Stochastic trace estimation and quantum typicality

a case study in interdisciplinary research

Tyler Chen

March 10, 2023

Collaborators

This is based on joint work with:

- Thomas Trogdon (University of Washington)
- Shashanka Ubaru (IBM Watson)
- Yu-Chen Cheng (Department of Data Science, Dana-Farber Cancer Institute)

and ongoing work with:

- Robert Chen
- Kevin Li
- Skai Nzeuton
- Yilu Pan
- Yixin Wang

A randomized estimator

Suppose $\mathbf{v}$ is a random vector with $\mathbb{E}[v_i] = 0$ and $\mathbb{E}[v_i v_j] = 1$.

Then,

$$\mathbb{E}[\mathbf{v}^\top \mathbf{A}\mathbf{v}] = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} \mathbb{E}[v_i v_j] = \sum_{i=1}^{n} A_{i,i} = \operatorname{tr}(\mathbf{A}).$$

A randomized estimator

Suppose $\mathbf{v}$ is a random vector with $\mathbb{E}[v_i] = 0$ and $\mathbb{E}[v_i v_j] = 1$.

Then,

$$\mathbb{E}[\mathbf{v}^\top \mathbf{A} \mathbf{v}] = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} \mathbb{E}[v_i v_j] = \sum_{i=1}^{n} A_{i,i} = \mathrm{tr}(\mathbf{A}).$$

This is colloquially called Hutchinson's trace estimator.

Hutchinson's estimator? (Stigler's law of eponymy)

Hutchinson's estimator? (Stigler's law of eponymy)

Hutchinson 1989 cites Girard 1987 in the abstract!

ABSTRACT

An unbiased stochastic estimator of tr(I–A), where A is the influence matrix associated with the calculation of Laplacian smoothing splines, is described. The estimator is similar to one recently developed by Girard but satisfies a minimum variance criterion and does not require the simulation of a standard normal variable. It uses instead simulations of the discrete random variable which takes the values 1, –1 each with probability 1/2. Bounds on the variance of the estimator, similar to those established by Girard, are obtained using elementary methods. The estimator can be used to approximately minimize generalised cross validation (GCV) when using

Analysis

Variance analyses appeared in these first papers. Often concentration inequalities are more useful/informative:

$$\mathbb{P}[|\operatorname{tr}(\mathbf{A}) - \mathbf{v}^\top \mathbf{A}\mathbf{v}| < \epsilon] \leq \ ??$$

Analysis

Variance analyses appeared in these first papers. Often concentration inequalities are more useful/informative:

$$\mathbb{P}[|\operatorname{tr}(\mathbf{A}) - \mathbf{v}^\top \mathbf{A}\mathbf{v}| < \epsilon] \leq \ ??$$

- Bai, Fahey, and Golub 1996: Hoeffding (some issues?)
- Avron and Toledo 2011: sub-Gaussian concentration
- Roosta-Khorasani and Ascher 2014: improved bounds
- Cortinovis and Kressner 2021: sub-Gamma concentration

Quantum typicality

In the formalism of quantum mechanics:

$$\text{observable} \iff \text{Hermitian matrix } \mathbf{A}\,, \qquad \text{state of system} \iff \text{unit vector } \mathbf{v}$$

When we measure an observable $\mathbf{A}$ in state $\mathbf{v}$, we get out one of the eigenvalues λ_i of $\mathbf{A}$ with probability proportional to $|\mathbf{u}_i^\top \mathbf{v}|^2$. Thus,

$$\mathbb{E}_{\text{qm}}[\text{measurement of } \mathbf{A} \text{ is state } \mathbf{v}] = \sum_{i=1}^{n} \lambda_i |\mathbf{u}_i^\top \mathbf{v}|^2 = \mathbf{v}^\top \mathbf{A} \mathbf{v} = \operatorname{tr}(\mathbf{A}\mathbf{v}\mathbf{v}^\top).$$

Quantum expectation value

General state state of the system is described unit-trace matrix

$$\boldsymbol{\rho} = \sum_{i=1}^{n} p_i \mathbf{v}_i \mathbf{v}_i^\top.$$

Loosely can be thought of as the "probability p_i of being in state v_i".

When we measure an observable $\mathbf{A}$ in state $\boldsymbol{\rho}$ we can view this as picking one of the state $\mathbf{v}_i$ and then measuring in that state:

$$\mathbb{E}_{\mathrm{qm}}[\text{measurement of } \mathbf{A} \text{ in state } \boldsymbol{\rho}] = \sum_{i=1}^{n} p_i \operatorname{tr}(\mathbf{A}\mathbf{v}_i\mathbf{v}_i^\top) = \operatorname{tr}(\mathbf{A}\boldsymbol{\rho}).$$

This is called the quantum expectation value (QEV).

Quantum typicality

When the system is equally likely to be in any possible state, $\boldsymbol{\rho} = n^{-1}\mathbf{I}$ so

$$\mathbb{E}_{\mathrm{qm}}[\text{measurement of } \mathbf{A} \text{ in state } \boldsymbol{\rho}] = n^{-1}\operatorname{tr}(\mathbf{A}).$$

Clearly when $\mathbf{v}$ is a uniformly random state ($\mathbf{v} \sim \mathrm{Unif}(\mathbb{S}^{n-1})$), then

$$\mathbb{E}_{\mathbf{v}}[\operatorname{tr}(\mathbf{A}\mathbf{v}\mathbf{v}^{\top})] = \operatorname{tr}(\mathbf{A}\mathbb{E}[\mathbf{v}\mathbf{v}^{\top}]) = \operatorname{tr}(\mathbf{A}n^{-1}\mathbf{I}) = n^{-1}\operatorname{tr}(\mathbf{A}).$$

That is, the expectation value of a measurement made in a random state is, on average, the QEV of $\mathbf{A}$.

Theoretical analyses

Are measurements made in a random state are typically near the QEV? That is, do they concentrate?

- studied from the outset of QM[1] (Schrödinger 1927; Neumann 1929)
- sub-Gaussian concentration via Levy's lemma (concentration of measure) (Popescu, Short, and Winter 2006; Gogolin 2010)

[1] Goldstein, Lebowitz, Mastrodonato, Tumulka, and Zanghì 2010.

Typicality algorithms

The use of random states[2] in algorithms is fairly old.

[2]Jin, Willsch, Willsch, Lagemann, Michielsen, and De Raedt 2021.

[3]Alben, Blume, Krakauer, and Schwartz 1975; Weaire and Williams 1976; Weaire and Williams 1977.

Typicality algorithms

The use of random states[2] in algorithms is fairly old.

Early work[3] with density of state as average of local density of states (even one sample often enough!)

[2] Jin, Willsch, Willsch, Lagemann, Michielsen, and De Raedt 2021.

[3] Alben, Blume, Krakauer, and Schwartz 1975; Weaire and Williams 1976; Weaire and Williams 1977.

Beyond quadratic trace estimation

The variance of $\mathbf{v}^\top \mathbf{A} \mathbf{v}$ is closely related to $\|\mathbf{A}\|_\mathsf{F}^2$.

Beyond quadratic trace estimation

The variance of $\mathbf{v}^\top \mathbf{A}\mathbf{v}$ is closely related to $\|\mathbf{A}\|_\mathsf{F}^2$. Decompose

$$\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\widetilde{\mathbf{A}}) - \mathrm{tr}(\mathbf{A} - \widetilde{\mathbf{A}}).$$

If $\|\mathbf{A} - \widetilde{\mathbf{A}}\|_\mathsf{F} \ll \|\mathbf{A}\|_\mathsf{F}$ and we can compute $\mathrm{tr}(\widetilde{\mathbf{A}})$ exactly, we have a variance-reduced estimator.

Beyond quadratic trace estimation

The variance of $\mathbf{v}^\mathsf{T}\mathbf{A}\mathbf{v}$ is closely related to $\|\mathbf{A}\|_\mathsf{F}^2$. Decompose

$$\operatorname{tr}(\mathbf{A}) = \operatorname{tr}(\widetilde{\mathbf{A}}) - \operatorname{tr}(\mathbf{A} - \widetilde{\mathbf{A}}).$$

If $\|\mathbf{A} - \widetilde{\mathbf{A}}\|_\mathsf{F} \ll \|\mathbf{A}\|_\mathsf{F}$ and we can compute $\operatorname{tr}(\widetilde{\mathbf{A}})$ exactly, we have a variance-reduced estimator.

Recently, Meyer, Musco, Musco, and Woodruff 2021 give an algorithm called Hutch++ which provably improves dependence on accuracy from $1/\epsilon^2$ to $1/\epsilon$.

- similar algorithms previously, but no nice analysis afaik.

Some interesting notes

Variance reduction via a control variate suggested in Girard 1987:

> La 2ème méthode est, elle, insensible à la translation précédente des valeurs propres. Mais s'il existe une matrice B de trace connue, dont les valeurs propres sont voisines de A, et si l'on sait calculer le produit By pour un y donné, on a intérêt à prendre comme estimateur:
>
> $$\frac{1}{n}\,\mathrm{tr}(A) \approx \frac{1}{n}\,\mathrm{tr}(B) + (w^t A w - w^t B w)/w^t w,$$
>
> puisque l'écart type de cet estimateur est la dispersion d(A−B) des valeurs propres de A−B.

Deflation suggested in Weiße, Wellein, Alvermann, and Fehske 2006.

Summary

	NLA/TCS	comp. physics	typicality
trace estimator	1985s[4]	folklore[5]	1930s[6]
variance bounds	folklore[7]	folklore	folklore[8]
concentration	1996,2011-now[9]	folklore?	2006[10]
low-rank approx.	1987,2010s-now[11]	?[12]	

[4] Girard 1987; Hutchinson 1989.

[5] Alben, Blume, Krakauer, and Schwartz 1975; Raedt and Vries 1989; Skilling 1989.

[6] Schrödinger 1927; Neumann 1929.

[7] Hanson and Wright 1971.

[8] Gemmer, Michel, and Mahler 2004.

[9] Bai, Fahey, and Golub 1996; Avron and Toledo 2011; Roosta-Khorasani and Ascher 2014; Cortinovis and Kressner 2021.

[10] Popescu, Short, and Winter 2006.

[11] Girard 1987; Lin 2016; Wu, Laeuchli, Kalantzis, Stathopoulos, and Gallopoulos 2016; Gambhir, Stathopoulos, and Orginos 2017; Meyer, Musco, Musco, and Woodruff 2021; Persson, Cortinovis, and Kressner 2022; Epperly, Tropp, and Webber 2023.

[12] Weiße, Wellein, Alvermann, and Fehske 2006; Morita and Tohyama 2020.

Outlook/questions

- Why is it called Hutchinson's estimator?
- Was it inevitble that these lines of research developed separately in parallel?
- What else are we overlooking in the literature?

Surveys

Kernel polynomial method: Weiße, Wellein, Alvermann, and Fehske 2006

Typicality: Goldstein, Lebowitz, Mastrodonato, Tumulka, and Zanghì 2010

Typicality algs: Jin, Willsch, Willsch, Lagemann, Michielsen, and De Raedt 2021

Randomized quadrature: Chen, Trogdon, and Ubaru 2021

References I

Alben, R. et al. (Nov. 1975). “Exact results for a three-dimensional alloy with site diagonal disorder: comparison with the coherent potential approximation”. In: *Physical Review B* 12.10, pp. 4090–4094.

Avron, Haim and Sivan Toledo (Apr. 2011). “Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix”. In: *Journal of the ACM* 58.2, pp. 1–34.

Bai, Zhaojun, Gark Fahey, and Gene Golub (Nov. 1996). “Some large-scale matrix computation problems”. In: *Journal of Computational and Applied Mathematics* 74.1-2, pp. 71–89.

Chen, Tyler, Thomas Trogdon, and Shashanka Ubaru (18–24 Jul 2021). “Analysis of stochastic Lanczos quadrature for spectrum approximation”. In: *Proceedings of the 38th International Conference on Machine Learning*. Vol. 139. Proceedings of Machine Learning Research. PMLR, pp. 1728–1739.

Cortinovis, Alice and Daniel Kressner (July 2021). “On Randomized Trace Estimates for Indefinite Matrices with an Application to Determinants”. In: *Foundations of Computational Mathematics*.

Epperly, Ethan N. , Joel A. Tropp, and Robert J. Webber (2023). *XTrace: Making the most of every sample in stochastic trace estimation*.

Gambhir, Arjun Singh, Andreas Stathopoulos, and Kostas Orginos (Jan. 2017). “Deflation as a Method of Variance Reduction for Estimating the Trace of a Matrix Inverse”. In: *SIAM Journal on Scientific Computing* 39.2, A532–A558.

Gemmer, Jochen, M. Michel, and Günter Mahler (2004). *Quantum Thermodynamics*. Springer Berlin Heidelberg.

Girard, Didier (1987). *Un algorithme simple et rapide pour la validation croisée généralisée sur des problèmes de grande taille*.

Gogolin, Christian (2010). “Pure State Quantum Statistical Mechanics”. MA thesis. Julius-Maximilians-Universität Würzburg.

Goldstein, Sheldon et al. (May 2010). “Normal typicality and von Neumann’s quantum ergodic theorem”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 466.2123, pp. 3203–3224.

Hanson, D. L. and F. T. Wright (June 1971). “A Bound on Tail Probabilities for Quadratic Forms in Independent Random Variables”. In: *The Annals of Mathematical Statistics* 42.3, pp. 1079–1083.

References II

Hutchinson, M.F. (Jan. 1989). “A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines”. In: *Communications in Statistics - Simulation and Computation* 18.3, pp. 1059–1076.

Jin, Fengpin et al. (Jan. 2021). “Random State Technology”. In: *Journal of the Physical Society of Japan* 90.1, p. 012001.

Lin, Lin (Aug. 2016). “Randomized estimation of spectral densities of large matrices made accurate”. In: *Numerische Mathematik* 136.1, pp. 183–213.

Meyer, Raphael A. et al. (Jan. 2021). “Hutch++: Optimal Stochastic Trace Estimation”. In: *Symposium on Simplicity in Algorithms (SOSA)*. Society for Industrial and Applied Mathematics, pp. 142–155.

Morita, Katsuhiro and Takami Tohyama (Feb. 2020). “Finite-temperature properties of the Kitaev-Heisenberg models on kagome and triangular lattices studied by improved finite-temperature Lanczos methods”. In: *Physical Review Research* 2.1.

Neumann, J. von (Jan. 1929). “Beweis des Ergodensatzes und des$H$-Theorems in der neuen Mechanik”. In: *Zeitschrift für Physik* 57.1-2. English translation `https://arxiv.org/abs/1003.2133`, pp. 30–70.

Persson, David, Alice Cortinovis, and Daniel Kressner (July 2022). “Improved Variants of the Hutch++ Algorithm for Trace Estimation”. In: *SIAM Journal on Matrix Analysis and Applications* 43.3, pp. 1162–1185.

Popescu, Sandu, Anthony J. Short, and Andreas Winter (Oct. 2006). “Entanglement and the foundations of statistical mechanics”. In: *Nature Physics* 2.11, pp. 754–758.

Raedt, Hans De and Pedro de Vries (June 1989). “Simulation of two and three-dimensional disordered systems: Lifshitz tails and localization properties”. In: *Zeitschrift für Physik B Condensed Matter* 77.2, pp. 243–251.

Roosta-Khorasani, Farbod and Uri Ascher (Sept. 2014). “Improved Bounds on Sample Size for Implicit Matrix Trace Estimators”. In: *Foundations of Computational Mathematics* 15.5, pp. 1187–1212.

Schrödinger, E. (1927). “Energieaustausch nach der Wellenmechanik”. In: *Annalen der Physik* 388.15, pp. 956–968.

Skilling, John (1989). “The Eigenvalues of Mega-dimensional Matrices”. In: *Maximum Entropy and Bayesian Methods*. Springer Netherlands, pp. 455–466.

Weaire, D and A R Williams (Sept. 1976). “New numerical approach to the Anderson localization problem”. In: *Journal of Physics C: Solid State Physics* 9.17, pp. L461–L463.

References III

Weaire, D and A R Williams (Apr. 1977). “The Anderson localization problem. I. A new numerical approach”. In: *Journal of Physics C: Solid State Physics* 10.8, pp. 1239–1245.

Weiße, Alexander et al. (Mar. 2006). “The kernel polynomial method”. In: *Reviews of Modern Physics* 78.1, pp. 275–306.

Wu, Lingfei et al. (2016). “Estimating the trace of the matrix inverse by interpolating from the diagonal of an approximate inverse”. In: *Journal of Computational Physics* 326, pp. 828–844.