Computing the Hypergeometric Function of a Matrix Argument

Plamen Koev
Massachusetts Institute of Technology

Joint work with Alan Edelman

SEA 06

Massachusetts Institute of Technology, July 10–14, 2006

Why ${}_pF_q(\cdot,\cdot,X)$?

- Distributions of

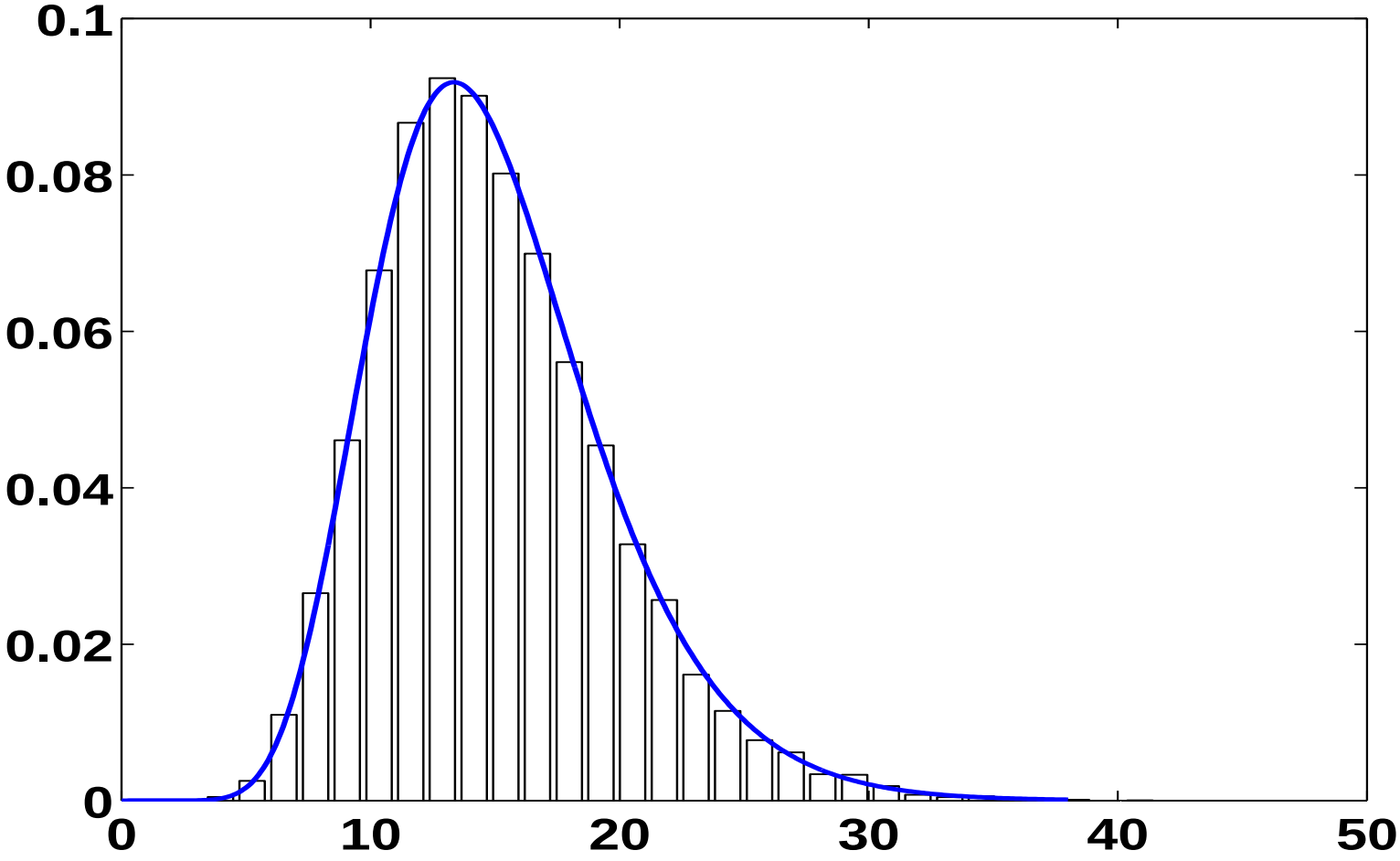
$$\lambda_{\min}, \lambda_{\max}, \det, \text{ etc.}$$

  of

$$\text{Wishart, Jacobi, Laguerre}$$

  expressed in terms of ${}_pF_q(\cdot,\cdot,X)$
- The distributions useful in:
  - Hypothesis testing (e.g., $\Sigma = I$, etc.)
  - Parameter estimation: $A \sim W_m(n, \sigma^2 I), \quad \sigma = ?$
- Applications in:
  - Population classification
  - Automatic target classification
  - Wireless communications
- Computing ${}_pF_q(\cdot,\cdot,X)$: 40-year-old open problem
  - Notorious complexity and slow convergence
  - Empirical methods inefficient

Distribution of $\lambda_{\max}$ of 4×4 Wishart with 7 DOF, $\Sigma = I$



- Exact vs Empirical with 20,000 replications

James 1964

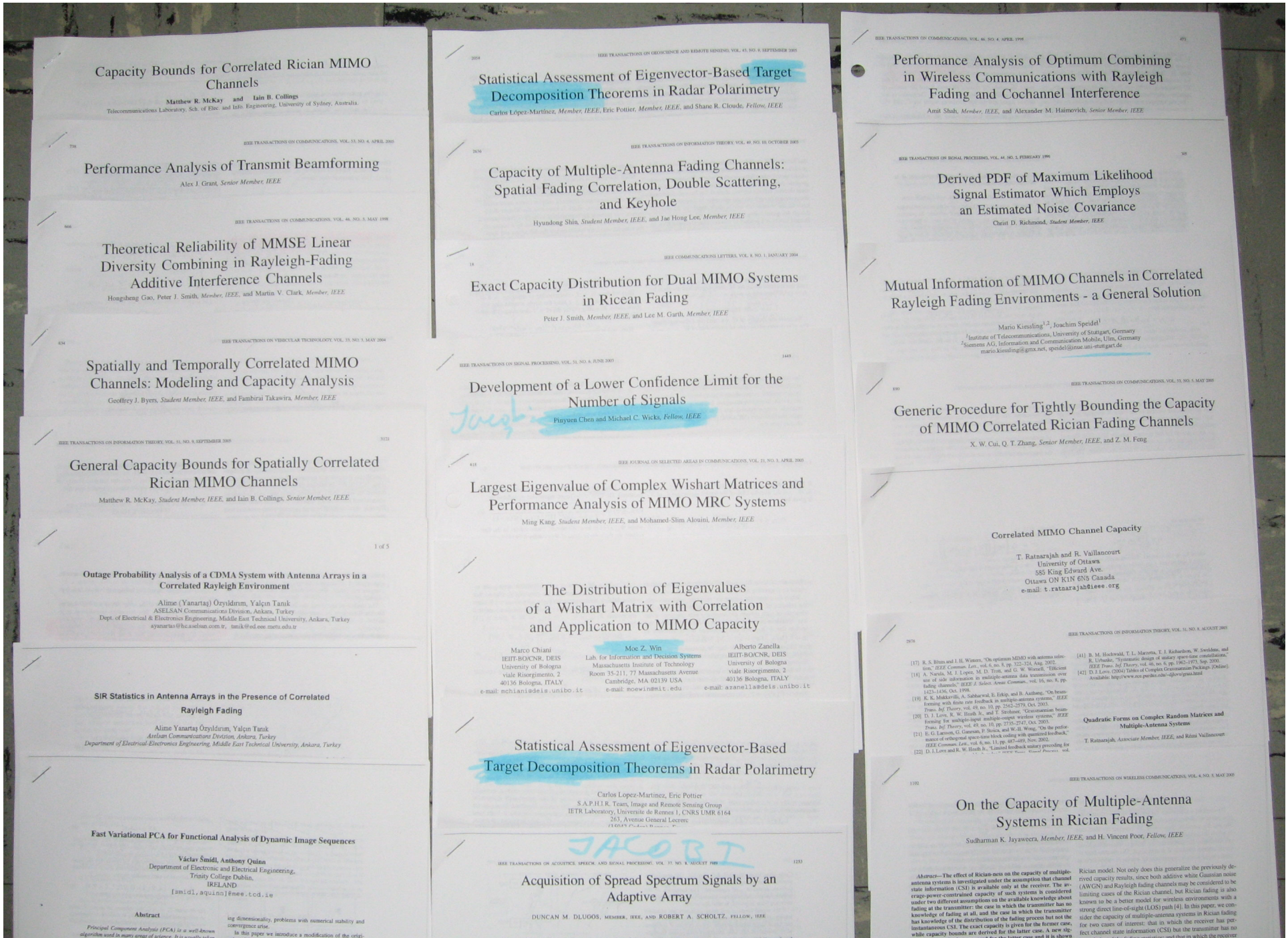
If $A \sim W_n(m, \Sigma)$ then

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

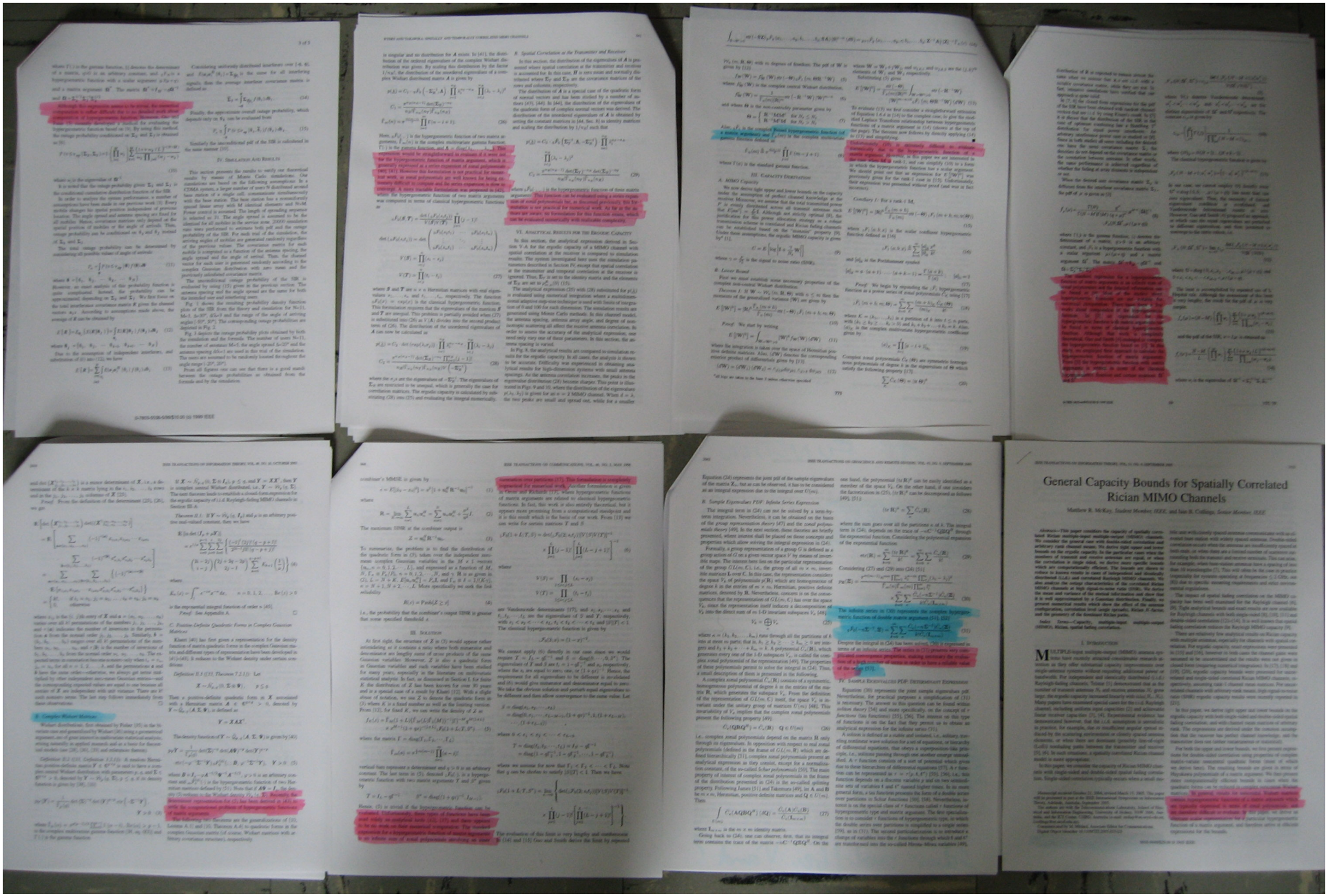
$$= x^{\frac{m}{2}} \cdot \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} p_\kappa \cdot x^k \cdot C_\kappa(\Sigma^{-1})$$

- Slow convergence $\Rightarrow$ $\infty \sim 50, 100, 150$
- $C_\kappa(X)$ – *Zonal Polynomial* – Really hard: $O(n^m)$ terms in each!
- Our Contribution: $O(n)$
- Impossible until now
  - Previous best algorithm ($n = 5$): 8 days
    (Gutiérrez, Rodriguez, Sáez, 2000)
  - New algorithm: $\frac{1}{100}$ second

Computing ${}_pF_q(\cdot,\cdot,X)$ is really hard!



Computing ${}_pF_q(\cdot,\cdot,X)$ is really hard!



Computing ${}_pF_q(\cdot, \cdot, X)$ 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 the case when M is rank-1

But really useful ...

$\lambda_{\max}$ in PCA

The Annals of Statistics
2001, Vol. 29, No. 2, 295–327

ON THE DISTRIBUTION OF THE LARGEST EIGENVALUE IN PRINCIPAL COMPONENTS ANALYSIS[1]

BY IAIN M. JOHNSTONE

Stanford University

Let $x_{(1)}$ denote the square of the largest singular value of an $n \times p$ matrix X, all of whose entries are independent standard Gaussian variates. Equivalently, $x_{(1)}$ is the largest principal component variance of the covariance matrix $X'X$, or the largest eigenvalue of a p-variate Wishart distribution on n degrees of freedom with identity covariance.

Consider the limit of large p and n with $n/p = \gamma \geq 1$. When centered by $\mu_p = (\sqrt{n-1} + \sqrt{p})^2$ and scaled by $\sigma_p = (\sqrt{n-1} + \sqrt{p})(1/\sqrt{n-1} + 1/\sqrt{p})^{1/3}$, the distribution of $x_{(1)}$ approaches the Tracy–Widom law of order 1, which is defined in terms of the Painlevé II differential equation

Automatic 3D Target Classification

Blazer

X_1

HMMWV

X_2

M1 A1 Abrams

X_3

Leopard

X_4

T62

X_5

Challenger

X_6

- X_i — $n \times 3$ matrices (given data)
- Observation: $X = (X_i + E)Q, \qquad e_{ij} \sim N(0, \sigma^2)$
- $X^T X$ — noncentral Wishart, eigs do not depend on Q
- Question: $i = ?$

$$L(i|X) \sim {}_1F_0(\cdot; X^T X)$$

- Reference: Michael Jeffris (MITRE), Proceedings of SPIE 2005

Computing ${}_pF_q(\cdot;\cdot;X)$

$$P(\lambda_{\max}(A) < x) \sim x^{\frac{m}{2}} \cdot {}_1F_1\left(\tfrac{m}{2}; \tfrac{n+m+1}{2}; -\tfrac{1}{2}x\Sigma^{-1}\right)$$
$$= x^{\frac{m}{2}} \cdot \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} p_\kappa \cdot x^k \cdot C_\kappa(\Sigma^{-1})$$

- Means computing zonal polynomials $C_\kappa(\Sigma^{-1})$
- $C_\kappa(\Sigma^{-1})$ depends only on the eigenvalues $x_1, x_2, \ldots, x_n$ of Σ^{-1}
- Illustrate $\beta = 2$ (complex); general β analogous

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

Computing $C_\kappa(X)$

IDEA: C_κ are $\chi(\mathrm{GL}_n(\mathbb{C}))$; $\chi(\mathrm{GL}_{n-1}(\mathbb{C}))$ induce $\chi(\mathrm{GL}_n(\mathbb{C}))$

Example:

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

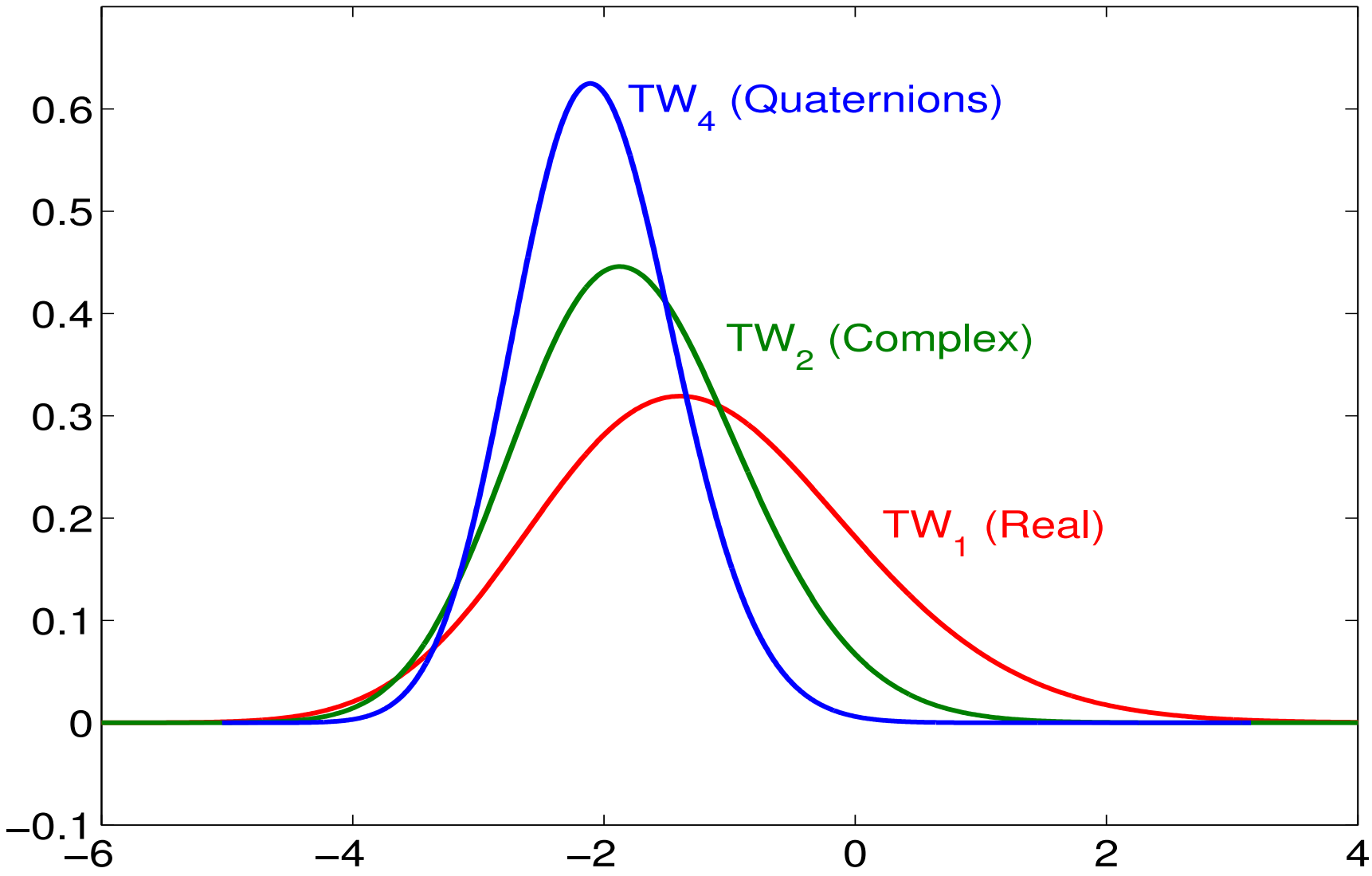
Algorithm:

$$\begin{aligned} s_1 &= x_1 \\ s_2 &= s_1 + x_2 && (= x_1 + x_2) \\ s_3 &= s_2 + x_3 && (= x_1 + x_2 + x_3) \\ &\vdots \\ s_{n-1} &= s_{n-2} + x_{n-1} && (= x_1 + x_2 + \cdots + x_{n-1}) \end{aligned}$$

$$C_{(1,1)}(X) = s_1x_2 + s_2x_3 + \cdots + s_{n-1}x_n$$

- **Cost: $O(n)$ versus $O(n^2)$**
- **In general: $O(n)$ versus $O(n^{|\kappa|})$**

Tracy–Widom Laws



Recall:

The Annals of Statistics
2001, Vol. 29, No. 2, 295–327

ON THE DISTRIBUTION OF THE LARGEST EIGENVALUE IN PRINCIPAL COMPONENTS ANALYSIS[1]

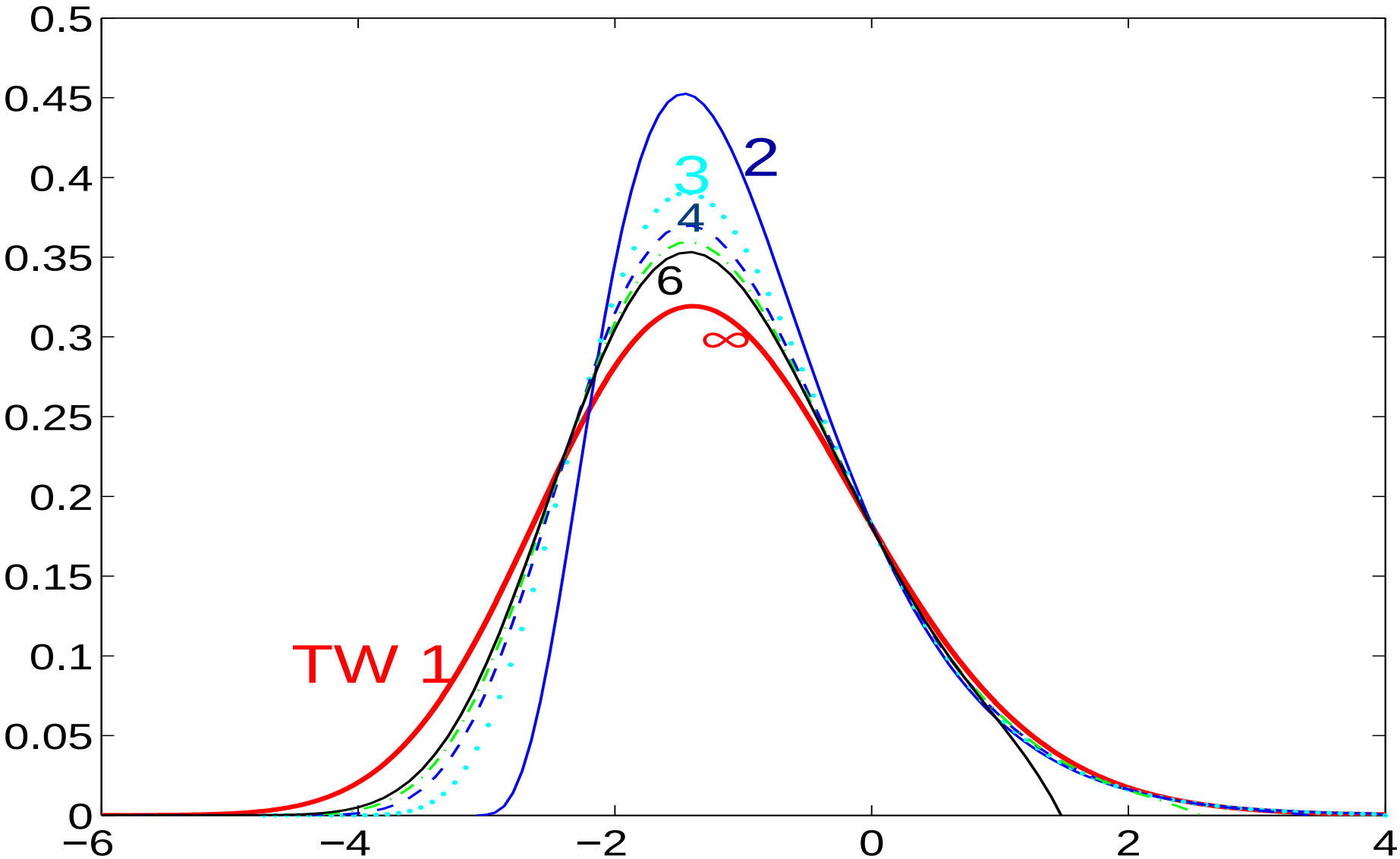
BY IAIN M. JOHNSTONE

Stanford University

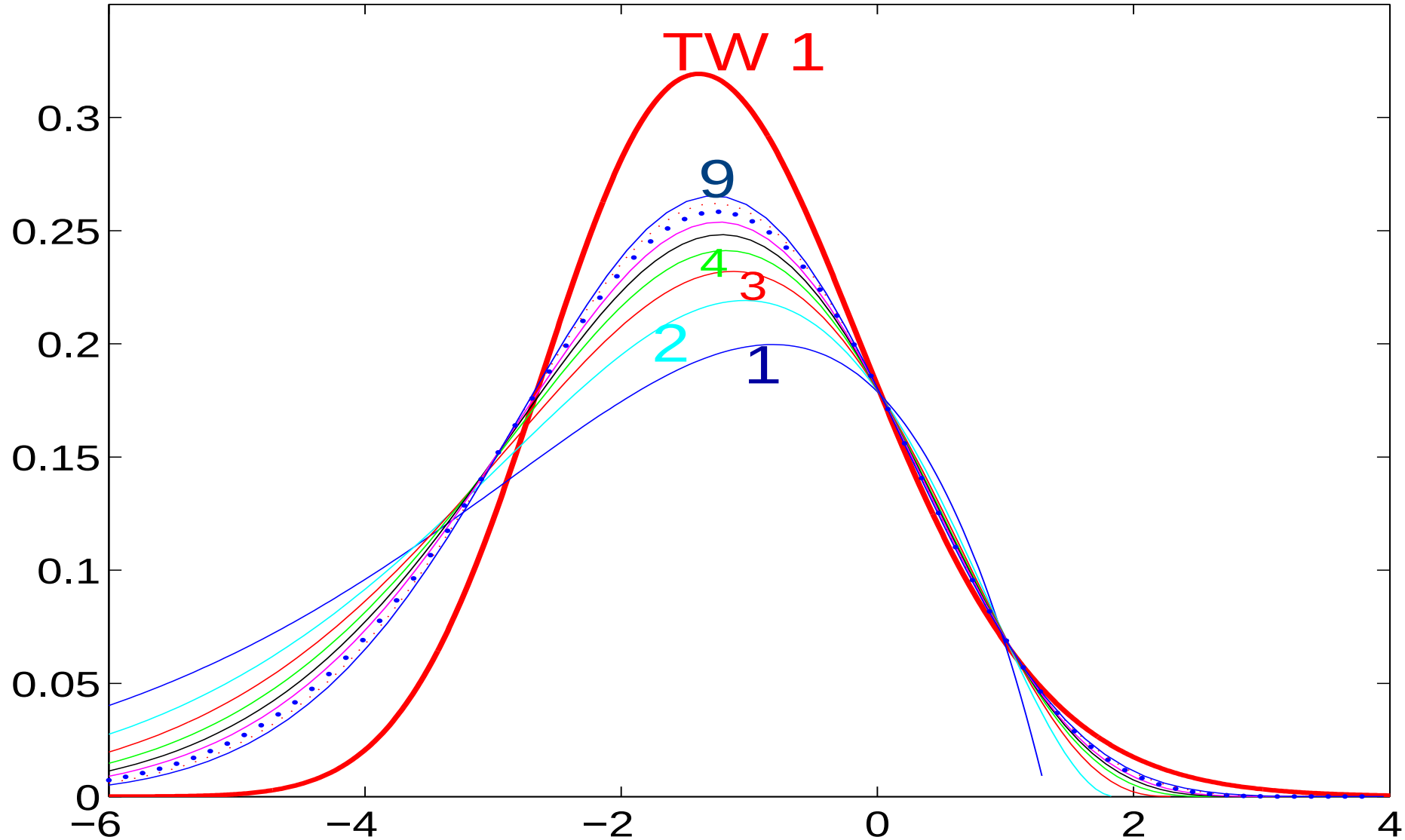
Let $x_{(1)}$ denote the square of the largest singular value of an $n \times p$ matrix X, all of whose entries are independent standard Gaussian variates. Equivalently, $x_{(1)}$ is the largest principal component variance of the covariance matrix $X'X$, or the largest eigenvalue of a p-variate Wishart distribution on n degrees of freedom with identity covariance.

Consider the limit of large p and n with $n/p = \gamma \geq 1$. When centered by $\mu_p = (\sqrt{n-1} + \sqrt{p})^2$ and scaled by $\sigma_p = (\sqrt{n-1} + \sqrt{p})(1/\sqrt{n-1} + 1/\sqrt{p})^{1/3}$, the distribution of $x_{(1)}$ approaches the Tracy–Widom law of order 1, which is defined in terms of the Painlevé II differential equation

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



Jacobi: $A_p \sim W_p(q, \Sigma)$, $B_p \sim W_p(n, \Sigma)$, $A_p - \lambda B_p$



$$p = 4k + 2,\ n/q = 2,\ n/p = 3,\ \lambda_{\max}(A_p B_p^{-1}) \to \lambda_{\max}(A_\infty B_\infty^{-1}) \sim \mathrm{TW}_1$$

Future work: Cooley–Tukey–type algorithm

- $(\mathrm{DFT})_{ij}$—characters of $\mathbb{Z}/n\mathbb{Z}$ $\longleftrightarrow$ C_λ—characters of $\mathrm{GL}_n(\mathbb{C})$
- Main identity

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

In matrix form:

$$\mathcal{C}_n = \mathcal{C}_{n-1} \cdot Y_n(x_n)$$

where for example

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

- Y_n structured ... MVM takes linear time
- Cost(New Alg) $\approx \sqrt{\text{Cost(Current Alg)}}$ ... just like FFT

Conclusions

- **New efficient algorithm for ${}_pF_q$: Takes seconds**
- **Works on $\Sigma = I$ <u>and</u> $\Sigma \neq I$**
- **Solves a 40-year-old problem**

- **Future Work:**
  - **Cooley–Tukey–like algorithm**

$$\text{Cost} = O(\sqrt{\text{Current cost}})$$

  - **Toolbox**

- **Paper in Math. Comp., MATLAB software, slides, all available from:**

`http://math.mit.edu/~plamen`