Research Statement

Mark Tygert

Disclaimer: The sole purpose of this research statement is to apply for jobs; I make no serious attempt to give a comprehensive historical review or assign appropriate credit to those who deserve the credit.

1 Who am I?

This section speaks in high-level generalities that can be skipped by readers who are more interested in technical details.

I consider myself to be an applied mathematician first and foremost and aim to remain so, despite the changing times. Given the breadth of my studies, this appellation might surprise some. The following two paragraphs explain my modus operandi in research and why applied mathematics is my home base and primary venue for publication.

Researchers typically solve particular problems or build general theories. I do both in every area in which I work. I start by solving specific, concrete problems, most often those that others need solved. And then, the solution to the particular problem typically turns out to generalize to abstract theories applicable well beyond the context of the original motivating applications.

The reason that the solutions to concrete particular problems generalize to become widely applicable abstract theories is obvious. My solutions always contribute new substantive mathematical content. Since mathematics at some level is purely formal, these mathematical contributions turn out to apply much more widely than I could have imagined. Thus, mathematics is core to my research and a constant source of pleasant surprise.

The following are four highlights from my career thus far:

2 Detecting differences while conditioning on or controlling for covariates

This section summarizes my foremost contribution to R&D in the most elementary terms possible.

Consider the following simplest special case of the general theory we eventually developed:

We would like a single scalar number to summarize the differences between n pairs of numbers (R_1, s_1), (R_2, s_2), ..., (R_n, s_n), where R_k is either 0 or 1 and s_k can take any real value from 0 to 1 (inclusive), for $k = 1, 2, \ldots, n$. Perfect calibration is when the expected value of R_k is equal to s_k, for $k = 1, 2, \ldots, n$. For instance, weather predictions are perfectly calibrated when it actually rains on $x\%$ of the days predicted to have an $x\%$ chance of rain, for all possible values of x.

The Kuiper metric is the absolute value of the sum of $(R_k - s_k)/n$, summing over only those k for which s_k falls in an interval. The interval is chosen such that the absolute value of the sum is greatest.

The reason for restricting to this worst-case interval is to minimize possible cancellation between positive and negative differences. The Kuiper statistic can take values ranging from 0 to 1 (inclusive).

The Kuiper metric is easy to calculate via cumulative statistics (due to the discrete version of the Fundamental Theorem of Calculus, giving the relation between finite differences and partial sums). I introduced this Kuiper statistic in [15], modifying and generalizing the classic statistics of Kuiper which test goodness-of-fit of distributional form.

The existing standard metrics, especially what is usually called the "ECE" or "ICI" (with the acronyms abbreviating several different phrases), have very poor properties. Indeed, the ECE and ICI are semi-parametric and require choosing bins for the bins in a histogram. Very often, the choice of bins can yield whatever result the chooser desires, allowing the choice to fudge the results.

We proved in [1] that the ECE (ICI) is asymptotically inconsistent or infinitely less statistically powerful than the Kuiper metric (in the limit of infinite sample size) and also demonstrated that the asymptotic limit is highly relevant to even the smallest sample sizes. Asymptotic inconsistency means that the ECE converges to the wrong answer even as the number of observations becomes arbitrarily large, if the resolution of the binning becomes commensurately fine. The ECE is unable to average away randomness in the observations even when there are infinitely many observations (at least if the resolution of the binning is sensible, not sacrificing infinitely much statistical power).

Thus, I replaced the standard metrics of calibration, which have such outrageously unacceptable statistical properties. Moreover, I replaced the standard graphical methods — so-called "reliability diagrams" or "calibration plots" — with appropriate equally simple yet vastly more reliable cumulative graphs. Reliability diagrams have been universally known to be unacceptably unreliable for decades.

And this construction for assessing calibration turns out to generalize to the analysis of differences between any two random variables being conditioned on a third (with the third known as a "covariate" or vector of covariates). Statisticians refer to the field of study of differences between two random variables as a function of a third as "analysis of heterogeneous treatment effects." Thus, I introduced a fully non-parametric complement to parametric regression quite generally. Earlier "non-parametric" methodologies were always actually semi-parametric and allowed analysts to fudge results (whether purposefully or unintentionally).

The abstraction permitted generalization to comparing a subpopulation to the full population, to comparing two disjoint subpopulations, to comparing two matched or paired subpopulations, to arbitrary real-valued responses, to calibration of non-binary ratings, to conditioning on (or controlling for) multiple covariates, to weighted samples, to the simultaneous calibration of multiple subpopulations, and beyond. Details are in [1], [3], [5], [14], [15], [16], [17], [18]. Many collaborators have been involved. A recent overview presented at the International Conference on Machine Learning to an audience of hundreds is available at `https://zenodo.org/records/15253140` in the form of a slide deck as well as a video recording.

The methods likely should become standard for analyzing randomized controlled trials and observational studies, at least when controlling for appropriate covariates. My current employer, Meta, conducts randomized controlled trials in the form of A/B tests and conducts observational studies in the form of case-control studies; the methods we developed aid in their analysis, complementing parametric regressions. Parametric modeling has the major disadvantage of requiring a statistical model (perhaps fit via machine learning) that may or may not be correct and that depends strongly on the choices of the analyst.

Moving forward, I really ought to collect together everything into a comprehensive monograph. And I should prepare a general-purpose, industrial-strength, comprehensive software package, unifying all the open-source software that accompanies the papers referenced in the previous paragraphs. Such activities may have to wait for a return to academia, which seems more likely to support such activities. In the meantime, Meta has deployed the methodologies throughout the company.

3 FastMRI

This section summarizes a couple-year sojourn into artificial intelligence for medical imaging.

We made some key contributions to clinically accepted acceleration of MRI, building on earlier techniques in deep learning from NYU and the optical illusion that adding noise can sharpen an image — see, for example, "Why does grain increase acutance?" at `https://photo.stackexchange.com/questions/110750/why-does-grain-increase-acutance`. Errors from our acceleration and diagnostically sound dithering are smaller than the machine errors already present in MRI. Siemens, Philips, GE, and AIRS currently market FDA-approved products based partly on our work, as summarized in our blog post at `https://ai.facebook.com/blog/fastmri-breakthrough-shows-ai-accelerated-mris-interchangeable-with-slow-traditional-mris` and in the short CBS News clip at `https://www.youtube.com/watch?v=9ncABdfkzuU` on YouTube.

Let me emphasize the "we" in this fastMRI. My contributions were necessary to the success, but far from sufficient, and at least a couple others made significantly greater contributions than I. This was a manifestly team project and I was in no way the most important member of the team.

4 Randomized algorithms for linear algebra

This section summarizes contributions to numerical linear algebra based on randomized methods.

Randomization recently revolutionized numerical methods for linear algebra. We have been contributing to many aspects of this movement, especially with regard to improvements important in practice. Thus far, randomized methods for dimensionality reduction via low-rank approximation have had the most practical impact among those to which we contributed. My main contribution was to prove rigorous guarantees that randomization yields nearly optimal accuracy computationally efficiently uniformly for all matrices, completely irrespective of the distributions of their singular values. These rigorous bounds attain for parameter settings that would be absurd in the classical analyses; for instance, a classical analysis might indicate that good accuracy attains only after thousands of iterations, whereas the randomized approach requires only a couple iterations. Needless to say, no one in their right mind would (or did) use such parameters settings prior to the proofs that they make sense and would always work correctly.

The basic idea underlying the randomized algorithms for low-rank approximation is that the dimension of the range of a low-rank matrix is low and so is efficient to calculate by applying the matrix to a few random vectors. Once a basis for the range of a good low-rank approximation of the matrix is known, classical algorithms from numerical linear algebra can straightforwardly and efficiently construct the low-rank approximation. Vladimir Rokhlin and Tamás Sarlós independently brought this pair of observations of theirs to prominence. Working up an effective implementation and proving mathematically rigorous guarantees of nearly optimal accuracy was then easy given my background, at least for any matrix in which the sum of the squares of nearly all singular values is negligible.

Unfortunately, the basic algorithm fails miserably when the matrix being approximated is noisy; "noisy" means that the sum of the squares of the tail of singular values is larger than the desired accuracy of the low-rank approximation. Noise from the tail of singular values pollutes the range computed by applying the matrix to random vectors. Fortunately, suppressing the tail of singular values relative to the singular values which dominate the range is possible by applying the matrix being approximated to each random vector a few times in succession, requiring no more than a few matrix-vector multiplications each. I observed this empirically first but was soon able to prove rigorous guarantees:

My best-regarded contribution to randomized algorithms was a cumbersome proof that the error of the approximation decays at least exponentially with the number of matrix-vector multiplications, with the coefficient in the exponential decay being a universal constant, uniform over all matrices. In one of the greatest works in numerical analysis — see [4] — Joel Tropp greatly simplified and tightened the proof, leveraging Israel Gelfand's spectral-radius formula. Due to ignorance, I had effectively replicated in [10] a proof of the spectral-radius formula. Curiously enough, the Musco brothers in [7] and many others later tightened the guarantees beyond Joel Tropp's, abandoning the latter's supremely elegant method for my cumbersome approach (much to my consternation). At any rate, these methods have become the standard for large-scale principal component analysis, including the default in the popular Python library, SciKit-Learn.

5 Efficient algorithms for special-function transforms

This section summarizes analogues of the fast Fourier transform for changes of basis involving basis functions other than the simple sinusoids of the usual Fourier transforms.

The usual Fourier transforms (including expansions into Fourier series) are linear combinations of sinusoids, with the coefficients in the linear combinations specifying the functions being transformed. However, there are many other similar spectral transforms that use other special functions instead. The second-most popular (after the original Fourier transforms) are probably spherical-harmonic transforms. Spherical-harmonic transforms take linear combinations of spherical harmonics or, after separating variables, associated Legendre functions. Other popular transforms take linear combinations of Bessel functions.

In all cases, there were large literatures proposing fast algorithms for the associated spectral transforms prior to any involvement from me. However, none of the proposed algorithms had demonstrated numerical stability, with most producing infinities and NaNs at even small sizes of transforms in implementations for standard finite-precision floating-point arithmetic. Arbitrary-precision arithmetic could stabilize the proposed algorithms, but at the cost of the fast, scalable algorithms becoming slow.

Effective, provably numerically stable analogues of the fast Fourier transform emerged in [11], [12], and [13]. There are two classes of methods. The careful organization of computations for especially large scales in [13] leveraged the Michielssen-Rokhlin generalization of the fast Fourier transform's butterfly, introduced by [6] and [8]. The other class leverages the Greengard-Rokhlin fast-multipole-method and its modern variants based on matrix compression, such as that of [9]. This latter class combines the fast orthonormal-basis-change of [2] with the classical fact that the values of special functions which satisfy three-term recurrence relations are the entries in the eigenvectors of certain tridiagonal matrices. (There is also a related, albeit much more complicated, less efficient generalization based on semi-separable matrices in [11]. The blame for starting with the senseless complication of semi-separable matrices lies with me. In my defense, though, I did discover the later simplification that now dominates.) The two classes have various pros and cons relative to each other; the common thread appears to be Vladimir Rokhlin's involvement.

References

[1] Imanol Arrieta-Ibarra, Paman Gujral, Jonathan Tannen, Mark Tygert, and Cherie Xu. Metrics of calibration for probabilistic predictions. *J. Mach. Learn. Res.*, 23:1–54, 2022. Available at `https://arxiv.org/abs/2205.09680`.

[2] Ming Gu and Stanley C. Eisenstat. A divide-and-conquer algorithm for the symmetric tridiagonal eigenproblem. *SIAM J. Matrix Anal. Appl.*, 16:172–191, 1995. Available at `https://zenodo.org/records/1236142`.

[3] Ido Guy, Daniel Haimovich, Fridolin Linder, Nastaran Okati, Lorenzo Perini, Niek Tax, and Mark Tygert. Measuring multi-calibration. Technical Report 2506.11251, arXiv, 2025. Available at `https://arxiv.org/abs/2506.11251`.

[4] Nathan Halko, Per-Gunnar Martinsson, and Joel Tropp. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011. Available at `https://arxiv.org/abs/0909.4061`.

[5] Isabel Kloumann, Hannah Korevaar, Chris McConnell, Mark Tygert, and Jessica Zhao. Cumulative differences between paired samples. Technical Report 2305.11323, arXiv, 2023. Available at `https://arxiv.org/abs/2305.11323`.

[6] Eric Michielssen and Amir Boag. Multilevel evaluation of electromagnetic fields for the rapid solution of scattering problems. *Microw. Opt. Technol. Lett.*, 7(17):790–795, 1994.

[7] Cameron Musco and Christopher Musco. Randomized block Krylov methods for stronger and faster approximate singular value decomposition. In Corinna Cortes, Neil Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett, editors, *Proceedings of the 28th International Conference on Neural Information Processing Systems*, pages 1396–1404. Curran Associates, 2015. Available at `https://arxiv.org/abs/1504.05477`.

[8] Michael O'Neil, Franco Woolfe, and Vladimir Rokhlin. An algorithm for the rapid evaluation of special function transforms. *Appl. Comput. Harmon. Anal.*, 28(2):203–226, 2010. Available at `https://www.sciencedirect.com/science/article/pii/S1063520309000888`.

[9] Per-Gunnar Martinsson and Vladimir Rokhlin. An accelerated kernel-independent fast multipole method in one dimension. *SIAM J. Sci. Comput.*, 29(3):1160–1178, 2007. Available at `https://amath.colorado.edu/faculty/martinss/Pubs/2007_hudson.pdf`.

[10] Vladimir Rokhlin, Arthur Szlam, and Mark Tygert. A randomized algorithm for principal component analysis. *SIAM J. Matrix Anal. Appl.*, 31(3):1100–1124, 2009. Available at `https://tygert.com/blanczos.pdf`.

[11] Vladimir Rokhlin and Mark Tygert. Fast algorithms for spherical harmonic expansions. *SIAM J. Sci. Comput.*, 27(6):1903–1928, 2006. Available at `https://tygert.com/sph2.pdf`.

[12] Mark Tygert. Fast algorithms for spherical harmonic expansions, II. *J. Comput. Phys.*, 227(8):4260–4279, 2008. Available at `https://tygert.com/spharmonic.pdf`.

[13] Mark Tygert. Fast algorithms for spherical harmonic expansions, III. *J. Comput. Phys.*, 229(18):6181–6192, 2010. Available at `https://tygert.com/butterfly.pdf`.

[14] Mark Tygert. Controlling for multiple covariates. Technical Report 2112.00672, arXiv, 2021. Available at https://arxiv.org/abs/2112.00672.

[15] Mark Tygert. Cumulative deviation of a subpopulation from the full population. *J. Big Data*, 8(117):1–60, 2021. Available at https://arxiv.org/abs/2008.01779.

[16] Mark Tygert. A graphical method of cumulative differences between two subpopulations. *J. Big Data*, 8(158):1–29, 2021. Available at https://arxiv.org/abs/2108.02666.

[17] Mark Tygert. Calibration of P-values for calibration and for deviation of a subpopulation from the full population. *Adv. Comput. Math.*, 49(70):1–22, 2023. Available at https://arxiv.org/abs/2202.00100.

[18] Mark Tygert. Cumulative differences between subpopulations versus body mass index in the behavioral risk factor surveillance system data. Technical Report 2411.07399, arXiv, 2024. Available at https://arxiv.org/abs/2411.07399.