Mathematics: The Tao of Data Science

Tamara Kolda^{1,2,3}

¹MathSci.ai, Dublin, California, United States of America,

²Sandia National Laboratories, National Technology and Engineering Solutions of Sandia, Livermore, California, United States of America,

³Department of Industrial Engineering and Management Sciences, Robert R. McCormick School of Engineering, Northwestern University, Evanston, Illinois, United States of America

Published on: Sep 30, 2020

DOI: <https://doi.org/10.1162/99608f92.66fb00a2>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Confucius once said, “Fish forget they live in water; people forget they live in the Tao” (Lin, 2007). Analogously, it may be easy for data scientists to forget they live in a world defined and permeated by mathematics.

1. Introduction

The two pieces, "[Ten Research Challenge Areas in Data Science](#)" by Jeannette M. Wing and "[Challenges and Opportunities in Statistics and Data Science: Ten Research Areas](#)" by Xuming He and Xihong Lin, provide an impressively complete list of data science challenges from luminaries in the field of data science. They have done an extraordinary job, so this response offers a complementary viewpoint from a mathematical perspective and evangelizes advanced mathematics as a key tool for meeting the challenges they have laid out. Notably, we pick up the themes of scientific understanding of machine learning and deep learning, computational considerations such as cloud computing and scalability, balancing computational and statistical considerations, and inference with limited data. We propose that mathematics is an important key to establishing rigor in the field of data science and as such has an essential role to play in its future.

2. Exploration of Mathematical Underpinnings

I define *mathematics* broadly to include any topic that relies heavily on abstraction and theory, including applied probability, statistics, theoretical computer science, mathematical signal processing, network science, and classical topics such as approximation theory, optimization, functional and harmonic analysis, differential equations, applied algebra, topology, and so on. I propose that mathematics is a scaffold to build the rigor needed to meet practical data science challenges that cannot be solved by trial and error, adding more computational power, or enlarging the training data. I highlight mathematics' role in a subset of the topics identified by Wing, He, and Lin. Due to limitations in space, I omit obvious (at least to me) clear connections to trustworthy AI, integrative analysis of different types and sources of data, statistical analysis of privatized data, causal inference for big data, and so on. Likewise, I will resist the temptation to expound on topics that Wing, He, and Lin did not have space to mention, including understanding the limits of inference, developing autonomous real-time control and online learning systems, and combating adversarial tampering in training or deployment. Instead, I focus on a subset of topics Wing, He, and Lin identified (listed in boldface) and talk about the mathematical connections in meeting those specific challenges.

2.1. Scientific understanding of learning, especially deep learning algorithms

Mathematics is everywhere, but if I *had to* pick one place where mathematics has a unique impact on data science, it is in Wing's challenge of establishing a scientific understanding of why learning methods, like deep learning, work so well and, more importantly, how to extend their success further. From a mathematical point of view, deep neural networks are a nonlinear mapping from input space to a reduced representation or a

classification, and so we study the properties of this and other mappings. Early work by Cybenko considered the approximation properties of one-layer neural networks and showed that they were universal approximators (Cybenko, 1989). More recent work by several different groups is characterizing the approximation capabilities of deep neural networks more completely; see, e.g. (Bölcskei et al., 2019). Recent work in the *Proceedings of the National Academy of Science (PNAS)* provides an accessible and fascinating mathematics explanation of the counterintuitive generalization performance of deep neural networks. The authors advocate for a “double-descent” bias-variance tradeoff curve to explain the performance of modern machine learning where more parameters yield better performance. They conjecture via mathematical arguments on the practical utility of overparameterized learning methods (Belkin et al., 2019). Mathematics is indispensable in the quest for scientific understanding of learning models.

2.2. Computing systems for data-intensive applications/Cloud-based scalable and distributed statistical inference

Both Wing and He and Lin reference the computing challenges in data science. A common approach to ‘solve’ problems in learning is to throw more processing power and data at the problems. However, algorithmic improvements have had arguably *more* impact than improvements in processors in many computationally intensive settings. Generally, these results rely on considerations of algorithm complexity, considering communication as well as computation (Ballard et al., 2014). For instance, a retrospective study on computing the matrix singular value decomposition (i.e., the key computation of principal component analysis) shows that algorithmic advances alone yield an order of magnitude speed-up on a single core and another order of magnitude of speed-up is gained because newer algorithms enable parallelism. The study compared old and new codes on modern hardware, and showed further that energy consumption was reduced by a factor of 40X (Dongarra et al., 2018). This situation is not unique. An earlier study found similar improvements in computational simulations, as shown in Figure 1, where we can see that algorithmic improvements yield several orders of magnitude more improvement than hardware advancements alone. Mathematically based algorithmic improvements are central to reducing computational needs. As Wing mentions, hardware designers should focus on the data. In scientific computing, the concept of *co-design* has been an ongoing effort to design hardware and algorithms together (Lucas et al., 2014). A challenge is to co-design hardware with data *and algorithms* in focus, as Google has arguably done with its tensor processing unit (TPU) (Cass, 2019). On the analysis front, many researchers are working on the convergence of optimization methods for deep neural networks; see, e.g. (Ward et al., 2019), and improvements here will reduce the enormous computational and energy needs for training neural networks.

Figure 1. A 2004 study of scientific computational simulations shows that computational hardware improvements alone yield only some speedups, and we propose that the same may be true for data science. Mathematical advances (algorithmic and modeling) account for several more orders of magnitude improvement. This image is specific to simulations of magnetic fusion energy. MHD = magnetohydrodynamics. From *A Science-Based Case for Large-Scale Simulations*, Vol. 2, Office of Science, U.S. Department of Energy, 2004.

2.3. Balance of Statistical and Computational Efficiencies

As He and Lin point out, with many large data sets, we must make the best use of data that is *locally* available via clever statistical methods. As an example, consider that some problems in network analysis are bedeviling in their complexity (e.g., requiring n^3 operations on a graph with n nodes) and intractable within the time limits needed to, for example, respond to web queries. Twitter, for instance, wants to suggest users for you to follow. Mathematically, one way to model this is to compute all pairs of dot products of vectors representing followers. Since Twitter has billions of users, this problem is nearly impossible. However, careful understanding of advanced techniques such as random projections (Johnson & Lindenstrauss, 1984) can yield sampling techniques that can be implemented efficiently on large-scale systems. More generally, random sampling can be adapted to yield provably near-optimal answers in sublinear time for real-world applications (Sharma et al., 2017). Leveraging rigorous sampling methods can reduce or eliminate computational roadblocks.

2.4. Inferring from Noisy and/or Incomplete Data

This challenge identified by Wing is both prevalent and far-reaching, and mathematical tools are key to addressing it. Around 2010, before deep learning’s dominance, *compressive sensing* was in vogue. Compressive sensing allows recovery of a signal with incomplete measurements by solving an optimization problem, and it has had a widespread impact in imaging science. Theory based on random matrices explains when and why the method works, and advances in optimization algorithms made it practical for accelerating applications such as Magnetic Resonance Imagery (MRI). For an overview, see Fornasier & Rauhut (2015). The medical implications are striking since far fewer measurements are needed; for instance, Siemens Healthineers advertises a “Compressed Sensing GRASP-VIBE” that can be used on patients who cannot hold their breath or follow breathing commands, and they provided impressive imagery that shows clear images using this technique (Chandarana et al., 2013). An example of the improvements they have is shown in Figure 2. Recently, deep learning is being used to further improve compressed sensing MRI, and these approaches will need further mathematical (as well as engineering) insights to verify their reliability (see, e.g., Zbontar et al., 2018).

Figure 2. Example of compressed sensing improvements in abdominal MRI that yield much sharper imagery. From <https://www.siemens-healthineers.com/en-us/magnetic-resonance-imaging/options-and-upgrades/clinical-applications/compressed-sensing-grasp-vibe>

More generally, mathematicians have long been intrigued with ideas of low-rank structure, in which case limited measures are sufficient for inference; for instance, recent mathematical work postulates why low-rank

is common in real-world data (Udell & Townsend, 2019).

3. Is Data Science a Discipline?

I'll add a few remarks on Wing's fascinating discussion on whether data science is a discipline. She had thoughtful questions about the role of the domain. I have a much simpler litmus test: let us call something a discipline when it has its own professional society! If we focus on the United States, statistics has the American Statistical Association, founded in 1839; engineering has the Institute of Electrical and Electronics Engineers, which traces its roots to 1884; mathematics has the American Mathematical Society, which traces its history to 1888; computer science has the Association for Computing Machinery, founded in 1947; and applied mathematics has the Society for Industrial and Applied Mathematics, founded in 1951. Does data science belong in one of the existing societies? If not, what is to be done? Something for the field (including mathematical data scientists) to contemplate.

4. A Seat at the Table for Mathematicians

I'll close with an appeal for a seat at the data science table for mathematicians. In December 2017, Ali Rahimi and Ben Recht famously declared that machine learning has become a form of alchemy “(not as a pejorative, but as a methodology that produce practical results)” (Hutson, 2018; Rahimi & Recht, 2017a; Rahimi & Recht, 2017b). They and many others allege that there is a need for increased rigor in AI, for understanding why algorithms work or don't work, for going beyond folklore, and for better reproducibility. Mathematicians can bring the required rigor to the field and hold the keys to future AI breakthroughs that trial and error cannot reveal. Mathematicians wield the tools of deep mathematical proofs and logical arguments, and more applied mathematicians put the abstract insights to the service of practical scenarios, adapting provably correct algorithms in idealized settings to real-world scenarios. Ideally, mathematicians can also engage with empirical findings and help to unravel and explain them. These facets are rarely emphasized in data science programs but deserve attention if we are truly to bring the needed rigor to the field so that more difficult challenges can be tackled.

Acknowledgments

The author thanks Misha Belkin, Jon Berry, Brett Larsen, Juan Meza, Justin Newcomer, and Rachel Ward, as well as the editor, Xiao-Li Meng, for helpful comments on this manuscript.

Disclosure Statement

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. Any

subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

Ballard, G., Carson, E., Demmel, J., Hoemmen, M., Knight, N., & Schwartz, O. (2014). Communication lower bounds and optimal algorithms for numerical linear algebra. *Acta Numerica*, 23, 1–155.

<https://doi.org/10.1017/s0962492914000038>

Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32), 15849–15854.

<https://doi.org/10.1073/pnas.1903070116>

Bölskei, H., Grohs, P., Kutyniok, G., & Petersen, P. (2019). Optimal approximation with sparsely connected deep neural networks. *SIAM Journal on Mathematics of Data Science*, 1(1), 8–45.

<https://doi.org/10.1137/18M118709X>

Cass, S. (2019). Taking AI to the edge: Google's TPU now comes in a maker-friendly package. *IEEE Spectrum*, 56(5), 16–17. <https://doi.org/10.1109/mspec.2019.8701189>

Chandarana, H., Feng, L., Block, T. K., Rosenkrantz, A. B., Lim, R. P., Babb, J. S., Sodickson, D. K., Otazo, R. (2013). Free-breathing contrast-enhanced multiphase MRI of the liver using a combination of compressed sensing, parallel imaging, and golden-angle radial sampling. *Investigative Radiology*, 48(1), 10–16.

<https://doi.org/10.1097/rli.0b013e318271869c>

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4), 303–314. <https://doi.org/10.1007/BF02551274>

Dongarra, J., Gates, M., Haidar, A., Kurzak, J., Luszczek, P., Tomov, S., & Yamazaki, I. (2018). The singular value decomposition: Anatomy of optimizing an algorithm for extreme scale. *SIAM Review*, 60(4), 808–865.

<https://doi.org/10.1137/17m1117732>

Fornasier, M., & Rauhut, H. (2015). Compressive sensing. In *Handbook of mathematical methods in imaging* (Vol. 1, pp. 205–256). Springer. https://doi.org/10.1007/978-1-4939-0790-8_6

Hutson, M. (2018, May 3). AI researchers allege that machine learning is alchemy. *Science*.

<https://doi.org/10.1126/science.aau0577>

Johnson, W., & Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. In *Conference in Modern Analysis and Probability (New Haven, Conn., 1982)* (Vol. 26, pp. 189–206). American

Mathematical Society. <https://doi.org/10.1007/BF02764938>

Lin, D. (2007). *The Tao of daily life*. Penguin.

Lucas, R., Ang, J., Bergman, K., Borkar, S., Carlson, W., Carrington, L., Chiu, G., Colwell, R., Dally, W., Dongarra, J., Geist, A., Haring, R., Hittinger, J., Hoisie, A., Klein, D. M., Kogge, P., Lethin, R., Sarkar, V., Schreiber, R., . . . Laros III, J. (2014). *DOE Advanced Scientific Computing Advisory Subcommittee (ASCAC) report: Top ten exascale research challenges*. Office of Scientific and Technical Information (OSTI).
<https://doi.org/10.2172/1222713>

Rahimi, A., & Recht, B. (2017a). An addendum to alchemy. *arg min* [blog]
<http://www.argmin.net/2017/12/05/kitchen-sinks/>

Rahimi, A., & Recht, B. (2017b). Reflections on random kitchen sinks. *arg min* [blog]
<http://www.argmin.net/2017/12/11/alchemy-addendum/>

Sharma, A., Seshadhri, C., & Goel, A. (2017). When hashes met wedges: A distributed algorithm for finding high similarity vectors. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 431–440), International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland. <https://doi.org/10.1145/3038912.3052633>

Udell, M., & Townsend, A. (2019). Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science*, 1(1), 144–160. <https://doi.org/10.1137/18M1183480>

Ward, R., Wu, X., & Bottou, L. (2019). AdaGrad stepsizes: Sharp convergence over nonconvex landscapes. In *Proceedings of Machine Learning Research: Vol. 97. Proceedings of the 36th International Conference on Machine Learning* (pp. 6677–6686). <http://proceedings.mlr.press/v97/ward19a.html>

Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M. J., . . . Lui, Y. W. (2018). fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv*. <https://doi.org/10.48550/arXiv.1811.08839>

©2020 Tamara Kolda. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.