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 \( \lambda_{\text{min}}, \lambda_{\text{max}}, \det, \text{etc.} \) of 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: \( A \sim W_m(n, \sigma^2I), \quad \sigma = ? \)

- 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 $\lambda_{\text{max}}$ of $4 \times 4$ Wishart with 7 DOF, $\Sigma = I$

- **Exact** vs **Empirical** with 20,000 replications
If $A \sim W_n(m, \Sigma)$ then

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

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

- Slow convergence $\Rightarrow \infty \sim 50, 100, 150$
- $C_{\kappa}(X) - \text{Zonal Polynomial} - \text{Really hard: } O(n^m) \text{ terms in each!}$
- **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!

But really useful ...
ON THE DISTRIBUTION OF THE LARGEST EIGENVALUE IN PRINCIPAL COMPONENTS ANALYSIS

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

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

$$L(i | X) \sim \text{1F0}(\vdots; X^T X)$$

- Reference: Michael Jeffris (MITRE), Proceedings of SPIE 2005
Computing $pF_q(\cdot; \cdot; X)$

\[ P(\lambda_{\text{max}}(A) < x) \sim x^{m/2} \cdot \sum_{\kappa} \sum_{k=0}^{\infty} p_{\kappa} \cdot x^k \cdot C_{\kappa}(\Sigma^{-1}) \]

- 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 $\Sigma^{-1}$
- Illustrate $\beta = 2$ (complex); general $\beta$ analogous

<table>
<thead>
<tr>
<th>Partition $\kappa$</th>
<th>$C_{\kappa}$</th>
<th>Number of terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) $x_1 + \cdots + x_n$</td>
<td>$O(n)$</td>
<td></td>
</tr>
<tr>
<td>(2) $\sum x_i x_j$</td>
<td>$O(n^2)$</td>
<td></td>
</tr>
<tr>
<td>(1, 1, 1) $\sum x_i x_j x_k$</td>
<td>$O(n^3)$</td>
<td></td>
</tr>
<tr>
<td>$\kappa \sum_T x^T$</td>
<td>$O(n^{</td>
<td></td>
</tr>
</tbody>
</table>
Computing $C_\kappa(X)$

**IDEA:** $C_\kappa$ are $\chi(\text{GL}_n(\mathbb{C}))$; $\chi(\text{GL}_{n-1}(\mathbb{C}))$ induce $\chi(\text{GL}_n(\mathbb{C}))$

**Example:**

\[
C_{(1,1)}(X) = \sum_{i<j} x_i x_j
= x_1 x_2 + (x_1 + x_2)x_3 + \cdots + (x_1 + \cdots + x_{n-1})x_n
\]

**Algorithm:**

\[
s_1 = x_1 \\
\begin{align*}
s_2 &= s_1 + x_2 \quad (= x_1 + x_2) \\
s_3 &= s_2 + x_3 \quad (= x_1 + x_2 + x_3) \\
\vdots \\
s_{n-1} &= s_{n-2} + x_{n-1} \quad (= x_1 + x_2 + \cdots + x_{n-1})
\end{align*}
\]

\[
C_{(1,1)}(X) = s_1 x_2 + s_2 x_3 + \cdots + s_{n-1} x_n
\]

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

TW_4 (Quaternions)

TW_2 (Complex)

TW_1 (Real)
ON THE DISTRIBUTION OF THE LARGEST EIGENVALUE IN PRINCIPAL COMPONENTS ANALYSIS

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.
$A_p \sim W_p(n, I); \quad n/p = 5; \quad (\lambda_{\text{max}}(A_p) - \mu_p)/\sigma_p \rightarrow TW_1$
Jacobi: $A_p \sim W_p(q, \Sigma), \ B_p \sim W_p(n, \Sigma), \ A_p - \lambda B_p$

$p = 4k + 2, \ n/q = 2, \ n/p = 3, \ \lambda_{\text{max}}(A_p B_p^{-1}) \rightarrow \lambda_{\text{max}}(A_\infty B_\infty^{-1}) \sim \text{TW}_1$
Future work: Cooley–Tukey–type algorithm

• \((\text{DFT})_{ij}\) — characters of \(\mathbb{Z}/n\mathbb{Z}\) \(\leftrightarrow\) \(C_\lambda\) — characters of \(\text{GL}_n(\mathbb{C})\)

• Main identity

\[
C_\kappa(x_1, x_2, \ldots, x_n) = \sum_{\lambda < \kappa} C_\lambda(x_1, x_2, \ldots, x_{n-1}) \cdot x_n^{\kappa - |\lambda|} \cdot f_{\lambda\kappa}^\alpha
\]

In matrix form:

\[
C_n = C_{n-1} \cdot Y_n(x_n)
\]

where for example

\[
Y_2(x) = \begin{bmatrix}
1 & x & x^2 & x^3 \\
1 & x & x^2 & x^3 & x^4 \\
1 & x & x^2 & x^3 & x^4 & x^5 \\
1 & x & x^2 & x^3 & x^4 & x^5 & x^6
\end{bmatrix}
\]

• \(Y_n\) structured ... MVM takes \text{linear} time

• \(\text{Cost}(\text{New Alg}) \approx \sqrt{\text{Cost}(\text{Current Alg})}\) ... just like FFT
Conclusions

• New efficient algorithm for $pF_q$: Takes seconds
• Works on $\Sigma = I$ and $\Sigma \neq I$
• 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