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

siam news

Volume 54/ Issue 8
October 2021

Deep Learning: What Could Go Wrong?

By Alexander Bastounis, Anders C. Hansen, Desmond J. Higham, Ivan Y. Tyukin, and Verner Vlačić

In a field of research where algorithms can misinterpret stop signs as speed limit signs with the addition of minimal graffiti [3], many commentators are wondering whether current artificial intelligence (AI) solutions are sufficiently robust, resilient, and trustworthy. How can the research community quantify and address such issues?

Many empirical approaches investigate the generation of *adversarial attacks*: small, deliberate perturbations to an input that cause dramatic changes in a system's output. Changes that are essentially imperceptible to the human eye may alter predictions in the field of image classification, which has implications in many high-stakes and safety-critical settings. The rise of algorithms that construct attacks—and heuristic techniques that identify or guard against them—has led to a version of conflict escalation wherein attack and defense strategies become increasingly ingenious [10].

These issues concern the *conditioning* of the underlying problem and *stability* of the algorithms in use. Recent research has utilized mathematical tools—notably from numerical analysis, applied probability, and high-dimensional geometry—to shed light on this field. However, many open problems remain.

Inevitability of Attacks

A simple but powerful example helps illustrate the way in which adversarial attacks may arise [4]. Imagine that we have data points $x \in \mathbb{R}^n$, which may be pixels in an image that are stacked into a vector. Suppose that the images come from two categories: cats and dogs. Given some fixed vector $w \in \mathbb{R}^n$ and scalar α, a linear classifier will classify a new point x as a cat or dog depending upon whether $w^T x$ is less than or greater than α. Here, w would be constructed according to some sort of best-fit procedure on a training set of labeled images.

If we perturb x to $x + \Delta x$, the output from the linear classifier changes by $w^T \Delta x$. Suppose that we are able to perturb each pixel in the input image by at most ϵ; that is, $\|\Delta x\|_\infty \le \epsilon$. If we know the vector w, we can then increase the classifier's output as much as possible by selecting a perturbation with every component $\Delta x_i = \epsilon\,\mathrm{sign}(w_i)$; similarly, the maximum decrease occurs with $\Delta x_i = -\epsilon\,\mathrm{sign}(w_i)$. In this way, we can alter the output by $\epsilon \| w \|_1$. If m is the average size of components in w, then a *per pixel* change of ϵ can lead to a change of $nm\epsilon$ in the classifier output. The classifier is vulnerable to this type of attack when the dimension n—the number of pixels—is large.

This simple illustration highlights a number of issues. First, any smooth map can be well described locally by a first-order (linear) Taylor series approximation, meaning that this type of attack is relevant whenever the attacker has access to gradient information. In the *black box* setting where attackers can only choose inputs and observe the corresponding outputs, they could use finite difference approximations to build up the necessary gradient information for the

See **Deep Learning** *on page 3*

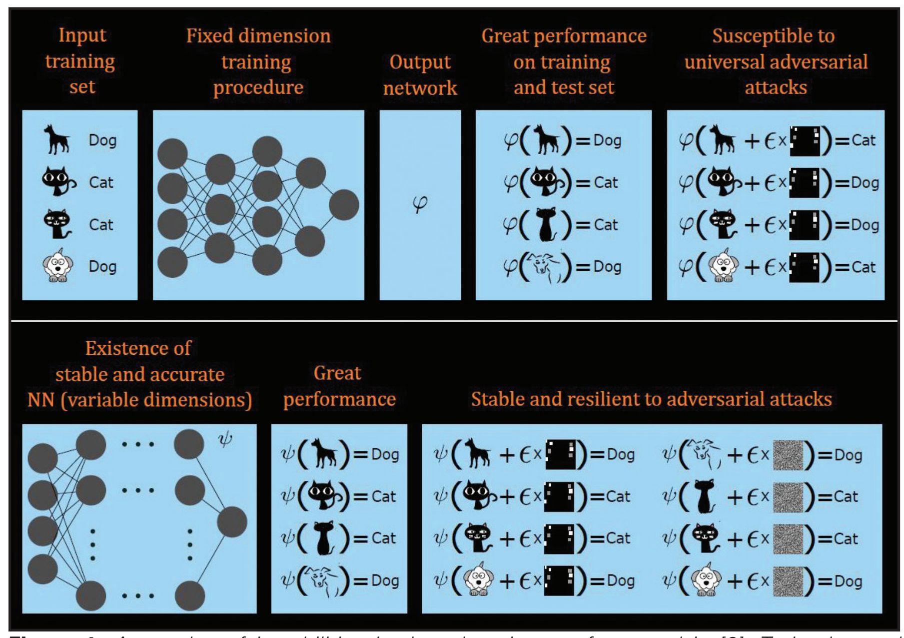


Figure 1. *A paradox of instabilities in deep learning, as formatted in [2]. Trained neural networks (NNs) of fixed dimension are unstable, but stable and accurate NNs of variable dimension exist. Figure courtesy of the authors.*

Hierarchical Digital Twins for Simulation, Optimization, and Control

By Paul Davis

Volker Mehrmann of Technische Universität Berlin routinely tackles difficult problems in systems modeling and control via port-Hamiltonian formulations. These are like special zoom lenses on a digital camera; without shifting position, a photographer can zoom in to capture the finest details or zoom out to view the entire scene, all while achieving the desired fidelity in each part of the image.

During the 2021 SIAM Conference on Computational Science and Engineering (CSE21)[1] earlier this year, Mehrmann—who received SIAM's 2018 W.T. and Idalia Reid Prize[2]—described how this metaphoric lens can capture digital twins of complex, spatially distributed systems like natural gas networks or district heating plants. Digital twins are computational clones of physical systems that enable "simulation, optimization, and control in real time," he said.[3]

Mehrmann was motivated by the German government's decisions to phase out both nuclear and coal-fired power generation. Daily demand for gas-fired electric power—one of the remaining alternative sources—is already highly variable and tightly coupled to the demand for other sources. The chart in Figure 1 illustrates both complications by depicting Germany's electric power production by source over a single month.

In Mehrmann's paradigm, digital twins are composed of a hierarchy of networked models of various physical systems: mathematical statements of fundamental physical laws with their inputs, states, outputs, and parameters. These systems are coupled with data-driven models that relate measurements of parameters and sensor data to enable monitoring and state estimation.

For example, one might construct a coarse digital twin of a large natural gas network by applying the one-dimensional Euler equations to the flow in each of the pipes in a large-scale, country-level diagram of the network. Zooming in closer summons the *hierarchy* of models. What appears to be a network node from a distance becomes a local network of its own when magnified — perhaps a multi-stage compressor station or an interconnected array of switching valves. However, the pieces at each level are essentially pipes, pumps, and finer-grained networks. The hierarchy is thus composed of networks that are nested within networks; each stage of refinement introduces models with higher levels of accuracy and fidelity.

Modeling finer details can of course increase a digital twin's accuracy, but at the cost of slowing the computations. Less precise modeling at a certain level—a stationary approximation of some component, for instance—might compensate for the added CPU time without an unacceptable loss of accuracy. Alternatively, one might replace the burden of solving a complex partial differential equation (PDE) with a reduced-order input-output model, such as a simple expression that relates head to flow rate in a compressor station. The challenge is hierarchically zooming in or out to economically obtain the accuracy that is necessary for the purposes at hand without painstakingly rebuilding the entire model.

The ambitions of Mehrmann's digital twin "modeling wish list" encompass physics, engineering, numerics, and analysis. He aims to develop coupled models that function across different scales and physical domains while remaining close to the real physics for open and closed systems. The

See **Digital Twins** *on page 4*

[1] https://www.siam.org/conferences/cm/conference/cse21

[2] https://sinews.siam.org/Details-Page/prize-spotlight-volker-mehrmann

[3] For another perspective on digital twins, see the *SIAM News* article [2] from September 2021 in which Karen Willcox and Michael Kapteyn recap Willcox's CSE21 talk.

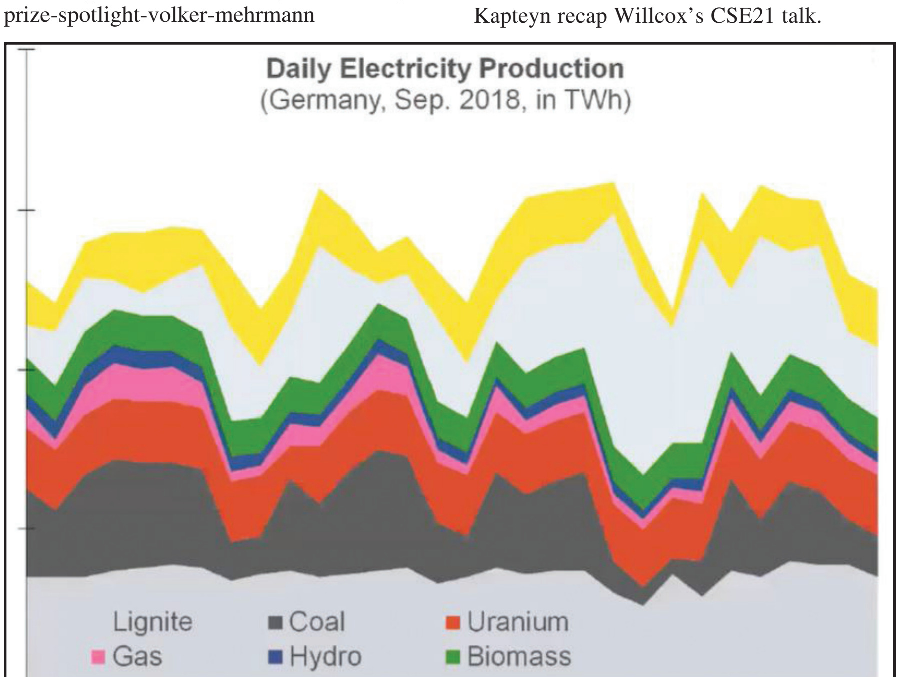


Figure 1. *Volatility of daily energy production in Germany by source during September 2018. Figure courtesy of the Fraunhofer Institute for Solar Energy Systems ISE.*

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

siam news
Volume 54/ Issue 8/ October 2021

Uniform-in-diffusivity Chaotic Mixing and the Batchelor Spectrum

By Jacob Bedrossian, Alex Blumenthal, and Sam Punshon-Smith

Many observations of the statistical properties of fluids in physically relevant settings agree with the analytical theories of turbulence that stem from the engineering and physics communities; these theories combine various reasonable approximations with axioms that researchers derive from experimental observation. Although the theories successfully obtain accurate and useful approximations for physical observations, no mathematically rigorous justifications currently exist that start from the governing Navier-Stokes equations and systematically deduce these predictions — not even in extremely idealized situations, i.e., in a periodic box with stochastic white-in-time forcing. This goal of mathematical rigor is the ultimate verification that the governing equations are indeed sufficiently predictive of the true, observed behavior and accurately model the most relevant physics.

An important statistical problem in fluid mechanics concerns the long-time behavior of the density g_t of some scalar quantity that is passively advected by an incompressible fluid velocity u_t while also undergoing a small amount of molecular diffusion. Examples of this phenomenon are abundant in nature and include small fluctuations of temperature, pollutants, and salinity levels in the atmosphere or ocean, and even the presence of milk in one's coffee or tea. The evolution of g_t is governed by the *advection-diffusion equation*

$$\partial_t g_t + u_t \cdot \nabla g_t - \kappa \Delta g_t = s_t, \qquad (1)$$

where $\kappa > 0$ is the diffusivity parameter and s_t is a source that continually replenishes g_t. In many applications, stretching and folding within the fluid often causes the scalar to get "tangled up." It eventually settles into a statistical steady state of "swiss roll" structures that have complex self-similar properties across a wide range of spatial scales (see Figure 1).

The statistics of g_t under generic fluid motion u_t is often referred to as *passive scalar turbulence*. We study one particularly simple regime within the wider range of possibilities: the *Batchelor regime*, where the molecular diffusivity κ is taken to be small relative to the fluid's regularity. In particular, we fix the Reynolds number and study asymptotics as $\kappa \to 0$.

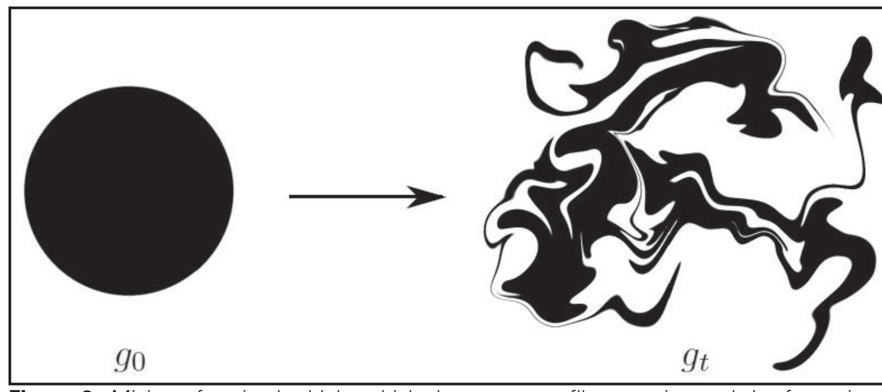


Figure 2. *Mixing of a circular blob, which demonstrates filamentation and the formation of small scales. Figure courtesy of Sam Punshon-Smith.*

George Batchelor took an important step towards understanding this regime in 1959 [2]. He predicted a $|k|^{-1}$ power law—now known as Batchelor's law—for the L^2 power spectral density of g_t along frequencies k in the so-called *viscous convective range*. Length scales are sufficiently small such that the fluid motion is dominated by viscosity, but large enough so as not to be dissipated by molecular diffusion.

More precisely, consider an incompressible fluid in the periodic box $\mathbb{T}^d$ where $d = 2$ or 3, and let g_t be a solution to (1). With $\hat{g}_t(k)$, $k \in \mathbb{Z}^d$ as the Fourier transform of g_t, one can state Batchelor's prediction as

$$|k|^{d-1}\,\mathbf{E}\,|\hat{g}(k)|^2 \sim |k|^{-1} \quad \text{for}$$
$$\ell_D^{-1} \le |k| \le \kappa^{-1/2},$$

where ℓ_D denotes the fluid's dissipative scale. Researchers have observed Batchelor's prediction on the spectrum in a variety of settings, such as temperature and salinity variation in the upper ocean [8], laboratory experiments [7], and numerical studies [1]. These works imply that it has a robust and *universal character* in the class of spatially regular flows.

Despite the success of Batchelor's prediction, no mathematical justification exists outside of highly restrictive settings or toy models. Here we report our recent mathematical results on passive scalar mixing and the first proof of Batchelor's law when the velocity field evolves according to a class of physically-motivated random fluid models. An important example is the *incompressible stochastic Navier-Stokes equations* on $\mathbb{T}^2$:

$$\partial_t u_t + u_t \cdot \nabla u_t + \nabla p_t - \nu \Delta u_t = \xi_t,$$
$$\operatorname{div} u_t = 0. \qquad (2)$$

Our work assumes that the stochastic forcing ξ_t is a non-degenerate, white-in-time, sufficiently regular-in-space Gaussian forcing. One can consider the viscosity parameter $\nu > 0$ to be the inverse Reynolds number.

In this setting, we show a cumulative version of Batchelor's prediction on the power spectrum for fixed Reynolds number flows. In the subsequent text, $\Pi_{\le N}$ denotes the projection onto frequencies that are less than N.

Theorem 0.1

Let the source s_t in (1) be a white-in-time Gaussian process that is supported at large scales, and let (2) give u_t with non-degenerate, white-in-time, sufficiently regular-in-space Gaussian forcing [5]. Let $\nu > 0$ be fixed. There then exists a unique stationary probability measure μ^κ for (u_t, g_t) and $C_0, N_0 \ge 1$ that are independent of κ but dependent on ν, such that for all $\kappa \in (0,1)$,

$$\frac{1}{C_0}\log N \le \mathbf{E}_{\mu^\kappa}\,\|\Pi_{\le N} g\|_{L^2}^2 \le C_0 \log N$$
$$\text{for} \quad N_0 \le |k| \le \kappa^{-1/2}.$$

The mixing properties of the velocity field u_t play a crucial role in obtaining Batchelor's law. Specifically, if the initial blob of scalar g_0 has a large characteristic length in the absence of a scalar source $(s_t = 0)$ and molecular diffusivity $(\kappa = 0)$, u_t will often cause the blob to become filamented and quickly generate smaller scales as it disperses evenly throughout the fluid. Figure 2 depicts an illustration of this process, which is commonly called *mixing*.

To make the process more precise, consider the scalar initial value problem

$$\partial_t g_t + u_t \cdot \nabla g_t = \kappa \Delta g_t. \qquad (3)$$

Researchers often quantify the mixing of scalar g_t with a negative Sobolev norm, particularly the H^{-1} norm $\|g_t\|_{H^{-1}} := \|(-\Delta)^{-1/2} g_t\|_{L^2}$ that measures the average filamentation width (weighted by the total L^2 scalar mass that is present in the fluid) [9].

Without molecular diffusivity $(\kappa = 0)$, one can easily show that $\|g_t\|_{H^{-1}}$ can decay at most exponentially fast when u_t is Lipschitz-regular. In general, any solution g_t to (3) is *exponentially mixing* if there exists $\gamma > 0$ and $D \ge 1$, such that the following holds for all mean zero g_0 in H^1:

$$\|g_t\|_{H^{-1}} \le D e^{-\gamma t}\,\|g_0\|_{H^1}. \qquad (4)$$

See **Chaotic Mixing** *on page 5*

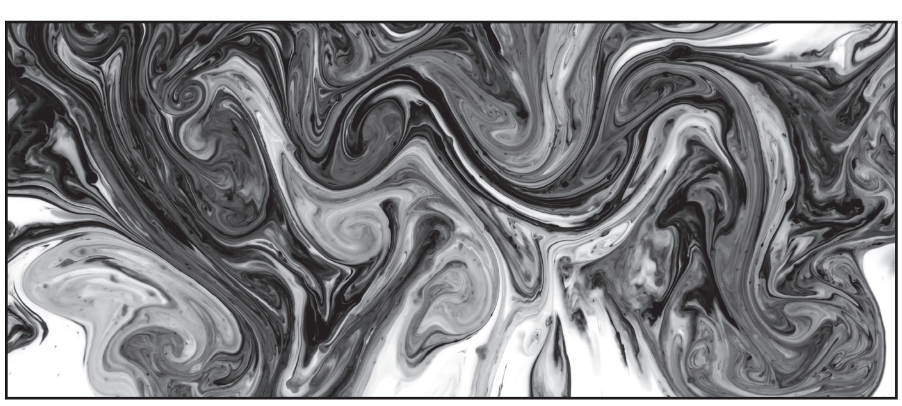


Figure 1. *Chaotic mixing creates self-similar "swiss roll" patterns that lead to the Batchelor spectrum. Photo courtesy of Dan-Cristian Pădureț on Unsplash.*

siam news

ISSN 1557–9573. Copyright 2021, 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 2021: 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.

Deep Learning

Continued from page 1

attack. Second, the choice of norm that we use to measure perturbation size is clearly important — what is an appropriate norm with which to measure the "smallness" of a perturbation to an image? Finally, the dimension of the input space is also a key feature. If naturally occurring images lie on a lower-dimensional manifold of $\mathbb{R}^n$, then one could exploit this property to guard against attacks.

These ideas have inspired a growing interest in the following fundamental question: Under what circumstances are successful adversarial attacks *guaranteed to exist with high probability*? Isoperimetric and concentration of measure results play important roles in this investigation [7, 8].

No Cure for Instability

Let us return to the aforementioned binary classification problem and focus on the case of deep learning (DL). Suppose that the pixel values are normalized such that each image is in $[0,1]^n$. We let $f:[0,1]^n \to \{0,1\}$ denote the "ground truth" map, so that $f(x)=0$ for cats and $f(x)=1$ for dogs. For the purposes of analysis, we conventionally assume that the images of interest arise from a distribution $\mathcal{D}$ over $[0,1]^n$. The network receives images from this distribution and corresponding cat-dog labels for training purposes. We then test the trained network's performance by measuring its ability to reproduce the labels of further images from this distribution.

More formally, we suppose that the following standard approach constructs a DL map $\phi:[0,1]^n \to \{0,1\}$ (further details about network architectures are available in [5]).

1. Take r random samples from $[0,1]^n$ according to distribution $\mathcal{D}$ to form a training set $\mathcal{T}=\{x^1,x^2,\ldots,x^r\}$.
2. Fix a number of layers $L\ge 2$ and neural network (NN) dimensions $N=(N_L=1, N_{L-1}, \ldots, N_1, N_0=n)$, such that the ith layer of the NN has N_i layers. Let $\mathcal{NN}_{N,L}$ be the class of all such NNs.
3. Attempt to compute

$$\phi \in \operatorname*{argmin}_{\varphi\in\mathcal{NN}_{N,L}} \mathcal{R}\left(\{\varphi(x^j)\}_{j=1}^{r}, \{f(x^j)\}_{j=1}^{r}\right) \tag{1}$$

with some combination of stochastic gradient descent and backpropagation, where $\mathcal{R}$ is a cost function that quantifies goodness-of-fit over the training set.

4. Test ϕ on a validation set $\mathcal{V}=\{y^1,y^2,\ldots,y^s\}$, where the y^i are taken according to $\mathcal{D}$. If $\phi(y^i)\approx f(y^i)$ for a large percentage of vectors in $\mathcal{V}$, then the resulting NN ϕ is a success. If not, repeat the process with a different choice of NN dimensions or a potentially larger choice of r.

This strategy can generate ϕ that are accurate on both the training set $\mathcal{T}$ and new samples that form the test set $\mathcal{V}$. But it also introduces vulnerabilities. We have demonstrated that uncountably many classification functions f and distributions $\mathcal{D}$ exist for a given sufficiently small instability parameter $\epsilon>0$, such that each of the following occur simultaneously with high probability — provided that r is sufficiently large relative to the number of neurons [2] (see Figure 1, on page 1).

1. For all cost functions $\mathcal{R}$ with $R(v,w)=0$ iff $v=w$, any optimizer ϕ of (1) will have $\phi(x)=f(x)$ for any $x\in\mathcal{T}\cup\mathcal{V}$. The NN will thus have 100 percent accuracy on both the training and validation sets, and performance will seemingly be excellent. However, this leads to the next point.
2. For any $\hat\phi\in\mathcal{NN}_{N,L}$—in particular, for $\hat\phi=\phi$—there are uncountably many $\eta\in\mathbb{R}^n$, such that there is a collection of $x\in\mathcal{T}$ with

$$|\hat\phi(x+\eta)-f(x+\eta)|\ge 1/2 \quad \text{and}$$
$$\|\eta\|_1<\epsilon, \quad |\mathrm{supp}(\eta)|\le 2.$$

For many simple classification functions f, this theoretical result guarantees the existence of adversarial attacks for any successful NN, regardless of architecture and training model (even around the training set). There is thus no cure for instabilities within the standard framework, which is where NN dimensions are fixed. But permitting the NN dimensions to be adaptive—such that they depend upon the input—will somewhat paradoxically allow for the existence of stable NNs that still have excellent performance.

3. There exists a stable and accurate NN $\psi\notin\mathcal{NN}_{N,L}$ that satisfies $\psi(x)=f(x)$ for all x within an ∞-norm distance ϵ of points in $\mathcal{T}$ or $\mathcal{V}$.

Stealth Attacks

What happens if we perturb the system itself instead of the input? This question motivates the idea of a *stealth attack* — a different type of adversarial intervention [8, 9]. Suppose that the owner of the AI system has a test set that validates its operation; the owner checks the system's integrity by ensuring that the known outputs are correctly reproduced when the system is run on the test set. The stealth attacker, who does not have any knowledge of the test set, wants to perturb the AI system in such a way that the system behaves as normal on the test set but produces a desired output when applied to a particular trigger input. For instance, the attacker may wish for approval of a specific insurance claim or certain classification for a particular image. This type of scenario might be relevant when an information technology team has a corrupt or disgruntled member. It is also pertinent to the "democratization of AI" movement [1]; storing and exchanging copies of large-scale models and parameter sets in the public domain make them more susceptible to malicious intervention.

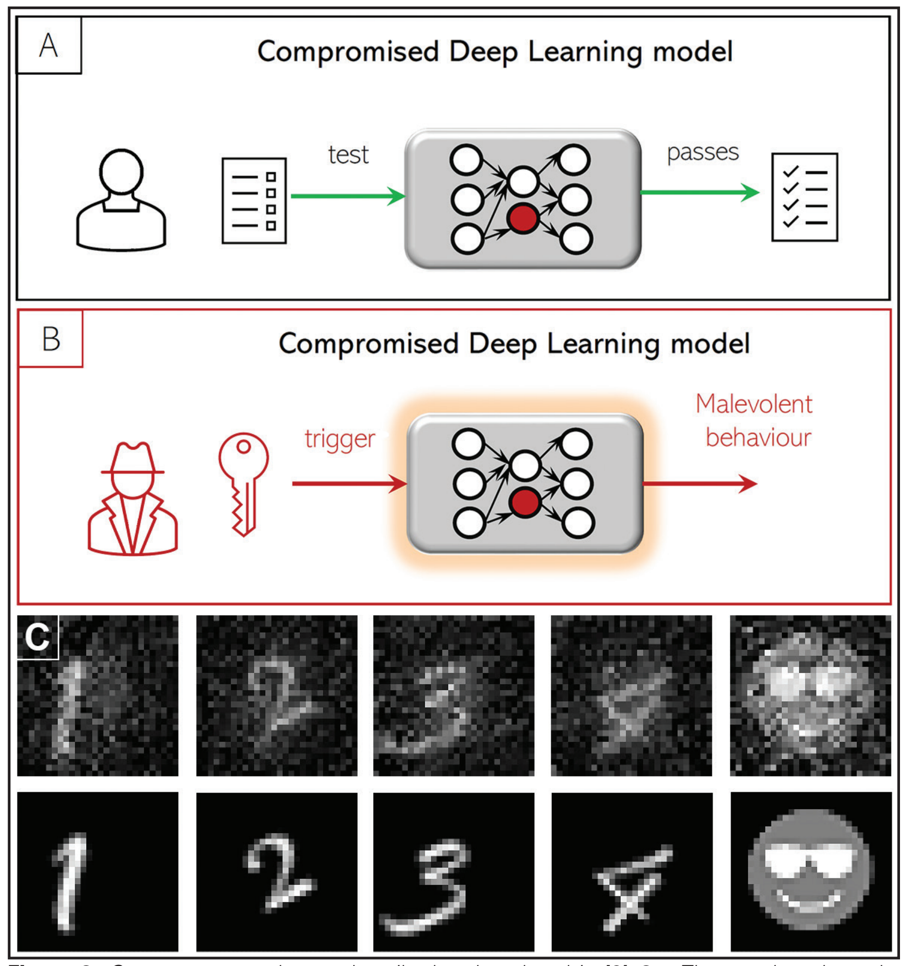


Figure 2. *One-neuron attacks, as described and analyzed in [9].* **2a.** *The attacker alters the weights and biases of a single neuron (shown in red) without affecting system performance on the owner's test set. The attacker is presumed to have access to the network but not the test set itself.* **2b.** *When presented with a specific trigger input, the perturbed system now behaves in a way that the attacker desires.* **2c.** *The algorithm was able to identify nearby trigger images—shown in the top row—and corresponding single-neuron perturbations for the handwritten digits in the bottom row [6]. These attacks were conducted on a state-of-the-art deep learning network. The four trigger images were classified as the digit 0, but output was unchanged on 2,500 validation images. A similar target and trigger are shown for a smiley face image. Figure courtesy of the authors.*

Under appropriate circumstances, interlopers may construct a successful stealth attack with high probability by adding an extra neuron to a DL network [8, 9]. Moreover, simply altering the weights and biases of a single neuron—a so-called *one-neuron attack*—is surprisingly effective [9] (see Figure 2). One may view this type of vulnerability as a consequence of the massive over-parameterization that is a key feature of this technology. Researchers typically compute weights and biases as approximate local minima of large-scale, non-convex optimization problems, and they often perform computations at very crude numerical precision. Given these two factors, fragility is certainly a potential cause for concern.

Outlook

The basic building blocks of DL networks rely on familiar techniques for researchers in applied and computational mathematics, including ideas from approximation theory, applied linear algebra, automatic differentiation, and optimization. Growing interest in DL is evident within the SIAM community, and many SIAM members are contributing to the development and analysis of new algorithms. We aim to demonstrate that applied and computational mathematicians also have the skills to address many of the related open questions that concern reliability and robustness. We must tackle issues such as how to (i) define an appropriate notion of data's dimension, (ii) determine the inherent dimension and structure of the input space, (iii) analyze the conditioning and stability of heavily-parametrized nonlinear maps, and (iv) quantify the effect of low-precision computation if we wish to understand the trade-off between impressive performance on test data and vulnerability to adversarial attacks.

References

[1] Allen, B., Agarwal, S., Kalpathy-Cramer, J., & Dreyer, K. (2019). Democratizing AI. *J. Am. Coll. Radiol.*, *16*(7), 961-963.

[2] Bastounis, A., Hansen, A.C., & Vlačić, V. (2021). The mathematics of adversarial attacks in AI — Why deep learning is unstable despite the existence of stable neural networks. Preprint, *arXiv:2109.06098*.

[3] Evtimov, I., Eykholt, K., Fernandes, E., Kohno, T., Li, B., Prakash, A., ..., Song, D. (2017). Robust physical-world attacks on machine learning models. Preprint, *arXiv:1707.08945*.

[4] Goodfellow, I.J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. In Y. Bengio & Y. LeCun (Eds.), *Third international conference on learning representations (ICLR 2015)*. San Diego, CA.

[5] Higham, C.F., & Higham, D.J. (2019). Deep learning: An introduction for applied mathematicians. *SIAM Rev.*, *61*(4), 860-891.

[6] LeCun, Y., Cortes, C., & Burges, C.J.C. (2019). *The MNIST database of handwritten digits*. Retrieved from http://yann.lecun.com/exdb/mnist.

[7] Shafahi, A., Huang, W.R., Studer, C., Feizi, S., & Goldstein, T. (2019). Are adversarial examples inevitable? In *Seventh international conference on learning representations (ICLR 2019)*. New Orleans, LA.

[8] Tyukin, I.Y., Higham, D.J., & Gorban, A.N. (2020). On adversarial examples and stealth attacks in artificial intelligence systems. In *2020 International joint conference on neural networks (IJCNN)* (pp. 1-6). Glasgow, UK: IEEE.

[9] Tyukin, I.Y., Higham, D.J., Gorban, A.N., & Woldegeorgis, E. (2021). The feasibility and inevitability of stealth attacks. Preprint, *arXiv2106.13997*.

[10] Yuan, X., He, P., Zhu, Q., & Li, X. (2019). Adversarial examples: Attacks and defenses for deep learning. *IEEE Trans. Neural Netw. Learn. Syst.*, *30*(9), 2805-2824.

Alexander Bastounis is a research associate at the University of Edinburgh who works in optimization, machine learning, and the foundations of computation. In 2019, he received a Leslie Fox Prize for Numerical Analysis. Anders C. Hansen is a professor of mathematics at the University of Cambridge, where he leads the Applied Functional and Harmonic Analysis Group. He is also a professor of mathematics at the University of Oslo. Desmond J. Higham is a professor of numerical analysis at the University of Edinburgh. He is a SIAM Fellow and editor-in-chief of SIAM Review. Ivan Y. Tyukin is a professor of applied mathematics at the University of Leicester. He is an editor of Communications in Nonlinear Science and Numerical Simulation. Verner Vlačić graduated in 2020 with a Ph.D. in applied mathematics and engineering from ETH Zürich, where he worked on optimization and neural network theory. He currently works in the tech industry.

NIH to Inject Healthy Bolus of Data to Sustain the Future of AI for Medical Discoveries

National Institutes of Health Common Fund's Bridge2AI Program Will Build Ethical, Inclusive, and Interpretable Data Sets

By Thomas M. Johnson and Grace C.Y. Peng

Computational modeling and artificial intelligence (AI) are integrating into medicine and biomedical research at a dizzying pace. Practitioners are utilizing AI in all aspects of healthcare, from the staging of lung cancer nodules on MRI images to workflow management in hospitals.

At the same time, the recording process of physiological measures from smartwatches, smartphones, and sensors in clothes, shoes, contact lenses, and toilet seats—any place where material can rub up against the body—is also rapidly expanding. These burgeoning technologies are incredibly promising but require well-defined data that are freely accessible and understandable by the intertwined web of "intelligent" machines, which reliably and efficiently analyze the data. Researchers can then potentially store, curate, and add these data to a collective knowledge base that will spawn the next remarkable medical advance.

Recognizing both the potential and problems of such an endeavor, the National Institutes of Health (NIH) is funding an innovative approach to get the data right. The NIH Common Fund's Bridge to Artificial Intelligence (Brige2AI) program[1] seeks to "build the bridge" over which data flows into the biomedical and behavioral research space in a useable form that works for scientists, engineers, physicians, and patients.

"We have a deluge of data, but they are not defined enough for machines to understand," Bruce Tromberg, Director of the National Institute of Biomedical Imaging and Bioengineering (NIBIB), said. NIBIB is one of the five NIH Institutes and Centers that are leading this enterprise. The Bridge2AI program acknowledges a call from the research community[2] for the generation of new data that are designed for biomedical discovery; such data will propel modern AI/machine learning (ML) models towards scientific discovery, advance a culture of ethical consideration around the data, and create a workforce that is skilled in this new method of scientific data creation.

The Bridge2AI program emphasizes the need to move away from "business as usual" in order to generate data that machines can understand. These data—described as "hypothesis agnostic"—shake free from human-generated hypotheses and inferences. One must possess clear knowledge of the data's origin, including all aspects that surround data collection (which may comprise multiple streams of measurements and signals from individuals, groups, communities, populations, and regions). The program hopes to instill a culture of ethical inquiry that is based on the grand challenge problems that motivate data collection (see Figure 1, on page 6). For example, how will we preserve privacy when creating digital twins of ourselves [4, 5]? How should machines handle consent and minimize errors when interpreting genomic data? How do social, cultural, racial, and gender issues affect the collection of wearable sensor data?

The Bridge2AI program acknowledges the fact that truly comprehensive data is only possible through contributions by and data collection from individuals with diverse backgrounds and life experiences. A central goal of the program is to create a sustainable culture change with diversity at the forefront.[3] Such clean, connected, and multi-modal data is then poised to enable AI to find the hidden signals that derive unbiased knowledge and provide insight into human health.

Bridge2AI data generation projects will "stitch together" ethically-sourced data from diverse perspectives that are motivated by biomedical and behavioral grand challenges. And the Bridge2AI Integration, Dissemination, and Evaluation (BRIDGE) Center will ensure that the program's spirit and culture change goals are infused throughout the life cycle of Bridge2AI-generated data. NIH program managers plan to utilize a nimble funding mechanism that provides increased flexibility for the reconfiguration and recombination of project elements to best meet the Bridge2AI program's needs. Dissemination of the data, tools, standards, and ethical considerations will include "jamborees" that foster a lively and creative learning environment within the broader scientific community.

The Bridge2AI Working Group hopes that the program will accelerate a culture change that will persist in the long term, even after the program is completed. The Bridge2AI culture change is also reflected in other programs across the broader NIH community, which is committed to the incorporation of ethical considerations throughout the scientific

See **Medical Discoveries** *on page 6*

[1] https://commonfund.nih.gov/bridge2ai

[2] https://acd.od.nih.gov/documents/presentations/12132019AI_FinalReport.pdf

[3] https://commonfund.nih.gov/bridge2ai/enhancingdiverseperspectives#PEDP

Digital Twins

Continued from page 1

formulation should discretize easily in space and time to support simulation, control, and optimization, and the models' analytic properties should include all of the essential credentials of existence, stability, etc. Finally, Mehrmann wants to reach these goals within a strong systems-theoretic framework.

This mathematical framework is built with the widely applicable formulation that is known as dissipative port-Hamiltonian (pH) systems. The word "port" evokes the connections through which model components exchange information with one another; "Hamiltonian" evokes energy, the *lingua franca* of these exchanges. Intuitively (albeit crudely), the resistors, inductors, and capacitors that comprise an RLC circuit are examples of pH elements. Energy in each element is a time integral of power, which is itself a product of the current through the element and voltage drop across it (flow multiplied by effort). Because of the resistor, the energy in each element obeys a dissipation inequality rather than an equality.

A power-conserving ensemble of these pH circuit elements, such as an ordinary RLC circuit, forms another pH system. When augmented by the constraints of Kirchhoff's Current and Voltage Laws—themselves a set of differential-algebraic equations (DAEs)—the simplest variety of pH DAE system emerges. Formalizing the underlying dual-space geometry, extending to infinite dimensions, and incorporating time-dependent PDEs then lay the foundation for the mathematical half of Mehrmann's paradigm shift: a hierarchical structure of pH DAE models that preserves their properties, including the dissipation inequality, through model reduction.

The computational requirements are similar in spirit and involve the preservation of key pH DAE properties during spatial and temporal discretization. Galerkin projection can discretize spatial variables as needed, and time discretization must preserve the pH conservation and dissipation requirements up to the discretization error.

Computational experiments have demonstrated some of the successes of pH DAE systems. For example, Mehrmann and his collaborators studied a four-level gas transport model hierarchy to adaptively manage the compromises between error tolerance and computational speed when constructing a digital twin. After determining sensitivities within the hierarchy as well as error estimates for the space and time discretizations, they fit a cost function that relates space and time discretization errors to CPU time. The "truth" was an expensive isothermal Euler equation model; the adaptive test offered three successive hierarchical stages of simplification. The adaptively constructed model reduced the four-hour runtime that was required by the original formulation by 80 percent [1, 5].

A similar effort with more complex gas networks demonstrated that an adaptive strategy can also construct an efficient digital twin to optimize the cost of compressor operation [4]. Another study displayed an effective tool for optimizing household power consumption in a district heating network whose capacity for generating power via waste incineration was limited [3].

How did Mehrmann's digital twin modeling wish list fare? He declared that "pH systems are great, and almost all wishes are fulfilled." Of course, much work remains. Mehrmann's to-do list includes the incorporation of real-time control and optimization, different physical domains, stochastic elements, stability analysis, error estimation, and new software, among other components.

References

[1] Domschke, P., Dua, A., Stolwijk, J.J., Land, J., & Mehrmann, V. (2018). Adaptive refinement strategies for the simulation of gas flow in networks using a model hierarchy. *Elect. Trans. Numer. Analys.*, *48*, 97-113.

[2] Kapteyn, M.G., & Willcox, K.E. (2021, September 1). Digital twins: Where data, mathematics, models, and decisions collide. *SIAM News*, *54*(7), p. 1.

[3] Krug, R., Mehrmann, V., & Schmidt, M. (2020). Nonlinear optimization of district heating networks. *Opt. Engin.*, *22*, 783-819.

[4] Mehrmann, V., Schmidt, M., & Stolwijk, J.J. (2018). Model and discretization error adaptivity within stationary gas transport optimization. *Vietnam J. Math.*, *46*, 779-801.

[5] Stolwijk, J.J., & Mehrmann, V. (2018). Error analysis and model adaptivity for flows in gas networks. *Analele Stiintifice ale Universitatii Ovidius Constanta, Seria Matematica*, *26*(2), 231-266.

Paul Davis is professor emeritus of mathematical sciences at Worcester Polytechnic Institute.

A New Biography of Kurt Gödel

BOOK REVIEW
By Ernest Davis

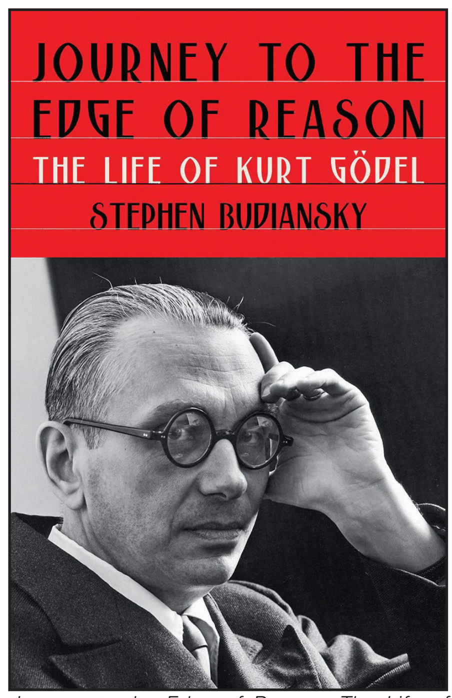
Journey to the Edge of Reason: The Life of Kurt Gödel. *By Stephen Budiansky. W.W. Norton & Company, New York, NY, May 2021. 368 pages, $30.00.*

Kurt Gödel's incompleteness theorem, which demonstrates the existence of mathematical propositions about natural numbers that one can neither prove nor disprove, is probably the best-known mathematical theorem of the 20th century. This theorem, together with Gödel's other seminal results—the inference system of first-order logic is complete, the axioms of arithmetic cannot be proved consistent, the axiom of choice is consistent with the Zermelo-Frankel axioms of set theory, and the continuum hypothesis is consistent with the ZFC axioms—firmly establish his position at the pinnacle of logicians.

Stephen Budiansky's new biography, *Journey to the Edge of Reason: The Life of Kurt Gödel*, is the sixth English-language biography of Gödel; among 20th-century mathematicians, only Alan Turing, Srinivasa Ramanujan, Emmy Noether, Bertrand Russell, and Norbert Wiener have more. Previous works about Gödel include biographical accounts; philosophical disquisitions; and expositions of Gödel's mathematical results, achievements in mathematical logic, and relation to the field's history.[1]

Gödel's life fell sharply into three parts: his childhood in Brünn, Austria-Hungary (1906-1924); his young adulthood at the University of Vienna (1924-1940); and his later years at the Institute of Advanced Studies (IAS) in Princeton, NJ (1940-1978). Biographies of Gödel naturally tend to follow this breakdown. But the first part of Budiansky's book only occasionally describes Gödel or his family; rather, it provides an entertaining historical and cultural portrait of the Austro-Hungarian Empire at the turn of the 20th century. Oddly, Budiansky barely mentions World War I — a reader who does not know what happened in Europe between 1914 and 1918 would hardly guess it from his account.

Gödel conducted most of his important mathematical work during his years in Vienna. Budiansky's narrative of these years includes an extended description of the Vienna Circle — a philosophical enterprise with the ambitious goal of formulating science as the logical analysis of sense data, inspired by the model of mathematics' construction from logic in *Principia Mathematica* by Alfred North Whitehead and Bertrand Russell. Gödel's doctoral advisor, Hans Hahn, was a founder of the Vienna Circle, and Gödel was one of very few students admitted as a regular member. He frequently attended gatherings, befriended some of the younger participants, and even presented his incompleteness theorem at a group meeting. However, the group did not contribute to Gödel's mathematics and he rejected its philosophy.



Journey to the Edge of Reason: The Life of Kurt Gödel. By Stephen Budiansky. Courtesy of W.W. Norton & Company.

Budiansky's book comprises careful, readable accounts of Gödel's most important results, particularly the incompleteness theorem and the consistency of the continuum hypothesis. He includes an appendix with a proof of the incompleteness theorem, though like most such presentations he omits the full demonstration that one can express the predicate "Provable(p)" in the language of arithmetic. Budiansky's coverage of other mathematical topics is sketchier. In a book for general readership, omitting some of Gödel's more technical accomplishments—such as his proof of the compactness theorem and contributions to intuitionistic logic—is certainly reasonable. It is less satisfactory for important colleagues and interactions to go unmentioned. For example, Alonzo Church—a leading logician in the U.S. during Gödel's time—and his many brilliant doctoral students were at Princeton University, only a few blocks away from the IAS. Budiansky barely mentions Church, even though Gödel had significant interactions with Church and his students Stephen Cole Kleene, John Barkley Rosser, and Dana Scott.[2]

Gödel visited the U.S. three times in the 1930s, but he was in Vienna when World War II broke out in September 1939. After Gödel obtained permission to leave Vienna, John von Neumann arranged for him to receive a visa to the U.S. and a position at the IAS. Gödel and his wife Adele traveled east, taking the Trans-Siberian Railway to Vladivostok, a steamer to Yokahama, an American ship to San Francisco via Honolulu, and a train to Princeton. He never again left the east coast of the U.S. Gödel spent the rest of his life at the IAS, working primarily on mathematical philosophy and publishing little. He had a small number of close friends there on whom he relied, especially Albert Einstein (the two walked to work together every morning) and mathematician and economist Oskar Morgenstern. Gödel wrote to his mother in Vienna warmly and frequently; they remained very close, and she and his older brother visited Princeton in 1958.

In general, Budiansky's book portrays mathematicians quite positively. The importance of Gödel's results—especially the incompleteness theorem—and the quality of his work were immediately and almost universally acknowledged, barring a handful of exceptions like Ernst Zermelo and Ludwig Wittgenstein. Throughout his life, Gödel's colleagues served as supportive and protective friends.

Though quiet, Godel was friendly, sociable, and even charming in his younger years, with a whimsical sense of humor. He was always immaculately dressed and groomed; according to fellow student Olga Taussky-Todd, Gödel was even "a bit ostentatious about his success in attracting the opposite sex." But he was not at all egotistical. In fact, when Morgenstern met Gödel's mother in Vienna after World War II, he was surprised to learn that she had no idea of her son's great successes. In 1960, when Paul Cohen completed the proof that the continuum hypothesis is independent of Zermelo-Fraenkel set theory (which Gödel had been working on for decades), Gödel was nothing but excited and enthusiastic in his response. He was not an elitist either culturally or socially; his artistic tastes were mostly lowbrow and his wife Adele—to the horror of his family and many of his friends—was an uneducated dancer.

Concurrent with Gödel's success was turmoil in both the outside world and within his own psyche. Budiansky's description of the rise of fascism in Austria is dramatic and harrowing. The government, led by Engelbert Dollfuss and Kurt Schuschnigg, grew steadily more authoritarian and anti-Semitic, until it was swept away by Hitler's Anschluss. The University of Vienna offered no refuge. "In 1923, the German Student Union demanded that all books in the library written by Jews be marked with a Star of David," Budiansky writes. "Vienna's *Neue Freie Presse* reported that on a stroll around the main building of the university, a visitor would encounter little but anti-Semitic posters and hate literature." In 1933, Gödel's colleague and friend Karl Menger wrote to Oswald Veblen about the circumstances. "[T]he situation at the university is as unpleasant as possible," he said. "Whereas I still don't believe that Austria has more than 45 percent Nazis, the percentage at the university is certainly 75 percent, and among the mathematicians I have to deal with, not far from 100 percent." Moritz Schlick, the chair of the Vienna Circle, was killed by a former student in 1936; though Schlick was neither Jewish nor politically active, the Nazis defended his murder as a justified reaction to his perverse philosophy. It must be noted, however, that there is no trace of either anti-Semitism or Nazi sympathy in Gödel's record.

See **Kurt Gödel** *on page 7*

[1] Existing biographies of Gödel include *Reflections on Kurt Gödel* by Hao Wang (1987), *Logical Dilemmas: The Life and Work of Kurt Gödel* by John W. Dawson, Jr. (1997), *Gödel: A Life of Logic* by John L. Casti and Werner DePauli (2000), *Incompleteness: The Proof and Paradox of Kurt Gödel* by Rebecca Goldstein (2005), and *Simply Gödel* by Richard Tieszen (2017).

[2] Thanks to Martin Davis for helpful information on this point.

Chaotic Mixing

Continued from page 2

The crucial ingredient to proving Theorem 0.1 is to show that solutions to (2) are exponentially mixing with a rate γ that is *uniform in diffusivity* κ. Several works collectively prove the following theorem [3, 4, 6].

Theorem 0.2

This theorem pertains to uniform-in-diffusivity, almost-sure exponential mixing [6]. *Let (u_t) solve (2). For all $\nu > 0$, a deterministic $\gamma > 0$ exists such that a random (almost surely finite) constant $D = D_\kappa(u, \omega)$ is present for all initial conditions u_0, noise paths ω, and $\kappa \in [0,1]$. This means that for every zero-mean $g_0 \in H^1$, (4) holds for all $t > 0$. The random constant D has second moment finite uniformly in κ:*

$$\mathbf{E}D^2 \leq C(u_0).$$

$C(u_0)$ only depends on the initial velocity u_0; in fact, it depends only on the L^2 norm of u_0 and its derivatives [6].

An important feature of this theorem is that (4) holds *almost surely*, so that nearly every fixed noise path ω gives an exponentially mixing velocity field u_t (which is deterministic after ω is fixed). While D is random and depends on the velocity field's initial condition, γ is deterministic and only depends on the structural parameters (e.g., the noise coloring, ν, etc.).

The proof of Theorem 0.2 uses a blend of mathematical ideas from random dynamical systems, spectral theory of Markov processes, and stochastic partial differential equations. Details and the theorem's proof are available in [3, 4, 6], and the proof of Batchelor's law (as in Theorem 0.1 from Theorem 0.2) is available in [5].

The proof of Batchelor's power law for scalars that are advected by spatially regular incompressible velocity fields (u_t)—such as (2) on the periodic box $\mathbb{T}^d$—provides some insight into the validity of Batchelor's law outside this strict, mathematically expedient setting. We now generally know when to expect the spectrum, even when a complete mathematical proof is not readily available.

Our results only scratch the surface of potential mathematically rigorous outcomes in this area. We believe that more sophisticated studies will prove our results for (2) when subjected to more realistic stochastic forcing — at least when the forcing is k-times differentiable in time and smooth in space. Similarly, studies of fluid that is driven by the stochastic motion of solid boundaries rather than somewhat non-physical body forcing would yield substantial value. However, much mathematical work must still be done before more complicated statistical problems—like Navier-Stokes at a high Reynolds number with stochastic forcing—are tractable. Indeed, although the proofs of Theorems 0.1 and 0.2 require many mathematical ideas, they are certainly among the easiest of such statistical questions in fluid mechanics. Nevertheless, we hope that our advancements will help spur new activity and eventually lead to larger progressions in these directions.

References

[1] Antonia, R.A., & Orlandi, P. (2003). Effect of Schmidt number on small-scale passive scalar turbulence. *Appl. Mech. Rev., 56*(6), 615-632.

[2] Batchelor, G.K. (1959). Small-scale variation of convected quantities like temperature in turbulent fluid part 1: General discussion and the case of small conductivity. *J. Fluid Mech., 5*(1), 113-133.

[3] Bedrossian, J., Blumenthal, A., & Punshon-Smith, S. (2018). Lagrangian chaos and scalar advection in stochastic fluid mechanics. *J. Euro. Math. Soc.*, to appear.

[4] Bedrossian, J., Blumenthal, A., & Punshon-Smith, S. (2019). Almost-sure exponential mixing of passive scalars by the stochastic Navier-Stokes equations. *Ann. Prob.*, to appear.

[5] Bedrossian, J., Blumenthal, A., & Punshon-Smith, S. (2019). The Batchelor spectrum of passive scalar turbulence in stochastic fluid mechanics at fixed Reynolds number. *Comm. Pure Appl. Math.*, to appear.

[6] Bedrossian, J., Blumenthal, A., & Punshon-Smith, S. (2021). Almost-sure enhanced dissipation and uniform-in-diffusivity exponential mixing for advection-diffusion by stochastic Navier-Stokes. *Prob. Theory Rel. Fields, 179*(3), 777-834.

[7] Gibson, C.H., & Schwarz, W.H. (1963). The universal equilibrium spectra of turbulent velocity and scalar fields. *J. Fluid Mech., 16*(3), 365-384.

[8] Grant, H.L., Hughes, B.A., Vogel, W.M., & Moilliet, A. (1968). The spectrum of temperature fluctuations in turbulent flow. *J. Fluid Mech., 34*(3), 423-442.

[9] Thiffeault, J.-L. (2012). Using multiscale norms to quantify mixing and transport. *Nonlin., 25*(2), R1-R44.

Jacob Bedrossian is a professor of mathematics at the University of Maryland, College Park. He earned his Ph.D. from the University of California, Los Angeles in 2011. Alex Blumenthal is an assistant professor of mathematics at the Georgia Institute of Technology. He earned his Ph.D. from New York University's Courant Institute of Mathematical Sciences in 2016. Sam Punshon-Smith is a member of the Institute for Advanced Study and an assistant professor of mathematics at Tulane University. He earned his Ph.D. from the University of Maryland, College Park in 2017.

Medical Discoveries

Continued from page 4

discovery pipeline. For example, the Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) Initiative[4] has been identifying neuroethical issues for consideration. More information is available in the BRAIN Neuroethics report.[5]

In addition to co-leading the Bridge2AI program, NIBIB serves a critical role in the promotion of novel technologies, including new methods that make these technologies useable and translatable. NIBIB supports the Center for Reproducible Biomedical Modeling,[6] which creates tools for reproducible mechanistic models. NIBIB also leads the Interagency Modeling and Analysis Group (IMAG),[7] which coordinates the Multiscale Modeling Consortium (MSM).[8] In 2019, IMAG and MSM held a pivotal meeting that combined mechanistic modeling with ML. The resulting papers [1, 6] set the stage for the integration of these two modeling modalities to ultimately create digital twins and increase patient safety by reducing medical errors [3]. In 2020, a new working group on Multiscale Modeling and Viral Pandemics[9] formed within the MSM. IMAG and MSM entities also worked to implement "Ten 'Not So' Simple Rules for Credible Practice of Modeling and Simulation" [2] on COVID-19 models.

As a bioengineer, Tromberg views mathematics as the cornerstone of the Bridge2AI program. "One of the unifying characteristics of biomedical engineers is our commitment to the idea that biological processes can be represented symbolically, by mathematical equations," he said in a video for Bridge2AI's kickoff.[10] "With enough measurements and the right equations, we believe that it's possible to understand and predict the behavior of any complex biologic system."

The mathematical and statistical communities that SIAM serves are well positioned to embrace the culture change that is promoted by the NIH Bridge2AI program, the NIH BRAIN Initiative, IMAG, and MSM. We at the NIH encourage readers to establish diverse teams that sustain equitable partnerships and work with data generators to create novel mathematical and statistical methods that address ethical issues surrounding data privacy, research consent, error reduction in scientific and clinical workflows, automatic data, and model annotation. We urge members of the SIAM community to collaborate with mechanistic modelers to design new modeling frameworks that complement AI/ML models. These methods will serve as powerful tools that rapidly accelerate the new future for scientific discovery and improved global health.

[4] https://braininitiative.nih.gov
[5] https://braininitiative.nih.gov/strategic-planning/acd-working-groups/brain-initiative%C2%AE-and-neuroethics-enabling-and-enhancing
[6] https://reproduciblebiomodels.org
[7] https://www.nibib.nih.gov/research-program/interagency-modeling-and-analysis-group-imag
[8] https://www.imagwiki.nibib.nih.gov
[9] https://www.imagwiki.nibib.nih.gov/working-groups/multiscale-modeling-and-viral-pandemics
[10] https://www.youtube.com/watch?v=yZ-4tDvT6lY

References

[1] Alber, M., Buganza Tepole, A., Cannon, W.R., De, S., Dura-Bernal, S., Garikipati, K., ..., Kuhl, E. (2019). Integrating machine learning and multiscale modeling — perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *Digital Med.*, *2*, 115.

[2] Erdemir, A., Mulugeta, L., Ku, J.P., Drach, A., Horner, M., Morrison, T.M., ..., Myers Jr., J.G. (2020). Credible practice of modeling and simulation in healthcare: Ten rules from a multidisciplinary perspective. *J. Translat. Med.*, *18*, 369.

[3] Makary, M.A., & Daniel, M. (2016). Medical error — the third leading cause of death in the US. *BMJ*, *353*(1), 353:i2139.

[4] Masison, J., Beezley, J., Mei, Y., Ribeiro, H., Knapp, A.C., Sordo Vieira, L., ..., Laubenbacher, R. (2021). A modular computational framework for medical digital twins. *Proc. Nat. Acad. Sci.*, *118*(20), e2024287118.

[5] Niederer, S.A., Sacks, M.S., Girolami, M., & Willcox, K. (2021). Scaling digital twins from the artisanal to the industrial. *Nat. Comput. Sci.*, *1*, 313-320.

[6] Peng, G.C.Y., Alber, M., Buganza Tepole, A., Cannon, W.R., De, S., Dura-Bernal, S., ..., Kuhl, E. (2021). Multiscale modeling meets machine learning: What can we learn? *Archiv. Comput. Meth. Engin.*, *28*(1), 1017-1037.

Thomas M. Johnson holds a Ph.D. in human genetics from the University of Michigan. He has worked at the National Institutes of Health (NIH) for 25 years as a postdoctoral research fellow, an extramural program manager, and the Deputy Director of the Office of Science Policy Analysis. He is currently a Senior Advisor and science writer at the NIH's National Institute of Biomedical Imaging and Bioengineering (NIBIB) in the U.S. Department of Health and Human Services. Grace C.Y. Peng, Ph.D., is the Director of Mathematical Modeling, Simulation, and Analysis at the NIBIB, where she has programmatic oversight of extramural activities in these areas. In 2003, she led the creation of the Interagency Modeling and Analysis Group (IMAG). She also holds leadership roles in the NIH Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) Initiative and the NIH Bridge to Artificial Intelligence (Brige2AI) program.

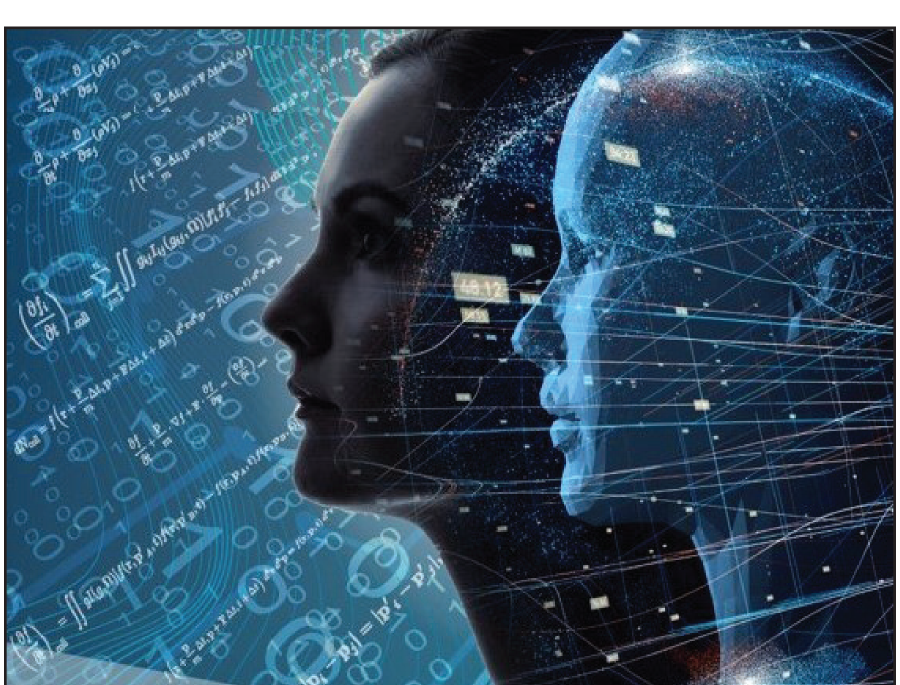


Figure 1. *The National Institutes of Health's Bridge to Artificial Intelligence (Bridge2AI) program supports the future of AI by promoting the creation of new biomedical and behavioral data that is ethically sourced and prepared from diverse perspectives. Researchers can combine this complex data with mechanistic mathematical models to create digital twins that yield novel insights into human health [4, 5]. Figure adapted from iStock/metamorworks and courtesy of the National Institute of Biomedical Imaging and Bioengineering.*

Toward Student-Centric Graduate Training

By Yara Skaf and Reinhard Laubenbacher

Career opportunities have never been better for people with Ph.D.s in the mathematical sciences. In addition to traditional employment in mathematics departments and non-academic engineering-based positions, a wide array of new careers in academia and beyond are now available due to the rapidly increasing use of mathematical and computational techniques in many fields. Medicine, education, and finance are prime examples of such fields. The great variety of requisite skills and backgrounds for these careers demands greater customization of student training programs. We intend to address possible ways in which academic departments can achieve student-centric program structures.

A 2018 report from the National Academies of Sciences, Engineering, and Medicine [2] offered several recommendations for reform, including more student-centric programs that account for students' career goals and interests and provide extensive advising. This advice is timely for the mathematical sciences community. Mathematics is unique; it is a research field in its own right that simultaneously provides a universal language and technology that now touch virtually every other scientific, technological, and social aspect of society. This pertinency leads to unprecedented career opportunities for highly trained mathematicians and mathematics educators, even as academic job prospects expand only moderately [3, 4]. At the same time, it challenges the mathematics community to adjust its training paradigms to keep up with these developments. Here we present some thoughts as to how such an adjustment might look. An expanded, more nuanced version of this article is available online [5].

The traditional model of doctoral education in mathematics effectively prepares trainees for tenure-track academic careers with a dual research/teaching mission (with various degrees of emphasis on research). But many Ph.D. students will ultimately pursue career paths that diverge from this trajectory and thus require different preparation. Programs that acknowledge this reality and adapt their training accordingly will better serve their students and allow them to take full advantage of the myriad exciting career opportunities inside and outside of academia. They will also attract a broader range of students who are interested in advanced mathematics training, thereby growing Ph.D. programs. Discussions about the effective training of doctoral students for modern careers are by no means new, and a robust body of literature exists on the topic [5].

Program Structure

A survey of the literature supports the following observations. First, most Ph.D. programs in the mathematical sciences take a "one-size-fits-all" approach, meaning that the program elements and requirements are essentially the same for every student. The vast majority of the top 50 Ph.D. programs in mathematics—as ranked by U.S. News & World Report [7]—still follow the traditional "algebra, analysis, topology" preliminary examination schedule that occupies the first two years of graduate school, with only minor variations. Second, the interests and career trajectories of graduate trainees extend far beyond tenured positions in academic institutions to a wide range of additional opportunities for which current training does not provide optimal preparation [1, 6]. To add perspective, we briefly contrast the key features of mathematics programs with two other fields that might provide useful ideas.

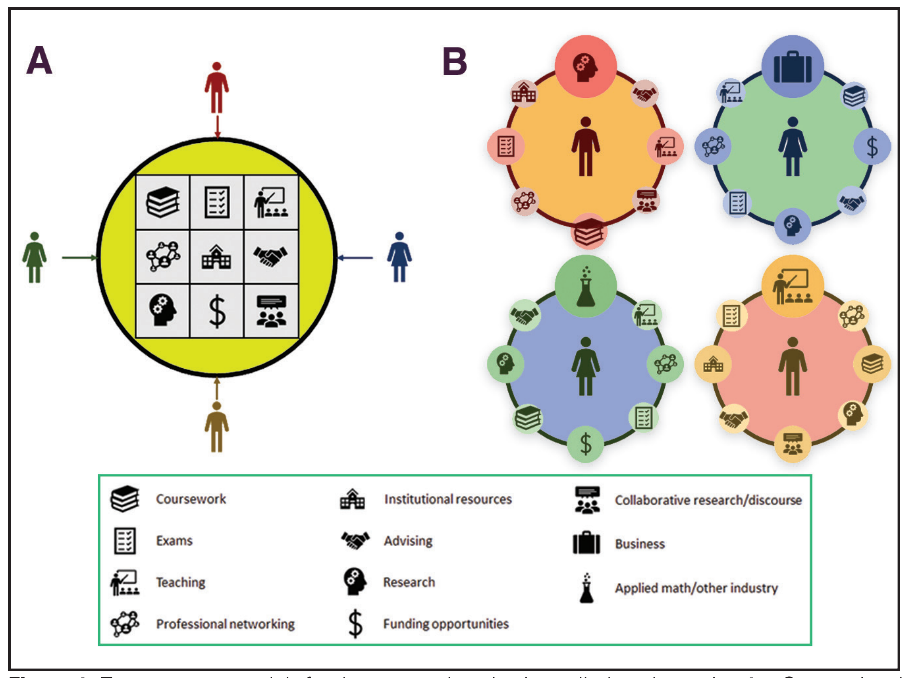


Figure 1. *Two separate models for doctorate education in applied mathematics.* **1a.** *Conventional approach to graduate education. Students access a predetermined set of training resources in a standard manner that is independent of any specific needs or interests.* **1b.** *Training activities are organized into four tracks around students' personal career objectives. Clockwise from top left: academic research, private sector, education, and engineering-based industry. Each track weights available resources differently. Figure courtesy of the authors and SIAM.*

Mathematics: In addition to dissertation research, standard requirements consist of preliminary examinations in several core subjects, qualifying examinations in a chosen focus area, and substantial coursework. Because stipends are predominantly funded through teaching endeavors, graduate students must tutor, grade, and otherwise assist with multiple undergraduate classes every semester — or teach an entire class themselves.

Biological sciences: In contrast to mathematics programs, most programs in the life sciences require minimal coursework and no preliminary exams. Instead, students must put together a thesis proposal. Mandatory classes are usually more like seminars or journal clubs in which students read and discuss research papers; these courses contain little to no new theory beyond what is taught in undergraduate curricula. Students immediately begin rotating through different faculty laboratories to select an advisor for their Ph.D. research project, which is well underway by the end of the first year. They are also expected to assist with grant proposals, prepare manuscripts, and present original research.

M.D./Ph.D. programs: M.D./Ph.D. programs train students to conduct translational research that generates or applies basic scientific results for improved medical care. As such, the Ph.D. portion of these programs tends to emphasize research activities almost exclusively. A typical M.D./Ph.D. program includes four years of biomedical science Ph.D. training, which is interspersed with four years of medical school. To enable this accelerated completion of a Ph.D. in four years rather than five or six, most programs have very few general requirements.

Many factors contribute to the significant differences between these three types of programs, including incoming students' preparedness for research projects. For example, most first-year graduate students in biology are capable of driving and executing significant research projects on their own simply because the theoretical foundation for biology is constructed long before graduate school. Another contributing factor is the difference in professional expectations. Graduates in the life sciences or M.D./Ph.D. programs are expected to have several peer-reviewed publications by graduation; be well prepared for the demands of professional positions; and possess the skills to design research projects, write papers and grant applications to support their research, and deliver presentations at conferences. In contrast, most mathematics programs do not strongly emphasize these skills.

A Proposal

Mathematics Ph.D. programs should strive for completely personalized training schedules that are built around the unique interests and career aspirations of each student. Such personalization should extend to coursework (or other means of acquiring expertise); professional training like teaching or internships in business, industry, or government; and the research

See **Graduate Training** *on page 10*

Kurt Gödel

Continued from page 5

Gödel himself dealt increasingly with obsessive-compulsive disorder, severe hypochondria, and an obsession with conspiracy theories. He experienced complete mental collapses in 1936 and 1970 that were marked by extreme paranoia and a fear of being poisoned; he only ate food that Adele spoon-fed him. In the end, Gödel effectively starved himself to death — he weighed just 65 pounds when he died.

As many scholars have noted, a straight line runs from some of Gödel's intellectual characteristics—even his intellectual virtues—to other aspects of his mental state. He was always obsessed with determining the reasons for things and philosophically believed that everything has a reason; this mindset can clearly slide into conspiracy theories. Gödel was also known for being exceptionally careful and wanting to get everything exactly right before publishing or announcing any results. In the context of his own health, this same caution turned into hypochondria. Yet his paranoia, which ultimately killed him, was much less of a constant in his psychological makeup and does not seem to connect in the same way.

Nevertheless, when we think of Kurt Gödel, let us remember the young man—handsome, charming, impressively brilliant—who saw more deeply into the nature of mathematical proofs than anyone had before and realized that a mathematical proposition could be encoded as a number — that an entire proof could be encoded as a number. He understood that "being provable" was just an arithmetic property like "being prime" and that, through an ingenious twist of legerdemain, he could thereby construct a proposition that mocked the hopes of mathematicians with its paradoxical assertion: "This sentence is not provable."

Ernest Davis is a professor of computer science at New York University's Courant Institute of Mathematical Sciences.

AN21 Panel Highlights Career Opportunities in Data Science

By Lina Sorg

The fields of data science and machine learning (ML) have grown tremendously in recent years, creating numerous career opportunities for applied mathematicians and computational scientists in a wide variety of industries. During the 2021 SIAM Annual Meeting,[1] which took place virtually in July, a panel of data scientists discussed their experiences in this burgeoning field. The panel, which was sponsored by the SIAM Industry Committee,[2] was chaired by Lalitha Venkataramanan (Schlumberger-Doll Research[3]) and consisted of Berton A. Earnshaw (Recursion Pharmaceuticals[4]), Stephanie Fitchett (Transamerica[5]), Stephen Jones (The Boeing Company[6]), and Nandi Leslie (Raytheon Technologies[7]). After an hour of general conversation, attendees entered a series of breakout rooms to directly interact with the speakers.

Earnshaw opened the session by acknowledging the blossoming nature of ML, computer vision, deep learning, artificial intelligence, and neural networks, all of which are advancing with unprecedented speed. "It's a fast-moving field," he said. "A great habit to get into is setting aside time to read and stay on top of what's going on in whatever field you're working on. That's a critical component for success in whatever your role ends up being." He also suggested that researchers join a journal club or reading group to stay abreast of current developments in their particular study areas. Some companies even sponsor employee participation in such groups, though plenty of virtual meetups are readily available as well.

Leslie seconded the importance of keeping up with the latest research and encouraged attendees to take courses in programming languages to make themselves more competitive. She commented that an applied mathematics education inherently supplies many of the skills that comprise the data science toolset; for example, computational science students routinely utilize ML approaches and classic modeling techniques that involve differential equations, stochastic processes, and topological concepts. Establishing a solid background in programming and coding allows students to effectively implement these concepts and increases their hiring desirability.

Some industry employers offer regular internships that provide students with hands-on experience while they are still earning their degrees. For instance, Transamerica had three interns this past summer — all of whom were graduate students in master's or Ph.D. programs. While some Transamerica interns come directly from data science programs, Fitchett noted that many have backgrounds in mathematics, statistics, computer science, and even physics, logic, and computational chemistry. "That's been very fun for me," she said.

Recursion Pharmaceuticals maintains a robust internship program and hosts three cohorts of interns every year. As the organization continues to expand, Earnshaw expects internship opportunities to grow as well. Interns typically tackle interesting problems that are carved out in convenient three- or four-month blocks, and openings are available on the company's career page.[8] Both the BIG Math Network[9] and SIAM's Career Resources page[10] advertise a host of job and internship listings as well, and Venkataramanan advised listeners to join SIAM for full access to career-related information.[11]

Conversation then turned to the hiring process. While internships are certainly valuable, a lack of internship experience does not immediately disqualify applicants from job openings. In fact, Jones focuses on candidates who have applied their academic knowledge to real-world problems in any setting, including volunteer work at a nonprofit or even data analysis for graduate students in different departments. "Any experience where you've shown your ability to learn and applied it to a practical situation is a plus," he said.

CAREERS IN MATHEMATICAL SCIENCES

While it is beneficial for researchers at all career stages to work with people who are experienced in their fields of interest, this type of involvement is particularly worthwhile for recent graduates. "A big driver for the position that I was looking for was to be part of a group of statisticians who were doing the same work that I was interested in doing," Jones said.

When it comes time to select a career path after graduation, Earnshaw urged students to refrain from dwelling too much on the typical "academia versus industry" debate and instead select the direction that feels right in the moment. Earnshaw was initially on an academic path and even held two postdoctoral appointments before ultimately pursuing industry and entrepreneurship. After completing his postdoc, he co-founded a company because it felt like a good challenge at the time. He finds his present employment at Recursion equally satisfying. "I have the opportunity to have a huge impact with my company's mission, using the skills that I developed during school to improve the world around me by hopefully improving patients' lives," Earnshaw said. He also reminded attendees that the choice between academia and industry is less of an either/or decision nowadays; pursing one does not prevent participation in the other. For instance, Earnshaw is currently an adjunct professor at the University of Utah.

A panel at the 2021 SIAM Annual Meeting, which took place virtually in July, addressed the many opportunities in data science for applied mathematicians and computational scientists, and offered advice for students and early-career professionals who wish to enter the field. Top row, left to right: Lalitha Venkataramanan (Schlumberger-Doll Research), Berton Earnshaw (Recursion Pharmaceuticals), and Stephanie Fitchett (Transamerica). Bottom row, left to right: Stephen Jones (The Boeing Company) and Nandi Leslie (Raytheon Technologies).

Earnshaw noted that the amount of on-the-job training for industry positions depends on the organization in question. For example, some larger companies have very specific onboarding experiences with mentors and rotation schedules. This process is less common for smaller companies, though many do assign mentors to bring new employees up to speed. "In a smaller company, there just aren't as many resources to make that a formal training process," Earnshaw said. "It probably varies very closely with the size and maturity of the company that you're with."

Next, panelists spoke about the balance of technical versus managerial work within the industry sector. "There are a number of opportunities in industry that are individual contributor-focused, and there are plenty of career opportunities that are management-focused," Leslie said. "It depends on your path." She affirmed that researchers who hope to serve as career individual contributors can spend 100 percent of their time on research and development (R&D) — even in a large, publicly traded company like Raytheon. However, one can also become an individual contributor with a business development role and work on business proposals that focus on capture management; this route requires both technical understanding of the state of the art in industry and technology as well as experience in business and management.

While some early-career professionals may choose to pursue entirely management-based careers, Leslie remarked that most data scientists begin as individual contributors. The first 10 years of a career in science, technology, engineering, and mathematics (STEM) tend to be mostly technical—at a national laboratory, they might center entirely on R&D—after which one can opt to transition to a business role.

To provide attendees with a feel for various data-science-based industry assignments, all panelists overviewed ongoing ventures at their respective companies. Fitchett is presently working on a fraud claims project that functions like an outlier detection problem and employs a model—built by two data scientists—that identifies unusual claims for accident insurance. Transamerica then sends the model to its business partners, who pick up alerts, investigate them in further detail, and occasionally forward them to the state for insurance fraud investigation.

Leslie's current endeavors, which focus on anomaly detection in the context of cyber intrusion, utilize mathematical approaches that are similar to Fitchett's methods. Because anomaly detection is a broad field, its techniques are applicable to multiple different areas.

In contrast, Jones is tackling a regression-like problem to understand the quality of a particular aircraft structure. His team created a sampling plan to examine the aircraft and developed a model to understand quality defects in the production system. The group then feeds the analysis to the engineering team, which ensures that the product can tolerate any quality issues that may arise. In this case, data science and statistical analysis comprise small parts of a larger project on the assembly of aircraft structure.

Earnshaw then discussed his work at Recursion, where researchers generate large amounts of data about several types of

See **Data Science** *on page 9*

[1] https://www.siam.org/conferences/cm/conference/an21
[2] https://www.siam.org/about-siam/committees/industry-committee
[3] https://www.slb.com
[4] https://www.recursion.com
[5] https://www.transamerica.com
[6] https://www.boeing.com
[7] https://www.rtx.com

[8] https://www.recursion.com/careers
[9] https://bigmathnetwork.org
[10] https://www.siam.org/careers/resources
[11] https://www.siam.org/students-education/programs-initiatives/thinking-of-a-career-in-applied-mathematics

Statistical Methods for Adaptive Interventions and Precision Medicine

By Michael R. Kosorok and Erica E.M. Moodie

The study of new medical interventions—or new sequences of existing interventions—is inextricably linked with the field of statistics, which extracts meaning from noise and provides researchers with a principled approach for understanding potentially large volumes of data. In recent years, the use of statistics for precision medicine—an area of medical treatment that tailors medical care via patient covariates that may be demographic, clinical, or biological—has seen tremendous development. Precision approaches also have applications in other areas, such as educational and behavioral interventions.

Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine,[1] edited by Michael R. Kosorok and Erica E.M. Moodie, was published by SIAM in 2015. The volume addresses both introductory and advanced topics on statistical methods for the estimation of precision medicine (also called dynamic treatment regimes or adaptive treatment strategies). It is split into two parts and considers issues that pertain to study design as well as data analysis. The book is accessible to a wide audience of statisticians and computer scientists with a focus on machine learning, although many chapters are also appropriate for epidemiologists and medical researchers with only modest statistical backgrounds.

The following text is a short excerpt from chapter two, entitled "DTRs and SMARTs: Definitions, Designs, and Applications." It has been modified slightly for clarity.

To address some of the shortcomings of analyzing observational data and provide prospective evidence of dynamic treatment regime (DTR) effects, much of the existing literature on clinical trial design examines DTRs [5-10]. These studies refer to the sequential multiple assignment randomized trial (SMART), wherein individuals are randomized multiple times and follow specific DTRs (see Figure 1). The intention of this type of design is to develop DTRs, estimate the outcomes for each regime in the trial, and select the most promising DTR to compare to standard of care in a follow-up randomized control trial [7]. This objective addresses the dimensionality challenge of not just answering the question "What treatment when?" but also determining how to best use tailoring variables and information before selecting treatment. Since it is unlikely that one will have optimized all of these components prior to conducting a SMART, the goal is to conduct a series of trials that build upon one another, thereby developing and refining promising DTRs and leading to a confirmatory trial that is similar to the multiphase experimental approach [1-4]. While this scenario is ideal, the time and cost of trials may limit the intent and lead practitioners to treat SMARTs as confirmatory trials.

FROM THE SIAM BOOKSHELF

Just as a randomized control trial is a fixed design that generally compares two or more treatments, a SMART is a fixed design that compares or constructs two or more treatment regimes. A SMART is thus a trial and a DTR is a guideline that is carried out by a physician. SMARTs aim to construct effective DTRs. The same individuals that begin a SMART are followed throughout multiple randomizations until the end of the trial, with fixed randomization probabilities and other trial operational characteristics. Therefore, SMARTs can address questions about the best treatment at certain points in time; the best sequences of treatments (or best modes of treatment delivery) depending on intermediate outcomes; the best intermediate outcomes to direct treatment; and the best ways to individualize sequences of treatments based on biological, diagnostic, and/or other patient information. As with any trial, there should be one primary objective. However, SMARTs may lead to more secondary and exploratory aims due to the tailoring of DTRs. In order for SMARTs to be feasible, the intermediate outcome must be available for assessment within a relatively short time period — likely not more than one year. This is mainly due to the scope and relevance of conducting a trial. Thus, for some diseases with long assessment periods (treatment of some breast or prostate cancers, for example), a SMART is not an appropriate choice. However, SMARTs may still be relevant when treating other comorbidities or mental health issues that are related to these types of diseases.

The most common SMART design includes two stages: an initial stage of randomization to one of two or more *first-stage* treatments, and a period of follow-up. At some point in time—or over a period of time—specialists assess the response to initial treatment and patient characteristics to subsequently re-randomize individuals for *second-stage* treatment. Depending on the intermediate outcome status, one may or may not be re-randomized to a treatment option. Upfront consent of sequential randomizations is recommended so that individuals are randomized to subsequent treatment once they are eligible. This allows for data usage until randomization for balance between the treatment assignments of responders and nonresponders. Conceptually, there is no difference between upfront or sequential randomization, and both can be handled accordingly through analysis; however, sequential randomization may allow for more balanced randomization and the identification of other potential tailoring variables at each step.

References

[1] Box, G.E.P., Hunter, W.G., & Hunter, J.S. (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building*. New York, NY: Wiley.

[2] Chakraborty, B. (2009). *A study of non-regularity in dynamic treatment regimes and some design considerations for multicomponent interventions* [Ph.D. thesis, Department of Statistics, University of Michigan]. University of Michigan Library.

[3] Collins, L.M., Chakraborty, B., Murphy, S.A., & Strecher, V.J. (2009). Comparison of a phased experimental approach and a single randomized clinical trial for developing multicomponent behavioral interventions. *Clin. Trials*, *6*(1), 5-15.

[4] Collins, L.M., Murphy, S.A., Nair, V.N., & Strecher, V.J. (2005). A strategy for optimizing and evaluating behavioral interventions. *Ann. Behav. Med.*, *30*(1), 65-73.

[5] Lavori, P.W., & Dawson, R. (2000). A design for testing clinical strategies: Biased adaptive within-subject randomization. *J. Roy. Stat. Soc. A*, *163*(1), 29-38.

[6] Lavori, P.W., & Dawson, R. (2004). Dynamic treatment regimes: Practical design considerations. *Clin. Trials*, *1*(1), 9-20.

[7] Murphy, S.A. (2005). An experimental design for the development of adaptive treatment strategies. *Stat. Med.*, *24*(10), 1455-1481.

[8] Thall, P.F., Millikan, R.E., & Sung, H.G. (2000). Evaluating multiple treatment courses in clinical trials. *Stat. Med.*, *19*(8), 1011-1128.

[9] Thall, P.F., Sung, H.G., & Estey, E.H. (2002). Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *J. Am. Stat. Assoc.*, *97*(457), 29-39.

[10] Thall, P.F., Wooten, L.H., Logothetis, C.J., Millikan, R.E., & Tannir, N.M. (2007). Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring. *Stat. Med.*, *26*(26), 4687-4702.

Enjoy this passage? Visit the SIAM Bookstore[2] *to learn more about Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine and browse other available SIAM titles.*

Michael R. Kosorok is a W.R. Kenan, Jr. Distinguished Professor of Biostatistics and a professor of statistics and operations research at the University of North Carolina at Chapel Hill. He received the 2019 Gottfried E. Noether Senior Scholar Award from the American Statistical Association (ASA) and is a Fellow of the ASA, American Association for the Advancement of Science, and Institute of Mathematical Statistics. Erica E.M. Moodie is a professor of biostatistics and a Canada Research Chair (Tier 1) in Statistical Methods for Precision Medicine at McGill University. She is the 2020 recipient of the CRM-SSC Prize in Statistics and an elected member of the International Statistical Institute. Moodie also holds a chercheur de merite career award from the Fonds de recherche du Québec – Santé.

[1] https://my.siam.org/Store/Product/viewproduct/?ProductId=27004511

[2] https://my.siam.org/Store

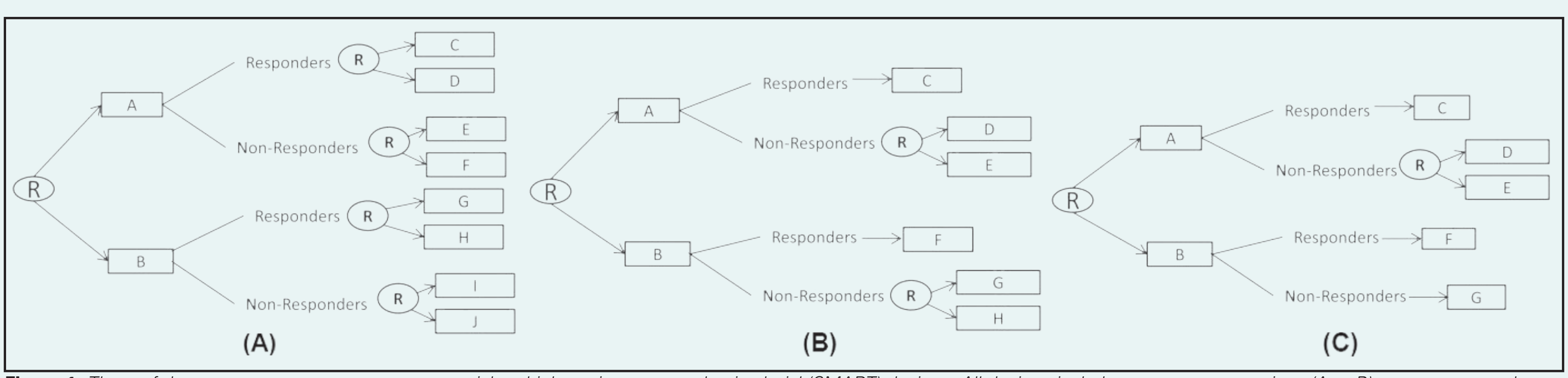


Figure 1. *Three of the most common two-stage sequential multiple assignment randomized trial (SMART) designs. All designs include two treatment options (A or B) at stage one and up to two treatment options depending on the intermediate outcome "responder" status.* **1a.** *A SMART where both responders and non-responders are re-randomized to treatment that depends on responder status.* **1b.** *A SMART where only non-responders are re-randomized.* **1c.** *A SMART where re-randomization depends on both responder status and initial treatment. Figure adapted from Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine.*

Data Science

Continued from page 8

cells under different conditions that mimic diseases. They use fluorescent microscopy to compare images of healthy cells, diseased cells, and diseased cells after the application of various drugs. "One primary data science and ML task is to develop ways to extract the relevant information from images in a way that makes sense when compared with all other possible representations of all other kinds of cells and diseases," Earnshaw said. After doing so, the team can answer questions about the cells' appearances and assemble a statistical method for the drugs and concentrations that might potentially treat certain diseases.

Despite the popularity of data science and ML, the speakers agreed that most companies still utilize conventional procedures in some capacity. Leslie noted that researchers frequently employ closed-form traditional approaches for comparison purposes or big-picture analyses. "ML is definitely getting a lot more of our attention, but it depends on the industry," she said. "People are still looking to graph theory and game theory, and many other approaches are still used."

Nevertheless, data science and ML's influence on modern-day society is substantial. Researchers now work on a number of problems that were not feasible before novel procedures like the data-driven and imperial-style approaches that accompany successful applications of ML. Biology is a prime example of a field that has profited immensely from data science because its complexity prevents closed-form solutions. "A lot of the techniques that I learned as an applied mathematician would not even work here," Earnshaw said, since many traditional applied mathematics tools cannot effectively handle the data's size.

Applied mathematicians, computational scientists, and other members of the STEM community will continue to benefit from the career opportunities that are associated with the growth of data science and related fields. Given the ever-increasing amounts of data in nearly every aspect of society, researchers who embrace data science practices and keep up with the state of the art will remain highly qualified for a multitude of industry positions.

Lina Sorg is the managing editor of SIAM News.

Minimizing the Length — and the Algebra

MATHEMATICAL CURIOSITIES
By Mark Levi

With the semester in full swing and some *SIAM News* readers likely teaching calculus, I thought of an amusing solution to the standard problem of minimizing the length of a fence that encloses three sides of a rectangular pasture of fixed area (the remaining side is bordered by a wall).

The Analog Computer

Figure 1 shows a mechanical "analog computer" from which one can simply read off a solution. The telescoping rectangle encloses incompressible two-dimensional fluid; the area is thus fixed. Sliders A and D can move without friction along the "wall" MN. All telescoping connections are frictionless, and a slippery rubber band is wrapped around the frame and affixed to sliders at A and D. Since the rubber band is slippery, *its tension T is the same throughout its length.*

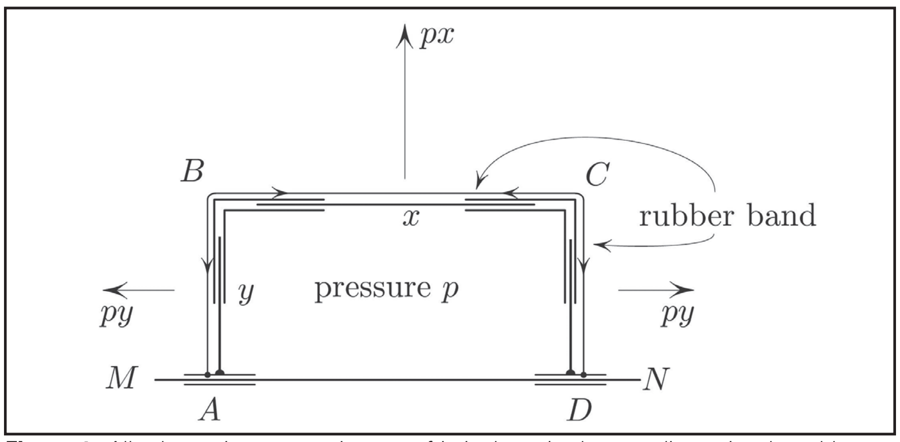


Figure 1. *All telescoping connections are frictionless. In the two-dimensional world, pressure force* $= p \cdot \text{length}$.

The Solution

The band tries to minimize the length, and the minimal length shape is in equilibrium. But then the vertical forces acting on the top side of the rectangle are in balance (see Figure 2):

$$px = 2T. \qquad (1)$$

Similarly, the left-right forces on a vertical side, say CD, are in balance as well:

$$py = T. \qquad (2)$$

Thus, $x = 2y$ and the problem is solved; the ratio of the optimal rectangle's sides is $1:2$.

Discussion

It is interesting to compare this admittedly nonrigorous solution to the textbook treatment in first-semester calculus. In our "solution," I replaced algebraic manipulation and differentiation with the work of inventing a mechanical device. I only meant this as a perhaps amusing complement to the "standard approach," rather than as a replacement. Incidentally, *the vanishing of the perimeter's derivative (with respect to one of the lengths) is equivalent to stating that the tension of the rubber band is the same along all three walls.*

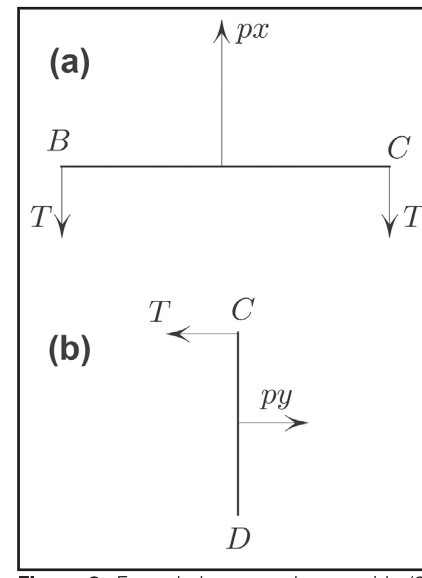


Figure 2. *Force balance on the top side (2a) and on the right side (2b). The band is slippery; its tension T is therefore the same throughout. Horizontal force at D vanishes because the slider is frictionless.*

The figures in this article were provided by the author.

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

Graduate Training

Continued from page 7

component. Students' individualized needs should guide the relative emphasis of major program components so that they can spend most of their time on activities that will best encourage their professional development. As an intermediate step, programs could therefore develop several distinct "tracks," each with a structure that is tailored toward a particular career trajectory (see Figure 1, on page 7). The forthcoming breakdown offers examples of each track.

Academia: This track is identical to the current paradigm that trains students for traditional careers in academia; few changes are required.

Teaching and Education: This track is intended for students who are interested in careers that emphasize education, such as faculty positions at liberal arts and community colleges, educational research, or the development of novel teaching methodologies in the private sector. Institutions should adjust the research component in this track to account for a reduced emphasis on original mathematical research in students' future careers. This path should also include appropriate internship and training opportunities in professional environments, e.g., teaching opportunities at local community colleges.

Non-academic Careers: This track prioritizes hands-on experience in the type of interdisciplinary work that is integral to non-academic careers. A portion of the courses in this track fall outside of the mathematics department to ensure that students gain adequate background in the application of mathematical methods. Students should also participate in semester- or year-long internships; online sites such as SIAM's Career Resources page[1] and the BIG Math Network[2] can provide guidance.

Academic Research: Mathematics Ph.D.s can now pursue research careers in academic settings outside of mathematics departments. These settings include a variety of departments and institutes that heavily rely on grants and contracts. For instance, biomedical and computational biology research provides many opportunities in data science and mathematical/computational modeling within medical schools and biomedical and biological research institutes. Though M.D./Ph.D. students seldom pursue graduate training in mathematics, more might choose to do so if programs were better able to adapt to the unique requirements of a four-year Ph.D. timeline.

Challenges

Students who take part in personalized programs must immediately start designing curricula that are right for them, making periodic adjustments as necessary. Intensive mentoring is crucial and should extend far beyond the efforts of current mathematics Ph.D. programs. Such programs also need to inform students of different career opportunities early in their training, and faculty mentors should know about the broader universe of available careers and resources. Student passage from one track to another must be flexible, and mentoring efforts should prepare students to transition to the next phase in their careers.

In a follow-up article to appear in the next issue of SIAM News, the authors will address some of the questions and challenges that pertain to the implementation of the program customizations in their proposal.

If readers have thoughts, questions, or suggestions about the aforementioned proposal, we encourage them to comment on the online version of this article or contact the authors directly at yara.skaf@ufl.edu and reinhard.laubenbacher@medicine.ufl.edu.

Acknowledgments: We are most grateful to the many members of the mathematics community who provided extensive feedback on an earlier version of this manuscript to greatly improve the final text.

References

[1] Langin, K. (2019, March 12). *In a first, U.S. private sector employs nearly as many Ph.D.s as schools do*. Science Careers. Retrieved from https://www.sciencemag.org/careers/2019/03/first-us-private-sector-employs-nearly-many-phds-schools-do.

[2] National Academies of Sciences, Engineering, and Medicine. (2018). *Graduate STEM education for the 21st century*. Washington, DC: The National Academies Press.

[3] National Center for Science and Engineering Statistics. (2019). *Survey of doctorate recipients*. National Science Foundation. Retrieved from https://nsf.gov/statistics/srvydoctoratework/#tools&tabs-3&infdsts&rSR&rWP&sd.

[4] Rieley, M. (June 2018). Big data adds up to opportunities in math careers. *Beyond the Numbers: Employ. & Unemploy.*, *7*(8).

[5] Skaf, Y., & Laubenbacher, R. (2021). Student-centric graduate training in mathematics: A commentary. Preprint, *arXiv:2109.07661*.

[6] Terry, M. (2019, March 20). *Report: When it comes to hiring PhDs, private sector and academia are about even*. BioSpace. Retrieved from https://www.biospace.com/article/report-when-it-comes-to-hiring-phds-private-sector-and-academia-are-about-even.

[7] U.S. News & World Report. (2018). *Best mathematics programs*. Retrieved from https://www.usnews.com/best-graduate-schools/top-science-schools/mathematics-rankings.

Yara Skaf is an M.D./Ph.D. student in the Department of Mathematics and the Laboratory for Systems Medicine in the Department of Medicine's Division of Pulmonary, Critical Care, and Sleep Medicine at the University of Florida. She is developing tools from topological data analysis and applying them to the analysis of electronic health records. Reinhard Laubenbacher is director of the Laboratory for Systems Medicine and a professor in the Department of Medicine's Division of Pulmonary, Critical Care, and Sleep Medicine at the University of Florida.

[1] https://www.siam.org/careers/resources
[2] https://bigmathnetwork.org/overview

Preserving the History of Applied Mathematics

By John P. Boyd

"To turn events into ideas is the function of history." – George Santayana

Applied mathematics has a history, but who will tell it? *SIAM News* periodically publishes obituaries of its esteemed members, but the honorees' accomplishments are communicated only through the brief summaries and humorous anecdotes of their memorialists. Mathematical and scientific concepts evolve and generate new ideas, but events turn into ideas as well. The mix of elaborations and calculations that embody the quirks and contingencies of individuals, culture, wars, fads, and politics makes it arguably impossible to fully understand the present and future without also having some grasp of the past.

If history is idealized as a smooth, continuous flow from plain to baroque, then vector calculus (universally applicable with three components) should have preceded quaternions (few applications, complicated theory, and four components) by half a century. Lord Kelvin noted that "Quaternions came from Hamilton after his really good work had been done, and though beautifully ingenious, have been an unmixed evil to those who have touched them in any way." But the graffiti that William Rowan Hamilton scratched on Broom Bridge on October 16, 1843 was the non-commutative multiplication law of quaternions. Vector calculus was shaped mostly by Josiah Willard Gibbs and Oliver Heaviside four decades later [1].

In the mindset that probability implies inevitability, Gibbs should have perished like 750,000 others of his generation in the American Civil War. But he had chronic lung disease and therefore never enlisted. He should have contracted tuberculosis, which killed his mother, and died by age 25; instead he lived into his 60s — just long enough to teach Edwin Bidwell Wilson. Since Gibbs refused all entreaties to write a book on vectors, his lecture notes of 1883 should have remained the secret doctrine of his small cohort of students. But Wilson, who was only 22 years old, filled the void with *Vector Analysis: A Text-book for the Use of Students of Mathematics and Physics, Founded Upon the Lectures of J. Willard Gibbs*; this was the first book on modern vector theory. Gibbs died unexpectedly of an intestinal blockage less than four years after Wilson took his class at Yale.

One view of mathematics is that it (metaphorically) evolves monotonically from archaea to the 12-primary-color, polarization-resolving eyes of the mantis shrimp. This viewpoint asserts that Jule Charney's theory of baroclinic instability in the ocean and atmosphere, which was solved by confluent hypergeometric functions, should have appeared at least five years *after* the simpler so-called "f-plane" theory whose solutions are merely hyperbolic functions. But history is indifferent to the prognostications of the learned; illogically, it was not until five years after Charney that Eric Eady took the Great Leap Backwards by omitting the so-called "beta effect," thereby showing that it was *significant* but *not essential* for baroclinic instability [3].

Meteorology was converted from an empirical science—a sort of botany-of-clouds—to a branch of physics and fluid mechanics around 1900. This happened in large part because Vilhelm Bjerknes put filial piety above his dreams of a brilliant physics career in Germany and returned to Stockholm to help his ailing father finish his life's work: a massive treatise read by none but its authors. His colleague Nils Ekholm, a pioneer aerodynamicist and survivor of the unsuccessful 1896 balloon expedition to the North Pole, encouraged Bjerknes to delve into weather.[1] A new branch of physics was born in an academic backwater from a man who was battered by years of depression and well into middle age.

The most unlikely of Bjerknes' many allies was neurobiologist Fridtjof Nansen. Though Nansen was unlearned in both mathematics and physical oceanography, he was a tireless and careful observer. Under the supervision of Bjerknes, Vagn Walfrid Ekman turned Nansen's puzzling but unchallengeable measurements of the Arctic Sea into explanations. And thus Nansen—far better at skiing and persuasion than algebra—was father to the mathematics behind the Ekman current spiral, Ekman boundary layer, Ekman pumping and Ekman suction, and the theory of "dead water." Tiny, remote Scandinavia dominated atmospheric physics for several decades mostly because Bjerknes happened to be Norwegian [2].

What might Évariste Galois have done if he had not been killed in a duel at age 20? Or Bernhard Riemann, who died of tuberculosis at 40? Contingency blasts holes again and again in the smooth evolution of mathematics. Unfortunately, misunderstanding the roles of contingency and personality is not the only cost of poor or nonexistent mathematical history.

Some organizations have begun making deliberate commitments to preserving the history and biography of women and underrepresented applied mathematicians. But even so, much is lost: Sylvia Skan of the Falker-Skan equation; Anne Nicolson of the Crank-Nicolson algorithm; Grace Vaisey, who earned Sir Richard Southwell his knighthood; Lorna Swain, who became a hydromechanics lecturer at the University of Cambridge in 1926; and Susan Martin, who rose from calculator girl to senior scientific officer over the course of 57 years at the National Physical Laboratory. Does anyone know the names of these pioneers?

Cicely Ridley was a pathfinder in numerical quantum chemistry and earned a doctorate in 1956 under Douglas Hartree. At age 30, she suspended her career to raise four children. Ridley returned to work as a grunt-level programmer/consultant at the National Center for Atmospheric Research (NCAR),[2] where she rapidly rose to NCAR's highest non-administrative rank of senior scientist. She was well known in atmospheric science as a co-creator of the Roble-Dickinson-Ridley computer code: the first general circulation model of the thermosphere. Cicely is a great role model, but she can no longer tell her story.

Many years ago, I was the summer research advisor for an African American undergraduate named Rudy Horne. Decades later as a tenured associate professor at Morehouse College, Horne was invited back to the University of Michigan to deliver a much-anticipated keynote lecture on Martin Luther King Jr. Day in 2018. He had spent four months as the mathematics consultant for the movie *Hidden Figures*, during which he educated the cast about the joy and practical uses of mathematics. In the film, Katherine Johnson's solution of a quadratic equation is in his handwriting. Horne could have told amazing stories about his experiences, but he died suddenly at age 49 before delivering the lecture.

History and biography are in a constant race against time to preserve documents, oral histories, and artifacts before death and landfills render them forever inaccessible. In this race to preserve and protect, many societies are competing by standing still. SIAM has superb editorial, production, and marketing staffs, as I know from personal experience. However, it does not have a book series for individual and collective biographies, collected papers of distinguished mathematicians, or applied mathematics history. Organizations like SIAM should establish committees to look after the history and biography of applied mathematics. Such committees would then appoint editors and encourage submissions in the aforementioned areas.

See **History** *on page 12*

[1] Bjerknes' first and only physics student, Nils Strindberg, was expected to someday metaphorically hang his own shield in the Hall of Weather Heroes, but Ekholm thought that ballooning needed a younger man and persuaded Strindberg to take his place as a weather observer on the 1897 expedition to the North Pole, which was lost with no survivors.

[2] I am grateful for Ridley's help with ordinary differential equation solvers for my thesis when she was still a programmer.

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****.*

Argonne National Laboratory

J.H. Wilkinson Postdoctoral Fellowship in Scientific Computing

The Mathematics and Computer Science Division of Argonne National Laboratory invites outstanding candidates to apply for the 2022 J.H. Wilkinson Postdoctoral Fellowship in Scientific Computing. The Wilkinson Fellowship is intended to encourage early-career scientists who are actively engaged in state-of-the-art research in scientific computing. The Fellowship provides an exceptional opportunity for innovative research in computational mathematics and scientific computing on advanced computing architectures.

Areas of interest span applied mathematics, scientific computing, and computational sciences and include algorithmic differentiation, computational fluid dynamics, data assimilation, data science, discretization, linear algebra, machine learning, meshing, modeling and simulation, solvers, multifidelity/multilevel/multiphysics/multiscale algorithms, quantum computing, operations research, optimization, statistics, stochastic systems, and uncertainty quantification. Additional areas of interest in applied mathematics, numerical software, and statistics can be found at **https://www.anl.gov/mcs/lans**.

We value and strive for diversity in backgrounds, experiences, and perspectives.

The benefits of the appointment are highly competitive. The appointment is for one year and may be renewed for another year. For full consideration, applications and letters should be submitted by December 1, 2021. For more details and to apply, see **https://www.anl.gov/mcs/wilkinson.**

Possible Proof of Goldbach's Conjecture

I realize that most readers of this periodical work in the field of applied mathematics. But Goldbach's Conjecture is so tantalizingly simple to state—yet so difficult to prove—that I believe a possible proof will be of some interest to mathematicians in a variety of fields.

The Conjecture states that each even positive integer ≥ 4 is the sum of two primes. Thus, e.g., $4=2+2$, $6=3+3$, $8=5+3$, $10=5+5$, $12=7+5$, and so forth. The Conjecture was first stated in 1742 and still has not been proved.

But I believe that I might have a proof, and in fact a very short one (two pages). It is the result of a strategy that, as far as I know, is original with me. It is as follows: Find the largest structure that contains all possibilities and shows important relationships between them. In the case of my possible proof, that structure is an infinite sequence of matrices — one for each even positive integer 2k, where each matrix contains all pairs of odd, positive integers that sum to 2k.

Interested readers should visit "A Possible Proof of Goldbach's Conjecture" on **occampress.com**.

— *Peter Schorer,* peteschorer@gmail.com

Addressing Climate Change, Boosting Environmental Resilience, and Advancing Clean Energy

Reflections of the SIAM Climate Task Force

By Alejandro Aceves, Hans Kaper, and Sven Leyffer

How can applied mathematics and computational science boost environmental resilience, drive clean energy innovation, and advance society's understanding of climate change and its impact on humans? These questions were at the heart of a series of meetings of the SIAM Climate Task Force, which was established by SIAM's Committee on Science Policy (CSP).[1] The goal of the task force was to create a report with guidelines for Congress and U.S. federal agencies that details the ways in which applied mathematics and computational science can positively impact climate science and mitigate the effects of climate change. The final report,[2] which the task force submitted to the CSP in August 2021, contains findings and recommendations in the following four broad categories.

Climate and Earth System Prediction

This topic includes the following subsets:

- Modeling, simulation, and optimization techniques
- Improved computational capabilities
- Innovation that is rooted in basic research, such as the development of statistical methods and computational algorithms for monumental volumes of diverse environmental data
- Support for research in areas like dynamical systems, optimization, and stochastic modeling to comprehend levels of possible prediction, characterize uncertainty, and improve forecasting capabilities.

The objective is to create robust models that help scientists understand climate change's impact on all forms of human activity, including agriculture and cities (see Figure 1). This research is inherently multidisciplinary and calls for interdisciplinary collaborations between the fields of mathematics, computational science, atmospheric and oceanic sciences, geosciences, social sciences, biology, and agricultural sciences.

Resilient Communities and Environmental Justice

Applied mathematics and computational science provide researchers with the tools to explore and evaluate scenarios to ultimately protect communities and critical infrastructure from extreme events, such as sustained periods of drought, shifting precipitation patterns, and rising ocean levels. Equally important is the use of these tools to implement and inform more equitable decision-making techniques that address existing economic and racial disparities and ensure that the impacts of climate change do not disproportionately affect communities that are already marginalized.

Algorithm and model development, data analysis, and computational simulation enable evidence-based choices despite the future's uncertainty. Applied mathematics and computational science can help society better understand the many consequences of climate change, including various community-level impacts, socioeconomic effects, and regional and global conflicts and migration patterns. To fully appreciate these complex interactions, one must establish collaborative research connections among mathematicians and social scientists, reach out to community stakeholders, and build networks with agencies that have not traditionally supported mathematical research, like the Federal Emergency Management Agency and the Environmental Protection Agency.

Clean Energy Innovation

Efforts to accelerate energy innovation to reduce emissions—which are of tremendous interest across the federal government—must incorporate contributions from applied mathematics and computational science. Intersections with energy production and use fall in two different contexts. The first is the reliable, resilient, and economic planning and operation of the future grid, including the integration of renewables. This effort will decarbonize our power delivery system and make it stronger, more flexible, and more secure. The second intersection pertains to the key role of applied mathematics and computational science in the development of new technologies, such as the optimization of materials for batteries and hydrogen storage, the design and placement of offshore wind, and the transformation of transportation. We can leverage the Exascale Computing Project[3] and build new partnerships through the Scientific Discovery through Advanced Computing program[4] to drive the innovation of clean energy.

Workforce Development

A prerequisite of the previous three recommendations is the existence of a mathematically-trained, multidisciplinary workforce. In order to address the complex challenges of climate change, instructors need to prepare applied mathematics and computational science students to work effectively on diverse, multidisciplinary, and convergent teams that include social scientists. The climate workforce of the future must possess skills in robust mathematics; computational science; statistics; and other science, technology, engineering, and mathematics (STEM) areas to innovate and address climate issues. It is therefore important for society to continue and grow federal investments that support mathematics, computational science, and interdisciplinary STEM education.

The SIAM Climate Task Force's report contains recommendations for the current administration, Congress, and federal research agencies like the National Science Foundation, Department of Energy, Department of Agriculture, National Oceanic and Atmospheric Administration, Department of Transportation, National Institutes of Health, Department of Defense, and other institutions that address climate and energy research. A recurring recommendation in the report is the establishment of cross-agency coordination to address the numerous challenges that are associated with climate change.

This report comes at a critical time, as it coincides with increased coverage of climate change in the press and the release of the Sixth Assessment Report[5] of the Intergovernmental Panel on Climate Change (IPCC)[6] [1]. Many of the task force's recommendations align with the IPCC report.

Acknowledgments: The SIAM Climate Task Force was assisted by Eliana Perlmutter of Lewis-Burke Associates LLC, SIAM's liaison in Washington, D.C.

References

[1] Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., …, Zhou, B. (Eds.). (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, U.K.: Cambridge University Press. In press.

Alejandro Aceves, former chair of the SIAM Activity Group on Nonlinear Waves and Coherent Structures, is a professor and former chair of the Department of Mathematics at Southern Methodist University. Hans Kaper, founding chair of the SIAM Activity Group on Mathematics of Planet Earth and editor-in-chief of SIAM News, is affiliate faculty in the Department of Mathematics and Statistics at Georgetown University. Sven Leyffer is a senior computational mathematician in the Mathematics and Computer Science Division at Argonne National Laboratory. He works on nonlinear optimization and currently serves as secretary of the International Council for Industrial and Applied Mathematics. All three authors were members of the SIAM Climate Task Force, with Leyffer acting as chair.

[1] https://www.siam.org/about-siam/committees/committee-on-science-policy-csp
[2] https://www.siam.org/Portals/0/Publications/Reports/SIAM_Climate_Task_Force_Report_with_Appendix.pdf?ver=2021-08-12-091101-927
[3] https://www.exascaleproject.org
[4] https://www.scidac.gov
[5] https://www.ipcc.ch/assessment-report/ar6
[6] https://www.ipcc.ch

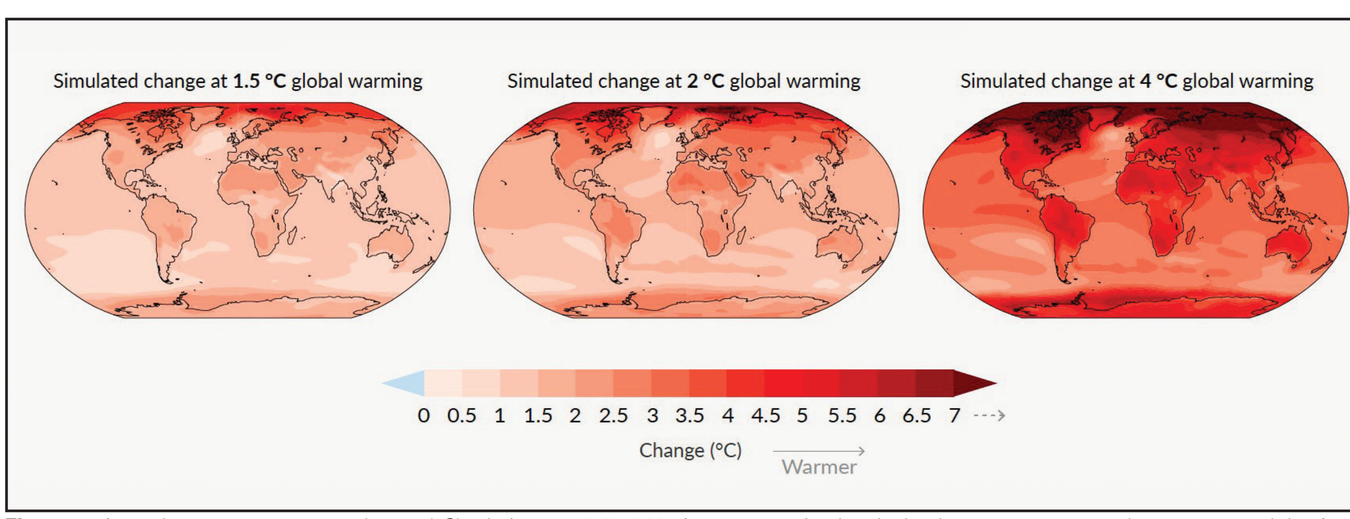


Figure 1. *Annual mean temperature change (°C) relative to 1850-1900. Across warming levels, land areas warm more than oceans and the Arctic and Antarctica warm more than the tropics. Figure courtesy of [1].*

History

Continued from page 11

Although a robust publishing program in history and biography would take some time to develop, it is certainly doable. SIAM already has a collection of downloadable oral histories, but only in numerical analysis. Prospective historical committees could solicit and encourage the archiving of oral histories, autobiographies, and other materials in all areas of applied mathematics. Formal and informal partnerships with university archives, research libraries, museums, and science history programs would also be beneficial.

Artifacts are as perishable as their makers. The aircraft carrier CV-6—the most highly decorated ship of World War II—would have made a wonderful floating museum. However, the carrier that survived innumerable bombs and a kamikaze hit was scrapped in 1956, defeated by a far more powerful adversary than the Kidō Butai: historical indifference. With this in mind, societies should aim to publicize relevant museum holdings. Cornell University's College of Engineering has a marvelous collection of pre-electronic calculating machines, and the Science Museum in London contains a differential analyzer (analog computer) that was built from a child's Meccano/Erector set. The museum also displays Lord Kelvin's 1876 ocean tide forecaster; in lieu of today's silicon integrated circuits, Kelvin's was a flock of brass gears and spheres that was used operationally for half a century. Furthermore, an example of one of the most powerful species of brass-and-electric-motor computers—the U.S. Torpedo Data Computer Mark IV—is still operational on the U.S. Navy submarine *Pampanito*, which is now a museum ship in San Francisco Bay. Perhaps mathematics organizations can serve as information pipelines between museums and prospective donors who want their mathematical artifacts to be displayed rather than hidden in a drawer by curators who are baffled by ellipsoids.

The recycling of warship names is a reminder that the past is deeply embedded in the future. The name of the CV-6 was born by six U.S. Navy ships before her and two more (to date) after her, a lineage that will presumably extend into the distant future. The true name of the ship that once stood alone against an empire—the last operational carrier in the Pacific Ocean—is now known in every part of the world thanks to a television voiceover on *Star Trek* that begins, "These are the voyages of the starship *Enterprise* . . . her continuing mission . . . to boldly go where no one has gone before." In the name of historical preservation, where will SIAM and other applied mathematics organizations boldly go?

References

[1] Crowe, M.J. (1967). *A history of vector analysis*. South Bend, IN: University of Notre Dame Press.
[2] Friedman, R.M. (1989). *Appropriating the weather: Vilhelm Bjerknes and the construction of a modern meteorology*. Ithaca, NY: Cornell University Press.
[3] Vallis, G.K. (2017). *Atmospheric and ocean fluid dynamics: Fundamentals and large-scale circulation*. New York, NY: Cambridge University Press.

John P. Boyd is a professor in the Department of Climate and Space Sciences and Engineering at the University of Michigan.