A Randomized Algorithm for Approximating the Log Determinant of a Symmetric Positive Definite Matrix

Eugenia-Maria Kontopoulou

in collaboration with

C. Boutsidis, P. Drineas, P. Kambadur & A. Zouzias

Department of Computer Science
Purdue University

CSESC 2017 Conference
Purdue, April 2017
Definition
Given a Symmetric Positive Definite matrix $A \in \mathbb{R}^{n \times n}$, compute (exactly or approximately) the log det($A$).

Application: Maximum likelihood estimations, Gaussian processes prediction, log det-divergence metric, barrier functions in interior point methods . . .

Straightforward Computation

1. Compute the Cholesky Factorization of $A$, and let $L$ be the Cholesky factor.
2. Compute the log-determinant of $A$ using $L$:

$$\log \det(A) = \log \det(L)^2 = 2 \log \prod_{i=1}^{n} L_{ii}.$$ 

Time Complexity: $O(n^3)$. 

Prohibitive for Large Data!!!!
Lemma
Let $A \in \mathbb{R}^{n \times n}$ be an SPD matrix. For any $\alpha$ with $\lambda_1(A) < \alpha$, define $B = A / \alpha$ and $C = I_n - B$. Then,

$$\log \det(A) = n \log(\alpha) - \sum_{k=1}^{\infty} \frac{\text{Tr}(\log(C^k))}{k}. $$

Algorithm 1

Input: $A \in \mathbb{R}^{n \times n}$, accuracy parameter $\epsilon > 0$, integer $m > 0$.

Output: $\hat{\log \det(A)}$, the approximation to the $\log \det(A)$.

1. Compute an estimate to the largest eigenvalue of $A$, $\tilde{\lambda}_1(A)$, using the Power Method.
2. $C = I_n - A / (7 \tilde{\lambda}_1(A))$
3. Create $p = \lceil 20 \log(2/\delta)/\epsilon^2 \rceil$ i.i.d random Gaussian vectors, $g_1, g_2, \ldots, g_p$.
4. Estimate $\Delta_2$ with a truncated Taylor Series type randomized trace estimator that computes:

$$\Delta_2 \approx \sum_{k=1}^{m} \left( \frac{1}{p} \sum_{i=1}^{p} g_i^\top C^k g_i \right).$$
Lemma

Let $\log \det(A)$ be the log det approximation of the above procedure. Then, we prove that with probability at least $1 - 2\delta$,

$$|\hat{\log \det(A)} - \log \det(A)| \leq 2\epsilon \Gamma$$

where $\Gamma = \sum_{i=1}^{n} \log \left(7 \cdot \frac{\lambda_1(A)}{\lambda_i(A)}\right)$ and $m \geq \lceil 7 \kappa(A) \log \left(\frac{1}{\epsilon}\right) \rceil$.

Running Time

$$\mathcal{O} \left( nnz(A) \cdot \left( \frac{m}{\epsilon^2} + \log n \right) \cdot \log \left( \frac{1}{\delta} \right) \right).$$
Relative Error Approximation

The Algorithm

Let $A \in \mathbb{R}^{n \times n}$ be an SPD matrix whose eigenvalues lie in the interval $(\theta_1, 1)$, for some $0 < \theta_1 < 1$. Let $C = I_n - A$. Then,

$$\log \det(A) = -\sum_{k=1}^{\infty} \frac{\text{Tr} \left( \log(C^k) \right)}{k \Delta}.$$ 

Algorithm 2

Input: $A \in \mathbb{R}^{n \times n}$, accuracy parameter $\epsilon > 0$, integer $m > 0$.

Output: $\hat{\log \det(A)}$, the approximation to the $\log \det(A)$.

1. $C = I_n - A$
2. Create $p = \lceil 20 \log(2/\delta)/\epsilon^2 \rceil$ i.i.d random Gaussian vectors, $g_1, g_2, \ldots, g_p$.
3. Estimate $\Delta$ with a truncated Taylor Series type randomized trace estimator as:

$$\Delta \approx \sum_{k=1}^{m} \left( \frac{1}{p} \sum_{i=1}^{p} g_i^\top C^k g_i \right).$$
Lemma

Let $\hat{\log \det(A)}$ be the log det approximation of the above procedure on inputs $A$ and $\epsilon$. Then, we prove that with probability at least $1 - \delta$, 

$$|\hat{\log \det(A)} - \log \det(A)| \leq 2\epsilon \cdot |\log \det(A)|$$

and $m \geq \left\lceil \frac{1}{\theta_1} \cdot \log\left(\frac{1}{\epsilon}\right) \right\rceil$.

Running Time

$$O\left(\frac{\log(1/\epsilon) \log(1/\delta)}{\epsilon^2 \theta_1} \cdot \text{nnz}(A)\right).$$
## Experiments I

**Dense Random Matrices**

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\log \det(A)$</th>
<th>$\log \det(A)$</th>
<th>$\log \det(A)$</th>
<th>$\log \det(A)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>exact</td>
<td>mean</td>
<td>std</td>
<td>exact</td>
</tr>
<tr>
<td>5000</td>
<td>-3717.89</td>
<td>-3546.920</td>
<td>8.10</td>
<td>2.56</td>
</tr>
<tr>
<td>7500</td>
<td>-5474.49</td>
<td>-5225.152</td>
<td>8.73</td>
<td>7.98</td>
</tr>
<tr>
<td>10000</td>
<td>-7347.33</td>
<td>-7003.086</td>
<td>7.79</td>
<td>18.07</td>
</tr>
<tr>
<td>12500</td>
<td>-9167.47</td>
<td>-8734.956</td>
<td>17.43</td>
<td>34.39</td>
</tr>
<tr>
<td>15000</td>
<td>-11100.9</td>
<td>-10575.16</td>
<td>15.09</td>
<td>58.28</td>
</tr>
</tbody>
</table>

**Table:** Parameters: $p = 60$, $m = 4$, $t = \log(\sqrt{4n})$. Ground truth computed via Cholesky. Mean and standard deviation reported over 10 repetitions.
Figure: Parameters: $p = 60$ and $t = 2 \log \sqrt{4n}$. Ground truth computed using Cholesky.
### Experiments II
#### Real Sparse Matrices
University of Florida Sparse Matrix Collection

<table>
<thead>
<tr>
<th>name</th>
<th>$n$</th>
<th>$nnz$</th>
<th>log det $A$</th>
<th>time (sec)</th>
<th>$m$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>exact</td>
<td>approx</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mean</td>
<td>std</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>exact</td>
<td>approx</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>thermal2</td>
<td>1228045</td>
<td>8580313</td>
<td>1.3869e6</td>
<td>1.3928e6</td>
<td>964.79</td>
</tr>
<tr>
<td>ecology2</td>
<td>999999</td>
<td>4995991</td>
<td>3.3943e6</td>
<td>3.403e6</td>
<td>1212.8</td>
</tr>
<tr>
<td>ldoor</td>
<td>952203</td>
<td>42493817</td>
<td>1.4429e7</td>
<td>1.4445e7</td>
<td>1683.5</td>
</tr>
<tr>
<td>thermomech_TC</td>
<td>102158</td>
<td>711558</td>
<td>-546787</td>
<td>-546829.4</td>
<td>553.12</td>
</tr>
<tr>
<td>boneS01</td>
<td>127224</td>
<td>5516602</td>
<td>1.1093e6</td>
<td>1.106e6</td>
<td>247.14</td>
</tr>
</tbody>
</table>

Table: Parameters: $p = 5$, $m = 1 : 5 : 150$ and select the one with best avg, $t = 5$). Ground truth computed via Cholesky. Mean reported over 10 repetitions.
Thank you!

Questions?
C. Boutsidis, P. Drineas, P. Kambadur, E. Kontopoulou, A. Zouzias (2016), ‘‘A Randomized Algorithm for Approximating the Log Determinant of a Symmetric Positive Definite Matrix’’, submitted to LAA