RandNLA approaches to estimate logarithm-based matrix functions

Eugenia-Maria Kontopoulou

SIAM Conference on Applied Linear Algebra
05/19/2021

Functions of Form

$$\boxed{f\left(\log\left(g\left(\mathbf{A}\right)\right)\right) = \gamma}$$

where $f(\cdot)$ is a matrix or scalar function, $g(\cdot)$ is a matrix function, $\mathbf{A} \in \mathbb{C}^{n\times n}$ is a PSD and $\gamma \in \mathbb{R}$.

Functions of Form

$$f\left(\log\left(g\left(\mathbf{A}\right)\right)\right) = \gamma$$

where $f(\cdot)$ is a matrix or scalar function, $g(\cdot)$ is a matrix function, $\mathbf{A} \in \mathbb{C}^{n \times n}$ is a PSD and $\gamma \in \mathbb{R}$.

Von Neumann Entropy

$$\mathcal{H}\left(\mathbf{A}\right) = -Tr\left[\mathbf{A}\log\left[\mathbf{A}\right]\right]$$

$$f(\mathbf{X}) = -Tr\left[\mathbf{X} \cdot \exp\left[\mathbf{X}\right]\right] : \mathbb{C}^{n \times n} \to \mathbb{R}$$
$$g(\mathbf{X}) = \mathbf{X} : \mathbb{C}^{n \times n} \to \mathbb{C}^{n \times n}$$

Functions of Form

$$\boxed{f\left(\log\left(g\left(\mathbf{A}\right)\right)\right) = \gamma}$$

where $f(\cdot)$ is a matrix or scalar function, $g(\cdot)$ is a matrix function, $\mathbf{A} \in \mathbb{C}^{n \times n}$ is a PSD and $\gamma \in \mathbb{R}$.

Von Neumann Entropy

$$\boxed{\mathcal{H}\left(\mathbf{A}\right) = -Tr\left[\mathbf{A}\log\left[\mathbf{A}\right]\right]}$$

$$f(\mathbf{X}) = -Tr\left[\mathbf{X} \cdot \exp\left[\mathbf{X}\right]\right] : \mathbb{C}^{n \times n} \to \mathbb{R}$$
$$g(\mathbf{X}) = \mathbf{X} : \mathbb{C}^{n \times n} \to \mathbb{C}^{n \times n}$$

Logarithm of Determinant

$$\boxed{\text{logdet}\left(\mathbf{A}\right) = \log\left(\det\left[\mathbf{A}\right]\right)}$$

$$f(x) = x : \mathbb{R} \to \mathbb{R}$$
$$g(\mathbf{X}) = \det\left[\mathbf{X}\right] : \mathbb{C}^{\mathrm{n} \times \mathrm{n}} \to \mathbb{R}$$

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, $\mathcal{H}(\mathbf{R})$. A Density Matrix is represented by the statistical mixture of pure states and has the form

$$\mathbf{R} = \sum_{i=1}^{n} p_i \mathbf{y}_i \mathbf{y}_i^\top = \mathbf{Y} \mathbf{\Sigma}_p \mathbf{Y}^\top \in \mathbb{R}^{n \times n},$$

where the vectors $\mathbf{y}_i \in \mathbb{R}^n$ 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 $\mathbf{R}$, $p_1, p_2, \ldots, p_n$ (e.g. using SVD).
2. Compute the Von-Neumann Entropy of $\mathbf{R}$ using p_i, $i = 1, \ldots, n$:

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

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

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

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

$$
\begin{aligned}
h(\mathbf{R}) &= \mathbf{R} \log \mathbf{R} \\
&= \mathbf{Y}\boldsymbol{\Sigma}_p\mathbf{Y}^\top \log(\mathbf{Y}\boldsymbol{\Sigma}_p\mathbf{Y}^\top) \\
&= \mathbf{Y}\boldsymbol{\Sigma}_p \log(\boldsymbol{\Sigma}_p)\mathbf{Y}^\top \\
&= \mathbf{Y}h(\boldsymbol{\Sigma}_p)\mathbf{Y}^\top
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{H}(\mathbf{R}) &= -\sum_i p_i \log p_i \\
&= -Tr\left[h(\boldsymbol{\Sigma}_p)\right] \\
&= -Tr\left[\mathbf{Y}\mathbf{Y}^\top h(\boldsymbol{\Sigma}_p)\right] \\
&= -Tr\left[\mathbf{Y}h(\boldsymbol{\Sigma}_p)\mathbf{Y}^\top\right] \\
&= -Tr\left[h(\mathbf{R})\right]
\end{aligned}
$$

Two Approaches

1. Using a Taylor expansion for the logarithm we can further manipulate $\mathcal{H}(\mathbf{R})$.
2. Estimating $h(\mathbf{R})$ with Chebyshev Polynomials.

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,

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

We estimate the trace of $\mathbf{R}(\mathbf{I} - u^{-1}\mathbf{R})^k$ using Gaussian trace estimator and Δ by truncation. The largest eigenvalue, u, is estimated using the power method with provable bounds.

Gaussian Trace Estimator Lemma
Power Method Lemma

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:

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

Theorem

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 *using the power method and let* $\widehat{\mathcal{H}(\mathbf{R})}$ *be the output of the algorithm above 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 $m = \lceil \frac{u}{\ell} \log(1/\epsilon) \rceil$*.*

Computation Time

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

In (Kon+20) further appears an analysis of the Taylor-based Algorithm for Hermitian density matrices.

Lemma

We can approximate $h(x) = x \log x$ *in the interval* $(0, u]$ *by*

$$f_m(x) = \sum_{t=0}^{m} \alpha_t \mathcal{T}_t(x),$$

where $\mathcal{T}_t(x) = \cos(t \cdot \arccos((2/u)x - 1))$*, the Chebyshev polynomials of first kind for* $t > 0$ *and,*

$$\alpha_0 = \frac{u}{2}\left(\log\frac{u}{4} + 1\right), \quad \alpha_1 = \frac{u}{4}\left(2\log\frac{u}{4} + 3\right), \quad \text{and} \quad \alpha_t = \frac{(-1)^t u}{t^3 - t} \text{ for } t \geq 2.$$

For any $m \geq 1$*,*

$$|h(x) - f_m(x)| \leq \frac{u}{2m(m+1)} \leq \frac{u}{2m^2},$$

for $x \in [0, u]$*. Then*

$$\widehat{\mathcal{H}(\mathbf{R})} = -Tr\left[f_m(\mathbf{R})\right] = -\frac{1}{p}\sum_{i=1}^{p} \mathbf{g}_i^\top f_m(\mathbf{R})\mathbf{g}_i$$

We estimate the trace using Gaussian trace estimator and we compute the scalars $\mathbf{g}_i^\top f_m(\mathbf{R})\mathbf{g}_i$ using the Clenshaw algorithm.

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: Create $p = \lceil 20 \log(2/\delta)/\epsilon^2 \rceil$ i.i.d random Gaussian vectors, $g_1, g_2, \ldots, g_p$.
4: Compute $\widehat{\mathcal{H}(R)}$ as:

$$\widehat{\mathcal{H}(R)} = -\frac{1}{p} \sum_{i=1}^{p} g_i^\top f_m(R) g_i.$$

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 using the power method and let* $\widehat{\mathcal{H}(\mathbf{R})}$ *be the output of the algorithm above on inputs* $\mathbf{R}$*,* m*, and* $\epsilon < 1$*; Then, with probability at least* $1 - 2\delta$*,*

$$\left|\widehat{\mathcal{H}(\mathbf{R})} - \mathcal{H}\left(\mathbf{R}\right)\right| \leq 3\epsilon\mathcal{H}\left(\mathbf{R}\right),$$

by setting $m = \sqrt{\frac{u}{2\epsilon\ell \ln(1/(1-\ell))}}$.

Computation Time

$$\mathcal{O}\left(\sqrt{\frac{u}{\ell \ln(1/(1-\ell))}} \cdot \frac{\ln(1/\delta)}{\epsilon^{2.5}} \cdot nnz(\mathbf{R}) + \ln(n) \cdot \ln(1/\delta) \cdot nnz(\mathbf{R})\right).$$

In (Kon+20) further appears an analysis of the Chebyshev-based Algorithm for Hermitian density matrices.

Assume that the density matrix $\mathbf{R} \in \mathbb{R}^{n \times n}$, has at most k non-zero probabilities, p_i. This means that at most k of its states are pure.

Issue & Solution

- x $n - k$ probabilities are zero $\rightarrow$ Chebyshev/Taylor approaches are not working.
- ✓ Project to a smaller full-dimension space $\rightarrow$ Random Projections.
- ✓ Fast construction of the random projector.

Construction of the Random Projector

- Gaussian Random Projector
- Sub-sampled Randomized Hadamard Transform
- Input Sparsity Transform
- Hartley Transform

Input: $R \in \mathbb{R}^{n \times n}$, integer $k \ll n$.
Output: $\widehat{\mathcal{H}(R)}$, the approximation to the $\mathcal{H}(R)$.

1: Construct the random projection matrix $\Pi \in \mathbb{R}^{n \times s}$.
2: Compute $\tilde{R} = R\Pi \in \mathbb{R}^{n \times s}$.
3: Compute and return the (at most) k non-zero singular values of $\tilde{R}$, $\tilde{p}_i$, $i = 1 \ldots k$.
4: Compute $\widehat{\mathcal{H}(R)}$ as:

$$\widehat{\mathcal{H}(R)} = \sum_{i=1}^{k} \tilde{p}_i \log \frac{1}{\tilde{p}_i}$$

Theorem

Let $\mathbf{R}$ *be a density matrix with at most* $k \ll n$ *non-zero probabilities and let* $\epsilon < 1/2$ *be an accuracy parameter. Then, with probability at least 0.9, the output of Algorithm 4 satisfies*

$$\left|p_i^2 - \tilde{p}_i^2\right| \leq \epsilon p_i^2$$

for all $i = 1 \ldots k$*. Additionally,*

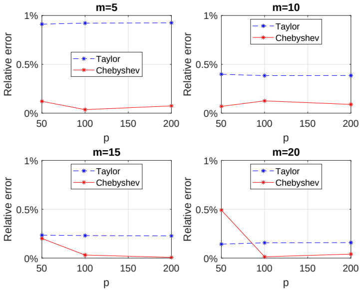
$$\left|\mathcal{H}(\mathbf{R}) - \widehat{\mathcal{H}(\mathbf{R})}\right| \leq \sqrt{\epsilon}\mathcal{H}(\mathbf{R}) + \sqrt{\frac{3}{2}}\epsilon.$$

Computation Time

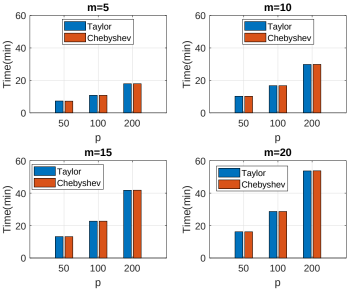
Algorithm 4 (combined with the Input Sparsity Transform) runs in time

$$\mathcal{O}\left(nnz(\mathbf{R}) + nk^4/\epsilon^4\right).$$

Matrix of size $30,000 \times 30,000$, $m = [5 : 5 : 20]$ and $u \approx \lambda_{max}$.



$p = [50 : 50 : 200]$

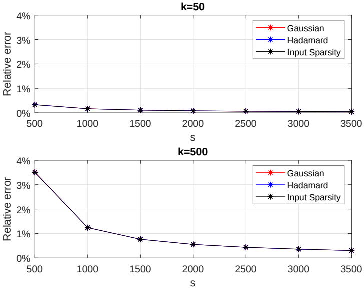


$p = \{50, 100, 200\}$

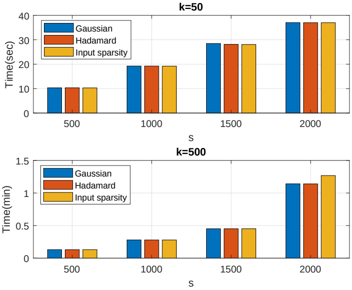
Notes

- Exact computation: 5.6 hours.
- Approximation of λ_{max}: 3.6 minutes.

Matrix of size $16,384 \times 16,384$ and $k = \{50, 500\}$.



$s = [500 : 500 : 3500]$



$s = [500 : 500 : 2000]$

Notes

- Exact computation for rank 50: 1.6 minutes
- Exact computation for rank 500: 20 minutes

[Kon+18] E. Kontopoulou, A. Grama, W. Szpankowski, P. Drineas, "**Randomized Linear Algebra Approaches to Estimate the Von Neumann Entropy of Density Matrices**, in Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT), pp. 2486-2490

[Kon+20] E. Kontopoulou, G. Dexter, A. Grama, W. Szpankowski & P. Drineas, "**Randomized Linear Algebra Approaches to Estimate the Von Neumann Entropy of Density Matrices**", in IEEE Transactions on Information Theory, to appear

The problem of $\operatorname{logdet}(\mathbf{A})$

Definition

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

Application: Maximum likelihood estimations, Gaussian processes prediction, logdet-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**:

$$\operatorname{logdet}(\mathbf{A}) = \operatorname{logdet}(\mathbf{L})^2 = 2 \log \prod_{i=1}^{n} l_{ii} = 2 \sum_{i=1}^{n} \log(l_{ii}).$$

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

Prohibitive for Large Data!!!!

Additive Error Approximation

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be an SPD matrix whose dominant eigenvalue is bounded by α. Define $\mathbf{C} = \mathbf{I}_n - \mathbf{A}/\alpha$. Then,

$$\text{logdet}(\mathbf{A}) \approx n \log(u) - \sum_{k=1}^{m} \frac{1}{k} \left(\frac{1}{s} \sum_{i=1}^{s} \mathbf{g}_i^\top \mathbf{C}^k \mathbf{g}_i \right)$$

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,

$$\text{logdet}(\mathbf{A}) \approx - \sum_{k=1}^{m} \frac{1}{k} \left(\frac{1}{s} \sum_{i=1}^{s} \mathbf{g}_i^\top \mathbf{C}^k \mathbf{g}_i \right)$$

Additive Error Algorithm/Lemma
Additive Error Experiments
Relative Error Algorithm/Lemma

[Bou+17] 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, pp.95-117.

Thank you!

Questions?

References I

(Bou+17) Christos Boutsidis et al. ``A Randomized Algorithm for Approximating the Log Determinant of a Symmetric Positive Definite Matrix''. In: *Linear Algebra and its Applications* 533 (2017), pp. 95–117.

(Kon+18) E. Kontopoulou et al. ``Randomized Linear Algebra Approaches to Estimate the Von Neumann Entropy of Density Matrices''. In: *2018 IEEE International Symposium on Information Theory*. 2018.

(Kon+20) E. Kontopoulou et al. ``Randomized Linear Algebra Approaches to Estimate the Von Neumann Entropy of Density Matrices''. In: *IEEE Transactions on Information Theory* to appear (2020).

(Tre11) L. Trevisan. *Graph Partitioning and Expanders.* Handout 7. 2011.

Appendix

In (Bou+17) appears the following lemma that builds on (Tre11) and guarantees a relative error approximation to the dominant eigenvalue:

Lemma

Let $\tilde{p}_1$ be the output of the Power Method algorithm with $q = \lceil 4.82 \log(1/\delta) \rceil$ and $t = \lceil \log \sqrt{4n} \rceil$. Then, with probability at least $1 - \delta$,

$$\frac{1}{6} p_1 \leq \tilde{p}_1 \leq p_1.$$

Back to Main Slides

Definition

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

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

where the $\mathbf{g}_i$'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* $1 - \delta$*,*

$$\left| Tr\left[\mathbf{A}\right] - \mathbf{G} \right| \leq \epsilon \cdot Tr\left[\mathbf{A}\right].$$

Back to Main Slides

The Clenshaw Algorithm

The Clenshaw algorithm is a recursive procedure that evaluates fast Chebyshev polynomials:

Input: Coefficients α_i, $i = 0, \dots, m$, matrix $R \in \mathbb{R}^{n \times n}$ and vectors $g \in \mathbb{R}^n$

Set $y_{m+2} = y_{m+1} = 0$
for $k = m, m-1, \dots, 0$ **do**
  $y_k = \alpha_k g + \frac{4}{u} R y_{k+1} - 2y_{k+1} - y_{k+2}$
end for

Output: $g^\top f_m(R) g = \frac{1}{2}\left(\alpha_0(g^\top g) + g^\top(y_0 - y_2)\right)$

Back to Main Slides

Mathematical Manipulation of $\operatorname{logdet}(\mathbf{A})$ 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 : $\mathbf{A} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^\top$.

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

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

Issues

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

Solution

Further manipulation of $Tr\left[\log(\mathbf{A})\right]$...

Lemma

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

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

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

Issues

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

Solution

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

Back to Main Slides

Input: $A \in \mathbb{R}^{n \times n}$, accuracy parameter $\epsilon > 0$, integer $m > 0$.
Output: $\widehat{logdet}(A)$, the approximation to the $\operatorname{logdet}(A)$.

1: Compute $\tilde{\lambda}_1(A)$, the estimation of the largest eigenvalue of A, using the power method.
2: Set $u = 7\tilde{\lambda}_1(A)$
3: $C = I_n - u^{-1}A$
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 $\widehat{logdet}(A)$ as:

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

Lemma

Let $\widehat{logdet}(\mathbf{A})$ *be the approximation of* $\operatorname{logdet}(\mathbf{A})$ *using the algorithm LogDetAdditive on inputs* $\mathbf{A}$, m *and* ϵ. *Then, we* **prove** *that with probability at least* $1 - 2\delta$,

$$\left|\widehat{logdet}(\mathbf{A}) - \operatorname{logdet}(\mathbf{A})\right| \leq 2\epsilon\Gamma$$

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

Running Time

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

Back to Main Slides

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: $\widehat{logdet}(A)$, the approximation to $\operatorname{logdet}(A)$.

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: Compute $\widehat{logdet}(A)$ as:

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

Lemma

Let $\widehat{logdet}(\mathbf{A})$ *be the approximation of* $\operatorname{logdet}(\mathbf{A})$ *using the algorithm LogDetRelative on inputs* $\mathbf{A}$*,* m *and* ϵ*. Then, we* **prove** *that with probability at least* $1-\delta$*,*

$$|\widehat{logdet}(\mathbf{A}) - \operatorname{logdet}(\mathbf{A})| \leq 2\epsilon \cdot |\operatorname{logdet}(\mathbf{A})|$$

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

Running Time

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

Implementation of *LogDetAdditive*

- C++
- OpenMPI 1.8.4
- Boost 1.55
- Elemental + OpenBLAS
- Eigen 3.2.4 (incl. BLAS, LAPACK)

Environment

- 60-core Intel Xeon E7-4890@ 2.7Ghz
- 1TB RAM

n	logdet (A)			time (secs)		
	exact	mean	std	exact	mean	std
5000	-3717.89	-3546.920	8.10	2.56	1.15	0.0005
7500	-5474.49	-5225.152	8.73	7.98	2.53	0.0015
10000	-7347.33	-7003.086	7.79	18.07	4.47	0.0006
12500	-9167.47	-8734.956	17.43	34.39	7.00	0.0030
15000	-11100.9	-10575.16	15.09	58.28	10.39	0.0102

Parameters: $p = 60$, $m = 4$, $t = \log(\sqrt{4n})$. Ground truth computed via Cholesky. Mean and standard deviation reported over 10 repetitions.

Experiments II

Real Sparse Matrices
University of Florida Sparse Matrix Collection

matrix name	n	nnz	logdet (A)			time (sec)		m
			exact	approx		exact	approx	
				mean	std		mean	
thermal2	1228045	8580313	1.3869e6	1.3928e6	964.79	31.28	31.24	149
ecology2	999999	4995991	3.3943e6	3.403e6	1212.8	18.5	10.47	125
ldoor	952203	42493817	1.4429e7	1.4445e7	1683.5	117.91	17.60	33
thermomech_TC	102158	711558	-546787	-546829.4	553.12	57.84	2.58	77
boneS01	127224	5516602	1.1093e6	1.106e6	247.14	130.4	8.48	125

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.

Back to Main Slides