\mathcal{O}(n)$ **time**

- **Will demonstrate the complex case** $\alpha = 1$
- **Then** $C_\kappa^1 = s_\lambda$**, the Schur function**

Examples of Schur/Jack functions

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

- New result: $\mathcal{O}(n)$
- Trick: Connection with group representations

Computing the Schur/Jack Function

- **Idea**: s_λ characters of irreducible representations of $\mathbf{GL}_n(\mathbb{C})$
- Combinatorial formula for s_κ

$$s_\kappa(x_1, \ldots, x_n) = \sum x^T$$

where summation is over all semistandard Young κ-tableaux T.

- SSYT (def): Fill Young diagram of κ, with $1, 2, \ldots, n$ strictly $\uparrow$ by columns, nonstrictly by rows.
- Example: $\kappa = (3, 3, 2, 1)$, $n = 6$

1	1	2
3	4	4
5	6	
6		

$$\longrightarrow x_1^2 x_2 x_3 x_4^2 x_5 x_6^2$$

- Observation: "n" can only be in a "bottom" box $\Rightarrow$

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

(the characters of $\mathbf{GL}_{n-1}(\mathbb{C})$ induce those of $\mathbf{GL}_n(\mathbb{C})$)

This is the first step

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

Summation over κ such that λ/μ is a horizontal strip

- **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
- **In general:** $\mathcal{O}(nN_\lambda)$ where $N_\lambda = \#\{\mu |\, \lambda/\mu$ is a horizontal strip$\}$
- **Next:** getting rid of N_λ

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 $\mathcal{O}(nN_\lambda)$**

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 |λ| ≤ M in reverse lexicographic order
      s_λ = s_λ + s_λ(i) x_n
```

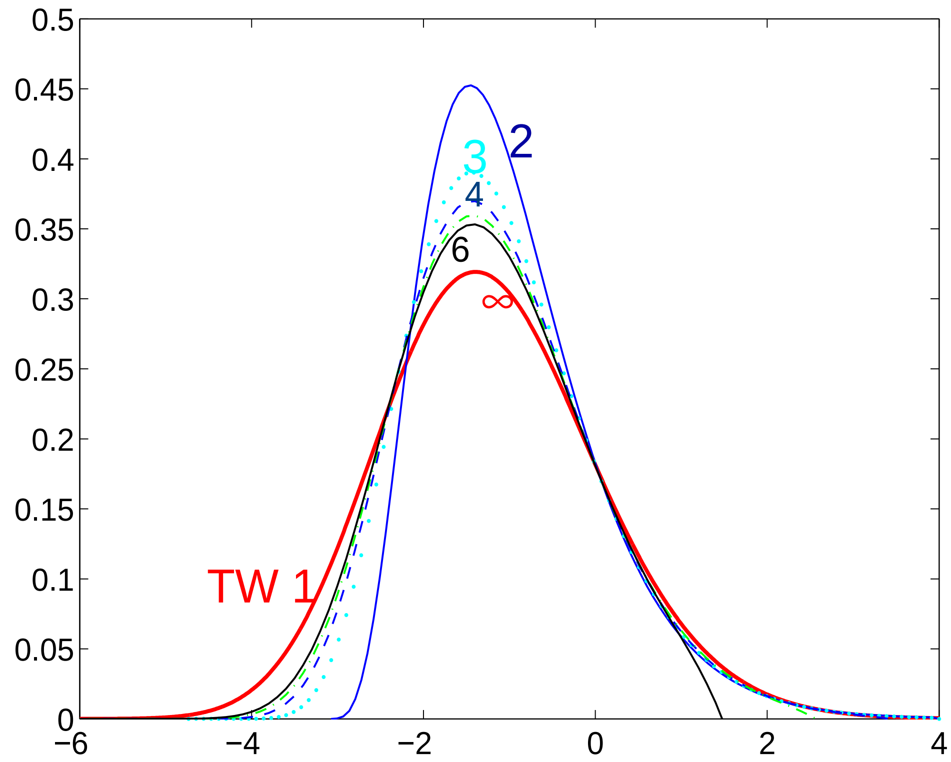
(where $\lambda^{(i)} \equiv (\lambda_1, \ldots, \lambda_i - 1, \ldots, \lambda_n)$)

- Final cost: $\lambda_1' n$ per each s_λ, optimal

Impact of New Algorithm

- $A_p \sim W_p(n, I); \quad n/p = 5$
  $(\lambda_{\max}(A_p) - \mu_p)/\sigma_p \to \text{TW}_1$

- **Finite eigenvalue distributions,**
  (with arbitrary covariances Σ!)



- **Enabled automatic 3D target classification**
- **Enabled population classification in genomics**
- **New theoretical results in random matrix theory**
- **Potential applications to wireless communications**
- **Included in MOCAPY biostatistics package**
- **Efficient C implementation at** `http://math.mit.edu/~plamen`

Conclusions

- New algorithms for ${}_pF_q$
- Papers and software at: `http://math.mit.edu/~plamen`
- Impact on important applications

Future work

- R, SAS implementations
- New algorithms based on saddle point approximations
- Automatic convergence detection
- FFT generalization to zonal polynomials

$$\begin{bmatrix} 1 & x & x^2(\alpha+1) & x^3(\alpha+1)(2\alpha+1) \\ & 1 & x & x^2(\alpha+1) \\ & & 1 & x \\ & & & 1 \end{bmatrix}$$

- Tracy–Widom finite inference