The Expected Determinant of the Random Gram Matrix and its Application to Information Retrieval Systems

Jacob Martin

October 22, 2007
Latent Semantic Indexing is a method of using Singular Value Decomposition to approximate high dimensional spaces for information retrieval.

Selective Latent Semantic Indexing is an improvement to Latent Semantic Indexing based on random gram determinants that Rod Canfield and I developed.
Latent Semantic Indexing is a method of using Singular Value Decomposition to approximate high dimensional spaces for information retrieval.

Selective Latent Semantic Indexing is an improvement to Latent Semantic Indexing based on random gram determinants that Rod Canfield and I developed.
Selective Latent Semantic Indexing

Efficiently estimates the singular values based on
- probability models, and
- matrix representations
in order to
- compute singular values faster, and
- improve information retrieval.
Outline

- Background
- Theoretical Results
- Computational Results
Outline

- Background
- Theoretical Results
- Computational Results
Outline

- Background
- Theoretical Results
- Computational Results
Counting Representation

Football  Basketball

Encode

d d d d d d d

\[
\begin{pmatrix}
t & 2 & 1 & 4 & 1 & 0 & 1 \\
2 & 1 & 5 & 2 & 3 & 0 & 0 \\
2 & 0 & 0 & 0 & 1 & 2 & 5 \\
2 & 0 & 0 & 0 & 0 & 0 & 3 & 1 \\
\end{pmatrix}
\]
Indicator Representation

Football  Basketball

\[
\begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}
\]
Singular Value Decomposition

\[
\begin{bmatrix}
2 & 1 & 4 & 1 & 0 & 1 \\
1 & 5 & 2 & 3 & 0 & 0 \\
0 & 0 & 0 & 1 & 2 & 5 \\
0 & 0 & 0 & 0 & 3 & 1 \\
\end{bmatrix}
= 
\begin{bmatrix}
\end{bmatrix}
\begin{bmatrix}
\end{bmatrix}
\begin{bmatrix}
\end{bmatrix}
\]

\[A = U\Sigma V^T\]
Singular Values

\[
A = U(\Sigma)V^T
\]
Latent Semantic Indexing

\[ A_2 = U_2 \Sigma_2 V_2^T \] is the best approximation among all rank 2 matrices

- Compresses
- Increases Information Retrieval performance (empirically)
Latent Semantic Indexing

$A_2 = U_2 \Sigma_2 V_2^T$ is the best approximation among all rank 2 matrices

- Compresses
- Increases Information Retrieval performance (empirically)
Latent Semantic Indexing

\[ A_2 = U_2 \Sigma_2 V_2^T \] is the best approximation among all rank 2 matrices

- Compresses
- Increases Information Retrieval performance (empirically)
Defining Success

Measure all *intratopic* (same) and *intertopic* (different) angles in the reduced space to calculate $O(\varepsilon)$-skewness

**Same Topics**

less than $O(\varepsilon)$

**Different Topics**

more than $1-O(\varepsilon)$
Reduced Rank SVD

\[ A_k = U_k \Sigma_k V_k^T \]

- **Question**: How can a rank reduction increase Information Retrieval performance (in terms of skewness)?
  - *Use probability models to investigate.*
Reduced Rank SVD

\[ A_k = U_k \Sigma_k V_k^T \]

- **Question**: How can a rank reduction increase Information Retrieval performance (in terms of skewness)?
- *Use probability models to investigate.*
Document Generation Model

Creating a Document

\{T\}

Topic Generator
Document Generation Model

Creating a Document

\{T\} \rightarrow \text{Topic Generator} \rightarrow \{T,L\} \rightarrow \text{Length Generator}
Document Generation Model

Creating a Document

\{T\} \quad \{T,L\}

Topic Generator \quad Length Generator \quad Term Generator

Sample L terms from T
Document Generation Model

Creating a Document

\{T\} \rightarrow \{T,L\}

Topic Generator \rightarrow \text{Length Generator} \rightarrow \text{Term Generator}

Sample L terms from T

Update Matrix

\begin{bmatrix}

\end{bmatrix}
Document Generation Model

Creating a Document

\{T\} \rightarrow \text{Topic Generator} \rightarrow \text{Length Generator} \rightarrow \text{Term Generator} \rightarrow \{T,L\} \rightarrow \text{Sample } L \text{ terms from } T

Update Matrix

[ ] [ ]
Document Generation Model

Creating a Document

\{T\} \quad \{T,L\}

Topic Generator \quad Length Generator \quad Term Generator

Update Matrix

\[
\begin{bmatrix}
| & \\
| & \\
\end{bmatrix}
\]
Document Generation Model

Creating a Document

\{T\} \rightarrow \text{Topic Generator} \rightarrow \{T,L\} \rightarrow \text{Length Generator} \rightarrow \text{Term Generator} \rightarrow \text{Sample } L \text{ terms from } T

Update Matrix

\begin{bmatrix}
1 & 1 & 1 & \cdots \\
1 & 1 & 1 & \cdots \\
1 & 1 & 1 & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
1 & 1 & 1 & \cdots \\
\end{bmatrix}
Pure Document Generation Models

A model is **pure** when each document is on a single topic

<table>
<thead>
<tr>
<th></th>
<th>(d_1)</th>
<th>(d_2)</th>
<th>(d_3)</th>
<th>(d_4)</th>
<th>(d_5)</th>
<th>(d_6)</th>
<th>(d_7)</th>
<th>(d_8)</th>
<th>(d_9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t_1)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_2)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_3)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_4)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_5)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_6)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_7)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(t_8)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(t_9)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
T_1 = \{t_1, t_2, t_3\}
\]

\[
T_2 = \{t_4, t_5, t_6\}
\]

\[
T_3 = \{t_7, t_8, t_9\}
\]
Pure Document Generation Models

A model is **pure** when each document is on a single topic

<table>
<thead>
<tr>
<th></th>
<th>(d_1)</th>
<th>(d_2)</th>
<th>(d_3)</th>
<th>(d_4)</th>
<th>(d_5)</th>
<th>(d_6)</th>
<th>(d_7)</th>
<th>(d_8)</th>
<th>(d_9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t_1)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_2)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_3)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_4)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_5)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_6)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(t_7)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(t_8)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(t_9)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Theorem (Papadimitriou et al.)**

*The rank-*\(k\)* SVD is *0-skewed with probability* \(1 - O\left(\frac{1}{m}\right)\) *when performed on any* \(t \times m\) *matrix drawn from a pure document generation model with* \(k\) *topics.*
Pure Document Generation Models

Perturbing the matrix $A$ with $\|E\|_F = O(\varepsilon)$

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
\end{bmatrix}
\quad +
\begin{bmatrix}
0 & 0 & 0 & 0 & \underline{1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \underline{1} & 0 \\
0 & 0 & 0 & \underline{1} & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & \underline{1} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \underline{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \underline{1} & 0 & 0 \\
\underline{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \underline{1} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \underline{1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]
Pure Document Generation Models

\[
\begin{bmatrix}
1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
\end{bmatrix}
\]

Theorem (Papadimitriou et al.)

If \( A \) is created from a pure model and \( \| E \|_F = O(\varepsilon) \) then the rank-\( k \) SVD on \( A + E \) is \( O(\varepsilon) \)-skewed.
Papadimitriou et al.

Theorem (Papadimitriou et al.)

*If $A$ is created from a pure model and $\|E\|_F = O(\varepsilon)$ then the rank-$k$ SVD on $A + E$ is $O(\varepsilon)$-skewed*

- **Same Topics**
  - less than $O(\varepsilon)$

- **Different Topics**
  - more than $1 - O(\varepsilon)$
However, things can go wrong...

\[
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 12 & 0 & 3 & 22 & 11 & 21 & 14 & 16 \\
0 & 0 & 0 & 5 & 60 & 8 & 7 & 0 & 20 & 41
\end{pmatrix}
\]

\[T'_2\text{\text{'}s\ eigenvalues} = \{6228.43, 1241.57\}\]

\[T'_1\text{\text{'}s\ eigenvalues} = \{2.62, 0.38\}\]

For a rank–2 decomposition, all singular values of $T_1$ are outweighed by the largest two singular values of $T_2$!
However, things can go wrong, even when we normalize the matrix!

\[
\begin{bmatrix}
0.71 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.71 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 1.00 & 0.00 & 0.05 & 0.94 & 0.84 & 1.00 & 0.57 & 0.36 \\
0.00 & 0.00 & 0.00 & 1.00 & 1.00 & 0.34 & 0.54 & 0.00 & 0.82 & 0.93
\end{bmatrix}
\]

\[T'_2\text{s eigenvalues} = \{2.4, 1.5\}\]

\[T'_1\text{s eigenvalues} = \{1.3, 0.5\}\]

For a rank–2 decomposition, all singular values of $T_1$ are outweighed by the largest two singular values of $T_2$!
Counterexamples to the theorem

Singular values depend on:

- topic probabilities,
- document lengths,
- topic sizes,
- representation!
Counterexamples to the theorem

Singular values depend on:

- topic probabilities,
- document lengths,
- topic sizes,
- representation!
Counterexamples to the theorem

Singular values depend on:
- topic probabilities,
- document lengths,
- topic sizes,
- representation!
Counterexamples to the theorem

Singular values depend on:
- topic probabilities,
- document lengths,
- topic sizes,
- representation!
How to fix this problem?

- Calculate the expected singular values according to the document generation model’s topics
How to fix this problem?

So, instead of retaining the largest

\[
\begin{bmatrix}
& \\
& \\
\end{bmatrix}
= 
\begin{bmatrix}
& \\
& \\
\end{bmatrix}
\begin{bmatrix}
& \\
& \\
\end{bmatrix}
\begin{bmatrix}
& \\
& \\
\end{bmatrix}
\]
How to fix this problem?

**Predict** and **select** the singular values *according to each topic*.

\[
\begin{bmatrix}
\end{bmatrix}
= \begin{bmatrix}
\end{bmatrix} \begin{bmatrix}
\end{bmatrix} \begin{bmatrix}
\end{bmatrix} \begin{bmatrix}
\end{bmatrix}
\]
Expectation\textsuperscript{1}

Predict singular values using the expectation operator $E$ (which is linear).

$$E(X + Y) = E(X) + E(Y)$$
$$E(cX) = cE(X)$$

\textsuperscript{1}In probability theory the expected value of a discrete random variable is the sum of the probability of each possible outcome of the experiment multiplied by the outcome value.
Singular Values

\[ A = U \Sigma V^T \]

\[ \Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \]
Singular Values

\[ A = U \Sigma V^T \]

\[ \Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix} \]
Singular Values

\[ \Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix} \]

\[ \{ \sigma_i \} = \sqrt{\text{eigenvalues}(AA^T)} \]
\[ = \sqrt{\text{eigenvalues}(A^TA)} \]
\[ = \sqrt{\text{roots of } \det(A^TA - \lambda I)} \] (1)
Singular Values

\[ \Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \]

\[ \{\sigma_i\} = \sqrt{\text{eigenvalues}(AA^T)} \]

\[ = \sqrt{\text{eigenvalues}(A^TA)} \]

\[ = \sqrt{\text{roots of } (\det(A^TA - \lambda I))} \]
Singular Values

\[
\{ \sigma_i \} = \sqrt{\text{roots of } \det(A^T A - \lambda I)}
\]

\[
\det(A^T A - \lambda I) = \begin{vmatrix}
(d_1 \cdot d_1 - \lambda) \\
\cdot \\
(d_m \cdot d_m - \lambda)
\end{vmatrix}
\]

\[
(-\lambda)^m + (-\lambda)^{m-1} \sum_{i=1}^{m} A^T A_{ii} \cdots + (-\lambda)^{m-t} \sum_{t \times t \text{ psm } \alpha} \det(\alpha)
\]
Singular Values

\[
\{ \sigma_i \} = \sqrt{\text{roots of } \det(A^T A - \lambda I)}
\]

\[
\det(A^T A - \lambda I) = \begin{vmatrix}
(d_1 \cdot d_1 - \lambda) & \cdots & \\
& \cdots & \\
& & (d_m \cdot d_m - \lambda)
\end{vmatrix}
\]

\[
(-\lambda)^m + (-\lambda)^{m-1} \sum_{i=1}^{m} A^T A_{ii} \cdots + (-\lambda)^{m-t} \sum_{t \times t \text{ psm } \alpha} \det(\alpha)
\]
Singular Values

\[ \{ \sigma_i \} = \sqrt{\text{roots of } \det(A^T A - \lambda I)} \]

\[ \det(A^T A - \lambda I) = \begin{vmatrix} (d_1 \cdot d_1 - \lambda) \\ \vdots \\ (d_m \cdot d_m - \lambda) \end{vmatrix} \]

\[ (-\lambda)^m + (-\lambda)^{m-1} \sum_{i=1}^{m} A^T A_{ii} \cdots + (-\lambda)^{m-t} \sum_{\text{t x t psm } \alpha} \det(\alpha) \]
Random Gram$^2$ Matrices

Let $w$ be a $t$-tall vector whose components $w_i$ are random variables (not necessarily independent)

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_t \end{bmatrix}.$$

Next, sample $n$ independent vectors $w^{(1)}, \ldots, w^{(n)}$; creating a $t \times n$ matrix $B$. Then the random gram matrix of size $n$ is

$$G_n = B^T B,$$

and has a distribution that depends on the underlying distribution of the random vector $w$.

$^2$Jørgen Pedersen Gram (1850-1916) was a Danish Mathematician.
Singular Values

\[
\det(A^T A - \lambda I) = (-\lambda)^m + (-\lambda)^{m-1} \sum_{i=1}^{m} A^T A_{ii} \cdots + (-\lambda)^{m-t} \sum_{t \times t \text{ psm } \alpha} \det(\alpha)
\]

\[
a_i = E(\det(G_i))
\]

\[
E(\det(A^T A - \lambda I)) = a_0 \binom{m}{0} (-\lambda)^m + a_1 \binom{m}{1} (-\lambda)^{m-1} + a_2 \binom{m}{2} (-\lambda)^{m-2} + \cdots + a_t \binom{m}{t} (-\lambda)^{m-t}
\]
Singular Values

\[ \det(A^T A - \lambda I) = (-\lambda)^m + (-\lambda)^{m-1} \sum_{i=1}^{m} A^T A_{ii} \cdots + (-\lambda)^{m-t} \sum_{t \times t \text{ psm } \alpha} \det(\alpha) \]

\[ a_i = E(\det(G_i)) \]

\[ E(\det(A^T A - \lambda I)) = a_0 \binom{m}{0} (-\lambda)^m + a_1 \binom{m}{1} (-\lambda)^{m-1} + a_2 \binom{m}{2} (-\lambda)^{m-2} + \cdots + a_t \binom{m}{t} (-\lambda)^{m-t} \]
Singular Values

\[
\text{det}(A^T A - \lambda I) = (-\lambda)^m + (-\lambda)^{m-1} \sum_{i=1}^{m} A^T A_{ii} \cdots + (-\lambda)^{m-t} \sum_{t \times t \text{ psm } \alpha} \text{det}(\alpha)
\]

\[
a_i = E(\text{det}(G_i))
\]

\[
E(\text{det}(A^T A - \lambda I)) = a_0 \binom{m}{0} (-\lambda)^m + a_1 \binom{m}{1} (-\lambda)^{m-1} + a_2 \binom{m}{2} (-\lambda)^{m-2} + \cdots + a_t \binom{m}{t} (-\lambda)^{m-t}
\]
E(det(G_n)) Recursion

\[ a_0 \binom{m}{0} (-\lambda)^m + a_1 \binom{m}{1} (-\lambda)^{m-1} + a_2 \binom{m}{2} (-\lambda)^{m-2} + \cdots + a_t \binom{m}{t} (-\lambda)^{m-t} \]

GOAL: efficient recursive calculation for the expected characteristic coefficients \( a_n = E(det(G_n)) \)

Characteristic coefficients:

- are used in computational chemistry
- give an algebraic solution of the inverse metric problem in general relativity.
E(\text{det}(G_n)) \) Recursion

\[ a_0 \binom{m}{0} (-\lambda)^m + a_1 \binom{m}{1} (-\lambda)^{m-1} + a_2 \binom{m}{2} (-\lambda)^{m-2} + \cdots + a_t \binom{m}{t} (-\lambda)^{m-t} \]

GOAL: efficient recursive calculation for the expected characteristic coefficients \( a_n = E(\text{det}(G_n)) \)

Characteristic coefficients:

- are used in computational chemistry
- give an algebraic solution of the inverse metric problem in general relativity.
Generating Functions

A generating function is a clothesline on which we hang up a sequence of numbers for display. ³

What is the generating function for a sequence of \( n \) ones?

\[
1, 1, 1, 1, \ldots
\]

\[
\frac{1-x^n}{1-x} = 1 + x^1 + x^2 + x^3 + \cdots + x^{n-1}
\]

\[
1 - x^n = (1 - x)(1 + x^1 + x^2 + x^3 + \cdots + x^{n-1})
\]

\[
= (1 + x^1 + x^2 + \cdots + x^{n-1}) - (x^1 + x^2 + x^3 + \cdots + x^n)
\]

\[
= 1 - x^n
\]

So,

\[
\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \quad (|x| < 1)
\]

E(\text{det}(G_n)) \text{ Exponential Generating Function}

Based on Euler’s\textsuperscript{4} infinite expansion

$$\exp \{ x \} = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

and the Leibniz\textsuperscript{5} formula for the determinant,

$$\text{det}(G_n) = \sum_{\sigma \in S_n} (-1)^{\text{sgn}(\sigma)} \prod_{i=1}^{n} (G_n)_{i,\sigma(i)},$$

we have proven the following exponential generating function equation for $a_n = E(\text{det}(G_n))$

$$\sum_{n=0}^{\infty} a_n \frac{x^n}{n!} = \exp \left\{ \frac{c_1 x}{1} - \frac{c_2 x^2}{2} + \frac{c_3 x^3}{3} - \ldots \right\}. \quad (2)$$

(the $c_i$’s will be defined shortly)

\textsuperscript{4}Leonhard Euler (1707-1783). A Swiss Mathematician and Physicist

\textsuperscript{5}Gottfried Leibniz: (1646-1716) A German Mathematician and Natural Philosopher
E(\det(G_n)) Recursion

\begin{equation}
\sum_{n=0}^{\infty} a_n \frac{x^n}{n!} = \exp \left\{ \frac{c_1 x}{1} - \frac{c_2 x^2}{2} + \frac{c_3 x^3}{3} - \ldots \right\}. \tag{3}
\end{equation}

Differentiate both sides to help discover a recursion...

\begin{equation}
\sum_{n=0}^{\infty} a_{n+1} \frac{x^n}{n!} = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!} \left( c_1 - c_2 x + c_3 x^2 - \ldots \right). \tag{4}
\end{equation}

By equating the coefficients of \( \frac{x^n}{n!} \) on the left and right hand side, we find:

\begin{align*}
a_0 &= 1 \\
ad n+1 &= \sum_j \binom{n}{j} (-1)^j j! a_{n-j} c_{j+1}
\end{align*}
What about the $c_i$???

Define,

$$M_{ij} = E(w_i w_j)$$

Then,

$$c_j = \text{trace}(M^j) = \lambda_1^j + \cdots + \lambda_t^j$$
Representations

\[
\begin{bmatrix}
d & d & d & d & d & d \\
t & 1 & 1 & 1 & 1 & 0 & 1 \\
t & 1 & 1 & 1 & 1 & 0 & 0 \\
t & 0 & 0 & 0 & 1 & 1 & 1 \\
t & 0 & 0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
\text{ vs }
\begin{bmatrix}
d & d & d & d & d & d \\
t & 2 & 1 & 4 & 1 & 0 & 1 \\
t & 1 & 5 & 2 & 3 & 0 & 0 \\
t & 0 & 0 & 0 & 1 & 2 & 5 \\
t & 0 & 0 & 0 & 0 & 3 & 1 \\
\end{bmatrix}
\]

\[c_j = \text{trace}(M^j)\]
\[a_0 = 1\]
\[a_{n+1} = \sum_j \binom{n}{j} (-1)^j j! a_{n-j} c_{j+1}\]

To get the expected characteristic coefficients, we just need to create \(M_{ij} = E(w_i w_j)\) according to the representation used.
0, 1 indicator function

\[ A_{ij} = \begin{cases} 
0, & \text{if } \text{term}_i \text{ is not present in } \text{doc}_j \\
1, & \text{if } \text{term}_i \text{ is present in } \text{doc}_j 
\end{cases} \]

\[ M_{ii} = \sum_{t_i \geq 1} \ldots \]
\[ = 1 - (1 - p_i)^l \]

\[ M_{ij} = \sum_{t_i \geq 1} \sum_{t_j \geq 1} \]
\[ = \sum_{t_i \geq 0} - \sum_{t_i = 0} - \sum_{t_j \geq 0} - \sum_{t_j = 0} \]
\[ = 1 - (1 - p_i)^l - (1 - p_j)^l - (1 - p_i - p_j)^l \]
$-1, 1$ indicator function

$$A_{ij} = \begin{cases} -1, & \text{if } \text{term}_i \text{ is not present in } \text{doc}_j \\ 1, & \text{if } \text{term}_i \text{ is present in } \text{doc}_j \end{cases}$$

$$M_{ij} = \begin{cases} 1, & \text{if } i = j \\ 1 - 4(1 - p_i - p_j)^l - 2(1 - p_i)^l - 2(1 - p_j)^l, & \text{if } i \neq j \end{cases}$$
Counting function

\[ A_{ij} = \text{number of times term}_i \text{ appears in doc}_j \]

\[ M_{ij} = \begin{cases} 
\sum_l (l(l-1)p_i^2 + lp_i) \cdot \text{prob}(l), & i = j \\
\sum_l (l(l-1)p_ip_j) \cdot \text{prob}(l), & i \neq j 
\end{cases} \]
Complexity Sketch

\[ a_0 \binom{m}{0} (-\lambda)^m + a_1 \binom{m}{1} (-\lambda)^{m-1} + a_2 \binom{m}{2} (-\lambda)^{m-2} + \cdots + a_t \binom{m}{t} (-\lambda)^{m-t} \]

\[ a_0 = 1 \]

\[ a_{n+1} = \sum_j \binom{n}{j} (-1)^j j! a_{n-j} c_{j+1} \]

- \( t \) coefficients to compute
- The \( c_j = \text{trace}(M^j) \) can be parallelized (hypercube) down to \( O(t) \)
Complexity Sketch

Nice properties of our algorithm:

- Simple and easily parallelized
- Doesn’t depend on the number of documents (full SVD is $O(m^2n + n^2m + n^3)$ with Golub-Reinsch algorithm)
- Reduced computation of characteristic coefficients to a recursion involving symmetric matrix powers.
- Accuracy increases as more documents are included
Singular Value Calculation Times

Read matrix, infer model, calculate expected coeffs, get roots

<table>
<thead>
<tr>
<th></th>
<th>32 x 500,000</th>
<th>32 x 1,000,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maple</td>
<td>580.63 secs</td>
<td>1211.59 secs</td>
</tr>
<tr>
<td>Selective</td>
<td>4.10 secs</td>
<td>10.47 secs</td>
</tr>
</tbody>
</table>

Table: Singular value calculation times of two random matrices for both Maple and Selective (not parallelized)
Information Retrieval Comparison

Compare retaining the largest

\[
\begin{bmatrix}
\end{bmatrix}
= 
\begin{bmatrix}
\end{bmatrix}
\begin{bmatrix}
\end{bmatrix}
\begin{bmatrix}
\end{bmatrix}
\]
Information Retrieval Comparison

To selecting the singular values according to each topic.

\[
\begin{bmatrix}
\end{bmatrix}
= 
\begin{bmatrix}
\end{bmatrix}
\begin{bmatrix}
\end{bmatrix}
\begin{bmatrix}
\end{bmatrix}
\]
Information Retrieval: (Real world sports articles)

Table: Comparison of our algorithm to traditional latent semantic indexing.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Algorithm</th>
<th>Intratopic Cosine</th>
<th>Intertopic Cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Selective</td>
<td>0.998</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>LSI</td>
<td>0.767</td>
<td>0.620</td>
</tr>
<tr>
<td>4</td>
<td>Selective</td>
<td>0.821</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>LSI</td>
<td>0.625</td>
<td>0.433</td>
</tr>
<tr>
<td>8</td>
<td>Selective</td>
<td>0.653</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>LSI</td>
<td>0.673</td>
<td>0.007</td>
</tr>
</tbody>
</table>
### Information Retrieval: (Probabilistic documents)

<table>
<thead>
<tr>
<th>Rank 2</th>
<th>Algorithm</th>
<th>Intratopic cosine</th>
<th>Intertopic cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Selective</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>LSI</td>
<td>0.46</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Table:** Average cosine over 100 different groups of 1000 random documents generated from a probabilistic document generation model containing two topics with probability 99% and 1%.
Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors
Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors
Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors
Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors
Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors
Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
  - Expected singular vectors
Future

- Compare IR performance across representations
- Relationships between singular values and:
  - Probability Models
  - Representations
- Generatingfunctionology for fast and accurate calculations
- Refine singular values
- Expected singular vectors
Conclusion

- New theoretical results concerning SVD
- More speed
- Better quality reduced rank approximation
- Might help with the “choice–of–rank conundrum”
Conclusion

- New theoretical results concerning SVD
- More speed
- Better quality reduced rank approximation
- Might help with the “choice–of–rank conundrum”
Conclusion

- New theoretical results concerning SVD
- More speed
- Better quality reduced rank approximation
- Might help with the “choice–of–rank conundrum”
Conclusion

- New theoretical results concerning SVD
- More speed
- Better quality reduced rank approximation
- Might help with the “choice–of–rank conundrum”
Contact Information

Jacob Martin
jacobgmartin@gmail.com
http://www.cs.tcd.ie/Jacob.Martin