Newsjournal of the Society for Industrial and Applied Mathematics sinews.siam.org

siam news

Volume 52/ Issue 2
March 2019

Learning from Their Mistakes: Self-Calibrating Sensors

By Benjamin Friedlander, Shuyang Ling, and Thomas Strohmer

The Internet of Things (IoT) contains billions of sensors that provide information for a large number of measurands. Many of these sensors are embedded in complex systems and can be deployed in remote locations. Careful calibration of such sensors, which is essential for optimal results, is often difficult or even impossible to achieve in practice. Indeed, the need for precise calibration of sensing devices—ranging from tiny sensors to space telescopes—manifests itself as a major roadblock in many scientific and technological endeavors beyond the IoT. In this context, calibration is an effort to correct for specific uncertainties or aberrations in the measurement process.

Consider the calibration of antenna arrays to correct gain/phase offsets in received data, a common problem in *direction-of-arrival estimation* when engineers attempt to calculate the direction of propagating waves based on data received from the antennae (see Figure 1) [6, 7]. Another instance is *blind deconvolution*, the issue of recovering a signal from its noisy convolution with a poorly-known or unknown point spread function [3]. This problem occurs in diverse fields, such as astronomical imaging and audio processing. Other instances arise in wireless communication [4], cryo-electron microscopy [12], and X-ray crystallography [11].

It is therefore highly desirable to equip sensors and systems with the capability for *self-calibration* using information collected by the system to simultaneously estimate the calibration parameters and perform the system's intended function (image reconstruction, signal estimation, target detection, etc.) [7, 9]. With some poetic license, we might say that sensors learn from their mistakes; instead of physically recalibrating themselves (usually an impractical or impossible task), they conduct a form of *virtual self-calibration* by correcting errors in the sensing process via carefully-constructed algorithms.

We can express many self-calibration problems in the following mathematical form:

$$y = A(\theta, x) + w, \qquad (1)$$

where y represents the measurements, $A(\theta,\cdot)$ is the sensing operator dependent on some unknown or imprecisely known parameters θ, x is the information we wish to recover, and w denotes additive noise. A may depend linearly or nonlinearly on θ, while the properties of sensing uncertainty θ depend largely on the application.

The generality of (1) renders it incapable of developing a rigorous and efficient framework for its solution. So we focus

See **Self-Calibrating Sensors** *on page 3*

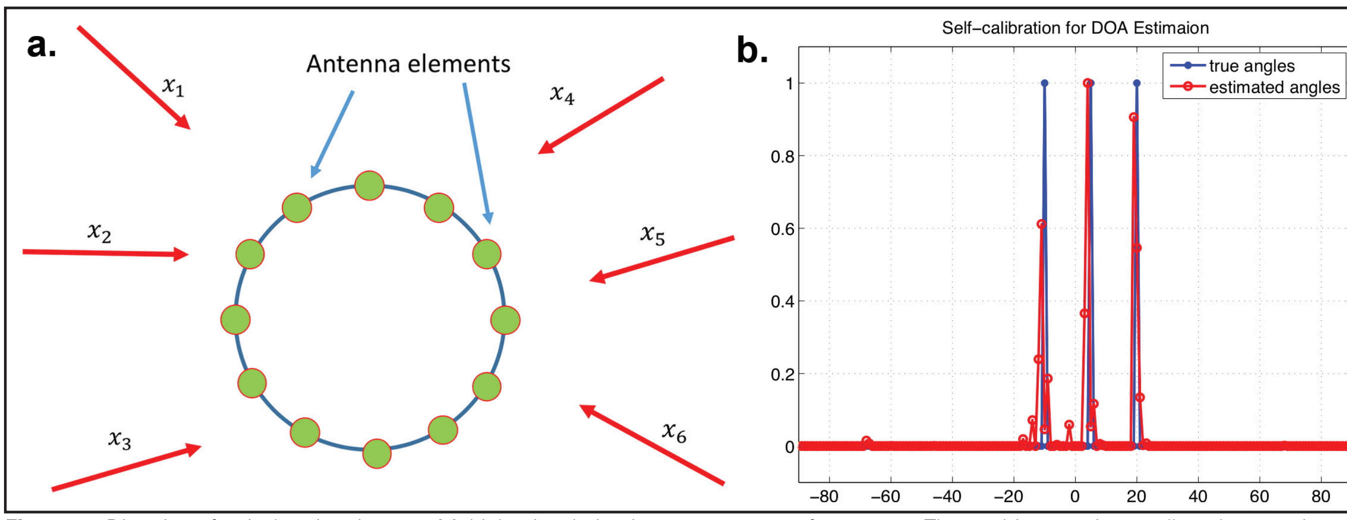


Figure 1. *Direction-of-arrival estimation.* **1a.** *Multiple signals impinge on an array of antennae. The goal is to estimate all angles x_i using the (noisy) data received at the array. Here we assume that the antenna gains are unknown.* **1b.** *A typical example of how one of the methods can solve this self-calibration problem. Figure 1a courtesy of authors and 1b courtesy of [6, 9].*

Nonprofit Org
U.S. Postage
PAID
Permit No 360
Bellmawr, NJ

siam
SOCIETY for INDUSTRIAL and APPLIED MATHEMATICS
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA

Algorithmic Trading in Competitive Markets with Mean Field Games

By Philippe Casgrain and Sebastian Jaimungal

Algorithms designed for automated trading on financial markets have existed for at least two decades but became ubiquitous with the creation of electronic exchanges. Because of their lightning-fast reaction times and ability to process huge quantities of data in real time, such algorithms are preferable to manual traders for intra-day trading.

Due to the speed and volume of information, trading decisions must be made without human intervention and designers must be conscious of market complexities. As all models are merely approximations, an ideal algorithm should learn from its environment and dynamically adapt its strategy. Some of the earliest mathematical work in algorithmic trading focused on the execution problem [1], but researchers have since devoted much time to areas like market-making, statistical arbitrage, and optimal tracking of stochastic targets [3].

The Limit Order Book and Price Impact

Market complexity stems from the millions of traders who continuously interact with one another to form prices. These interactions showcase themselves in the dynamics of the *limit order book* (*LOB*), which contains the outstanding collection of *limit orders* (*LOs*) that traders are willing to buy and sell assets at. Incoming *market orders* (*MOs*) are matched with the best available prices (see Figure 1) and gradually chip away at the LOB. The combined actions of posting (and cancelling) LOs and executing MOs move prices according to general supply and demand for the asset in question. Although individual trader impact is minuscule, the accumulation of all traders' actions is significant.

The Market Model

We develop a very general market model where a large population N of intelligent heterogenous agents trades against each other in a market with latent factors [4]. Our model is inspired by studies that address single-agent problems accounting for order flow and latent factors [5], as well as multiple homogeneous agents without latent factors [2, 6]. These agents' actions, along with exogenous factors accounting for other traders' actions, drive the (controlled) asset price process $S^\nu = (S_t^\nu)_{t\in[0,T]}$. For simplicity, all agents trade continuously at rates $\{(\nu_t^j)_{t\in[0,T]}\}_{j=1}^N$. The midprice satisfies the stochastic differential equation

$$dS_t^\nu = \Big(A_t + \underbrace{\frac{\lambda}{N}\Sigma_{j=1}^N \nu_t^j}_{\text{Order-Flow Impact}}\Big)\, dt + dM_t, \qquad (1)$$

where $A = (A_t)_{t\in[0,T]}$ represents the mean excess drift on the asset price and $M = (M_t)_{t\in[0,T]}$ is a martingale representing an exogenous noise source. The term $\frac{\lambda}{N}\sum_{j=1}^N \nu_t^j$ acccounts for the effect of net order flow on price; excess buy or sell pressure pushes prices up or down respectively.

Agents' strategies are adapted only to asset price S, total order flow $\overline{\nu}^{(N)} = \sum_{j=1}^N \nu_t^j$, and their own holdings $Q^j = (Q_t^j)_{t\in[0,T]}$. Thus, A and M are invisible to agents and may contain latent factors in the market. In addition, other agents' inventories are invisible to any one given agent. As part of the stochastic game's solution, agents must filter the excess drift $\hat{A}_t = \mathbb{E}[A_t \,|\, \mathcal{F}_t]$ from the visible filtration.

Agents trade over the interval $[0,T]$ and aim to maximize their own objective functional

See **Algorithmic Trading** *on page 2*

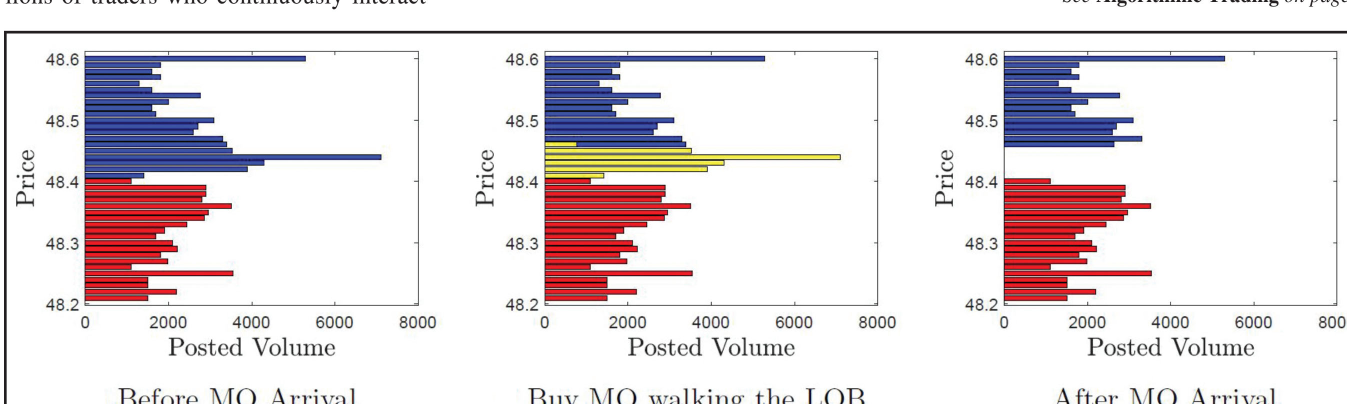
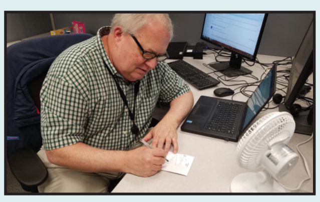


Figure 1. *The limit order book (LOB) of Intel Corporation stock at 10:36 on March 26, 2018, as a buy market order (MO) for 21,000 arrives. Blue bars represent the available volume of sell orders at shown price, red bars indicate the available volume of buy orders at shown price, and yellow bars designate limit orders (LOs) matching incoming MOs. Figure courtesy of Sebastian Jaimungal.*

siam news
Volume 52/ Issue 2/ March 2019



Algorithmic Trading

Continued from page 1

$$H_j(\nu^j,\nu^{-j}) = \mathbb{E}\Bigg[\underbrace{-\int_0^T (S_u^\nu - a\nu_u^j)\,\nu_u^j du}_{\text{Cash from Trading}} + \underbrace{Q_T^{j,\nu^j}\left(S_T^\nu - \Psi Q_T^{j,\nu^j}\right)}_{\text{Liquidation Cost}} - \underbrace{\phi\int_0^T (Q_u^{j,\nu^j})^2 du}_{\text{Model Risk}}\Bigg], \quad (2)$$

where $\nu^{-j} := \{\nu^1,\dots,\nu^{j-1},\nu^{j+1},\dots,\nu^N\}$ denotes the actions of all agents except agent $-j$. This objective represents a combination of three quantities: the accumulated cash from trading, a liquidation cost for holding inventory at time T, and a running penalty that accounts for model risk. The relative importance of these penalties is controlled by $\Psi,\phi \geq 0$. *All* agents' actions affect the objective through the asset price process S_t^ν.

The goal is to obtain a Nash equilibrium, the collection of strategies such that

$$H_j(\nu^j,\nu^{-j,*}) \leq H_j(\nu^{j,*},\nu^{-j,*}), \quad (3)$$

for all $j \in \{1,\dots,N\}$ and $\nu^j \in \mathcal{A}_j$, where $\mathcal{A}_j$ is the set of admissible strategies for agent $-j$. No agent can improve by unilaterally deviating from the Nash equilibria.

The Mean Field Game Approximation

Obtaining the Nash equilibria for the finite player game is difficult. As an alternative, we take the limit $N \to \infty$ to obtain a mean field game (MFG) and apply a version of the optimal MFG strategy to the finite player game. While the MFG problem itself is still challenging to solve, we demonstrate that one can apply convex analysis tools [4] rather than dynamic programming techniques or the stochastic Pontryagin maximum principle, as is typically done in MFG problems. For the MFG limit, [4] uses the following approach: (i) take the mean field trading rate $\overline{\nu}$ as given; (ii) maximize each agent's strictly concave objective functional by setting its Gâteaux derivative to zero; (iii) derive a system of forward-backward stochastic differential equations (FBSDEs), which induces the Gâteaux derivative to vanish:

$$\begin{cases} -d(2a_k\nu_t^{j,*}) = \left(\mathbb{E}^{\mathbb{P}^k}\left[A_t^k + \lambda_k^T\overline{\nu}_t^* \mid \mathcal{F}_t^j\right] - 2\phi_k q_t^{j,\nu^{j,*}}\right)dt - d\mathcal{M}_t^j, \\ 2a_k\nu_T^{j,*} = -2\Psi_k q_T^{j,\nu^{j,*}}; \end{cases} \quad (4)$$

(iv) solve the FBSDEs; and (v) average over all agents' optimal strategy and equalize it to the initial mean field trading rate $\overline{\nu}$. In other words, we obtain a fixed point on the space of controls that simultaneously optimizes each agent's objective functional, resulting in a Nash equilibrium.

Carrying out this program, [4] proves that the expression

$$\overline{\nu}_t = (2a)^{-1}(g_{1,t} + g_{2,t}\,\overline{q}_t^\nu), \quad (5)$$

where $g_{2,t}$ is a deterministic matrix-valued function, yields the optimal mean field strategy. In the above,

$$g_{1,t} = \int_t^T \mathbb{E}\left[\xi_t^{-1}\xi_u\hat{A}_u \mid \mathcal{F}_t\right]du, \quad (6)$$

where ξ_t is a stochastic, matrix-valued process. All subpopulations are interlinked and cannot be factorized.

The term $g_{2,t}\,\overline{q}_t^\nu$ in (5) pulls the mean field inventories towards zero and corresponds to an *optimal execution* component of the strategy. The term $g_{1,t}$ incorporates predictions about the filtered future latent states and corresponds to the strategy's *statistical arbitrage* component. Estimating the model parameters and filtering requires setup of a *machine learning* problem to "learn" the behaviour of prices and latent states. Figure 2 depicts a graphical model of the discretisation of (1), and [5] demonstrates parameter estimation via an expectation-maximisation and forward-backward algorithm.

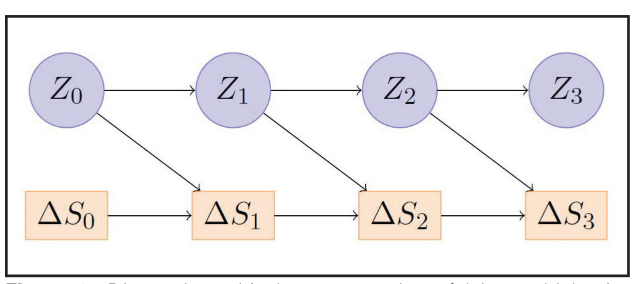


Figure 2. *Directed graphical representation of (observable) price changes ΔS_t and (unobservable) latent states Z_t for the continuous time model in (1). Figure adapted from [5].*

For an individual agent $-j$ in subpopulation $-k$, we obtain the Nash equilibrium at

$$\nu_t^j = \overline{\nu}_t^k + \frac{1}{2a_k}h_{2,t}^k\left(q_t^{j,\nu^j} - \overline{q}_t^{k,\nu}\right), \quad (7)$$

where $h_{2,t}^k < 0$ is a subpopulation-specific deterministic function. Thus, the individual trades at a rate that pulls his/her inventory towards the subpopulation's mean field inventory. We prove that such strategies form an ϵ-Nash equilibrium when applied to the finite player game; i.e., agents may improve by unilaterally deviating from the MFG strategy, but only by an amount ϵ with $\epsilon \to 0$ as $N \to \infty$ [4].

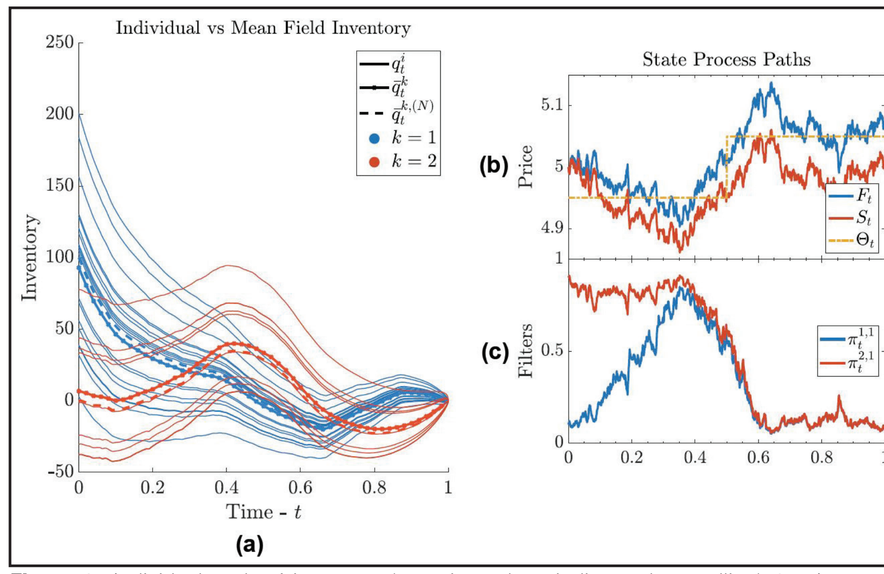


Figure 3. *Individual traders' inventory (negative values indicate short selling).* **3a.** *Inventory paths: the blue population is more urgent than the orange population. Broken lines represent subpopulation averages and dotted lines represent the corresponding mean fields.* **3b.** *Price path: midprice process S_t, unimpacted midprice F_t (subtracting order flow), and latent Markov chain Θ_t.* **3c.** *Posterior probability: estimated posterior probability of $\Theta_t = 4.95$. Figure courtesy of Philippe Casgrain and Sebastian Jaimungal.*

A Simulated Example

Figure 3 shows how two subpopulations with differing goals and beliefs interact and react to the market. The midprice follows a pure jump process that mean-reverts, where a latent two-state Markov chain Θ_t modulates the mean-reversion level. All agents aim to fully liquidate their holdings by time $T = 1$, have different urgencies in doing so, and agree on possible mean-reversion levels. The two subpopulations, however, differ on the prior distribution of the latent Markov chain's initial state and on their urgency in unwinding. The results illustrate the strategies' complexities; one group liquidates its position while the other performs statistical arbitrage, and both account for each other's impact.

Many interesting questions remain. For example, how do agents account for uncertainty in their selected models? And how can we correct the strategy to factor in finite population size? Future exploration pertaining to the ways in which traders resolve the bid-ask spread, nonlinear trading impact, and/or trade of nonlinear contracts would also be valuable.

References

[1] Almgren, R., & Chriss, N. (2001). Optimal execution of portfolio transactions. *J. Risk, 3*, 5-40.

[2] Cardaliaguet, P., & Lehalle, C.-A. (2018). Mean field game of controls and an application to trade crowding. *Math. Fin. Econ., 12*(3), 335-363.

[3] Cartea, Á., Jaimungal, S., & Penalva, J. (2015). *Algorithmic and High-frequency Trading*. Cambridge, U.K.: Cambridge University Press.

[4] Casgrain, P., & Jaimungal, S. (2018). Algorithmic trading with partial information: A mean field game approach. Preprint, *arXiv:1803.04094*.

[5] Casgrain, P., & Jaimungal, S. (2018). Trading algorithms with learning in latent alpha models. *Math. Fin.*

[6] Huang, X., Jaimungal, S., & Nourian, M. (2015). Mean-field game strategies for a major-minor agent optimal execution problem. *SSRN Electronic J.*

Philippe Casgrain recently obtained his Ph.D. in mathematical finance from the Department of Statistical Sciences at the University of Toronto. Sebastian Jaimungal is a professor of mathematical finance in the Department of Statistical Sciences at the University of Toronto. He is the current chair of the SIAM Activity Group on Financial Mathematics and Engineering, an editorial board member of the SIAM Journal on Financial Mathematics, and a managing editor for Quantitative Finance.

siam news

ISSN 1557–9573. Copyright 2019, all rights reserved, by the Society for Industrial and Applied Mathematics, SIAM, 3600 Market Street, 6th Floor, Philadelphia, PA 19104-2688; (215) 382-9800; siam@siam.org. To be published 10 times in 2017: January/February, March, April, May, June, July/August, September, October, November, and December. The material published herein is not endorsed by SIAM, nor is it intended to reflect SIAM's opinion. The editors reserve the right to select and edit all material submitted for publication.

Advertisers: For display advertising rates and information, contact Kristin O'Neill at marketing@siam.org.

One-year subscription (nonmembers): Electronic-only subscription is free. $73.00 subscription rate worldwide for print copies. SIAM members and subscribers should allow eight weeks for an address change to be effected. Change of address notice should include old and new addresses with zip codes. Please request address change only if it will last six months or more.

AMS-SIAM Norbert Wiener Prize in Applied Mathematics

SIAM and the American Mathematical Society (AMS) jointly award the Norbert Wiener Prize in Applied Mathematics for an outstanding contribution to applied mathematics in the highest and broadest sense. The recipient must be a member of one of these two societies. The prize was established in 1967 in honor of Norbert Wiener and endowed by a fund from the Department of Mathematics at the Massachusetts Institute of Technology. The endowment was further supplemented by a generous donor. The 2019 Norbert Wiener Prize is awarded to Marsha Berger (New York University) and Arkadi Nemirovski (Georgia Institute of Technology).

Berger is being recognized for her fundamental contributions to adaptive mesh refinement (AMR) and Cartesian mesh techniques by automating the simulation of compressible flows in complex geometry.

AMR algorithms can improve the accuracy of a partial differential equation's solution by locally and dynamically resolving a simulation's complex features. Berger helped invent AMR. She introduced the block-structured approach to AMR in her Ph.D. thesis and later developed the Berger-Oliger algorithm and the Berger-Colella algorithm with Joseph Oliger and Phillip Colella respectively. Berger provided the mathematical foundations, algorithms, and software that allowed the solution of many otherwise intractable simulation problems, including those related to blood flow, climate modeling, and galaxy simulation. She is part of the team that created Cart3D, a NASA code based on her AMR algorithms that is extensively used for aerodynamic simulations and was instrumental in understanding the Space Shuttle Columbia disaster.

Berger received her Ph.D. in computer science from Stanford University in 1982. She conducted postdoctoral research at New York University's Courant Institute of Mathematical Sciences and is currently a Silver Professor of Computer Science and Mathematics in the institute's Computer Science Department, where she has been since 1985.

Berger's honors include membership in the National Academy of Sciences, the National Academy of Engineering, and the American Academy of Arts and Sciences. She is also a Fellow of SIAM. Berger was the 2004 recipient of the Institute of Electrical and Electronics Engineers Computer Society's Sidney Fernbach Award, and was part of the team that won NASA's 2002 Software of the Year Award for its Cart3D software.

Upon learning of her receipt of the Norbert Wiener Prize, Berger expressed her delight and extended her gratitude to colleagues. "What a thrill to learn that I will be one of the recipients of the 2019 Norbert Wiener Prize," she said. "One of the main enjoyments of my research is developing tools that others can use to solve real problems in aerodynamics, tsunami modeling, etc. This has been possible because of the collaborators I have been fortunate to meet, starting with Phil Colella and Antony Jameson, and later Randy LeVeque and Michael Aftosmis, along with a number of postdocs."

"I am particularly pleased that this kind of research is being recognized," she continued. "The AMR and Cartesian grid projects have both required the creation of new techniques in mathematics and computer science. They were decade-long efforts during which my colleagues and I developed theory and algorithms while paying attention to important practical aspects of their use in realistic geometries. Complicated algorithms have complicated implementations, and accuracy, robustness, and performance are all essential parts of the research."

Nemirovski is being honored for his fundamental contributions to high-dimensional optimization and discovery of key phenomena in the theory of signal estimation and recovery.

A powerful and original developer of the mathematics of high-dimensional optimization, Nemirovski—along with David Yudin—invented the ellipsoid method that Leonid Khachiyan used to show (for

See **Norbert Wiener Prize** *on page 5*

Self-Calibrating Sensors

Continued from page 1

instead on some important special cases. A seemingly simple yet surprisingly useful case is the model

$$y = D(\theta)Ax + w, \tag{2}$$

where the matrix A and vector y are known but the diagonal matrix $D(\theta)$ and vector x are not. As before, w is additive noise and thus unknown as well, though we may have statistical information about it. We can also interpret the goal of finding x as solving a linear system where an *unknown* gain $D_i(\theta)$ rescales the ith row of the system matrix A.

One may assume that we have reached a level of simplicity with (2) that is too trivial to analyze mathematically and no longer useful in practice. However, as is often the case in mathematics, an outwardly simple model can be deceptive. On the one hand, a rigorous analysis of the aforementioned diagonal calibration model requires nontrivial mathematical tools; on the other, this "simple" model arises in numerous important applications [9].

One way to tackle this situation is by solving a nonlinear least squares problem

$$\min_{\theta,x} \|D(\theta)Ax - y\|^2. \tag{3}$$

This scenario is far more intricate than linear least squares, where a well-defined solution often exists. The objective function's nonconvexity may result in many local minima. In fact, recovering x is not necessarily an easy task even if θ is completely known, especially when the linear system is underdetermined. If D depends linearly on θ, then (3) becomes a bilinear problem, which should make its solution easier; alas, bilinear problems are non-deterministic polynomial-time-hard in general. We will describe how recent advances in optimization can resolve these issues in many cases of interest.

Compressive sensing [2, 5] has become a game changer in modern signal processing. Using convex optimization, we can exploit the signal's sparsity and accelerate the sensing process tremendously. Unlike additive perturbations in the measurement matrix, sparse signal reconstruction from compressive measurements is sensitive to multiplicative perturbations. The linear dependence of D on θ results in a *bilinear compressive sensing* problem: how do we estimate the calibration parameter θ and sparse x from their *bilinear* measurements? For example, after proper discretization in direction-of-arrival estimation, the location of nonzero entries of x represents the direction of waves and $D(\theta)$ denotes the antennas' unknown gains [6, 7].

We have developed a convenient method called *SparseLift* to tackle such bilinear

Figure 2. *Blind deconvolution.* **2a.** *Blurred image of a flying swan with unknown blurring kernel.* **2b.** *Recovered image via regularized gradient descent method. Original unblurred image (not pictured) courtesy of Steve Byland.*

compressive sensing problems [9]. Linear algebra reveals that the recovery of two vectors (θ, x) is equivalent to estimating a sparse rank-1 matrix $\theta x^\top$ in this bilinear inverse problem. Due to the bilinearity, the measurement y is actually a *linear* function on the matrix space. We denote this linear map as $\mathcal{A}$ and estimate $\theta x^\top$ by exploiting the sparsity of the "lifted" rank-1 matrix $\theta x^\top$ via ℓ_1-regularization:

$$\min \|Z\|_1, \quad \text{s.t.} \quad \mathcal{A}(Z) = y, \tag{4}$$

where $\|Z\|_1$ is the ℓ_1-norm of the vectorized Z. Ideally we have $\theta x^\top = Z$, but in practice we compute the leading left and right singular vectors of Z to extract θ and x from Z (up to a scalar). Indeed, SparseLift recovers (θ, x) successfully if the number of constraints equals $\tilde{\Omega}(\|\theta x^\top\|_0)$, where $\|\theta x^\top\|_0$ is the cardinality of nonzero entries in $\theta x^\top$. Many variations and extensions of (4) naturally exist.

Besides direction-of-arrival estimation with unknown antenna gains, we can express various other problems—either directly or after some proper transform—in the form of (2). The blind deconvolution problem is perhaps the most widely-known example. How can we reconstruct two signals (f, g) from their convolution $y = f * g$? This highly ill-posed inverse problem pervades many areas of science and technology, such as image deblurring, wireless communication [1], and spike detection in neuroscience. The blind deconvolution problem is equivalent to the self-calibration model (2) in the frequency domain. If we take the Fourier transform of f, then $\hat{y} = \hat{f} \circ \hat{g}$, where $\hat{f}$ represents the Fourier transform of f and "$\circ$" signifies entrywise multiplication. Now this exactly fits (2) by setting $D(\theta) = \hat{f}$ and $A = \widehat{W}$, where $g = Wx$ is the image/function, W is a dictionary to represent g (e.g., wavelet basis), and x denotes the coefficients (possibly sparse or approximately sparse).

We propose a simple gradient-descent-based algorithm to this blind deconvolution problem [8]. The algorithm consists of two steps: construction of a suitable initial guess via the spectral method and application of the gradient descent method to the objective function. The proposed algorithm is robust in the presence of noise and comes with rigorous convergence guarantees under relatively mild assumptions.

Figure 2 depicts an example. Figure 2a portrays convolution of an image with an unknown motion blurring kernel, while 2b shows the reconstructed image via the aforementioned method [8]. This approach outperforms convex alternatives in computation time and sampling complexity. The framework also allows us to solve joint blind deconvolution and demixing problems that can arise in multiuser communication scenarios of the IoT [10, 13].

While we have barely scratched the surface of self-calibration in this article, we hope we have communicated its versatility as a playground for mathematicians — full of difficult theoretical problems, interesting numerical challenges, and cutting-edge applications.

References

[1] Ahmed, A., Recht, B., & Romberg, J. (2014). Blind deconvolution using convex programming. *IEEE Trans. Inform. Theory, 60*(3), 1711-1732.

[2] Candès, E.J., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory, 52*(2), 489-509.

[3] Chaudhuri, S., Velmurugan, R., & Rameshan, R. (2014). *Blind Image Deconvolution: Methods and Convergence*. New York, NY: Springer.

[4] Ding, Z., & Li, Y. (2001). *Blind Equalization and Identification*. New York, NY: CRC Press.

[5] Donoho, D.L. (2006). Compressed sensing. *IEEE Trans. Inform. Theory, 52*(4), 1289-1306.

[6] Friedlander, B., & Strohmer, T. (2014). Bilinear compressed sensing for array self-calibration. In *2014 48th Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, CA.

[7] Friedlander, B., & Weiss, A.J. (1991). Self-calibration for high-resolution array processing. In S. Haykin (Ed.), *Advances in Spectrum Analysis and Array Processing* (Vol. II) (pp. 349-413). Upper Saddle River, NJ: Prentice Hall.

[8] Li, X., Ling, S., Strohmer, T., & Wei, K. (2018). Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *Appl. Comput. Harmon. Anal.*

[9] Ling, S., & Strohmer, T. (2015). Self-calibration and biconvex compressive sensing. *Inv. Prob., 31*(11), 115002.

[10] Ling, S., & Strohmer, T. (2017). Regularized gradient descent: A nonconvex recipe for fast joint blind deconvolution and demixing. *Inform. Infer. J. IMA*., iax022.

[11] Philipp, H.T., Ayyer, K., Tate, M.W., Elser, V., & Gruner, S.M. (2012). Solving structure with sparse, randomly-oriented x-ray data. *Opt. Exp., 20*(12), 13129-13137.

[12] Singer, A. (2018). Mathematics for cryo-electron microscopy. Preprint, *arXiv:1803.06714*.

[13] Wunder, G., Boche, H., Strohmer, T., & Jung, P. (2015). Sparse signal processing concepts for efficient 5G system design. *IEEE Acc., 3*, 195-208.

Benjamin Friedlander is a professor of electrical engineering at the University of California, Santa Cruz, studying statistical signal processing and its applications to wireless communications and array processing. Shuyang Ling is a Courant Instructor at the Courant Institute of Mathematical Sciences at New York University. His research interests focus on the mathematics of data science, including signal processing, inverse problems, optimization, and machine learning. Thomas Strohmer is a professor of mathematics at the University of California, Davis. His research interests include applied harmonic analysis, mathematics of information, signal and image processing, and machine learning.

Obituary: Zeev Schuss

By Amit Singer and David Holcman

Zeev Schuss, professor emeritus of applied mathematics at Tel Aviv University (TAU), passed away on July 29, 2018. Zeev was born in Poland in 1937. He graduated from the Tel Aviv Academy of Music in 1963 with a degree in composition, conducting, and theory, and earned a degree in mathematics two years later. Zeev received his Ph.D. in mathematics from Northwestern University in 1970, working with Avner Friedman on classical analysis of partial differential equations (PDEs). He went on to become a professor at TAU and chaired the Department of Applied Mathematics from 1993 to 1995.

During the course of his career, Zeev published over 200 papers in pure and applied math, chemistry, physics, engineering, and biology. He also wrote six books in applied mathematics, published by Springer and Wiley. He supervised dozens of M.Sc. and Ph.D. students in the aforementioned disciplines, many of whom hold positions at prestigious institutions around the world.

It took time for Zeev to define the meaning of applied mathematics for himself. After attaining his Ph.D., he decided to move in a more applied direction. The starting point was a class that he took with Henry McKean about stochastic processes and the asymptotics of differential equations that McKean had developed with Bernard Matkowsky. Zeev then discovered Sivaramakrishna Chandrasekhar's 1943 review of thermal activation escape from an attractor. He was able to obtain a formal asymptotic formula as a solution to the exit problem in n dimensions. News of this work spread following a talk Zeev delivered at New York University's Courant Institute of Mathematical Sciences in the 1970s. Although his result was not considered mathematically rigorous, Zeev recognized its newness and proceeded in this direction for the next 40 years. It was this type of mathematics that he wanted to develop and apply to the sciences, engineering, technology, etc. Rather than discover formal proofs, he aimed to define and accurately execute new computations, as well as find new physical mechanisms through modeling and novel explanations from calculations.

Closed formulas, obtained by asymptotic approximation to solutions of PDEs, are among the most robust and efficient tools for uncovering physical laws. These formulas deal precisely with manipulation of infinities and are thus very relevant in understanding the studied systems' refined properties. Zeev made several significant contributions by applying asymptotics to rare events, such as thermally activated escape from an attractor in physics and chemistry, and loss of lock in signal tracking.

With his collaborators, friends, and students, Zeev developed new tools to analyze data about selectivity of ionic channels (the selection of ions in a channel pore that is only a few atomic diameters in size). Another of his innovations was the narrow escape theory (escape of a stochastic particle from a narrow window), which inspired many scientific communities of applied mathematicians, physicists, biophysicists, and computational biologists, in addition to the fourth episode of the television series *Fargo*. Zeev continued to work through retirement, authoring five of his six books during this time.

Zeev Schuss (1937-2018). He is pictured here with Bob Eisenberg's granddaughter. Image courtesy of Bob Eisenberg.

Throughout his life, Zeev demonstrated that there are no standard academic paths for a career in applied mathematics. This is a lesson for all scholars: do what you like, because nobody knows where the next revolution in applied mathematics lies.

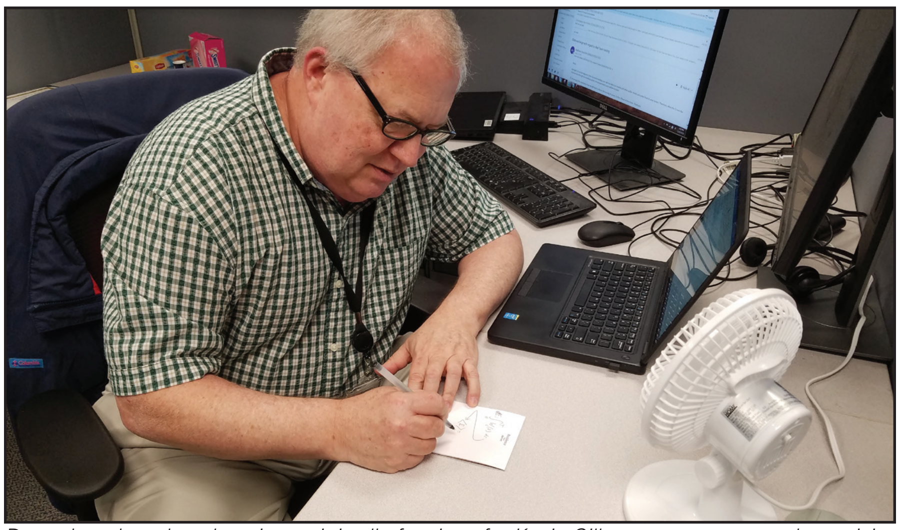
Amit Singer is a professor of mathematics in the Program in Applied and Computational Mathematics at Princeton University. David Holcman is Director of Research at École Normale Supérieure in Paris, France and a fellow of Churchill College in Cambridge, U.K.

Ex Numeris: Confessions of a Common Mathematician (with apologies to Anne Fadiman)

By Kevin Gillette

I am a common mathematician, thoroughly mathematical by training, temperament, and trade. I've trod weary footpaths through numerous industrial sectors, plying my skills in areas as diverse as banking, transportation, telecommunications, retail, and even venture capital. My appetite for mathematical application is quite ecumenical; although I focus on optimization and discrete mathematics, I embrace *any* opportunity to contribute to the world amenity through judicious employment of mathematical principles and practices. Yes, I am a common mathematician.

It was not always this way. My original intent upon entering university was to become a biochemist. Like many scientists before and after me, I aimed (in no particular order) to discover a cure for cancer; find a solution to Type 1 diabetes; win a Nobel Prize in Chemistry; and *not* set fire to, blow up, or otherwise damage any laboratories or chemical facilities. As luck would have it, I achieved none of these goals. Indeed, my migration from chemistry to mathematics was precipitated by an unfortunate "incident" in the laboratory of my then-advisor, James Collman, when a combination of my maladroitness and ignorance inadvertently caused a reaction vessel to catch fire. The details, I believe, are unnecessary [ahem].



Deep thoughts, deep learning — it is all of a piece for Kevin Gillette, a common mathematician at Accenture Federal Services. Photo credit: Joon Yoon.

Suitably abashed, I abandoned my goal of becoming a chemist of any stripe. Instead, I accepted a lateral "promotion" within Stanford University's School of Humanities and Sciences and switched my major to mathematics. In those days (the late 1970s and early 1980s), one could receive either a B.A. in (pure) mathematics or a B.S. in (applied) mathematical sciences. Being of a pragmatic bent, I chose the latter, which sparked my interest in operations research. Excess credits from both advanced placement classes in high school and an aggressive freshman course load afforded me a chance to simultaneously pursue my bachelor's and master's degrees. I leapt at the opportunity to save time and money while scoring two degrees within four years at an expensive school like Stanford.

CAREERS IN MATHEMATICAL SCIENCES

Upon graduating with both degrees, I faced the perennial question that many young SIAM members encounter: What comes next? I had always fancied being a college professor, and contemplated working towards a Ph.D. in math or a related discipline. But just how much PTA (pain/torture/agony) would I endure along the way, between the whims and vicissitudes of advisors and faculty and the subsequent POP (publish or perish) mentality that pervades academia? Achieving tenure requires suffering for one's art, and I've never savored the idea of being a starving artist, irrespective of the art form (yes, mathematics is indeed an art form).

My practical and productive upbringing steered me towards an industrial career. I had happily completed an undergraduate internship at Bank of America and evidently left a good impression on the team; they invited me to join them full-time upon graduation, which I did. Mathematical finance is now all the rage, and ample job opportunities—and commensurate salaries—exist for mathematicians who enjoy working with Ito calculus and stochastic differential equations. In the early 1980s, when such techniques were still relatively inchoate, the practicing mathematician had two basic paths: actuarial work or general consulting within the company. My team's title, "Management Sciences," aptly described its role. Our projects on behalf of the bank included issues like float management (cash that the bank can use for investments before it posts against account ledgers), portfolio diversification, and risk scoring. I spent 18 months working on these types of problems. My marriage and a desire for different work caused me and my new bride to relocate.

I moved to Dallas, Texas, where I continue to reside. My second industrial setting was transportation — specifically American Airlines, headquartered in the Dallas/Fort Worth area. I was invited to join their Operations Research department. Interesting work abounded there, including network optimization problems (crew assignments to flights, repair of crew and passenger assignments during weather complications, etc.); inventory-theoretic problems (rotable parts inventory stockpiles); queueing-theoretic problems (gate assignments in real time); and simulations (which replicated how complicated airports like Dallas/Fort Worth or Chicago O'Hare operated under different gating protocols, call center resource distribution, and so forth). I also applied my tradecraft as a systems support analyst in the Operations Engineering group. This team comprised the lion's share of American Airlines' aeronautical engineers and performed mission analyses for both existing and prospective aircraft fleet types, flight planning system studies and data grooming, and weight-and-balance examination for individual flights. I became an in-house expert on map projection equations and techniques for calculating route distances and headings — fairly quotidian calculations, but vitally important to get right the first time.

Since then, I have served in a wide variety of positions and contexts. I worked for nearly five years as a *de facto* senior engineer with MCI, where statistical analysis was exceedingly important. Following my tenure, I engaged in numerous contract programming and analytical assignments. One of them was a small (two-person) venture capital experiment where I vetted hard-science investment opportunities (chemistry, physics, geophysics, life sciences, etc.) for a select list of investors. This job required that I read through and scour nearly 50 peer-reviewed journals each month to get a feel for current research.

I also spent four years working at Blockbuster LLC prior to its ultimate demise. The need for statistical knowledge once again became paramount, as our business was primarily devoted to product placement and assortment. My team met the challenge of demand forecasting with a variety of techniques, most of them standard (principal component analysis, clustering, autoregressive integrated moving average, and so on).

All of this experience served as prelude to my current position as a principal in the analytics practice at Accenture Federal Services, a domestic subsidiary

See **Ex Numeris** *on page 7*

A New Mathematical Field Answers Old Questions

Reverse Mathematics: Proofs from the Inside Out. *By John Stillwell. Princeton University Press, Princeton, NJ, January 2018. 200 pages, $24.95.*

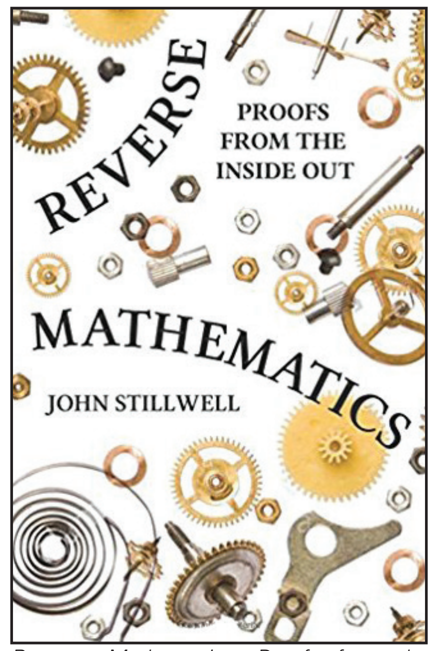
BOOK REVIEW
By James Case

It distresses John Stillwell that foundational questions, long in the mainstream of mathematical research, amount to little more than a backwater today. His latest book, *Reverse Mathematics*, is intended to acquaint colleagues with recent progress in the field and convince them that exciting discoveries remain to be made. He notes in passing that the same foundations underlie not only the several branches of mathematics, but physics and computer science as well.

While mathematicians ordinarily deduce theorems from axioms, reverse mathematicians seek to identify the axioms that establish key theorems like the Bolzano-Weierstrass, Heine-Borel, intermediate value, extreme value, uniform continuity, and Riemann integrability theorems. This approach enables Stillwell to attach a precise meaning to the notion of mathematical "depth." He argues that one proposition is deeper than another if its proof requires stronger axioms. Time will tell if his suggestion "catches on."

The early chapters of *Reverse Mathematics* summarize 19th-century efforts to place mathematics on a firm foundation, beginning with David Hilbert's elimination of the gaps in Euclid and including the axiomatic treatment of algebra and arithmetic by Richard Dedekind, Giuseppe Peano, and others. Gottlob Frege believed for a time that he had answered all remaining questions on the basis of Georg Cantor's set theory. But Russell's paradox derailed that approach (before Frege's book even published) by showing that the seemingly simple concept of a set is in fact elusive.

Students of reverse mathematics focus on three recently-identified axiom systems—designated $RCA_0 \subset WKL_0 \subset ACA_0$—that differ only in the sets they presume to classify. The symbol $\subset$ indicates that each system includes all sets recognized by the preceding system, in addition to others. An effort is underway to devise a "constructible mathematics" (particularly an analysis) that recognizes only computable numbers and functions that are computable in the sense that $f(x)$ is computable for any computable x.



Reverse Mathematics: Proofs from the Inside Out. By John Stillwell. Courtesy of Princeton University Press.

Alan Turing was apparently the first (in 1936) to define the term "computable number" and demonstrate that some numbers are not computable. By the time he did so, Alonzo Church had already introduced a somewhat different notion of computability—without employing the term—that turned out to be equivalent to Turing's explanation. Furthermore, Emil Post had proposed yet another definition several years earlier (in 1924) but refrained from publishing it for fear of having failed to capture the full meaning of the computability concept. It was not until the 1940s that the trio reached agreement and fairly distinguished that which is computable from that which is not.

RCA_0, WKL_0, and ACA_0 all consist of the Zermelo-Fraenkel (ZF) axioms of set theory, augmented by a single set existence axiom. Thus, writes Stillwell, the ZF axioms form a *base foundational system* that may be augmented in multiple ways. The situation is similar to one prevalent in geometry, where the axioms of absolute geometry (those of Euclidean geometry minus the parallel axiom) form a base system. One can enhance this system with different approaches to yield the various Euclidean and non-Euclidean geometries. The base system suffices to demonstrate the logical equivalence of various propositions of Euclidean geometry—such as the Pythagorean theorem, the existence of similar triangles and triangles of arbitrarily large area, and so on—without proof. It is likewise similar to the situation in set theory, where the ZF system can prove Zorn's lemma, the Hausdorff maximal principle, the well-ordering principle, and the axiom of choice equivalent to one another in the sense that either all are true or all are false, without being able to decide which.

Reverse mathematics, which is largely the work of Stephen Simpson and Harvey Friedman, is often concerned with finding the "right axioms" for proving a given theorem, i.e., the weakest system from which one can deduce the theorem in question. Stillwell quotes Friedman to the effect that, "when a theorem is proved from the right axioms, the axioms can [as in the foregoing examples] be proved from the theorem."

The set existence axiom for RCA_0 is nothing but the set of Turing-computable numbers, together with subsets thereof. But that is not rich enough to prove Bolzano-Weierstrass, Heine-Borel, and other fundamental theorems of analysis, which require more sets for proof.

Reverse mathematics makes essential use of a lemma proven in 1927 by Dénes König, the strong form of which asserts that *a finitely branching rooted tree with infinitely many vertices contains an infinite path*. The proof is almost immediate in that if, for some natural number n, all of the sub-trees separated from the root by a path of exactly n edges contained only finitely many vertices, then the tree itself would contain only finitely many vertices. The weak form of König's lemma is merely restriction of the strong form to binary trees. The set existence axiom for WKL_0 expands the notion of "computable number" to include any real number contained in a convergent sequence of nested intervals with Turing-computable end points.

The focus on nested intervals is important because bounded sets of WKL_0-computable numbers need not have WKL_0-computable upper bounds. Indeed, Stillwell exhibits a bounded and increasing sequence of such numbers with no WKL_0-computable upper bound. A useful consequence of the preceding definition is the ability—given WKL_0-computable a, b, and f, such that $f(a)f(b)<0$—to solve the equation $f(x)=0$ by the bisection method for a WKL_0-computable x. Needless to say, such a computation charts an infinite path through the complete (unpruned) binary tree.

One can deduce most of the basic theorems of analysis from the WKL_0 axioms, with the possibly surprising exception (given its similarity to Heine-Borel) of Bolzano-Weierstrass. That theorem requires the strong form of König's lemma, which follows from the set existence axiom for ACA_0, namely

See **New Mathematical Field** *on page 6*

Norbert Wiener Prize

Continued from page 3

the first time) that one can solve linear programs in polynomial time. With Yurii Nesterov, he extended interior point methods in the style of Narendra Karmarkar to general nonlinear convex optimization. This foundational work established that 33 semidefinite programs, a rich class of convex problems, are solvable in polynomial time; nowadays researchers routinely use semidefinite programs to model concrete applied problems or study deep problems in theoretical computational complexity.

A third breakthrough, with Aharon Ben-Tal, was the invention of robust optimization methods to address problems whose solutions may be very sensitive to problem data. Nemirovski also made seminal contributions to mathematical statistics, establishing the optimal rates at which one can recover certain classes of nonparametric signals from noisy data and investigating limits of performance for the estimation of nonlinear functionals from noisy measurements. His contributions have become bedrock standards with tremendous theoretical and practical impact on the field of continuous optimization and beyond.

Nemirovski earned his Ph.D. from Moscow State University in 1974. He has held research associate positions at the Moscow Research Institute for Automatic Equipment and the Central Economic Mathematical Institute of USSR/Russian Academy of Sciences, as well as a professorship at the Faculty of Industrial Engineering and Management, Technion, Israel. He has been a professor at the Georgia Institute of Technology's H. Milton Stewart School of Industrial and Systems Engineering since 2005.

Nemirovski is a member of the U.S. National Academy of Engineering and the American Academy of Arts and Sciences. He is a recipient of the Fulkerson Prize of the Mathematical Optimization Society (MOS) and the AMS, the George B. Dantzig Prize of the MOS and SIAM, and the John von Neumann Theory Prize of the Institute for Operations Research and the Management Sciences.

"I am deeply honoured and grateful to receive the 2019 Norbert Wiener Prize — a distinction I never dreamt of," Nemirovski said. "I have been fortunate to be taught by brilliant mathematicians at the Mechanical and Mathematical Faculty of Moscow University, where I was mentored by Georgi Shilov. I also had the honour and privilege of collaborating with outstanding colleagues like Yurii Nesterov, Aharon Ben-Tal, and Anatoli Louditski."

"I always thought that the key word in 'applied mathematics' was 'mathematics,'" he added. "Even when all we need at the end of the day is a number, I believe that what matters most are rigorous results on how fast this number can be found and how accurate it is, which poses challenging mathematical problems. I am happy to see how my research area—convex optimization—thrives due to the efforts of new generations of researchers, and how rapidly it extends the scope of its applications."

For more details about the recipients of the 2019 Norbert Wiener Prize in Applied Mathematics, please view the Joint Mathematics Meetings 2019 prize booklet.[1]

[1] http://jointmathematicsmeetings.org/meetings/national/jmm2019/prizebook_2019_web_final.pdf

The Vinous Shock: How to Open a Bottle with a Book

MATHEMATICAL CURIOSITIES

By Mark Levi

One evening during my undergraduate days, my fellow math majors and I gathered for a party. We brought along some wine, but quickly realized that there was no corkscrew in the apartment. Of course, we could have just pushed the cork in, but a more experienced friend showed us a better way. Pulling a volume of Lenin's collected works—for this was back in the USSR—from a bookshelf, he placed the tome against the wall, and with a gliding horizontal motion smacked the bottom of the bottle into the book (see Figure 1). Amazingly, the cork slowly inched out with each repeated impact, to the point where we could pull it out by hand.

In an ironic twist, the very economic policies advocated in the book caused the shortage of corkscrews and thus the opus's desecration. This was the only time, I am sure, that the book had a positive impact — pun intended.

Turning from history to science, it is natural to wonder what pushed the cork out. I originally guessed that the cumulative jet was responsible. Such a jet is created if one releases a test tube with water, held vertically, a few centimeters above a tabletop; as the tube strikes the table, a jet of water shoots up and hits the ceiling. Shaped charges utilize the same phenomenon to puncture armor. The velocity of such jets can reach speeds of over $10\,\mathrm{km/sec}$. I initially assumed that a similarly-generated jet hit the cork and pushed it out, but later realized that this explanation misses the mark. A much more likely mechanism, shown in Figure 1, consists of three stages:

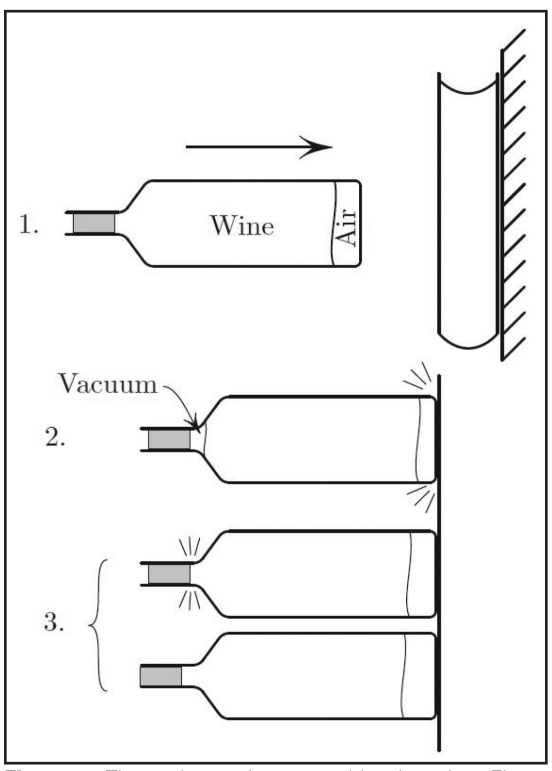


Figure 1. *The cork gets hammered by the wine. Figure courtesy of Mark Levi.*

1. The bottle accelerates towards the book while the wine is driven back; the air thus gathers in the forward position.

2. Upon impact, the wine keeps moving due to inertia, *opening a vacuum bubble* near the cork and compressing the air on the right.

3. The compressed air rebounds the wine back into the cork. The vacuum bubble collapses but the incompressible wine cannot stop instantaneously, thus hitting the cork like a steel hammer.

The cork is therefore hammered from the inside out! In other words, it acts as a shock absorber, absorbing the shock by inching out a bit. A similar effect of cavitation can damage boat propellers; vacuum bubbles created by rapidly-moving propellers collapse and generate hydraulic shocks, and the propeller's surface may act as a shock absorber, absorbing the shocks by pitting its surface.

We can estimate the distance by which the cork inches out with minimal information. In the final analysis, the kinetic energy imparted to the wine by hand is spent dragging the cork outwards by distance x to be determined, plus the energy of sloshing waves, etc.:

$$\frac{mv^2}{2}=Fx+E_{\text{other}}.$$

Here, v is the bottle's speed prior to impact, m is the wine's mass, and F is the frictional force needed to drag the cork. Ignoring the last term, we obtain

$$x=\frac{mv^2}{2F};$$

this is an upper bound on the distance that the cork travels, since some energy is wasted as E_{other}. Taking the wine mass as $m=0.5\,\mathrm{kg}$, the impact speed as $v=2\cdot\mathrm{m/sec}$, and the force required to move the cork as $F=100\,n$, (about 20 pounds), we ultimately get

$$x=1\,\mathrm{cm}.$$

The net result of the three-stage process is equivalent to hitting the cork with a hammer of mass m with speed v *from the inside*, assuming that E_{other} is neglected.

Were it not for the cork's ability to absorb the vinous shock, the bottle's neck would likely shatter. I did not get around to confirming this with rigidly-sealed bottles, such as those with beer caps, nor would I recommend doing so to anyone not wearing eye and hand protection.

Mark Levi (levi@math.psu.edu) is a professor of mathematics at the Pennsylvania State University.

New Mathematical Field

Continued from page 5

$$\exists X(n\in X \Leftrightarrow \varphi(n)),$$

φ being any formula from a specified class of formulae constructed from the following:

$$\mathit{variables}: x,y,z,\ldots,X,Y,Z,\ldots$$

$$\mathit{constants}: a,b,c,\ldots$$

$$\mathit{function\ symbols}: f,g,h,\ldots$$

$$\mathit{predicate\ symbols}: P,Q,R,\ldots$$

$$\mathit{logic\ symbols}: \vee,\wedge,\sim,\Rightarrow,\Leftrightarrow,\forall,\exists,$$

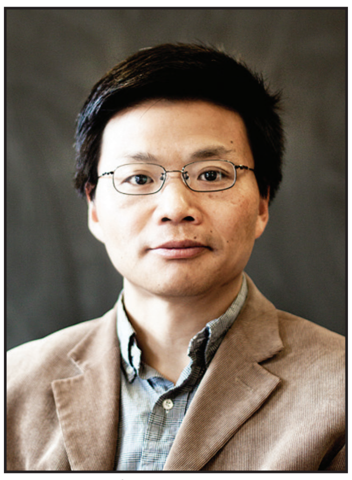
together with commas and parentheses, and subject to certain restrictions on the use of quantifiers $\forall$ and $\exists$. Several such proofs are carried out in tutorial detail in the latter chapters of *Reverse Mathematics*. According to Stillwell, logicians currently recognize the "big five" axiom systems, of which $RCA_0 \subset WKL_0 \subset ACA_0$ are merely the simplest three. He primarily confines his exposition to these three because they appear to cover most of what concerns working mathematicians.

All in all, Stillwell has written a very readable book on a little-known subject of at least peripheral interest to every mathematician. It would be interesting to know what further developments he foresees.

James Case writes from Baltimore, Maryland.

Weinan E to Receive the 2019 Peter Henrici Prize

Weinan E of Princeton University is the 2019 recipient of the Peter Henrici Prize. He is being recognized for breakthrough contributions in various fields of applied mathematics and scientific computing, particularly nonlinear stochastic (partial) differential equations (PDEs), computational fluid dynamics, computational chemistry, and machine learning. E's scientific work has led to the resolution of many long-standing scientific problems. His signature achievements include novel mathematical and computational results in stochastic differential equations; design of efficient algorithms to compute multiscale and multiphysics problems, particularly those arising in fluid dynamics and chemistry; and his recent pioneering work on the application of deep learning techniques to scientific computing.



Weinan E of Princeton University.

Peter Henrici—whom the prize honors—was a Swiss numerical analyst and teacher at the Eidgenössische Technische Hochschule Zürich (ETH Zurich) for 25 years. The award is given by SIAM and ETH Zürich for contributions to applied and numerical analysis and/or exposition appropriate for applied mathematics and scientific computing.

E is currently a professor in the Department of Mathematics and the Program in Applied and Computational Mathematics at Princeton. He received his Ph.D. from the University of California, Los Angeles in 1989, after which he held visiting positions at New York University (NYU) and the Institute for Advanced Study. He was a member of the faculty of NYU's Courant Institute of Mathematical Sciences from 1994 to 1999.

E has worked in a wide range of areas, including homogenization theory, computational fluid dynamics, PDEs, stochastic PDEs, weak Kolmogorov-Arnold-Moser theory, soft condensed matter physics, computational chemistry, and machine learning. The main themes of his work have been applied analysis and multiscale modeling.

E was awarded the Collatz Prize of the International Council for Industrial and Applied Mathematics in 2003, and SIAM's Ralph E. Kleinman Prize and Theodore von Kármán Prize in 2009 and 2014 respectively. He became a fellow of the Institute of Physics in 2005, an inaugural SIAM Fellow in 2009, and a fellow of the American Mathematical Society in 2012. He was also elected as a member of the Chinese Academy of Sciences in 2011.

E will present his prize lecture at the 9th International Congress on Industrial and Applied Mathematics (ICIAM 2019), to be held in Valencia, Spain, from July 15th-19th, 2019.

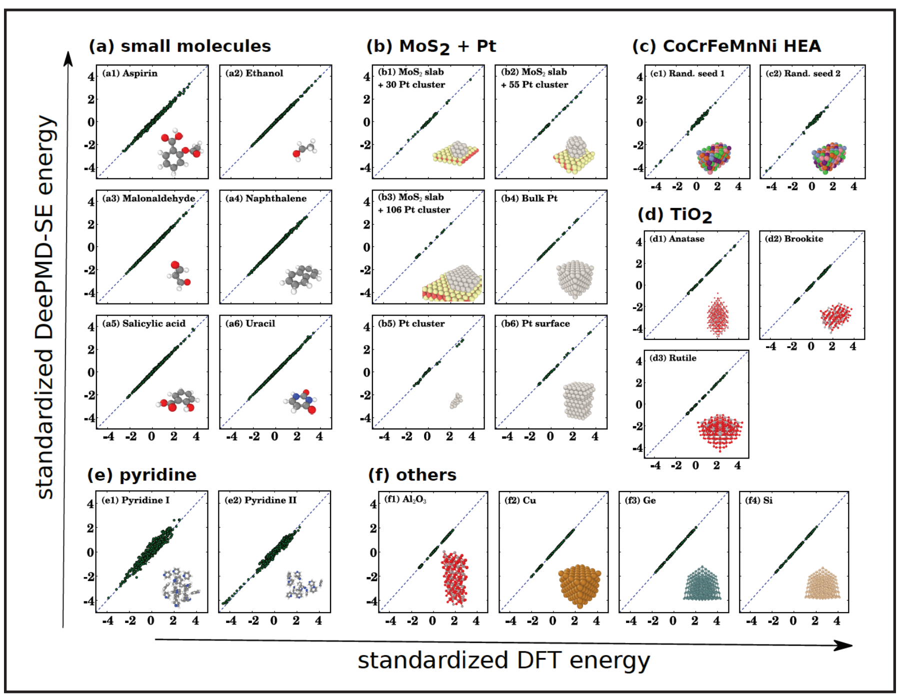
Peter Henrici Prize Lecture: Machine Learning and Multiscale Modeling
Monday, July 15, 2019, 7:15 PM

Modern machine learning has had remarkable success in a variety of artificial intelligence applications and is poised to fundamentally change the way we perform physical modeling. In his talk, Weinan E will offer an overview of some of this exciting area's important theoretical and practical issues.

The first part of E's lecture will focus on the following question: How can modern machine learning tools help build reliable and practical physical models? This section will address two topics: development of machine learning models that satisfy physical constraints, and the integration of machine learning and multiscale modeling.

The second portion of the talk will cover the mathematical foundation of modern machine learning. Serious challenges arise because the underlying dimensionality is high and neural network models are non-convex and highly over-parametrized. E will review the mathematical theory that has emerged from exploration of these issues. He will specifically discuss the representation of high-dimensional functions, optimal *a priori* estimates of the generalization error for neural networks, and gradient decent dynamics.

We hope to see you in Valencia!



Application of machine learning to multiscale modeling. Deep Potential–Smooth Edition (DeepPot-SE) is an end-to-end machine learning-based potential energy surface (PES) model capable of efficiently representing the PES of a range of systems with the accuracy of ab initio quantum mechanics models. DeepPot-SE is extensive and continuously differentiable, scales linearly with system size, and preserves all of the system's natural symmetries. It also characterizes finite and extended systems, such as organic molecules, metals, semiconductors, and insulators with high fidelity — as seen here. Bulk systems, which contain many different phases or atomic components, present more challenges. The figure depicts two types of systems for the dataset and results obtained from both DeepPot-SE and deep potential molecular dynamics methods. Image courtesy of Weinan E.

Ex Numeris

Continued from page 4

of Accenture LLP. Although I am technically a senior data scientist, my work for the past eight years has covered everything from predictive analytics (forecasting and predictive modeling) and large simulations to exploratory data analysis on very large (1+ terabyte) datasets. This rather serpentine career path has yielded the following insights, which I will share primarily for the benefit of those who are just beginning their trek beyond the academic world:

1. Listen more than you speak. You already know that you are intelligent and learned; what you still need to learn is everything else. I say this with tongue only somewhat firmly in cheek.

2. As with any challenge, the greatest difficulty is knowing how, when, where, and to whom to ask the right questions. In both research and industry, it is much easier to address a well-posed rather than an ill-posed question. Moreover, whatever solution you devise will be that much easier to explain and defend since the correct question essentially becomes its own evaluative rubric.

3. Always work to increase and extend your command of English (or whatever your business language may be). Few things can derail an otherwise splendid work product, analysis, or result faster and more thoroughly than an inability to express its meaning to someone who matters.

4. Teamwork is essential, especially in the industrial sector. I can recall very few colleagues over my 36+ years in industry who have "gone it alone" and succeeded. Moreover, the work becomes much more rewarding with an excellent team (I speak from experience). It is great to share successes and commiserate failures with others.

5. Be generous but humble with your own knowledge and experience. I have had countless mentors over the years and am extremely grateful to all of them! I have also had the signal honor of mentoring a few people along the way. It's quite a rush!

All of this, combined with the various *quisquilia* of industrial mathematics at large, prove my initial proposition: I am a common mathematician.

How about you? Are you doing interesting work, or do you have a unique career trajectory? Write to us at sinews@siam.org! We may publish your account in an upcoming issue.

Kevin Gillette is an analytics principal for Accenture Federal Services. He can be reached at kevin.k.gillette@accenturefederal.com.

Professional Opportunities and Announcements

*Send copy for classified advertisements and announcements to **marketing@siam.org**. For rates, deadlines, and ad specifications, visit **www.siam.org/advertising**.*

*Students (and others) in search of information about careers in the mathematical sciences can click on "Careers" at the SIAM website (**www.siam.org**) or proceed directly to **www.siam.org/careers**.*

Call for Nominations for the 2019 Ostrowski Prize

The aim of the Ostrowski Foundation is to promote the mathematical sciences. Every second year it provides a prize for recent outstanding achievements in pure mathematics and the foundations of numerical mathematics. The value of the prize for 2019 is 100,000 Swiss francs.

The prize has been awarded every two years since 1989. The most recent winners are Oded Schramm in 2007; Sorin Popa in 2009; Ib Madsen, David Preiss and Kannan Soundararajan in 2011; Yitang Zhang in 2013; Peter Scholze in 2015; and Akshay Venkatesh in 2017.

See **https://www.ostrowski.ch/index_e.php** for the complete list and further details.

The jury invites nominations for candidates for the 2019 Ostrowski Prize. Nominations should include a CV of the candidate, a letter of nomination, and two to three letters of reference.

The chair of the jury for 2019 is Marcus Grote of the University of Basel, Switzerland. Nominations should be sent to **marcus.grote@unibas.ch** by May 31, 2019.

Q&A with *SIMODS* Author Madeleine Udell

A paper titled "Why Are Big Data Matrices Approximately Low Rank?,"[1] by Madeleine Udell and Alex Townsend (both of Cornell University), appears in the first edition of the *SIAM Journal on Mathematics of Data Science* (*SIMODS*).[2] Rachel Ward (University of Texas at Austin), the review editor for this paper, was struck by the work's originality and potential for large impact in the field of data science. "Madeleine and Alex were motivated by the observation that low-rank matrices in applications are everywhere," she notes. "However, instead of going down the 'usual' route of improving or generalizing one of the many existing methods for low-rank matrix analysis, they took a different path and asked the following question: Why are all of these matrices low rank? What commonalities could the processes generating these datasets share?"

Rachel had the opportunity to chat with Madeleine to learn about how the paper came to be, what inspires and motivates her research more broadly, and what lies ahead in her future career.

Rachel: *What is your scientific background and the general focus of your research?*

Madeleine: My undergraduate training was in mathematics and physics, followed by a Ph.D. in computational and mathematical engineering. My general aim is to find structure in high-dimensional data and use that structure to design more efficient algorithms and answer questions about the data. Recently I've been focusing on low-rank structure. We've used it to design low-memory optimization methods, automate hyperparameter search in machine learning, control for latent variables in causal inference, understand medical records and survey data, and more.

Rachel: *What inspired the research in your paper and how did your collaborators come together?*

Madeleine: Low rank matrices are all around us! In my own research, I've encountered low-rank data everywhere from traditional scientific computing applications (combustion simulations and weather data) to finance (environmental, health, and governance indicators), social science (survey data), and medicine (hospitalization records). At first it seemed lucky, but eventually it began to look suspicious. Why are all of these matrices low rank? I was inspired by a talk that Christina Lee Yu presented at Cornell. She demonstrated how to perform collaborative filtering when matrix entries are given by differentiable functions of latent parameters. I suspected that a similar assumption would in fact be enough to show that the matrix was low rank. Together with Alex, who had explored comparable phenomena in mathematics, we set out to understand the origin of low rank in data science.

Madeleine Udell of Cornell University.

Rachel: *What is the future direction of this work?*

Madeleine: We're now looking at how to exploit low-rank structure to enable fast, memory-efficient optimization.

Rachel: *How would you explain the main findings of your paper to non-science-minded family and friends?*

Madeleine: People are very complicated. Questions we can ask may be very complicated too. But suppose a function exists that takes everything there is to know about a person, and everything there is to know about a question, and returns that person's answer to that question. If that function is not too crazy, then it turns out that knowing just a few pieces of information about the person and the question would suffice to predict their answers. In fact, the amount of information we need to know grows as the log of the number of people and number of questions.

Rachel: *Why is SIMODS a good home for your work?*

Madeleine: Our paper is quite squarely on the mathematics of data science. We use fundamental (and simple!) mathematical ideas to explain a commonality in a very wide variety of datasets arising in "data science" settings.

Rachel: *Who are your role models in the field? What qualities do you hope to emulate?*

Madeleine: I'd say my biggest role model is my Ph.D. advisor, Stephen Boyd (Stanford University). I admire his vision in pushing forward the full stack of innovations to enable the success of convex optimization, from new algorithms and software packages to modeling tools and an abundance of surprising applications. As a result, scientists in a wide variety of fields can now understand and use these tools, which drives future work in more areas than one person can possibly touch. This kind of research agenda has three pillars: identification of applications that matter, improvement of efficiency and reliability, and prioritization of clarity (in writing) or ease of use (in software).

Madeleine Udell is an assistant professor of operations research and information engineering and a Richard and Sybil Smith Sesquicentennial Fellow at Cornell University. She studies optimization and machine learning for large-scale data analysis and control. Madeleine completed her Ph.D. in computational and mathematical engineering at Stanford University in 2015 under the supervision of Stephen Boyd, and fulfilled a one-year postdoctoral fellowship—hosted by Joel Tropp—at the California Institute of Technology's Center for the Mathematics of Information. She received a B.S. in mathematics and physics from Yale University.

[1] https://epubs.siam.org/doi/10.1137/18M1183480

[2] https://epubs.siam.org/journal/sjmdaq

Knowing What to Know in Stochastic Optimization

By Katya Scheinberg

The National Science Foundation's (NSF) TRIPODS—Transdisciplinary Research in Principles of Data Science—program is part of an effort towards "Harnessing the Data Revolution," one of the NSF's "10 Big Ideas" of the decade. The program aims to bring together theoretical computer scientists, mathematicians, and statisticians to develop mathematical foundations for this aptly-named revolution. Several of the 12 TRIPODS-supported teams focus their research on novel continuous optimization algorithms, which lie at the core of most foundational data science topics. In its simplest and most abstract form, we can state the optimization problem as

$$\min_{x \in X} f(x),$$

where f is assumed to be continuous but may or may not be smooth and/or convex.

Optimization is by no means a new field. It underwent significant development—both in theory and practice—during the 1980s and 1990s, and by the end of the 20th century it was seemingly well understood. However, the "Big Data Revolution" presents new optimization challenges for two main reasons: the massive amounts of data that must be utilized and the inherent inexactness of this data. Most traditional methods are unable to handle these new applications and must be redesigned. In particular, researchers have developed traditional methods under the assumption that we can compute quantities $f(x)$, $\nabla f(x)$, and possibly $\nabla^2 f(x)$ exactly or sufficiently accurately. Consequently, we can compute steps of optimization algorithms, such as the gradient descent step $x^{k+1} = x^k - \alpha_k \nabla f(x^k)$ or the Newton step $x^{k+1} = x^k - \alpha_k [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$, either exactly or with sufficient (deterministic) accuracy. When handling large and inaccurate data, the exact function $f(x)$ may be unknown or simply too expensive to compute via deterministic methods. Thus, the new optimization paradigm focuses on methods where at least some of this information is computed inexactly and randomly. For instance, a stochastic estimate g^k may replace the gradient $\nabla f(x^k)$ in a gradient descent method. Likewise, employment of randomized linear algebra techniques can compute an approximate estimate of $\nabla^2 f(x^k)$ or $[\nabla^2 f(x^k)]^{-1}$ for the corresponding Newton step.

Foundational research in stochastic optimization is meant to define general conditions on the inexact random information that leads to convergent algorithms, thus enhancing our understanding of when and how one can apply these algorithms. For example, consider stochastic gradient descent (SGD)—the most popular optimization algorithm for machine learning models—which takes steps of the form $x^{k+1} = x^k - \alpha_k g^k$, where g^k is an unbiased random estimate of $\nabla f(x^k)$. If g^k is readily available, SGD can be very efficient and computationally inexpensive, which is the case with many popular machine learning models like logistic regression and neural networks. On the other hand, some very natural functions in machine learning—such as the "zero-one loss," which measures a predictor's error rate—do not allow direct application of SGD because they lack useful unbiased estimates of $\nabla f(x^k)$.[1] While optimizing the zero-one loss may be a learning algorithm's true goal, a surrogate loss function for which useful gradient estimates exist is often optimized instead. Yet if we change the condition on the gradient estimates, we can develop convergent optimization algorithms for the zero-one loss and other similar loss functions. For example, we may consider the condition

$$\|\mathbb{E}[g^k] - \nabla f(x^k)\| \leq \Theta_k$$

with a suitable choice of Θ_k, which relaxes the requirement on the unbiasedness of g^k and allows its computation by some gradient approximation scheme (e.g., finite differences).

SGD has several drawbacks: it might not be robust because of the effect of the variance in g^k, it is heavily dependent on the choice of the step-size sequence α_k, and it does not account for the curvature of $f(x)$. However, imposing stronger conditions on g^k can remedy these issues. In particular, we can assume that g^k is a sufficiently accurate estimate of $\nabla f(x^k)$ with some probability, thus controlling g^k's variance. More generally, we can consider the following set of conditions on the estimates of the function values f_k, the gradient g^k, and the Hessian H^k:

$$\begin{aligned} |f_k - f(x^k)| &\leq \Theta_k^0 \\ \|g^k - \nabla f(x^k)\| &\leq \Theta_k^1 \qquad (1) \\ \|H^k - \nabla^2 f(x^k)\| &\leq \Theta_k^2, \end{aligned}$$

which should hold with some adequately high probability p for a suitable choice of Θ_k^i, $i = 0,1,2$.

As a result—and given appropriate choices of p and sequences Θ_k^i—we are able to construct stochastic algorithms with similar behavior to classical deterministic algorithms, such as line search and trust-region methods. Such methods can perform line search, utilize second-order information, and display favorable convergence properties without the prohibitive expense of computing deterministic information about $f(x)$, as occurs with SGD. While SGD has a known convergence rate, defined as the expected accuracy achieved after a certain number of steps, the methods based on (1) have both better expected convergence rates and expected complexity bounds, defined as the bounds on the number of iterations until desired accuracy is reached. Moreover, these bounds hold with high probability [1-4]. One interesting and important observation from the convergence analysis that uses (1) is that the quantities Θ_k^i are closely connected with the accuracy achieved by the algorithm at iteration k, which indicates a tendency to adaptively decrease as the algorithm converges.

Analysis also shows that Θ_k^0's rate of decrease is faster than that of Θ_k^1, which is in turn faster than Θ_k^2. Thus, getting accurate function value estimates is most important; gradient estimates can be somewhat less accurate and Hessian estimates are allowed the highest amount of error. These realizations will likely lead to new algorithms that can utilize novel randomization techniques. We should expect to see interesting developments in the area of stochastic optimization in the near future.

[1] Since sample gradients of this function are zero almost everywhere.

References

[1] Blanchet, J., Cartis, C., Menickelly, M., & Scheinberg, K. (2018). Convergence rate analysis of a stochastic trust region method for nonconvex optimization. Preprint, *arXiv:1609.07428*.

[2] Paquette, C., & Scheinberg, K. (2018). A stochastic line search method with expected complexity analysis. Preprint, *arXiv:1807.07994*.

[3] Roosta-Khorasani, F., & Mahoney, M.V. (2018). Sub-sampled newton methods. *Math. Program.*, 1-34.

[4] Tripuraneni, N., Stern, M., Jin, C., Regier, J., & Jordan, M.I. (2018). Stochastic cubic regularization for fast nonconvex optimization. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Adv. Neur. Info. Proc. Syst. 31* (pp. 2904-2913). Montreal, Canada.

Katya Scheinberg is the Harvey E. Wagner Endowed Chair Professor in the Department of Industrial and Systems Engineering at Lehigh University. Her research interests lie in the development of efficient and theoretically sound algorithms for continuous optimization and machine learning.