Towards Randomized Algorithms for the Estimation of Log-Determinants and Von-Neumann Entropies

Eugenia-Maria Kontopoulou

Department of Computer Science
Purdue University

TCS Seminars

Purdue, October 2017

Definition

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

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 $\mathbf{A}$, and let $\mathbf{L}$ be the Cholesky factor.
2. Compute the log-determinant of $\mathbf{A}$ using $\mathbf{L}$:

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

Time Complexity: $\mathcal{O}(n^3)$.

Prohibitive for Large Data!!!!

Mathematical Manipulation of *logdet* $(\mathbf{A})$ I

Theorem

Any symmetric (hermitian) matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$ $(\mathbb{C}^{n \times n})$:

1. has only real eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_n$;
2. has orthogonal eigenvectors, $\mathbf{U}$;
3. is always diagonalizable : $\mathbf{A} = \mathbf{U}\,\Lambda\,\mathbf{U}^\top$.

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

$$\begin{aligned} \mathit{logdet}\,(\mathbf{A}) &= \mathit{logdet}\left(\mathbf{U}\,\Lambda\,\mathbf{U}^\top\right) \\ &= \log(\det(\Lambda)) \\ &= \log\left(\prod_{i=1}^{n} \lambda_i\right) \\ &= \sum_{i=1}^{n} \log(\lambda_i) \\ &= \mathbf{Tr}\,(\log(\mathbf{A})) \end{aligned}$$

Issues

1. Computing the trace is easy.
2. Computing $\log \mathbf{A}$ costs $\mathcal{O}\left(n^3\right)$.

Solution

Further manipulation of $\mathbf{Tr}\,(\log(\mathbf{A}))$...

Mathematical Manipulation of *logdet* $(\mathbf{A})$ II

Lemma

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ *be a symmetric matrix whose eigenvalues lie in the interval* $(-1, 1)$. *Then*

$$\log\left(\mathbf{I}_n - \mathbf{A}\right) = -\sum_{k=1}^{\infty} \frac{\mathbf{A}^k}{k}.$$

$$\begin{aligned}
\mathbf{Tr}\left(\log(\mathbf{A})\right) &= \mathbf{Tr}\left(\log(\mathbf{I}_n - \mathbf{I}_n + \mathbf{A})\right) \\
&= \mathbf{Tr}\left(\log(\mathbf{I}_n - \underbrace{(\mathbf{I}_n - \mathbf{A})}_{\mathbf{C}})\right) \\
&= \mathbf{Tr}\left(\log(\mathbf{I}_n - C)\right) \\
&= \mathbf{Tr}\left(-\sum_{k=1}^{\infty} \frac{\mathbf{C}^k}{k}\right) \\
&= -\sum_{k=1}^{\infty} \frac{\mathbf{Tr}\left(\mathbf{C}^k\right)}{k}
\end{aligned}$$

Issues

1. Computing the series.
2. Computing $\mathbf{Tr}\left(C^k\right)$.

Solution

1. Truncate the Taylor Series...
2. Trace Estimators!!!

Definition

A Gaussian trace estimator for a symmetric positive-definite matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is

$$\mathbf{G} = \frac{1}{p} \sum_{i=1}^{p} g_i^\top \, \mathbf{A} \, g_i,$$

where the g_i's are p independent random vectors whose entries are i.i.d. standard normal variables.

Lemma

Let $\mathbf{A}$ *be an SPD matrix in* $\mathbb{R}^{n \times n}$*, let* $0 < \epsilon < 1$ *be an accuracy parameter, and let* $0 < \delta < 1$ *be a failure probability. Then for* $p = \lceil 20 \log(2/\delta) \epsilon^{-2} \rceil$*, with probability at least* $1 - \delta$*,*

$$|\mathbf{Tr}(\mathbf{A}) - \mathbf{G}| \leq \epsilon \cdot \mathbf{Tr}(\mathbf{A}).$$

[1] H. Avron & S. Toledo (2010),"*Randomized Algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix*" (link)

Algorithm 1

Input: $A \in \mathbb{R}^{n\times n}$ with eigenvalues lie in $(\theta_1, 1)$ where $\theta_1 > 0$, accuracy parameter $\epsilon > 0$, integer $m > 0$.

Output: $\widehat{\log\det}(A)$, 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 $\widehat{\log\det}$ as:

$$\widehat{\log\det}(A) = \sum_{k=1}^{m} \left(\frac{1}{p} \sum_{i=1}^{p} g_i^\top \; C^k \; g_i \right).$$

Lemma

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

$$\left|\widehat{\log\det}(\mathbf{A}) - \log\det(A)\right| \leq 2\epsilon \cdot \left|\log\det(\mathbf{A})\right|$$

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

Running Time

$$\mathcal{O}\left(\frac{\log(1/\epsilon)\log(1/\delta)}{\epsilon^2\theta_1} \cdot nnz(\mathbf{A})\right).$$

Von-Neumann Entropy of Density Matrices I

Definition

Given a Density Matrix $\mathbf{R} \in \mathbb{R}^{n \times n}$, compute (exactly or approximately) the Von-Neumann entropy, $\mathcal{H}(R)$. A Density Matrix is represented by the statistical mixture of pure states and has the form

$$\mathbf{R} = \sum_{i=1}^{n} p_i \psi_i \psi_i^\top = \boldsymbol{\Psi} \boldsymbol{\Sigma}_p \boldsymbol{\Psi}^\top \; \in \; \mathbb{R}^{n \times n},$$

where $\psi_i \in \mathbb{R}^n$ represent the pure states of a system and are pairwise orthogonal and normal, while p_i's correspond to the probability of each state and satisfy $p_i > 0$ and $\sum_{i=1}^{n} p_i = 1$.

Application: Information theory, quantum mechanics, ...

Von-Neumann Entropy of Density Matrices II

Straightforward Computation

1. Compute the singular values of $\mathbf{R}$, $s_1, s_2, \ldots, s_n$.
2. Compute the Von-Neumann Entropy of $\mathbf{R}$ using s_i, $i = 1, \ldots, n$:

$$\mathcal{H}(\mathbf{R}) = -\sum_{i=1}^{n} s_i \log s_i.$$

Time Complexity: $\mathcal{O}(n^3)$.

Prohibitive for Large Data!!!!

Mathematical Manipulation of $\mathcal{H}(\mathbf{R})$ I

Assume the function $h(x) = x \log x \in \mathbb{R}$.

$$
\begin{aligned}
h(\mathbf{R}) &= \mathbf{R} \log \mathbf{R} \\
&= \Psi \Sigma_p \Psi^\top \log(\Psi \Sigma_p \Psi^\top) \\
&= \Psi \Sigma_p \log(\Sigma_p) \Psi^\top
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}(\mathbf{R}) &= -\sum_{pi} p_i \log p_i \\
&= -\mathbf{Tr}\left(h(\Sigma_p)\right) \\
&= -\mathbf{Tr}\left(\Psi \Psi^\top h(\Sigma_p)\right) \\
&= -\mathbf{Tr}\left(\Psi h(\Sigma_p) \Psi^\top\right) \\
&= -\mathbf{Tr}\left(\mathbf{R} \log \mathbf{R}\right) \\
&= -\mathbf{Tr}\left(h(\mathbf{R})\right)
\end{aligned}
$$

Using a Taylor expansion for the logarithm we can further manipulate $\mathcal{H}(\mathbf{R})$.

Mathematical Manipulation of $\mathcal{H}\left(\mathbf{R}\right)$ II

Lemma

Let $\mathbf{R} \in \mathbb{R}^{n \times n}$ be a symmetric positive definite matrix with unit trace and whose eigenvalues lie in the interval $[\ell, u]$, for some $0 < \ell \leq u \leq 1$. Then,

$$\mathcal{H}\left(\mathbf{R}\right) \approx \log u^{-1} + \underbrace{\sum_{k=1}^{\infty} \frac{\mathbf{Tr}\left(\mathbf{R}(\mathbf{I} - u^{-1}\mathbf{R})^{k}\right)}{k}}_{\Delta}.$$

We estimate the trace of $\mathbf{R}(\mathbf{I} - u^{-1}\mathbf{R})^{k}$ using Gaussian trace estimator and Δ by truncating the Taylor series.

Algorithm 2

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

Output: $\widehat{\mathcal{H}(R)}$, the approximation to the $\mathcal{H}(R)$.

1. Compute $\hat{p}_1$, the estimation of the largest singular value of R, using power method.
2. Set $u = \min\{1, 6\hat{p}_1\}$
3. $C = I_n - u^{-1}R$
4. Create $p = \lceil 20 \log(2/\delta)/\epsilon^2 \rceil$ i.i.d random Gaussian vectors, $g_1, g_2, \ldots, g_p$.
5. Compute $\widehat{\mathcal{H}(R)}$ as:

$$\widehat{\mathcal{H}(R)} = \log u^{-1} + \frac{1}{p} \sum_{i=1}^{p} \sum_{k=1}^{m} \frac{g_i \; R \; C^k \; g_i}{k}.$$

Lemma

Let $\mathbf{R}$ *be a density matrix such that all probabilities* p_i, $i = 1 \ldots n$ *satisfy* $0 < \ell \leq p_i$. *Let* u *be computed as in Algorithm 2 and let* $\widehat{\mathcal{H}(\mathbf{R})}$ *be the output of Algorithm 2 on inputs* $\mathbf{R}$, m, *and* $\epsilon < 1$*; Then, with probability at least* $1 - 2\delta$,

$$\left|\widehat{\mathcal{H}(\mathbf{R})} - \mathcal{H}(\mathbf{R})\right| \leq 2\epsilon\mathcal{H}(\mathbf{R}),$$

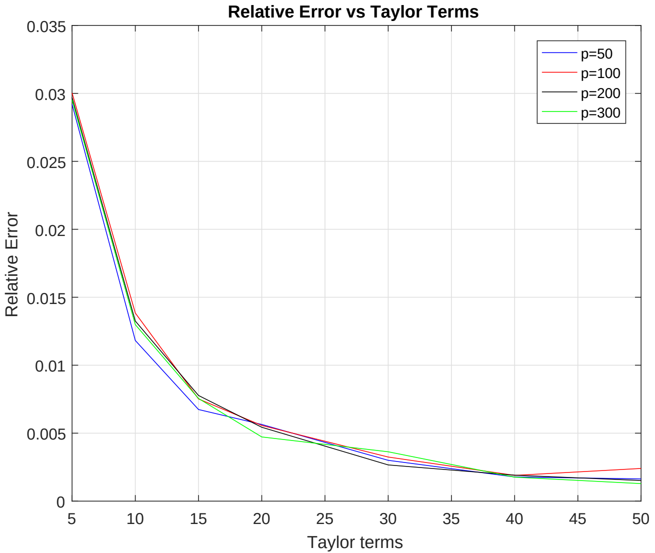
by setting $m = \left\lceil \frac{u}{\ell} \log(1/\epsilon) \right\rceil$.

Computation Time

$$\mathcal{O}\left(\frac{u}{\ell} \cdot \frac{\log(1/\epsilon)\log(1/\delta)}{\epsilon^2} \cdot nnz(\mathbf{R}) + \log n \cdot \log(1/\delta) \cdot nnz(\mathbf{R})\right).$$

Experiment

Matrix of size 5000×500, $m = \{5, 10, 15, 20, 30, 40, 50\}$, $p = \{50, 100, 200, 300\}$ and $u = 2 \cdot \lambda_{max}$.



Thank you!

Questions?

- C. Boutsidis, P. Drineas, P. Kambadur, E. Kontopoulou, A. Zouzias, *``A Randomized Algorithm for Approximating the Log Determinant of a Symmetric Positive Definite Matrix"*, in Linear Algebra and its Applications, 533:95-117, 2017,
- H. Avron. S. Toledo, *``Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix´´*, in Journal of the ACM, 58(2):8, 2011
- F. Roosta-Khorasani, U. Ascher, *``Improved bounds on sample size for implicit matrix trace estimators´´*, ArXiv, 2013