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

Eugenia-Maria Kontopoulou & Petros Drineas

Computer Science, Purdue University

LogDet Problem

Given an SPD matrix $A \in \mathbb{R}^{n \times n}$, compute (exactly or approximately) $\log\det(A)$.

Additive Error Approximation

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be an SPD matrix. For any α with $\lambda_1(\mathbf{A}) < \alpha$, define $\mathbf{B} = \mathbf{A}/\alpha$ and $\mathbf{C} = \mathbf{I}_n - B$. Then,

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

Algorithm 1

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

1. Compute an estimate to the largest eigenvalue of $\mathbf{A}$,$\tilde{\lambda_1}(\mathbf{A})$, using the Power Method.
2. $\mathbf{C} = \mathbf{I}_n - \mathbf{A}/(7\tilde{\lambda_1}(\mathbf{A}))$.
3. Create $p = \lceil 20\log(2/\delta)/\epsilon^2 \rceil$ i.i.d random Gaussian vectors, $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_p$.
4. Estimate $\sum_{k=1}^{\infty} \frac{\mathbf{Tr}\left(\log\left(\mathbf{C}^k\right)\right)}{k}$ with a truncated Taylor Series type randomized trace estimator that computes $\sum_{k=1}^{m}\left(\frac{1}{p}\sum_{i=1}^{p} \mathbf{g}_i^\top \mathbf{C}^k \mathbf{g}_i\right)$.

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

$$|\widehat{\log\det}(\mathbf{A}) - \log\det(A)| \le 2\epsilon\Gamma$$

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

Efficiency and Applications

Intensive Computational Kernel:

- Exact computation complexity $\mathcal{O}(n^3)$, **prohibitive for**

  Big Data!!!
- Approximation using Algorithm 1:

$$\mathcal{O}(nnz(A) \cdot (\mathbf{A}) \cdot (m\epsilon^{-2} + \log(n))).$$

Real world applications: Multivariate Statistics (e.g. Computation of Maximum Likelihood), Spatial-Temporal (e.g. GIS, GPS e.t.c.), Data Mining (e.g. Classification of data) e.t.c.

More RandNLA Techniques

Krylov Subspace Methods: Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a starting guess matrix $\mathbf{X} \in \mathbb{R}^{n \times s}$, we want to use the block Krylov space $\mathcal{K}_q(\mathbf{A}\mathbf{A}^\top, \mathbf{A}\mathbf{X})$ to approximate the **left singular vector space** of $\mathbf{A}$.
We prove:

- Spectral & Frobenius bounds for the distance between the approximate and the actual space.
- Quality measurements of the bounds relative to the best low-rank approximation.

Sparse Principal Component Analysis: We relax the Sparse PCA problem to a **problem with convex constraints (but non-convex objective)**:

$$\underset{\mathbf{w}}{\text{maximize}} \;\; \mathbf{w}^\top \mathbf{X}^\top \mathbf{X}\, \mathbf{w}$$
$$\text{subject to } \|\mathbf{w}\|_1 \le \sqrt{k},\; \|\mathbf{w}\|_2 \le 1,\; \mathbf{w} \in \mathbb{R}^n.$$

We **design** a two-phase algorithm that first approximates the relaxed problem and then uses a randomized rounding strategy to sparsify the approximation.

Relative Error Approximation

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

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

Algorithm 2

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

1. $\mathbf{C} = \mathbf{I}_n - \mathbf{A}$
2. Create $p = \lceil 20\log(2/\delta)/\epsilon^2 \rceil$ i.i.d random Gaussian vectors, $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_p$.
3. Estimate $\sum_{k=1}^{\infty} \frac{\mathbf{Tr}\left(\log\left(\mathbf{C}^k\right)\right)}{k}$ with a truncated Taylor Series type randomized trace estimator that computes $\sum_{k=1}^{m}\left(\frac{1}{p}\sum_{i=1}^{p} \mathbf{g}_i^\top \mathbf{C}^k \mathbf{g}_i\right)$

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 $1 - \delta$,

$$|\widehat{\log\det}(\mathbf{A}) - \log\det(A)| \le 2\epsilon \cdot |\log\det(\mathbf{A})|.$$

Experiments

Table 1: Real-World data from University of Florida Sparse Matrix Collection and C++ Parallel Implementation. (**Names:** thermal2, ecology2, ldoor, thermomech_TC, boneS01)

n	$\log\det(\mathbf{A})$			time (sec)		m
	exact	approx		exact	approx	
		mean	std		mean	
1228045	1.3869e6	1.3928e6	964.79	31.28	31.24	149
999999	3.3943e6	3.403e6	1212.8	18.5	10.47	125
952203	1.4429e7	1.4445e7	1683.5	117.91	17.60	33
102158	-546787	-546829.4	553.12	57.84	2.58	77
127224	1.1093e6	1.106e6	247.14	130.4	8.48	125

We **prove** that with probability at least $3/4$ the sparse principal component is close to the actual principal component and its 2-norm is close to 1. We **demonstrate** applications on real world data, e.g. Genomics, Text Clustering.

Citation

- 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*, under review at Journal of Linear Algebra and its Applications. ArXiv: https://arxiv.org/abs/1503.00374
- P. Drineas, I. Ipsen, E. Kontopoulou, M. Magdon-Ismail (2016), *Structural Convergence Results for Low-Rank Approximations from Block Krylov Spaces*, submitted to SIAM Journal on Matrix Analysis and Applications.ArXiv: https://arxiv.org/abs/1609.00671
- K. Fountoulakis, A. Kundu, E. Kontopoulou, P. Drineas (2016), *A Randomized Rounding Algorithm for Sparse PCA*, under review at ACM Transactions on Knowledge Discovery from Data. ArXiv: https://arxiv.org/abs/1508.03337