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

siam news

Volume 54/ Issue 2
March 2021

Imaging Science Special Issue

*In this **special issue**, explore current research in the field of imaging science and various applications of mathematical imaging.*

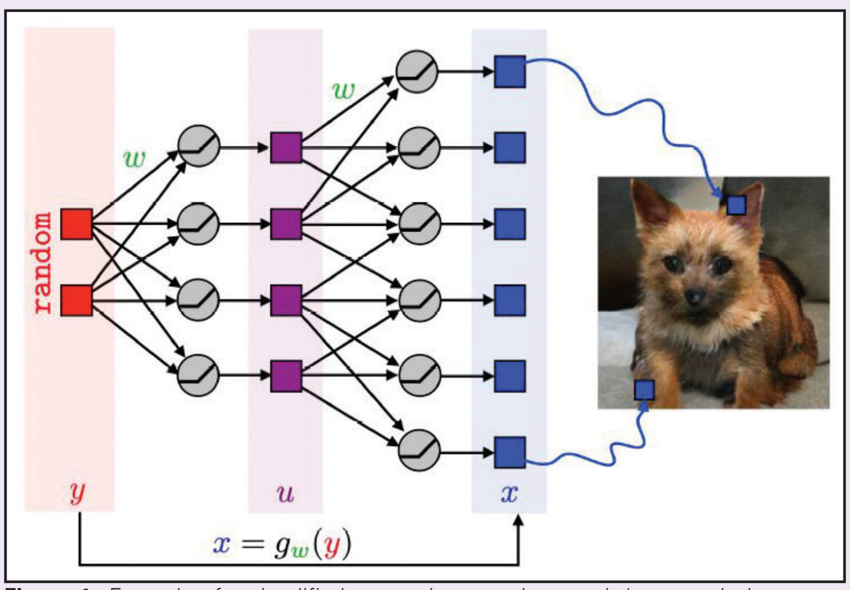


Figure 1. *Example of a simplified generative neural network (a network that generates such complex images actually has more layers). Image courtesy of Gabriel Peyré, code courtesy of [1].*

The deep learning revolution began with the resolution of supervised classification problems. It currently faces new challenges for the unsupervised generation of text, images, and videos. In an article on page 7, Gabriel Peyré explains how researchers can use the theory of optimal transport to formulate and solve this class of problems.

A Denoiser Can Do Much More than Just Clean Noise

Regularization by Denoising

By Yaniv Romano, Michael Elad, and Peyman Milanfar

Nearly all image processing tasks require access to some "approximate" notion of the images' probability density function. This problem is generally intractable, especially due to the high dimensions that are involved. Rather than directly approximating this distribution, the image processing community has consequently built algorithms that—either explicitly or implicitly—incorporate key features of the unknown distribution of natural images. In particular, researchers have proposed very efficient denoising algorithms (i.e., algorithms that remove noise from images, which is the simplest inverse problem) and embedded valuable characteristics of natural images in them. The driving question is thus as follows: How can we systematically leverage these algorithms and deploy their implicit information about the distribution in more general tasks?

Consider a noisy image observation $\mathbf{y} = \mathbf{x} + \mathbf{v}$, where $\mathbf{x}$ is an unknown image that is corrupted by zero-mean white Gaussian noise $\mathbf{v}$ of a known standard deviation σ. We use f to denote an image denoising function — a mapping from $\mathbf{y}$ to an image of the same size $\hat{\mathbf{x}} = f(\mathbf{y})$, such that the resulting estimate will be as close as possible to the unknown $\mathbf{x}$. This innocent-looking problem has attracted much attention over the past 50 years and sparked innovative ideas across different fields, including robust statistics, harmonic analysis, sparse representations, nonlocal modeling, and deep learning. Indeed, denoising engines are now at the core of the image processing pipeline in any smartphone device, surveillance system, and medical imaging machine.

The recent development of sophisticated and well-performing denoising algorithms has led researchers to believe that current methods have reached the ceiling in terms of noise reduction performance. This belief comes from the observation that substantially different algorithms lead to nearly the same denoising performance;

See **Denoiser** *on page 2*

Plug-and-Play: A General Approach for the Fusion of Sensor and Machine Learning Models

By Charles A. Bouman, Gregery T. Buzzard, and Brendt Wohlberg

Regularized or Bayesian inversion has revolutionized our ability to reconstruct images from incomplete data. For example, suppose that we want to reconstruct an image x from a vector of sensor measurements y, given by

$$y = Ax + w,$$

where A is a linear forward model and w is additive white Gaussian noise with variance σ^2. The regularized reconstruction then comes from

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2} \|y - Ax\|^2 + h(x) \right\},$$

where $h(x)$ is a term that encourages a "regular" solution.

But how should we choose the regularizing function $h(x)$? If we select $h(x) = -\log p(x)$, where $p(x)$ is an assumed prior distribution, $\hat{x}$ then becomes the Bayesian maximum a posteriori (MAP) reconstruction. Other reasonable choices for $h(x)$ include the total variation or Markov random field cost functions. However, the simplistic nature of these analytical priors—which do not always accurately represent the true distribution of real image collections—often limits the quality of the resulting MAP reconstructions.

Over the last decade, image denoisers such as block-matching and 3D filtering—and more recently, convolutional neural network denoisers—have demonstrated that dramatic improvements in denoising performance are possible with the use of increasingly complex image operations. These advanced denoising algorithms effectively model the distribution of real images but do not utilize any explicit cost function $h(x)$. This raises the following question: How can we fuse the traditional models of regularized inversion with the implicit models of modern denoising algorithms?

Plug-and-play (PnP) methods answer this question by providing a framework for fusing traditional sensor models with black-box models. These black-box models can range from advanced denoising algorithms that are used as priors to more general "agents" that are typically trained via machine learning methods, like deep neural networks.

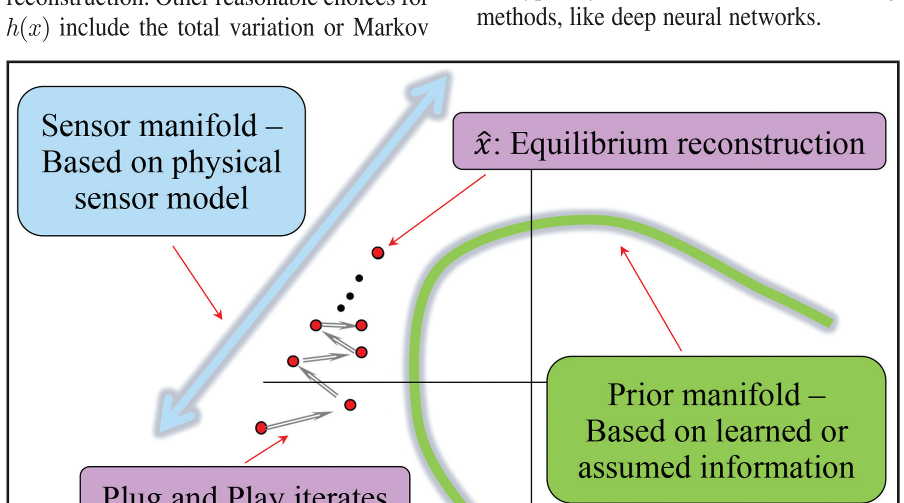


Figure 1. *The plug-and-play (PnP) solution balances the goals of fitting sensor data and finding a plausible answer to the problem. The alternating application of a forward model and "plug-in" denoiser result in a sequence that converges to a reconstruction equilibrium. Figure courtesy of the authors.*

Model Fusion with Plug-and-Play

We can express the MAP reconstruction in the simpler and more general form of

$$\hat{x} = \underset{x}{\operatorname{argmin}} \{ f(x) + h(x) \}, \qquad (1)$$

where $f(x) = \frac{1}{2\sigma^2}\|y - Ax\|^2$ is the sensor term and $h(x)$ is again the regularizing prior model term.

Figure 1 graphically illustrates this equation with a sensor manifold that corresponds to small values of $f(x)$ and a prior manifold that corresponds to small values of $h(x)$. The MAP reconstruction is thus at a location that minimizes the distance to both manifolds, making it maximally consistent with the data and prior.

The important special case of image denoising occurs when $A = I$. In this case, the observations y consist of x plus additive Gaussian white noise and the MAP reconstruction is given by $\hat{x} = H(y)$, where

$$H(y) = \underset{x}{\operatorname{argmin}} \left\{ \frac{1}{2\sigma^2} \|y - x\|^2 + h(x) \right\}.$$

The key insight of PnP is that the denoiser $H(y)$ is also the proximal map of $h(y)$. That is, H is an operator that takes a step to reduce h while maintaining proximity to the input point.

Interestingly, we can use the well-known alternating direction method of multipliers (ADMM) algorithm to solve our MAP optimization problem by alternately applying H along with a second forward model proximal map that comes from

$$F(v) = \underset{x}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2\sigma^2} \|x - v\|^2 \right\}.$$

The ADMM algorithm for solving (1) is then given by the following iteration:

See **Plug-and-Play** *on page 4*

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

siam news
Volume 54/ Issue 2/ March 2021

Denoiser

Continued from page 1

it has been corroborated by theoretical studies that aimed to derive denoising performance bounds. These insights led researchers to conclude that improving image denoising algorithms may be a task with diminishing returns, or to put it more bluntly: a dead end.

Surprisingly, a consequence of this realization is the emergence of a new and exciting area of research: the leveraging of denoising engines to solve other, far more challenging inverse problems. Examples of such problems include image deblurring, super-resolution imaging, inpainting, demosaicing, and tomographic reconstruction. The basis for achieving this goal resides in the formulation of an inverse problem as a general optimization task that seeks to solve

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} l(\mathbf{y},\mathbf{x}) + \lambda R(\mathbf{x}). \qquad (1)$$

The term $l(\mathbf{y},\mathbf{x})$ is called the likelihood and represents $\mathbf{x}$'s faithfulness to measurement $\mathbf{y}$. For example, $l(\mathbf{y},\mathbf{x}) = \|\mathbf{x}-\mathbf{y}\|_2^2$ in the case of image denoising and $l(\mathbf{y},\mathbf{x}) = \|H\mathbf{x}-\mathbf{y}\|_2^2$ in the case of image deblurring, for which we assume that $\mathbf{y} = H\mathbf{x} + \mathbf{v}$ with a linear blurring operator H. The term $R(\mathbf{x})$ represents the prior, or regularizer, that aims to drive the optimization task towards a unique or stable solution; one typically cannot achieve such a solution via the likelihood term alone. The hyperparameter λ controls the regularization strength.

For the denoising problem, the choice of $\lambda = 0$ in (1) leads to a trivial solution for which $\hat{\mathbf{x}} = \mathbf{y}$. This solution reveals the crucial role of $R(\mathbf{x})$; loosely speaking, an ideal prior should penalize the appearance of noise in $\hat{\mathbf{x}}$ while preserving edges, textures, and other internal structures in the unknown $\mathbf{x}$. This intuition has motivated the formulation of important image denoising priors, such as Laplacian smoothness, total variation, wavelet sparsity, enforcement of nonlocal self-similarity, Gaussian mixture models, Field of Experts models, and sparse approximation.

How can we leverage a given powerful denoising machine $f(\mathbf{x})$ to handle other image processing problems? The Plug-and-Play priors (PPP) framework[1] is an innovative, systematic approach for treating a wide class of inverse problems via denoising engines [2]. PPP's key novelty is the observation that one can use denoising algorithms as "black box" solvers, which in turn define general image priors. The framework achieves this by introducing an auxiliary image $\mathbf{z}$ to (1) that decouples the denoising task from the likelihood term:[2]

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x},\mathbf{z}} l(\mathbf{y},\mathbf{x}) + R(\mathbf{z}) + \frac{1}{2\mu}\|\mathbf{x}-\mathbf{z}\|_2^2. \qquad (2)$$

We can minimize the above objective with alternating optimization techniques. For example, consider a deblurring problem with $l(\mathbf{y},\mathbf{x}) = \|H\mathbf{x}-\mathbf{y}\|_2^2$. When treating $\mathbf{z}$ as fixed, the minimization of (2) with respect to $\mathbf{x}$ involves solving a simple linear system of equations — a sharpening step. When optimizing (2) with respect to $\mathbf{z}$ while $\mathbf{x}$ is fixed, we obtain a denoising problem that treats the sharpened image $\mathbf{x}$ as the noisy input. We can interpret the hyperparameter μ as the noise level in the candidate estimate $\mathbf{x}$.

Inspired by the PPP rationale, the framework of Regularization by Denoising (RED) [1] takes a different route and defines an *explicit* regularizer $R(\mathbf{x})$ of the form

$$R(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T(\mathbf{x} - f(\mathbf{x})).$$

Put simply, the value of the above penalty function is low if the cross-correlation between the candidate image $\mathbf{x}$ and its denoising residual $\mathbf{x} - f(\mathbf{x})$ is small, or if the residual itself is small. RED brings a modern interpretation of the classic Laplacian regularizer $R(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T(\mathbf{x} - W\mathbf{x})$, for which W is a fixed and predefined smoothing operator, like a Gaussian filter. In striking contrast to the classic Laplacian prior, RED replaces the naïve filter W with a state-of-the-art image adaptive denoising filter that is defined by a black box function f.

What are the mathematical properties of the RED prior? Can we hope to compute its derivative? Recall that scientists often formulate state-of-the-art denoising functions as optimization problems; therefore, computing the derivative of f will likely be highly nontrivial. Surprisingly, research has shown that RED's penalty term is differentiable and convex under testable conditions, and its gradient is simply the residual $\mathbf{x} - f(\mathbf{x})$ [1]. As a result, for a convex likelihood $l(\mathbf{y},\mathbf{x})$—as in the deblurring example—the optimization problem

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} l(\mathbf{y},\mathbf{x}) + \lambda\mathbf{x}^T(\mathbf{x} - f(\mathbf{x}))$$

is convex as well, thus guaranteeing global convergence to the optimum. One can flexibly treat this task with a wide variety of first-order optimization procedures, as the gradient is simple to obtain and necessitates only a single activation of the denoiser. In its formal form, RED requires the chosen denoiser to meet some strict conditions, including local homogeneity, differentiability, and Jacobian symmetry. From an empirical standpoint, however, RED-based recovery algorithms seem to be highly stable and capable of incorporating any denoising algorithm as a regularizer—from the simplest median filtering to state-of-the-art deep learning methods—and treating general inverse problems very effectively.

The PPP and RED frameworks pose new and exciting research questions. The gap between theory and practice has inspired the development of a series of new variations for RED's prior, as well as novel numerical algorithms. Provable convergence guarantees further support these new methods, broadening the family of denoising machines that one can use to solve general inverse problems. Another exciting line of research seeks a rigorous connection between RED and PPP, with the hope that such an understanding will lead to improved regularization schemes and optimizers. In terms of machine learning aspects, RED solvers formulate novel deep learning architectures by replacing the traditional nonlinear activation functions—like rectified linear units or sigmoid functions—with well-performing denoising algorithms. This approach offers new ways for researchers to train data-driven solvers for the RED functional, with the hope of ultimately achieving superior recovery in fewer iterations than the analytic approach.

This article is based on Yaniv Romano's SIAM Activity Group on Imaging Science Early Career Prize Lecture at the 2020 SIAM Conference on Imaging Science,[3] which took place virtually last year. Romano's presentation is available on SIAM's YouTube Channel.[4]

References

[1] Romano, Y., Elad, M., & Milanfar P. (2017). The little engine that could: regularization by denoising (RED). *SIAM J. Imaging Sci.*, *10*(4), 1804-1844.

[2] Venkatakrishnan, S.V., Bouman, C.A., & Wohlberg, B. (2013). Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing* (pp. 945-948). Austin, TX: IEEE.

Yaniv Romano is an assistant professor in the Departments of Electrical Engineering and Computer Science at the Technion – Israel Institute of Technology. He was a postdoctoral researcher in the Department of Statistics at Stanford University and completed his Ph.D. in electrical engineering at the Technion. Romano's research focuses on selective inference, predictive inference, machine learning, and computational imaging. Michael Elad is a professor in the Computer Science Department at the Technion – Israel Institute of Technology. His research focuses on inverse problems, sparse representations, and machine learning in the wider context of signal and image processing applications. Elad has served as editor-in-chief of the SIAM Journal on Imaging Sciences since 2016 and was elected as a SIAM Fellow in 2018. Peyman Milanfar is a principal scientist/director at Google Research, where he leads the Computational Imaging team. He was a professor of electrical engineering at the University of California, Santa Cruz from 1999-2014 and Associate Dean for Research at the Baskin School of Engineering from 2010-2012. He is a Distinguished Lecturer of the IEEE Signal Processing Society and a Fellow of the IEEE.

[1] See page 1 for an article by Charles Bouman, Gregery Buzzard, and Brendt Wohlberg about Plug-and-Play.

[2] Here we present a simplified version of the original PPP objective by replacing the hard constraint $\mathbf{x} = \mathbf{z}$ with a penalty; the original PPP relied on augmented Lagrange and the alternating direction method of multipliers.

[3] https://www.siam.org/conferences/cm/conference/is20

[4] https://www.youtube.com/watch?v=jtu-kKDM1Yo

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.

Printed in the USA.

Obituary: Stanley C. Eisenstat

By Howard Elman and Avi Silberschatz

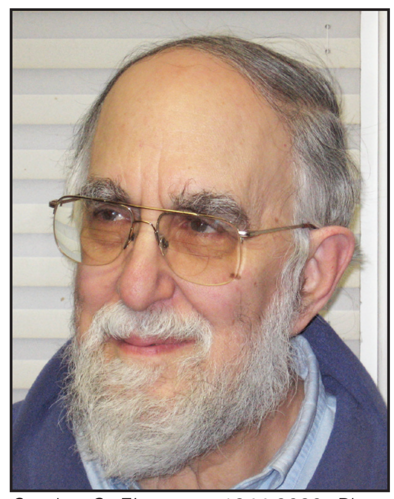
Stanley C. Eisenstat, an internationally renowned computational mathematician and computer scientist, passed away from a pulmonary embolism in New Haven, Conn., on December 17, 2020. He was 76 years old.

Stan was born in 1944 in New York City. He received his bachelor's degree in mathematics from the Case Institute of Technology (which later merged with Western Reserve University to form the present-day Case Western Reserve University). While at Case, Stan took his first graduate course in numerical analysis with his longtime colleague and collaborator, Martin Schultz. He completed his master's degree and Ph.D. at Stanford University under the direction of John Herriot and Cleve Moler.

In 1971, Stan joined the faculty of the Department of Computer Science at Yale University. He remained at Yale for nearly 50 years. Stan's employment had a somewhat auspicious start, as he began working at Yale before submitting his Ph.D. thesis and immediately found his new research activities more compelling than the mechanics of turning in his dissertation. Fortunately, the Yale administration forced him to complete his degree to remain on the faculty.

During his long and distinguished research career, Stan made fundamental contributions to algorithms in numerical analysis — with an emphasis on numerical linear and nonlinear algorithms. In virtually every case, he was responsible for developing new ways of thinking to construct algorithms. For example, Stan was heavily involved in the development of inexact Newton methods for nonlinear algebraic systems, iterative methods and preconditioning methods for linear systems of equations, algorithms and mathematical software for sparse direct methods, and fast and robust algorithms for eigenvalue and singular value decompositions.

Stan was both a strong theoretician and a world-class software implementor. He served as a key contributor to the Yale Sparse Matrix Package, a widely used software package for the solution of linear sparse systems of equations (MATLAB later adopted its sparse matrix technology). Stan also developed the so-called "Eisenstat trick," which enables the implementation of preconditioners based on incomplete factorization with essentially no overhead cost. His work was known for its brevity and pointedness — nearly all of his important publications are less than 20 pages long. A few examples include a seminal paper with Ron Dembo and Trond Steihaug on inexact Newton methods (nine pages) [1], a study with Ming Gu on rank-one updates of eigenvalue problems that led to the development of stable divide-and-conquer algorithms for computing the singular value decomposition (11 pages) [4], and several influential papers on iterative methods that are only four pages in length [2, 3].


Stanley C. Eisenstat, 1944-2020. Photo courtesy of Dana Angluin.

Throughout his career, Stan served as a mentor and *éminence grise* for multiple generations of young scientists—including students, postdoctoral researchers, and junior faculty—who passed through Yale under the aegis of the Yale Research Center for Scientific Computing (which Stan co-directed with Martin Schultz). People left meetings with Stan feeling that his comments had shaped some of their most important research contributions, and the majority of his mentees enjoyed distinguished careers. Stan was thus uniformly respected for the clarity and insight of his advice.

Stan played a pivotal role in the growth of Yale's Department of Computer Science. He was devoted to undergraduate education in the department and served as the director of undergraduate studies for many years. Stan designed and taught fundamental courses entitled "Data Structures and Programming Techniques" and "Introduction to Systems Programming and Computer Organization," and he was fiercely dedicated to both the program and its students.

Stan joined SIAM as a graduate student in 1967 and remained a lifelong member. He became a SIAM Fellow in 2018 and was cited "for development and analysis of fast computational algorithms for linear and nonlinear systems of equations." In 1997, he received the SIAM Activity Group on Linear Algebra Best Paper Prize. Stan also served on the editorial boards of three important journals for many years: the *SIAM Journal on Scientific Computing*, *SIAM Journal on Matrix Analysis and Applications* (*SIMAX*), and *Journal of the Association for Computing Machinery*. At the time of his passing, he was completing his fourth term as associate editor on the *SIMAX* Editorial Board. Daniel Szyld, the former editor-in-chief of *SIMAX*, praised Stan's work for the journal. "On several occasions he improved the proof of a theorem or reframed the results to make them more general or applicable to a larger class of problems," Szyld said. "In more than one case, the authors insisted on listing Stan as a co-author in a revision that included his new results or insights."

Stan is survived by his wife Dana Angluin, his son David, and his daughter Sarah. He will be deeply missed by everyone who knew him.

References

[1] Dembo, R.S., Eisenstat, S.C., & Steihaug, T. (1982). Inexact Newton methods. *SIAM J. Numer. Anal., 19*(2), 400-408.

[2] Eisenstat, S.C. (1981). Efficient implementation of a class of preconditioned conjugate methods. *SIAM J. Sci. Stat. Comput., 2*(1),1-4.

[3] Eisenstat, S.C. (1983). A note on the generalized conjugate gradient method. *SIAM J. Numer. Anal., 20*(2), 358-361.

[4] Gu, M., & Eisenstat, S.C. (1994). A stable algorithm for the rank-one modification of the symmetric eigenproblem. *SIAM J. Matrix Anal. Appl., 15*(4), 1266-1276.

Howard Elman is a Professor of Computer Science and director of the Applied Mathematics Program at the University of Maryland, College Park. He earned his doctorate at Yale University under the direction of Stan Eisenstat and Martin Schultz. Avi Silberschatz is the Sidney J. Weinberg Professor of Computer Science at Yale University. Prior to coming to Yale in 2003, he served as vice president of the Information Sciences Research Center at Bell Labs. He previously held an endowed professorship at the University of Texas at Austin, where he taught until 1993.

Deep Learning in Scientific Computing: Understanding the Instability Mystery

By Vegard Antun, Nina M. Gottschling, Anders C. Hansen, and Ben Adcock

Deep learning (DL) is causing profound changes in society. It has inspired unprecedented advances in historically challenging problems, such as image classification and speech recognition. And now, perhaps inevitably, it is markedly affecting scientific computing.

Yet DL has an Achilles' heel. Current implementations can be highly *unstable*, meaning that a certain small perturbation to the input of a trained neural network can cause substantial change in its output. This phenomenon is both a nuisance and a major concern for the safety and robustness of DL-based systems in critical applications—like healthcare—where reliable computations are essential. It also raises several questions. Why does this instability occur? Can it be prevented? And what does it mean for scientific computing, a field in which *accuracy* and *stability* are paramount? Here we consider these questions in the context of inverse problems, an area of scientific computing where DL has shown significant promise.

Instabilities in Image Classification

The story of instabilities begins with image classification. Researchers first observed these instabilities in 2013 upon the introduction of an algorithm that *fooled* a trained neural network classifier [10]. Given a fixed input image x with label p, the algorithm computes a small perturbation r, such that the image $x+r$—while indistinguishable from x to the human eye—is misclassified with label $q \neq p$. Figure 1 depicts several examples of this effect. Though the perturbations are barely visible, each one prompts the classifier to fail in a dramatic way.

The study of *adversarial perturbations* (or *adversarial attacks*) on classification problems has since become an active subfield of machine learning research [11]. Scientists have constructed real-world adversarial perturbations in applications that range from image classification and speech recognition to surveillance, self-driving vehicles, and automated diagnosis.

Figure 1. *Adversarial perturbations in image classification. Perturbed images $x+r$ and the labels produced by the classifier are shown here. The network correctly classifies the unperturbed images. Figure reproduced from [7].*

Deep Learning for Inverse Problems

Although quite different from classification problems, inverse problems—specifically inverse problems in imaging—comprise an area in which DL methods have made particularly rapid progress. Numerous studies have reported superior DL performance over current state-of-the-art techniques in various image reconstruction tasks, including medical imaging modalities like magnetic resonance imaging (MRI) and X-ray computed tomography [1, 5, 8, 12]. Such optimism is perhaps best exemplified by the following quote from *Nature Methods* [9], which reports on recent work [12]: *"AI transforms image reconstruction. A deep-learning-based approach improves the speed, accuracy, and robustness of biomedical image reconstruction."*

The simplest type of inverse problem—but one that is often sufficient in practice—is the discrete linear problem:

$$\text{Given measurements } y = Ax + e \in \mathbb{C}^m \text{ of } x \in \mathbb{C}^N, \text{ recover } x. \tag{1}$$

Here, $x \in \mathbb{C}^N$ is the (vectorized) unknown image, $A \in \mathbb{C}^{m \times N}$ represents the measurement process, and $e \in \mathbb{C}^m$ is the noise. Because of physical constraints, this problem is often highly undersampled in practice—the number of measurements m is generally much smaller than the image size N—and therefore challenging. Typical DL approaches seek to overcome this issue by learning a neural network $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ that produces accurate reconstructions $\Psi(Ax+e) \approx x$ for relevant image classes. This process is facilitated by a set of *training data*

$$\mathcal{T} = \{(x^j, y^j) : j = 1, \ldots, K\},$$

which consists of typical images x^j (e.g., MRI scans of different brains) and their measurements $y^j = Ax^j + e^j$.

Researchers have proposed multiple different DL approaches to solve (1). However, growing evidence indicates that many of these approaches are also *unstable*. Figures 2 and 3 (on page 5) provide examples of this effect. In both cases, a small perturbation causes a significant degradation in the reconstruction's quality. While Figure 2 is based on a worst-case perturbation (similar to the case of classification problems), Figure 3 indicates that purely random perturbations can sometimes elicit substantial effects. In contrast, state-of-the-art (untrained) sparse regularization methods [1] are typically far less susceptible to perturbations.

The Universal Instability Theorem

While DL approaches perform very well on some image reconstruction tasks, many methods appear to do so at the price of instability. The *universal instability theorem* sheds light on this issue [4]. Let $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ be a continuous reconstruction map for (1), and suppose that there are two vectors $x, x' \in \mathbb{C}^N$ for which

$$\|x - x'\| > 2\eta \quad (x \text{ and } x' \text{ are far apart}), \tag{2}$$

$$\|Ax - Ax'\| \leq \eta \quad (\text{the measurements of } x \text{ and } x' \text{ are similar}), \tag{3}$$

and

$$\|\Psi(Ax) - x\| + \|\Psi(Ax') - x'\| < 2\eta \quad (\Psi \text{ recovers } x \text{ and } x' \text{ well}) \tag{4}$$

See **Instability Mystery** *on page 5*

Principles and Trends in Mathematical Imaging

By Carola-Bibiane Schönlieb, Hongkai Zhao, Gabriele Steidl, and Michael B. Wakin

The fascinating and emerging field of imaging science is situated at the interface of physics, electrical engineering, computer science, and mathematics. Its broad applications reach from photography to biomedical, seismic, and astronomical imaging. Mathematical imaging pertains to the development and analysis of mathematical models and methods that process imaging data in efficient, explainable ways. Effectively coping with the application at hand requires tools from diverse fields of mathematics that often interact in interesting manners, conversely influencing the mathematical theory in these fields. Harmonic analysis, partial differential equations (PDEs) and related variational methods, stochastics, and differential geometry are just some of the fields that are relevant to imaging science.

In the early 19th century, mathematician Joseph Fourier realized that one can represent every periodic function as a superposition of sines and cosines. This basic idea became a cornerstone of signal and image processing. As a result, the principle of analyzing and modifying functions by approximating them as linear combinations of appropriate elements from a dictionary was generalized in harmonic analysis and approximation theory. Scientists have thus begun widely utilizing wavelets and their sophisticated directional counterparts, such as curvelets and shearlets, in the multiresolution analysis of images. Indeed, the language of wavelets originated from several dialects, including square integrable group representations, bandpass filters, and windowed Fourier transforms.

Quite recently, researchers have successfully implemented nonlinear eigenfunction systems—based on variational methods like total variation regularization—for imaging tasks. This process has revealed even more challenging questions in nonlinear operator theory and computation. Furthermore, appropriate sparsity assumptions on images have made sub-Nyquist imaging systems possible, ultimately resulting in the emergence of compressive imaging systems, sparse dictionary learning techniques, and super-resolution algorithms.

Experts have long used PDEs in image restoration, particularly for edge- and coherence-enhancing nonlinear diffusion of grayscale values. The development of operator splitting methods in large-scale optimization in the 1960s essentially advanced the evolution to related variational image reconstruction methods. In the 1980s, researchers applied operator splitting methods to solve monotone inclusion equations in convex analysis; nearly 20 years later, they successfully adopted these approaches in both imaging and machine learning.

Because real data is always noisy and often results from a random measurement process—as with photonic imaging—the field of statistics is another important area of mathematical imaging. Statistics helps formalize the solutions of many image processing and analysis tasks as estimation problems for which one can provide confidence bounds and quantify uncertainty. The use of Bayesian models is therefore significant in image restoration applications.

Researchers have also incorporated nonlocal image models, graph-based models, and mixture models into imaging techniques (see Figure 1). Optimal transport-based techniques are highly interesting for certain imaging tasks, including barycenter computation. These procedures again originated from an old mathematical problem—Gaspard Monge's 18th-century effort to transport mass at minimal costs and Leonid Kantorovich's relaxed formulation in the 1930s—and now have an established measure-theoretic foundation upon which imaging scientists can build. Novel developments are related to multimarginal optimal transport and generalized Schrödinger bridges.

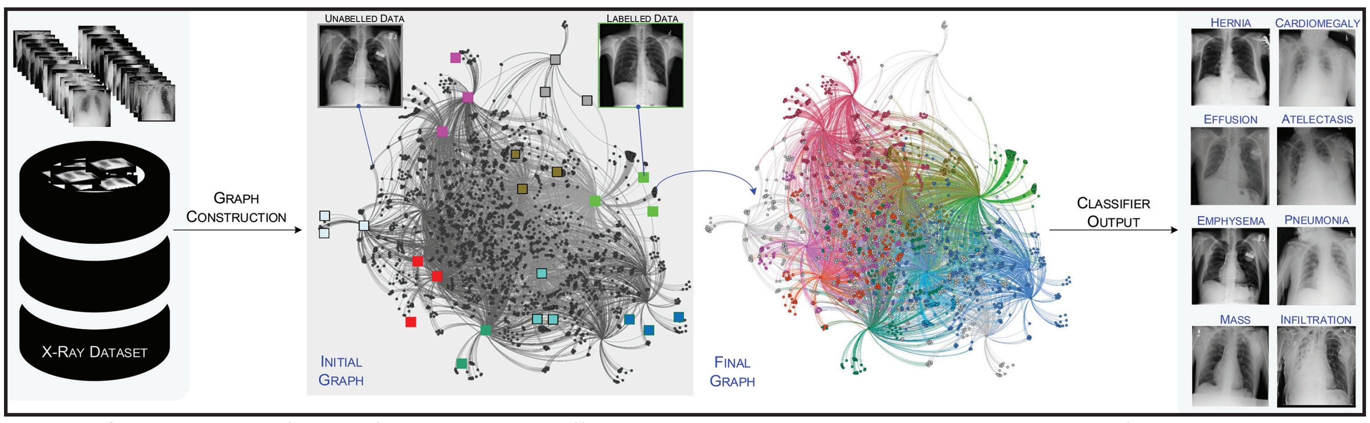


Figure 1. *Graph-based classification of chest X-rays into different pathologies. Based on only a very small number of labeled X-rays per class, researchers aim to classify the rest. To do so, they map the X-rays into a high-dimensional feature space where known labels are propagated to the unlabeled X-rays by the graph Laplacian. Figure courtesy of [1].*

See **Mathematical Imaging** *on page 6*

Plug-and-Play

Continued from page 1

Repeat{

$$x \leftarrow F(v-u) \qquad (2)$$

$$v \leftarrow H(x+u) \qquad (3)$$

$$u \leftarrow u+x-v \qquad (4)$$

}.

We obtain the PnP algorithm by simply replacing the original proximal map $H(y)$ with a novel black-box operator and running the new algorithm. Therefore, "plugging in" a black-box or learned denoiser $H(x)$ yields a new algorithm with the same outer loop but a fresh interpretation.

Again, Figure 1 (on page 1) illustrates the intuition behind PnP. Alternating applications of the plug-in operator $H(x)$ and the forward model proximal map $F(x)$ move the solution between the sensor and data manifolds in a zig-zag sequence that converges to a fixed point under appropriate hypotheses. When $H(x)$ is a black-box denoiser, this fixed point no longer minimizes a cost function; however, one can view it as reaching an equilibrium.

In this sense, PnP is a meta-algorithm — it takes existing algorithms for function minimization and converts them into new algorithms that use more general input-output maps. The basic idea of PnP [3, 4] has been applied to a wide variety of problems in several application domains with excellent results.

Multi-Agent Consensus Equilibrium

A shortcoming of PnP is that it is a solution without a problem. The original ADMM algorithm was designed to minimize the MAP cost function; but after replacing some components with black-box operators, there is no longer any cost function to minimize.

To address this issue, we introduce equilibrium methods that determine a system of equations to solve rather than a function to minimize. The basic form of consensus equilibrium (CE) stems from the converged solutions of the updates in (2)-(4). When converged, it must be true that $x^*=v^*$. This substitution yields CE equations that define the problem that the PnP algorithm solves [1]:

$$x^* = F(x^* - u^*)$$

$$x^* = H(x^* + u^*).$$

Since H is a denoiser and x is the reconstructed image, we can interpret u in this context as noise that is removed in the operation $x^*=H(x^*+u^*)$.

When $H(x)$ is a general black-box operator and not a proximal map, this system of equations no longer determines a cost function's minimum. However, the CE equations do determine a well-defined equilibrium condition. So we see that the goal of PnP methods is not to solve the optimization problems of traditional regularized inversion. Instead, they aim to solve more general and flexible sets of equilibrium equations.

Multi-agent consensus equilibrium (MACE) generalizes PnP to the case of more than two agents. It defines a stacked operator of agents $\mathbf{F}$, along with a consensus operator $\mathbf{G}$ that computes the average of its inputs. These operators are given by

$$\mathbf{F}(\mathbf{w}) = \left[F_1(w_1), \ldots, F_K(w_K)\right]^T$$

and

$$\mathbf{G}(\mathbf{w}) = \left[\frac{1}{K}\sum_k w_k, \ldots, \frac{1}{K}\sum_k w_k\right]^T,$$

where each agent $F_k(w_k)$ is intuitively designed to move the solution closer to some desired goal. The MACE equations then take the simple form of

$$\mathbf{F}(\mathbf{w}^*) = \mathbf{G}(\mathbf{w}^*).$$

Figure 2 presents an overview of MACE's role by separating the ideas into four categories: criterion versus algorithm and cost functions versus agents. MACE completes the matrix by providing criteria for the formulation of problems that are based on agent equilibrium rather than simply on cost function minimization.

In summary, PnP is a framework that incorporates modern black-box operators into regularized inversion problems. And MACE delivers a problem criterion in the form of equilibrium equations that the PnP algorithm solves. The aforementioned PnP and MACE methods are just the first steps in a range of new techniques that fuse traditional models with emerging machine learning and algorithmic models. Code that illustrates these methods is available in [2].

This article is based on Charles A. Bouman's SIAM Activity Group on Imaging Science Best Paper Prize Lecture at the 2020 SIAM Conference on Imaging Science,[1] which took place virtually last year. Bouman's presentation is available on SIAM's YouTube Channel.[2]

Acknowledgments: Charles A. Bouman and Gregery T. Buzzard were partially supported by NSF CCF-1763896. Brendt Wohlberg was supported by Los Alamos National Laboratory's Laboratory Directed Research & Development Program under project number 20200061DR.

References

[1] Buzzard, G.T., Chan, S.H., Sreehari, S., & Bouman, C.A. (2018). Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium. *SIAM J. Imag. Sci., 11*(3), 2001-2020.

[2] Buzzard, G.T., Wohlberg, B., & Bouman, C.A. (2021). *PnP and MACE reference implementation*. Software library available at https://github.com/gbuzzard/PnP-MACE.

[3] Sreehari, S., Venkatakrishnan, S.V., Wohlberg, B., Buzzard, G.T., Drummy, L.F., Simmons, J.P., & Bouman, C.A. (2016). Plug-and-play priors for bright field electron tomography and sparse interpolation. *IEEE Trans. Comput. Imag., 2*(4), 408-423.

[4] Venkatakrishnan, S.V., Bouman, C.A., & Wohlberg, B. (2013). Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing* (pp. 945-948). Austin, TX: IEEE.

1 https://www.siam.org/conferences/cm/conference/is20
2 https://www.youtube.com/watch?v=P94MFLW18fU

Charles A. Bouman is the Showalter Professor of Electrical and Computer Engineering and Biomedical Engineering at Purdue University and a member of the National Academy of Inventors. His research inspired the first commercial model-based iterative reconstruction system for medical X-ray computed tomography. Gregery T. Buzzard is a professor of mathematics at Purdue University, where he served as department head from 2013-2020. His research focuses on analysis and algorithms for adaptive measurement and reconstruction. Brendt Wohlberg is a senior research scientist in the Theoretical Division at Los Alamos National Laboratory. His research centers on foundations and algorithms for imaging inverse problems.

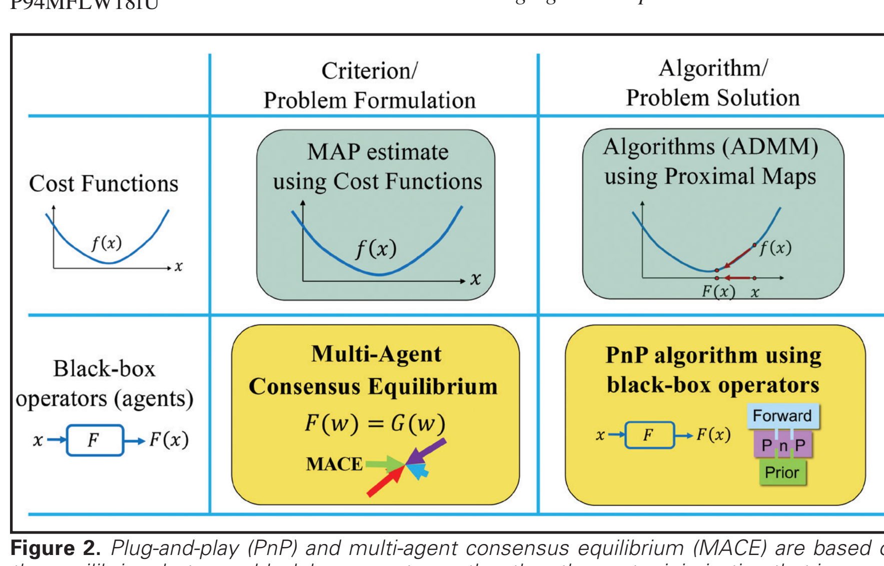


Figure 2. *Plug-and-play (PnP) and multi-agent consensus equilibrium (MACE) are based on the equilibrium between black-box operators rather than the cost minimization that is associated with traditional regularized inversion methods. MACE provides the criterion that the PnP algorithm solves. Figure courtesy of the authors.*

Instability Mystery

Continued from page 3

for some $\eta > 0$. The theorem then states the following:

(a) Instability. There is a closed, non-empty ball $\mathcal{B}_y \subset \mathbb{C}^m$ centered at $y = Ax$, such that the local ε-Lipschitz constant at any $\tilde{y} \in \mathcal{B}_y$ satisfies

$$L^{\varepsilon}(\Psi, \tilde{y}) := \sup_{0<\|z-\tilde{y}\|\le\varepsilon} \frac{\|\Phi(z)-\Phi(\tilde{y})\|}{\|z-\tilde{y}\|} \ge \frac{1}{\eta}\Big(\|x-x'\| - 2\eta\Big), \qquad \forall \varepsilon \ge \eta.$$

Because the Lipschitz constant measures the effect of perturbations, this result states that any map that *overperforms*—i.e., accurately recovers two vectors x and x' (4) even though their measurements are similar (3)—must also be unstable. This implies that a delicate tradeoff exists between accuracy and stability, with the quest for too much accuracy (i.e., attempting to extract more from the data than is reasonable) leading to poor stability.

The prior result helps explain why DL can become unstable. Simply put, *DL approaches often have no mechanisms for protecting against overperformance*. Recall that a typical training goal is to obtain a small *training error*, i.e., $\Psi(y^j) \approx x^j$ for $j = 1, \ldots, K$. However, if the training set contains two elements x, x' with $\|x - x'\| \gg 2\eta$ and $\|Ax - Ax'\| \le \eta$, successful training will necessarily cause instabilities. As the training set is often large and A often has a large null space (e.g., when $m \ll N$), this situation can arise in many potential ways.

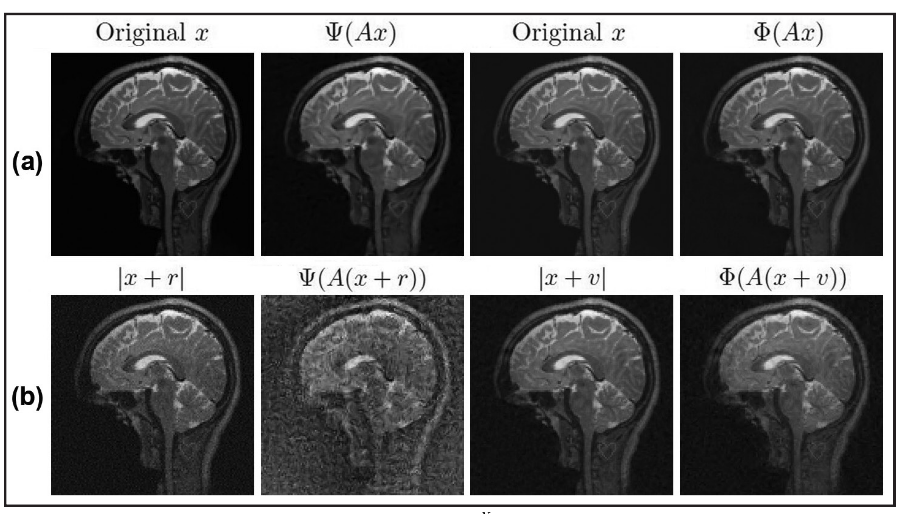


Figure 2. *The AUTOMAP network $\Psi : \mathbb{C}^m \to \mathbb{C}^N$ [12] for MRI is unstable to worst-case perturbations, whereas the sparse regularization method Φ is not.* **2a.** *Original image x and unperturbed reconstructions.* **2b.** *Worst-case perturbations r and v for the respective methods and their effects. Poor conditioning is not responsible for this instability since the condition number $\mathrm{cond}(AA^*) = 1$. Figure courtesy of [2].*

(b) False negatives. There is a $z \in \mathbb{C}^N$ with $\|z\| \ge \|x - x'\|$; an $e \in \mathbb{C}^m$ with $\|e\| \le \eta$; and closed, non-empty balls $\mathcal{B}_x, \mathcal{B}_e$ centered at x and e respectively, such that

$$\|\Psi(A(\tilde{x}+z)+\tilde{e}) - \tilde{x}\| \le \eta, \qquad \forall \tilde{x} \in \mathcal{B}_x,\ \tilde{e} \in \mathcal{B}_e. \tag{5}$$

False positives also arise in an analogous way. One can interpret this property by viewing x as a "healthy" brain image and z as a "tumor." It asserts that Ψ may falsely reconstruct a healthy brain x given measurements of an unhealthy brain $x + z$. It also implies that *instabilities are not rare events*. If e is a random vector (with mild assumptions on its distribution), then the fact that (5) occurs in a ball means that

$$\mathbb{P}(\|\Psi(A(x+z)+e) - x\| \le \eta) \ge c > 0$$

for some $c > 0$. Therefore, purely random perturbations can create false negatives (and positives) with nontrivial probability, as seen in Figure 3.

False Negatives and Threading the Accuracy-Stability Needle

What to do? It is of course elementary to create a stable network. The zero network would do the job but obviously produce many false negatives. The difficulty comes with simultaneously ensuring both stability *and* performance; Figure 4 highlights this issue. The network was trained on images that are comprised of ellipses and is quite stable in practice. Yet if a small detail that was not in the training set is inserted, the network washes it out almost entirely. The 2019 FastMRI[1] challenge has also reported similar effects on practical MRI datasets, with networks failing to reconstruct small but physically-relevant image abnormalities [3]. It is also worth noting that encouraging stability during training is not easy. Common methods like *adversarial training*, *random sampling patterns*, and *enforcing consistency* fail to protect against overperformance and thus remain susceptible to the universal instability theorem [4]. Overall, determining the best approach to walking the tightrope between accuracy and stability remains a significant open problem.

Limits of Deep Learning in Scientific Computing

The universal instability theorem is an example of a methodological boundary. Historically, scientific progress is often shaped by the presence or absence of such boundaries. Theoretical computer science, for example, developed with a thorough understanding of its limitations thanks to Gödel and Turing's fundamental work on non-computability. Numerical analysis has boundaries such as the Dahlquist and Butcher barriers in numerical ordinary differential equations, stability of Gaussian elimination, performance of the simplex method, and so forth.

Given the tradition for trial-and-error approaches in DL research—often accompanied by grand performance claims—such boundaries are more important now than ever. Neural networks are substantially more complex than the traditional tools of scientific computing. Critical assessment of new DL methods is needed, and further theoretical insights into accuracy-stability tradeoffs are essential for navigating the development of these new methods. To do so, we must ask a guiding question: What are the limits of DL in scientific computing?

Acknowledgments: Nina M. Gottschling acknowledges support from the U.K. Engineering and Physical Sciences Research Council grant EP/L016516/1. Anders C. Hansen thanks NVIDIA for a GPU grant in the form of a Titan X Pascal and acknowledges support from a Royal Society University Research Fellowship and the 2017 Philip Leverhulme Prize. Ben Adcock acknowledges support from the Pacific Institute for the Mathematical Sciences CRG in High-dimensional Data Analysis, Simon Fraser University's Big Data Initiative "Next Big Question" Fund, and the Natural Sciences and Engineering Research Council through grant R611675.

References

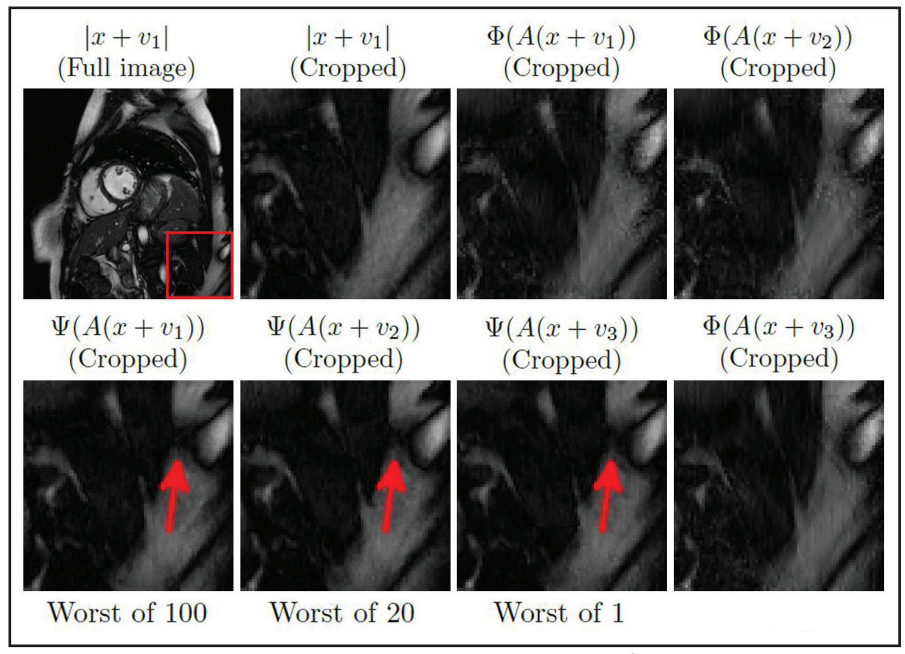


Figure 3. *The deep MRI network [8] is unstable with respect to Gaussian noise and produces false image regions. 100 Gaussian noise vectors $w_j \in \mathbb{C}^N$ were computed, then the eyeball metric was used to pick the one (subsequently labeled v_1) for which $\Psi(A(x+w_j))$ yields the largest artifact. This process was then repeated with 20 noise vectors, as well as one new noise vector, to give perturbations v_2 and v_3 respectively. The red arrows indicate that Ψ introduces a false dark area. Poor conditioning does not cause this instability since $\mathrm{cond}(AA^*) = 1$ as before. Conversely, the sparse regularization method Φ accurately recovers the image without the false region. Figure courtesy of [4].*

[1] Adcock, B., & Hansen, A.C. (2020). *Compressive Imaging: Structure, Sampling, Learning*. Cambridge University Press (in press).

[2] Antun, V., Renna, F., Poon, C., Adcock, B., & Hansen, A.C. (2020). On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc. Natl. Acad. Sci.*, *117*(48), 30088-30095.

[3] Cheng, K., Calivá, F., Shah, R., Han, M., Majumdar, S., & Pedoia, V. (2020). Addressing the false negative problem of deep learning MRI reconstruction models by adversarial attacks and robust training. In T. Arbel, I.B. Ayed, M. de Bruijne, M. Descoteaux, H. Lombaert, & C. Pal (Eds.), *Proceedings of the Third Conference on Medical Imaging with Deep Learning* (pp. 121-135). Montreal, Canada: PMLR.

[4] Gottschling, N.M., Antun, V., Adcock, B., & Hansen, A.C. (2020). The troublesome kernel: Why deep learning for inverse problems is typically unstable. Preprint, *arXiv:2001.01258*.

[5] Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., & Knoll, F. (2018). Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.*, *79*(6), 3055-3071.

[6] Jin, K.H., McCann, M.T., Froustey, E., & Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.*, *26*(9), 4509-4522.

[7] Moosavi-Dezfooli, S., Fawzi, A., Fawzi, O., & Frossard, P. (2017). Universal adversarial perturbations. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 86-94). Honolulu, HI: IEEE.

[8] Schlemper, J., Caballero, J., Hajnal, J.V., Price, A., & Rueckert, D. (2017). A deep cascade of convolutional neural networks for MR image reconstruction. In M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap, & D. Shen (Eds.), *Information Processing in Medical Imaging* (Vol. 10265). In *Lecture Notes in Computer Science* (pp. 647-658). Springer.

[9] Strack, R. (2018). Imaging: AI transforms image reconstruction. *Nature Methods*, *15*(5), 309.

[10] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I.J., & Fergus, R. (2013). Intriguing properties of neural networks. Preprint, *arXiv:1312.6199*.

[11] Wiyatno, R.R., Xu, A., Dia, O., & de Berker, A. (2019). Adversarial examples in modern machine learning: A review. Preprint, *arXiv:1911.05268*.

[12] Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., & Rosen, M.S. (2018). Image reconstruction by domain-transform manifold learning. *Nature*, *555*(7697), 487-492.

Vegard Antun is a postdoctoral fellow in applied mathematics at the University of Oslo. His research is centered on deep learning-based techniques for scientific computing, with a particular focus on inverse problems and imaging. Nina M. Gottschling is a third-year Ph.D. student in the Applied Functional and Harmonic Analysis group in the Department of Applied Mathematics and Theoretical Physics at the University of Cambridge. Her research interests include harmonic analysis, applied functional analysis, numerical analysis, optimisation theory, and mathematical physics. Anders C. Hansen is a reader in mathematics at the University of Cambridge and a professor of mathematics at the University of Oslo. His research interests include foundations of computational mathematics, applied functional and harmonic analysis, artificial intelligence, and data science. Ben Adcock is an associate professor of mathematics at Simon Fraser University. His research interests include numerical analysis, mathematics of data science, approximation theory, and applied and computational harmonic analysis.

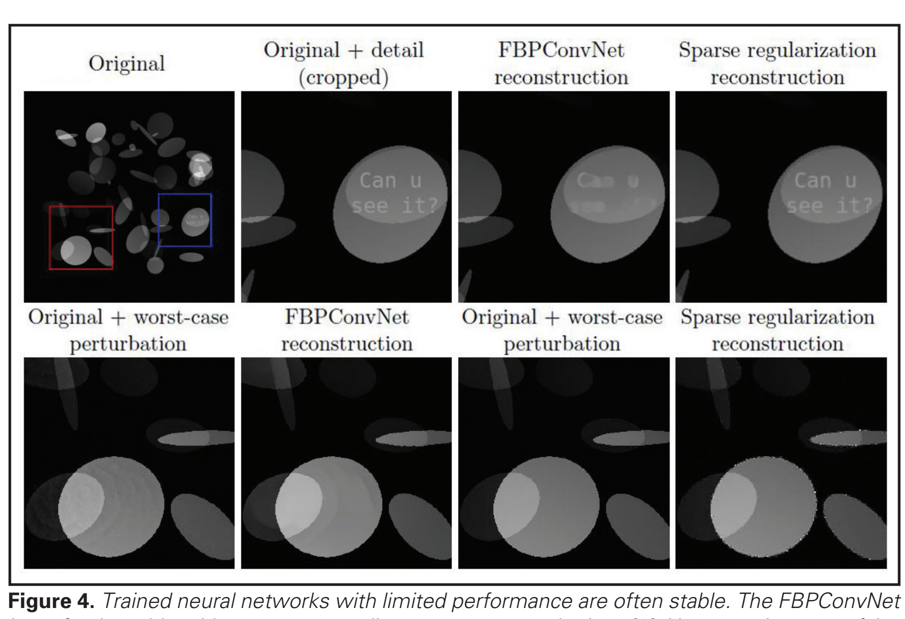


Figure 4. *Trained neural networks with limited performance are often stable. The FBPConvNet is perfectly stable with respect to small, worst-case perturbations [6]. However, it creates false negatives when recovering other details because it is trained on images that consist only of ellipses. Conversely, the deep MRI network in Figure 3 can accurately recover such details but is unstable. A standard sparse regularization method recovers the details and remains stable to worst-case perturbations of the same magnitude. This perturbation is less visible since it affects the image's dark regions. Figure courtesy of the authors.*

[1] https://fastmri.org

Mathematical Imaging

Continued from page 4

Many of the predominant mathematical problems in imaging are best formulated as inverse problems. Inverse problems involve the reconstruction of an unknown physical quantity from indirect measurements and may arise in tomographic imaging (like magnetic resonance imaging, computed tomography and positron emission tomography scans, and optical tomography), wave imaging (like ultrasound or seismic imaging), and hybrid imaging (like photoacoustic tomography). Most of these types of inverse problems are ill-posed and thus require appropriate mathematical treatment to recover meaningful solutions. Mathematical concepts from functional analysis, statistics, and numerical analysis play an important role.

New imaging techniques and hardware—i.e., novel developments in photoacoustic tomography, optical tomography, and lensless imaging—have also shaped mathematical imaging. Recent advances in quantum imaging suggest the possibility of overcoming "Rayleigh's curse," which is a statistical limit on resolution, via specially-designed quantum optical systems. Diffusion tensor imaging in medical scenarios and electron backscatter tomography in materials science provide "images" that are manifold-valued, thus rendering differential geometry tools essential. Dynamical imaging—the treatment of videos and multimodal images—pertains to questions in optical flow (see Figure 2), image metamorphosis, and registration. The concept of metamorphosis particularly endows the space of images with a nonlinear Riemannian structure, which one can use in applications like diffeomorphism estimation by minimizing the path energies of corresponding geodesics.

Finally, the emergence of machine learning and impressive performance of deep neural networks (DNNs) in many imaging applications have fueled a significant amount of research in data-driven methods. Discovering the mathematical principles behind the seemingly simple concept of neural networks is a tremendous undertaking. Understanding the design, training, and performance of DNNs will again draw from nearly every area of mathematics and surely lead to even more insights and theories that feed back into these fields. Many interesting questions remain, as it appears that standard DNNs are mostly insufficient for solving imaging problems — especially in areas like medical and scientific imaging where training data is scarce. Although these data are embedded in high dimensions, they actually stay near a low-dimensional manifold; this fact calls for deep geometric learning of big data to explore and exploit the underlying geometry in many applications. Generalizing neural networks to arbitrary geometric domains like graphs and manifolds and improving the accuracy, efficiency, and interpretability of the learning process therefore remains a challenge.

New developments in machine learning have the potential to solve data analysis and processing tasks that were previously unthinkable. At the same time, researchers must confront the limitations of these techniques—such as data bias, instabilities, and computational challenges—when applying them to practical imaging problems. This contrast between promise and practice calls for novel developments in mathematical imaging that could potentially help combine mathematical modeling and analysis with data-driven components. The question of how to best do this, however, remains unanswered.

The broad and diverse range of mathematical topics in imaging science demonstrates the true richness of this area for mathematical research and explains the subject's lure for mathematicians from various fields within and beyond academia. New trends and developments in mathematical imaging are also inspiring newfound career perspectives for early-career researchers, as the emergence of data science within imaging has created employment opportunities in many industrial sectors.

References

[1] Aviles-Rivero, A.I., Papadakis, N., Li., R., Sellars, P., Fan, Q., Tan, R.T., & Schönlieb, C.-B. (2019). GraphXNET – chest X-ray classification under extreme minimal supervision. In *International conference on medical image computing and computer assisted intervention*. Shenzen, China.

[2] Balle, F., Beck, T., Eifler, D., Fitschen, J.H., Schuff, S., & Steidl, G. (2019). Strain analysis by a total generalized variation regularized optical flow model. *Inverse Prob. Sci. Engin., 27*(4), 540-564.

Carola-Bibiane Schönlieb, Hongkai Zhao, Gabriele Steidl, and Michael B. Wakin respectively serve as chair, vice chair, program director, and secretary of the SIAM Activity Group on Imaging Science. Carola-Bibiane Schönlieb is a professor of applied mathematics in the Department of Applied Mathematics and Theoretical Physics at the University of Cambridge, where she leads the Cambridge Image Analysis Group. Her current research focuses on variational methods, partial differential equations and machine learning for image analysis, image processing, and inverse imaging problems. Hongkai Zhao is a professor of mathematics at Duke University. His research interests are in computational and applied mathematics, including numerical analysis and scientific computing, inverse problems, and image processing. Gabriele Steidl is a professor at the Technical University of Berlin. She has previously held positions as an associate professor at the Technical University of Darmstadt and a full professor at the University of Mannheim and the University of Kaiserslautern. Michael B. Wakin is a professor of electrical engineering at the Colorado School of Mines. His research interests include signal and data processing via sparse, low-rank, and manifold-based models.

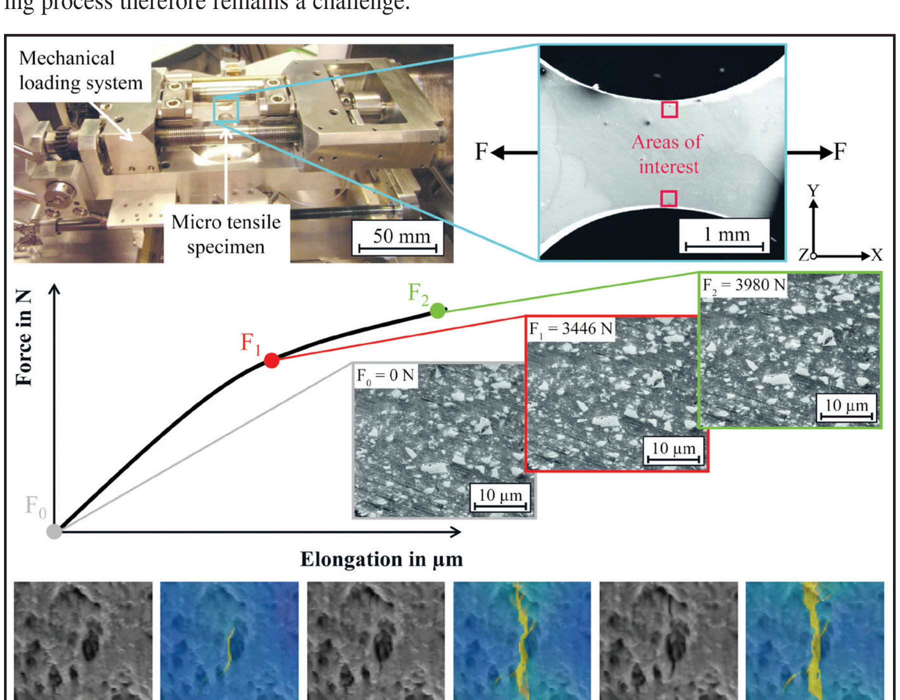


Figure 2. *Crack detection in materials during tensile tests by optical flow variational methods. Even cracks that are not visible to the human eye (gray value images) can be detected. Figure courtesy of [2].*

Optimal Transport for Generative Neural Networks

By Gabriel Peyré

Since 2012, deep neural networks (DNNs) have been revolutionizing machine learning. Although the concept is far from new, DNNs have enabled spectacular advances in the recognition of text, sounds, images, and videos in recent years. Perhaps even more surprising is that one can also use these neural networks in an unsupervised manner to automatically generate "virtual" text or images, which are often called "deep fakes." Here I will draw a link between the learning of generative neural networks and the theory of optimal transport. This method, which was framed by Gaspard Monge in the 18th century and reformulated by Leonid Kantorovich in the 20th century, is now a tool of choice for tackling important problems in data science.

Generative Neural Networks

Rather than use neural networks to analyze images, researchers can employ them "backwards" to generate images [3]. These generative neural networks find applications in special effects, video games, and artistic creation. For example, Figure 1 (on page 1) depicts the structure of a generative network g_w that depends on weights w. The layers play mirror roles when compared to the architecture of classical discriminating neural networks. Indeed, while networks for discriminating tasks take high-dimensional data (such as an image) and output a low-dimensional representation that is useful for classification, the exact opposite occurs in generative networks. A user can generate an image $x = g_w(y)$ from a "latent" vector y that is composed of a small number of values, which are typically drawn randomly.

Training such networks is an unsupervised problem. Consider a large collection of n training images $\{z_1, z_2, ..., z_n\}$. The goal is to select the weights w of the network g_w's neurons so the generated "fakes" resemble the training set images as closely as possible. One produces these fake images $(x_i)_i$ by randomly drawing the input latent values y_i and applying the network to these inputs to obtain $x_i = g_w(y_i)$. The training optimization problem is therefore

$$\min_w \text{Distance}(\{g_w(y_1), ..., g_w(y_n)\}, \{z_1, ..., z_n\}).$$

We thus wish to define a suitable notion of distance between two sets of points.

Monge's Optimal Transport

Monge formulated the optimal transport problem in 1781 for military applications [5]. He sought to determine the most economical method of transferring objects from a set of sources $\{x_1, ..., x_n\}$ to a set of destinations $\{z_1, ..., z_n\}$; for Monge, it was a matter of moving soil from cuttings to create embankments for the protection of soldiers. But this scenario finds a multitude of applications. When training generative networks, for example, the fake images that the network generates are the sources and the database's images are the destinations.

Researchers thus seek a permutation s of $\{1, ..., n\}$ so that each point x_i is linked to a single point z_{s_i}. Figure 2 displays a simple example of such a permutation with $n = 6$ elements. Monge's problem then involves finding the permutation that minimizes the sum of the transport costs. He decided that the cost of transportation between a source x and a destination z is equal to the Euclidean distance $\|x - z\|$ between the two points. However, one can choose another cost — i.e., a traveling time or the price requirement for gasoline when driving a truck. We must then solve the problem

$$\text{Distance}(\{x_1, ..., x_n\}, \{z_1, ..., z_n\}) \overset{\text{def.}}{=} \min_{\text{permutation } s} \|x_1 - z_{s_1}\| + \|x_2 - z_{s_2}\| + ... + \|x_n - z_{s_n}\|.$$

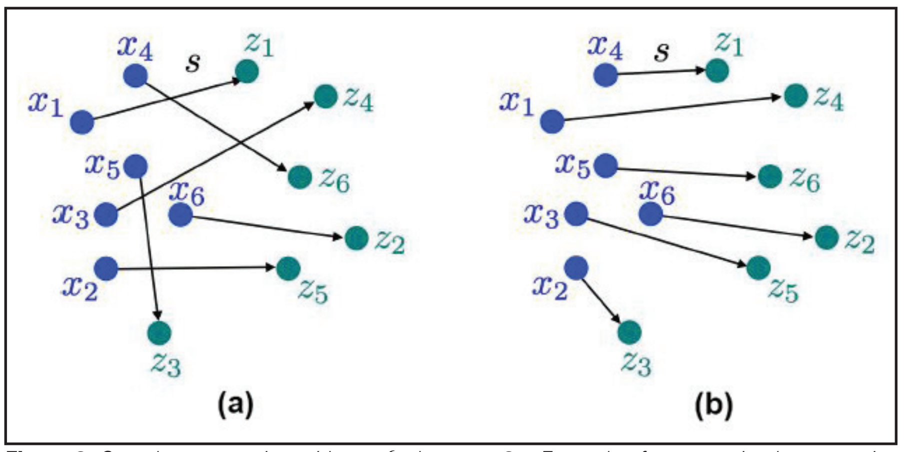


Figure 2. *Sample permutation with $n = 6$ elements.* **2a.** *Example of a non-optimal permutation s.* **2b.** *The corresponding optimal permutation. Image courtesy of Gabriel Peyré.*

This problem's solution provides an optimal assignment between the points but also defines the notion of distance between the sets of points in question.

The difficulty of calculating this distance is that the total number of permutations that must be tested is very large; indeed, there are $n! = n \times (n-1) \times (n-2) \times ... \times 2 \times 1$ possibilities. For example, there are $6! = 720$ possible permutations in Figure 2. We can test them all and select the best one, as depicted in Figure 2b. The difficulty is that there are more than 10^{100} possibilities for $n = 70$ — compared to the 10^{79} atoms in the universe. And when training neural networks, n is even bigger.

Monge was unable to provide an efficient method for solving this problem [5]. Instead, Kantorovich identified a new formulation for the optimal transport problem in 1942 [4]. His formulation allows scientists to divide each source into several parts; for instance, one can split a source into two equal parts, each with a weight $1/2$. This division of production that simplifies the optimization problem is also natural for Kantorovich's problem, which attempted to model and plan production in economics. Kantorovich received the Nobel Prize for Economic Sciences for this idea in 1975. In 1947, George Dantzig introduced the simplex algorithm [2], which makes it possible for

See **Optimal Transport** *on page 8*

A Western Sunrise

MATHEMATICAL CURIOSITIES

By Mark Levi

In a distant galaxy, a lone planet orbits its sun; all other celestial bodies are far away. The planet's axis is not tilted relative to the ecliptic (the plane of its orbit).

For many days, a creature living on the planet's equator has been watching the sun rise in the east and set in the west. One evening, the sun sunk as usual below the western ocean and the light began to fade — but then it stopped fading, the sky brightened, and the sun rose back up from where it just set, from the *west*! The sun then described the arc overhead and set in the *east*. After a brief period of darkness, the sun reappeared in the east, made its overhead arc, and set in the west — just as in previous days. How did this happen?

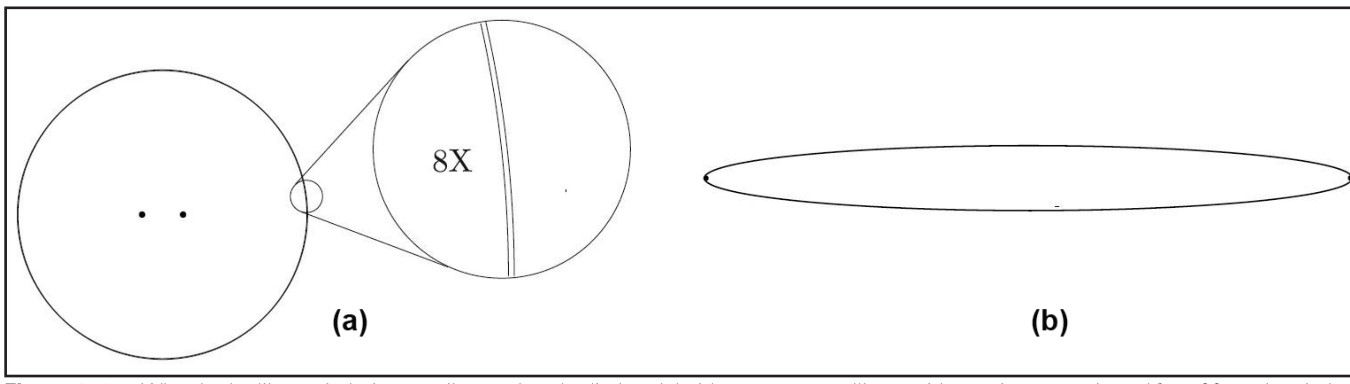


Figure 2. 2a. *What looks like a circle is actually two barely distinguishable curves: an ellipse with semiaxes* $a=1$ *and* $b=.99$ *and a circle of radius* $(a+b)/2$. *The ellipse's foci are surprisingly far apart at the distance* $\approx .282$. **2b.** *The foci for an ellipse with aspect ratio* $1{:}10$ *are 99.5 percent of the way from the center to the vertex.*

A Solution

Let us first consider the planet with zero axial spin. To the creature on such a planet, the sun will travel east—as in Figure 1—and make one revolution in the sky per year. For eccentric orbits, this eastward progression is very uneven: fast for a short time and slow most of the time. This spike occurs when the angular velocity ω_{orbital} of the planet's position vector relative to the sun spikes during the close passage to the sun, near the perihelion. The key to the puzzle's solution is that during the quick passage between points A and B in Figure 1, the sun travels by more than 180° in the sky — and does so rapidly.

Let us now spin our planet on its axis with angular velocity ω_{spin} in the same direction as travel around the sun, i.e., counterclockwise in Figure 1, just like Earth. The sun—as seen in the sky—will thus acquire angular velocity ω_{spin} east to west (the conventional direction), in addition to the preexisting ω_{annual} in the opposite direction. In the planet's sky, the sun will therefore travel with angular velocity

$$\omega_{\text{sun}} = \omega_{\text{spin}} - \omega_{\text{orbital}}, \qquad (1)$$

measured from east to west. With a sufficiently eccentric orbit (see Figure 1), ω_{orbital} experiences a spike that can dominate ω_{spin} and makes ω_{sun} change sign. This occurrence manifests as the reversal of the sun's motion in the sky. With sufficient eccentricity, the sun's backward travel by $>180^\circ$ for the spinless planet happens so quickly that the spin does not appreciably change this angle. It is thus possible for the sun to make a west-to-east trip in the sky by more than 180° once a year.

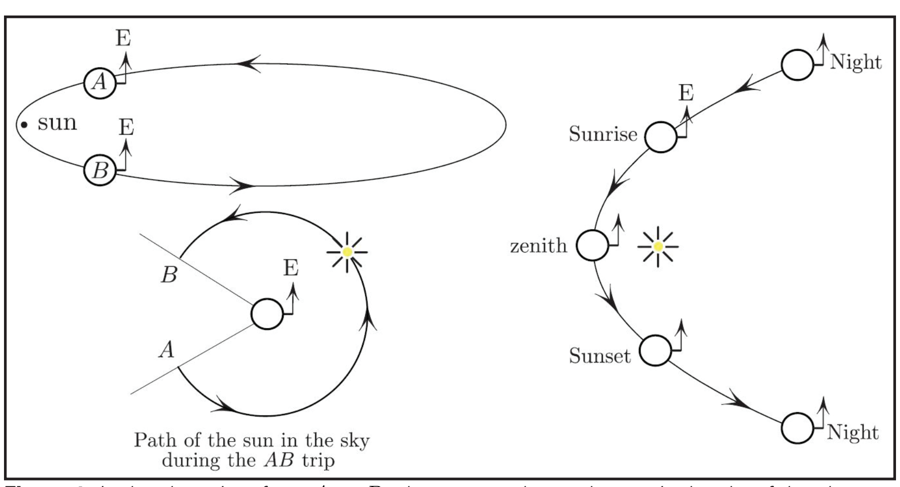


Figure 1. *In the short time from* A *to* B, *the sun travels east by* $>\pi$ *in the sky of the observer for the planet with no axial spin. The arrows point east.*

Back to Earth

For those of us on Earth, $\omega_{\text{spin}} >> \omega_{\text{orbital}}$ so that ω_{sun} does not change sign — although it does vary. A very slight slowdown of the sun's travel from east to west occurs as we get closer to the sun. This year we passed the perihelion on January 2; the sun advanced the slowest in our sky around that day, but of course not slow enough to move backwards.

Sliding Days

Another interesting consequence of the eccentricity of Earth's orbit is the "sliding" of the daylight time for a few days around each solstice. For example, at 40° latitude—roughly the latitude of State College, Pa.—there is a period of about three weeks around December 21 when *both* sunrise and sunset occur later and later with each passing day. The daylight interval slides from day to day and becomes the shortest on December 21, after which it continues sliding while also elongating. Around January 4 or so, the sunrise times reverse and begin getting earlier with each passing day.

On Focus

It is interesting to note the close proximity of the foci of an elongated ellipse to the vertices of the ellipse, i.e., the closeness of the planet as it passes by the sun. Figure 1 illustrates this phenomenon with anatomical correctness: the semiaxes $a=1$ and $b=0.25$ and the distance in question is <0.032 (this distance is $a-c$, with $c=\sqrt{a^2-b^2}$ as the distance from the focus to origin). The focus is over 96 percent of the way from the center of the ellipse to its vertex. For the $1{:}10$ aspect ratio, the foci seem to almost lie on the ellipse in the resolution of Figure 2.

As a flip side of this coin, consider a nearly round ellipse. The distance between the foci for such an ellipse is large relative to $a-b$: it is $2\sqrt{a^2-b^2} = 2\sqrt{(a+b)(a-b)} > 2\sqrt{2b}\sqrt{(a-b)}$. Figure 2 illustrates the way in which a tiny deformation causes a relatively great split of the center into foci. A circular room serves as a good whispering gallery if the speaker and listener are symmetrically positioned relatively close to the center, as in Figure 2. This is so because the circle is almost indistinguishable from an ellipse with these foci.[1]

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.

[1] The ellipse is in fact close to the circle in the C^2-norm, which is what matters for the focusing properties.

Optimal Transport

Continued from page 7

one to efficiently solve large-scale transport problems. Its numerical complexity when solving an optimal transport problem between n points is of the order of n^3, which is much lower than $n!$. The simplex algorithm is at the heart of a large number of industrial systems that must optimize the adequacy between means of production and consumption. Researchers can also use it to train generative neural networks. Further details on optimal transport theory, efficient algorithms, and their application to data science are available in [7].

Adversarial Networks

A difficult aspect of applying optimal transport to create generative networks is choosing the transport cost between two images. One could calculate the Euclidean distance between the images' pixels, but this method does not work well because it fails to account for the geometry of the objects that are present in the images. In 2014, Ian Goodfellow and his collaborators introduced a more successful idea [3]. In this approach, a second network f—called an adversary network—plays a discriminative role. While generator g aims to create fake images that look real, f plays the role of an opponent that must recognize true and fake images. The joint training of these two networks corresponds to what one may call a zero-sum game. John Nash studied this concept [6]; like Kantorovich, he too received the Nobel Prize in Economic Sciences in 1994.

These recent advances [3] have made it possible for researchers to obtain convincing results in image generation. Figure 3 depicts results from the calculation of "paths" of images between dogs and cats [1]. This "animation" generates a continuous path $x(t) = g_w((1-t)y_0 + ty_1)$ for $t \in [0,1]$, which is a linear interpolation in latent space between two fixed vectors y_0 and y_1.

This article is based on Gabriel Peyré's joint plenary address at the 2020 SIAM Annual Meeting[1] and the 2020 SIAM Conference on Imaging Science,[2] which were co-located and took place virtually last year. Peyré's presentation is available on SIAM's YouTube Channel.[3]

Acknowledgments: I thank Gwenn Guichaoua for her attentive proofreading, as well as Sébastien Racanière and Vincent Barra for their corrections.

References

[1] Brock, A., Donahue, J., & Simonyan, K. (2019). Large scale GAN training for high fidelity natural image synthesis. In *Proceedings of the 2019 international conference on learning representations (ICLR 2019)*. New Orleans, LA.

[2] Dantzig, G.B. (1990). Origins of the simplex method. In S.G. Nash (Ed.), *A history of scientific computing* (pp. 141-151). New York, NY: Association for Computing Machinery.

[3] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., …, Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems (NIPS 2014)* (pp. 2672-2680). Montreal, Canada.

[4] Kantorovich, L. (1942). On the transfer of masses. *Doklady Akademii Nauk*, *37*(2), 227-229.

[5] Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences* (pp. 666-704).

[6] Nash, J.F. (1950). Equilibrium points in *n*-person games. *Proc. Nat. Acad. Sci.*, *36*(1), 48-49.

[7] Peyré, G., & Cuturi, M. (2019). Computational optimal transport. *Found. Trends Mach. Learn.*, *11*(5-6), 355-607.

Gabriel Peyré is a senior researcher at the Centre National de la Recherche Scientifique (CNRS) and a professor at the École Normale Supérieure. His research is focused on optimal transport methods for machine learning.

[1] https://www.siam.org/conferences/cm/conference/an20

[2] https://www.siam.org/conferences/cm/conference/is20

[3] https://www.youtube.com/watch?v=Y6Qndy6rsio

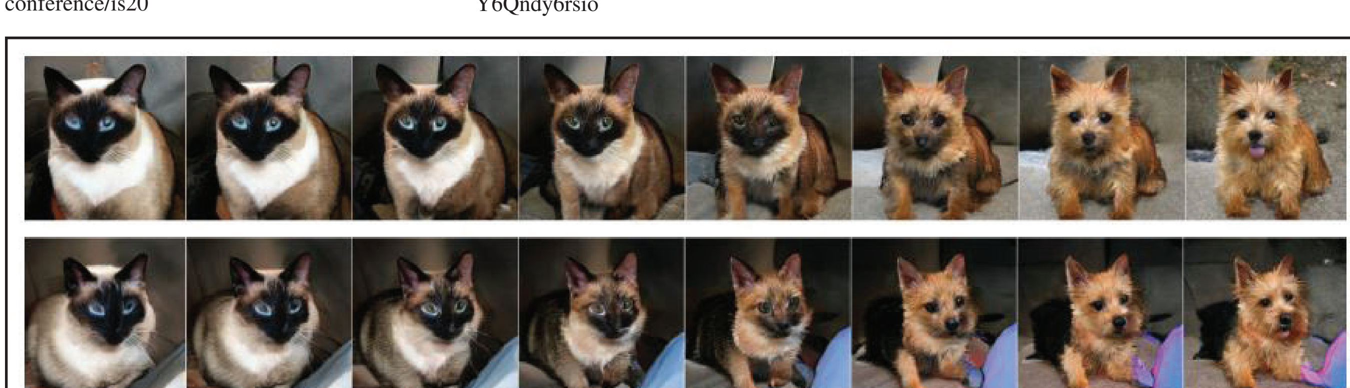


Figure 3. *Two examples of "deep fakes" — virtual images that interpolate between cats and dogs. Image courtesy of Gabriel Peyré.*

Transitioning from Academia to the Healthcare Industry

Embracing a Career Change and Interdisciplinary Practical Thinking

By Anuj Mubayi

I consider myself to be an experienced health decision analyst and mathematical modeler. Appropriate interdisciplinary mathematical modeling techniques for complex systems can remove the mystery, randomness, and powerlessness from various real-life scenarios, including deadly and destructive infectious outbreaks. Because these methods are important for public health preparedness, they are in high demand in today's workforce.

My interests and thinking closely align with a sentiment of renowned mathematician Carl Friedrich Gauss: "*Surely it is not knowledge, but learning; not owning but earning; not being there, but getting there that gives us the greatest pleasure.*" I obtained my Ph.D. in applied mathematics from Arizona State University (ASU), where I studied stochastic processes and dynamical systems. A strong foundation in these approaches deepened my understanding of computational modeling and data analytical methods, which I routinely apply in my work in the public health and social science fields. My research applications focus on the ecology of infectious diseases, epidemiology, and health economics, and the questions that drive my interests lie at the intersection of public health, medicine, and the life and social sciences. These questions are shaped by the use of mathematical modeling to control infectious diseases and have helped advance society's understanding of the complexities of new and emerging outbreaks. They have also led to a timely evaluation of interventions, drastically helping to mitigate mortality and morbidity rates. I am particularly interested in exploring infectious disease dynamics over multiple temporal and spatial scales and levels of organization, and conducting economic analyses for health interventions for diverse populations from a number of different perspectives.

After more than a decade of experience in academia, I recently transitioned to the healthcare industry and joined *PRECISIONheor*,[1] which is a member of a family of companies of Precision Medicine Group[2] — a leader in the field of medical sciences. I am an associate director in PRECISIONheor's Advanced Modeling Group, where I lead a highly qualified team of applied mathematicians, data analysts, and health economists who provide key insights into the healthcare sector. I also hold visiting faculty positions in the Department of Mathematics at Illinois State University (ISU), ASU's College of Health Sciences, and the Prevention Research Center in Berkeley, Calif.

During my time as an academician, I directly mentored over 105 students — more than half of whom were underrepresented minorities. I strongly believe that the pursuit of knowledge and understanding is enriched by an environment wherein people of diverse backgrounds learn from each other and participate in free and genuine exchanges of ideas. As I continue my career in a new sector, I remain committed to teaching, mentoring, and advising in numerous local and national programs for minority students and colleges, including ISU's Intercollegiate Biomathematics Alliance[3] and Purdue University's National Alliance for Doctoral Studies in the Mathematical Sciences.[4] I will also continue to conduct seminars at various scientific events around the world and collaborate directly with the leading governmental, educational, and healthcare sectors in Colombia, Ecuador, India, Peru, Portugal, and the U.S. to broadly and effectively put disease-related models into action.

In academia, individuals naturally gain valuable experience with critical thinking, the scientific method, technical writing, conceptual modeling, and independent research. These skills are all beneficial in industry settings. However, anyone transitioning to industry must also be adept at leadership, entrepreneurship, interpersonal relations, and project management on a short time frame. When making this move, it is important to become familiar with the necessary methods and tools for the position and organization in question, and know how to remain productive while simultaneously learning from peers.

CAREERS IN MATHEMATICAL SCIENCES

Although I was initially skeptical, my transition from academia to industry was extremely smooth. The type of work I do remains somewhat similar — I am still conducting research, partaking in interdisciplinary collaborations, developing mathematical modeling methods, applying data analysis techniques, and mentoring. I was fortunate to find a perfect fit within a company that directly aligns with my interests. In my experience, PRECISIONheor is like a semi-academic institution in that it promotes research publications in scientific journals, professional development through technical conferences, and hands-on experience with rigorous mathematical modeling methods. Most of the industrial projects with which I have been involved incorporate mathematical modeling components and have applications in the pharmaceutical and healthcare industries. The key difference between my academic and industry life is teaching, though I can still teach part time through my adjunct affiliations at ISU and ASU.


Although he has now transitioned to an industry position, Anuj Mubayi remains active in academic networks, mentoring programs, and teaching projects. Photo courtesy of Anuj Mubayi.

PRECISIONheor is a world leader in the generation of strategic, innovative, and credible evidence that supports the development and commercialization of novel healthcare innovations. The company is renowned for its unparalleled expertise in evidence synthesis, economic modeling, and real-world data analysis, as well as its ability to deliver timely academic insights that answer the demands of all stakeholders — from payers to policymakers. I chose to transition from academia to industry because I was looking for a new challenge and different responsibilities. I wanted to evolve my skill set from theoretical to practical, with direct implications for society, and have since realized that the change from academia to industry affects the style of work and thinking more than the actual tasks themselves. While working with cohorts and real data, I have found that industry positions generally seem to allow more flexibility to grow and evolve than academia. For example, I enjoy a high degree of freedom when selecting projects, planning research methodologies, and entertaining the possibility of publication.

A typical day in the life of a research scientist in the healthcare industry involves multiple components, including attending project meetings with team members, preparing materials for client update meetings, and working on individual projects. During meetings, we discuss a project's technicality, generate presentation documents, and

See **Healthcare Industry** *on page 12*

[1] https://www.precisionheor.com
[2] https://www.precisionmedicinegrp.com

[3] https://about.illinoisstate.edu/iba/about/research-mentors
[4] https://mathalliance.org/mentor/anuj-mubayi


Anuj Mubayi delivers a hands-on lecture during the NSF-funded Partnerships for Enhanced Engagement in Research (PEER) Project Training Workshop at El Salvador's Universidad Francisco Gavidia in January 2017. Photo courtesy of Anuj Mubayi.

Jim Simons' Road from Mathematics to Market Maven

The Man Who Solved the Market: How Jim Simons Launched the Quant Revolution. *By Gregory Zuckerman. Portfolio/Penguin Random House, New York, NY, November 2019. 384 pages, $30.00.*

BOOK REVIEW
By James Case

Jim Simons has enjoyed a long and varied career in pure mathematics, code-breaking, and finance. He is perhaps best known for founding the Simons Foundation, which supports research in mathematics and the basic sciences. As of June 2020, his net worth allegedly exceeded $23 billion. In *The Man Who Solved the Market: How Jim Simons Launched the Quant Revolution*, Gregory Zuckerman describes the twists and turns along the road that ultimately led to Simons' success. Because Simons has long maintained a notoriously low profile—and those who have worked for him have been almost equally reticent—piecing together his story was a difficult task for Zuckerman.

Simons was born in 1938, near the end of the Great Depression, to middle-class parents in Brookline, Mass. The family moved to nearby Newton in time for him to attend the highly regarded Newton North High School. After watching his father, who was trapped in a job he disliked and always a bit short on cash, Simons concluded that he wanted enough wealth to be the master of his own fate. He also decided to study math at the Massachusetts Institute of Technology (MIT).

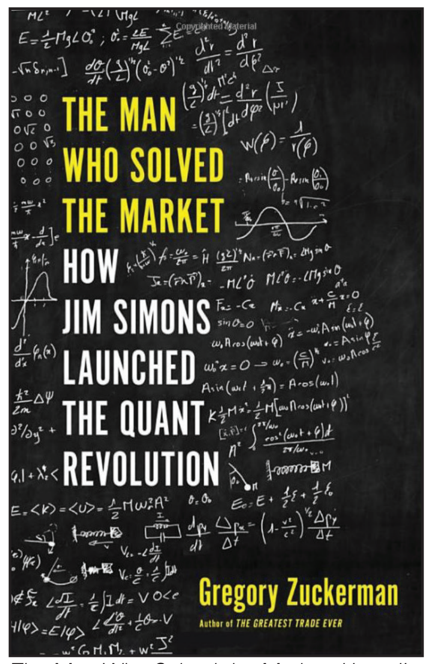
Simons was an inconsistent student who did outstanding work in courses he liked while neglecting those he did not. However, he experienced an epiphany of sorts in the calculus class that introduced him to Stokes' theorem. There he witnessed for the first time the way in which algebra, geometry, and analysis could combine symbiotically. Soon fellow students were asking Simons for help in the subject. "I just blossomed," he later told a friend. Though he considered himself less talented than some of his classmates, Simons felt fully capable of making a substantial contribution.

After graduating with a B.A. in mathematics in 1958, Simons felt the need for a little adventure. He and two friends embarked on a motor scooter trip that they dubbed "Buenos Aires or bust." Though the group never made it to Argentina, two of the three persevered long enough to reach Bogotá, Colombia, where an MIT classmate welcomed them into his family home. Simons and his friend luxuriated in creature comforts until it was time to return to MIT for graduate school.

Before long, Simons' advisor at MIT suggested that he transfer to the University of California, Berkeley, where he could study geometry under Shiing-Shen Chern. But when Simons arrived at Berkeley, he found that Chern was on sabbatical elsewhere. Simons was therefore obliged to work with others, including his eventual thesis advisor Bertram Kostant. He completed his dissertation about holonomy groups on Riemannian manifolds in 1961.

While in graduate school, Simons married his 18-year-old girlfriend, became a father, and gained his first investing experience. After deciding that stock prices moved too slowly to hold his interest, Simons followed a broker's advice and bought soybean futures (financial contracts that promise the delivery of a good or service at a pre-determined price and date). These rose almost immediately from $2.50 to $3.00 per bushel, leaving him with a paper profit of several thousand dollars. When more experienced friends urged Simons to realize the profit by selling immediately, he ignored their advice and barely broke even when he finally did sell. Nevertheless, he was hooked.

Upon leaving Berkeley, Simons accepted a three-year teaching fellowship at MIT. But he grew restless after a single year of teaching and decided to return to Bogotá and start a floor-covering business with a friend. With financial backing from the friend's family—along with a paltry sum that Simons and his father scraped together—the duo opened a factory that produced vinyl floor tile and PVC piping. Once the business was established, Simons took a teaching position at Harvard. Though he was a popular teacher, he earned no more than any other postdoc. And because he borrowed money to invest in the flooring business, Simons had to moonlight at a local junior college just to make ends meet. Not surprisingly, his own research suffered.


The Man Who Solved the Market: How Jim Simons Launched the Quant Revolution. By Gregory Zuckerman. Courtesy of Portfolio/Penguin Random House.

To double his income and kickstart his research, Simons accepted a job with the Princeton, N.J., division of the Institute for Defense Analyses (IDA). In an effort to attract the best talent, the Defense Department encouraged staff members to divide their time fairly evenly between code-breaking activities for the government and their own personal research. The organization was a beehive of ideas. Simons proved to be an adept listener with a knack for recognizing his colleagues' better ideas and devising algorithms to test them. In this setting, he soon became a star code breaker.

At the same time, Simons' academic research began to prosper. His 1968 paper titled "Minimal Varieties in Riemannian Manifolds" extended the solution of Plateau's problem through six dimensions and conjectured a counterexample in dimension seven, which has since been verified. The paper attracted a good deal of attention and was rich in applications; it is still cited on a regular basis. This single paper established Simons as a leading geometer.

Meanwhile, Simons' interest in investment had returned. He and three IDA colleagues published an internal classified paper entitled "Probabilistic Models for and Prediction of Stock Market Behavior," which promised annual gains of 50 percent. He also persuaded his IDA boss and the institute's best programmer to join him in an investment company, which was to be called iStar. However, Simons was unable to assemble sufficient funding to make the project a reality.

All of this was happening during the Vietnam War. Students everywhere were protesting and attempting to remove military installations of every description—including the IDA near Princeton University—from their campuses. After Simons contributed a mildly anti-war opinion piece to a local newspaper, he was interviewed by a *Newsweek* reporter who coaxed even more subversive thoughts out of him. He was fired from his IDA position as a result.

Almost immediately, Simons was offered the chairmanship of the Mathematics Department at Stony Brook University. The school was only 11 years old and the state of New York wanted to turn it into a "Berkeley of the East." The university had already hired Nobel Prize-winning physicist Chen Ning Yang, and the president was looking for an aggressive leader to build a world-class math department. Under Simons' leadership, the Stony Brook Mathematics Department became a mecca for young mathematicians.

In 1976 at the age of 37, Simons received the American Mathematical Society's Oswald Veblen Prize in Geometry for his work on minimal varieties and a paper he wrote with Chern that introduced a new class of geometric invariants. The award cemented his place among the first rank of active geometers. In fact, Chern-Simons theory has since generated tens of thousands of citations in mathematics and mathematical physics literature.

Though he was comfortable with his acclaimed research and power as department chairman, Simons grew restless again. When a larger firm acquired the flooring company in Colombia, Simons and his fellow shareholders enjoyed windfall profits. His dormant penchant for investing was reawakened. In 1978, Simons resigned from his position at Stony Brook to found an investment company called Monemetrics that would develop algorithms to identify unusually profitable short-term investment opportunities.

Monemetrics was not very successful and Simons was forced to shut it down after only a few years. However, this experience did nothing to shake his conviction that one could algorithmically tame the market. But it was not until he and Howard Morgan founded Renaissance Technologies (RT) in 1982 that real progress commenced. Simons' role in the new firm was primarily that of a leader, and he assembled his workforce much like the IDA had assembled its code-breaking crew. He hired bright

See **Market Maven** *on page 11*

Exploring Symmetry in Chaos

The much-anticipated second edition of Symmetry in Chaos: A Search for Pattern in Mathematics, Art, and Nature, by Michael Field and Martin Golubitsky, was published by SIAM in 2009. It features numerous new illustrations, addresses recent progress in the mathematics that underlies symmetric chaos, and serves as a follow-up to the first edition, which was released by Oxford University Press in 1992.

The book is written for a general audience and illustrates the ways in which classical symmetry and modern chaotic dynamical systems can interact to produce a set of striking images. It explains the relevant mathematical background, provides a detailed description of how the images are produced, and describes several implications of the mixture of symmetry and chaos.

The following excerpt begins a discussion of the production process for the many images of symmetric chaos. This text comes from chapter 1, "Introduction to Symmetry and Chaos," and is modified slightly for clarity.

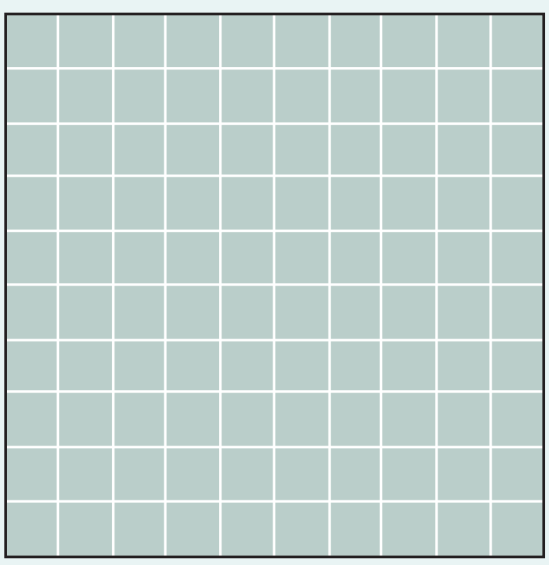


Figure 1. *Pixels on a* 10×10 *grid.*

FROM THE SIAM BOOKSHELF

Pixel Rules

If we intentionally confuse pixels on the screen with points (in the plane), then the rules that make our pictures are similar to the arithmetic rules we use. However, unlike the doubling rule, we do not want our rules on repeated application to grow without bound (otherwise, points would soon leave the computer screen). We think of a pixel rule as a rule that has pixels as input and pixels as output. The pixel rule may depend on some complicated mathematical formula, but for the moment we wish to keep the arithmetic hidden. To make a black and white picture, we assume that the screen is black. We choose one pixel and turn it on — the corresponding point on the screen will then be white. Then we invoke our pixel rule, beginning with the first pixel as input, and obtain a new pixel that we turn on. Finally, we repeat this rule over and over again until we decide to stop. The whole process is called *iteration*. In this scheme, there is no reason why one pixel cannot be visited more than once.

As an example of a very simple pixel rule, choose one pixel from the screen, say the top left. We define a pixel rule by requiring that we always select the top left pixel as output, regardless of the pixel we choose from the screen as input. However many times we apply the rule, we never see more than two pixels lit on the screen: the initial pixel and the top left pixel.

Next we look at a slightly more complicated pixel rule. Following Figure 1, suppose that the monitor screen has 100 pixels arranged in a 10×10 grid. Choose a pixel $\mathcal{P}$ from the screen and the direction left. The pixel rule has two parts: if you can, move one pixel in the direction you are going; if you cannot, turn right one quarter of a turn. The picture that will result from this pixel rule is easy to describe. There is an initial segment moving left from the initial point $\mathcal{P}$ to the boundary of the grid, followed by a never-ending circumnavigation of the boundary in the clockwise direction (see Figure 2).

Even though the rules we describe here are rather simple, there are one or two interesting features that we want to single out for special mention.

First of all, note that the first part of the pixel sequence is different from its long-term behavior. In particular, the pixels on the initial line segment—labelled L in Figure 2—are never revisited. We say that this part of the pixel sequence represents the initial behavior and often use the term *transient* to describe the initial behavior. The transient behavior is seen at the beginning but not in the long term. The part of the pixel sequence that begins at the boundary represents the *long-term* behavior.

A second important observation about this example is that the long-term behavior repeats *ad infinitum*. We refer to this characteristic as *periodicity*. Since there are 36 pixels on the perimeter, this pixel rule repeats itself every 36 iterates (ignoring the initial transient).

Indeed, if we apply any pixel rule enough times, at least one pixel will eventually be revisited. To see why this is so, suppose that there are 100 pixels on the screen (any large number will do equally well). After 100 iterates we have "lit" 101 pixels, so at least one pixel must have been "lit" twice (this argument is an example of the *pigeonhole* principle: if 101 letters are to be put in 100 pigeonholes, at least one pigeonhole must contain at least two letters). It follows that if the rule we used to create our pictures was actually a pixel rule, then after an initial transient we would have to find periodic behavior. In general, our picture rules do not lead to this simple kind of periodic behavior, and color can be used to understand this point.

Coloring by Number

We now say more about how we color our figures. The basic idea is quite simple. Start with a mathematical formula that generates a picture, such as Figure 3. Choose an initial point and apply the rule a large number of times, typically between 20,000,000 and 100,000,000. Ignore the transient part of the pixel sequence that is produced (in practice, we only count pixel hits after the first 1,000 applications of the rule). Record the number of times each pixel is hit and color the pixel according to the value of that number. This process is no more than *coloring by number*. The actual colors are chosen according to which colors best bring out the underlying structure. Figure 3 shows the result of coloring a figure with five-fold symmetry after 667,000,000 iterations on a $3,000 \times 3,000$ pixel grid. Since there are 9,000,000 pixels, it follows by the pigeonhole principle that some pixels must have been hit more than once. In practice, many pixels are hit more than once; the color band in Figure 3 shows the colors assigned to pixels based on the number of times they have been hit. As we usually do, we leave the pixel black if it has not been hit. We color white shading to yellow if the pixel has been hit between one and 10 times, yellow if the pixel has been hit between 11 and 30 times, yellow shading to red if the pixel has been hit between 31 and 270 times, and so on, ultimately ending up with navy blue if the pixel has been hit at least 2,370 times (the maximum number of hits on an individual pixel was 42,534).

Thus far, we have confused pixels and points on the screen and regarded our mathematical formula as a pixel rule. However, when making a large number of applications of our rule, we really must distinguish the underlying arithmetical rule from a pixel rule. To see why this is so, recall that a pixel rule begins with a transient and then behaves periodically. A consequence is that the only sensible choice of coloring for pixels that are chosen using a pixel rule would be one color for the transient pixels (those visited only once) and another color for the pixels that are visited periodically. If we look at the colorings of Figure 3, we see that the picture represents a process that is far from periodic.

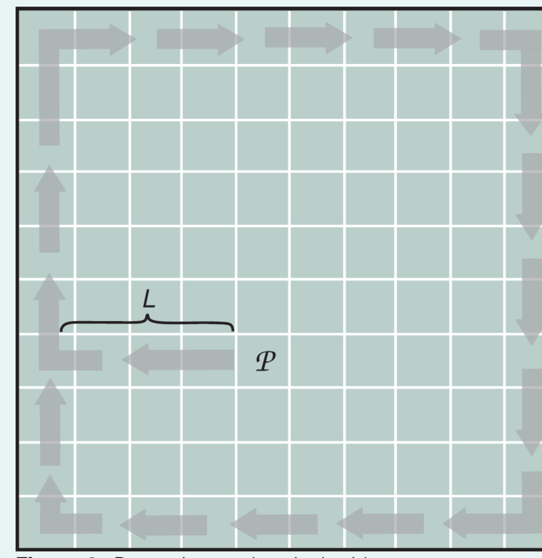


Figure 2. *Dynamics on the pixel grid.*

Enjoy this passage? Visit the SIAM bookstore[1] to learn more about Symmetry in Chaos[2] and browse other SIAM titles.

Michael Field is a professor in the Department of Mechanical Engineering at the University of California, Santa Barbara. He is currently working on theoretical problems in non-convex optimization and machine learning using ideas that originate in dynamics and symmetry. Martin Golubitsky is a distinguished professor of mathematics at the Ohio State University. He is the founding editor-in-chief of the SIAM Journal on Applied Dynamical Systems and a past president of SIAM.

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

[2] https://my.siam.org/Store/Product/viewproduct/?ProductId=1011

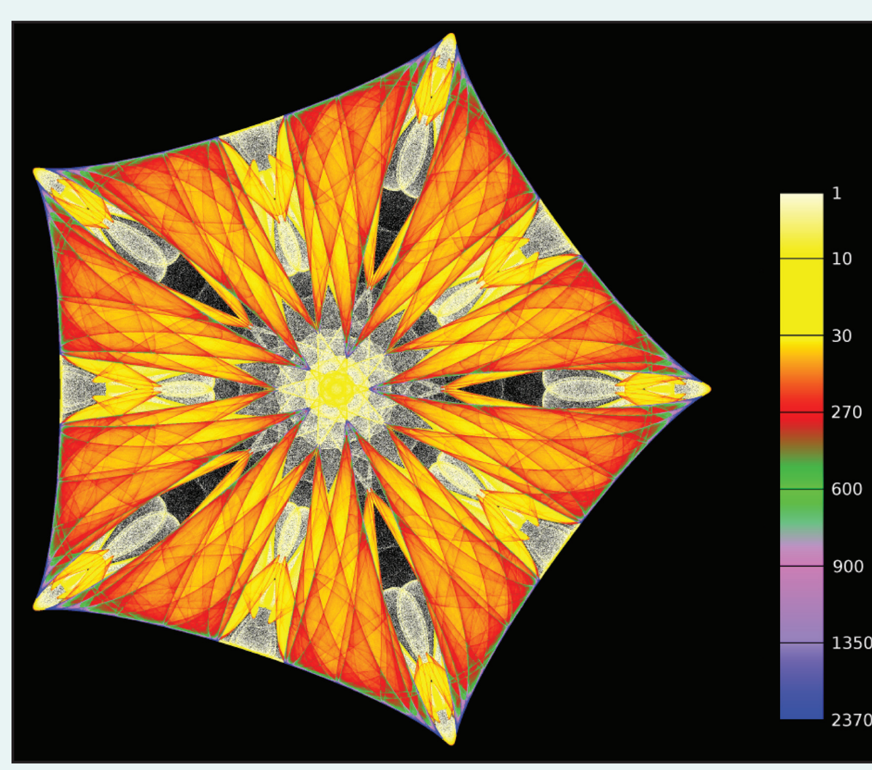


Figure 3. *Emperor's cloak: pentagonal symmetry and a color bar.*

Market Maven

Continued from page 10

young mathematicians, provided them with the best programming support that money could buy, and awaited results.

Robert Mercer and Peter Brown were two of RT's most important hires. The pair was directly involved in IBM's work to translate natural languages using artificial intelligence techniques, and Simons was eager to explore whether similar methods could identify exploitable patterns in market data. Mercer and Brown, perhaps enticed by the prospect of genuine wealth, agreed to leave IBM for RT.

Researchers at RT were not looking for fundamental principles of market behavior; instead, they searched for exploitable patterns in data. For instance, they noticed that large asset price movements—both upward and downward—are frequently followed by substantial movements in the opposite direction. If an asset price goes down, one can purchase that asset before the price rebounds and sell it afterwards for an immediate profit. If a price goes up, the asset can profitably be "sold short."

But RT did not become a legendary cash cow on the strength of one such observation. Profitability required a steady flow of similar findings, each of which was tested with care before considered "investment grade." The hedge fund required terabyte (and later petabyte) data storage capacity, along with immense computing power to analyze their treasure trove. Most of the patterns that RT employees discovered were more complex and less enduring than the foregoing example; such windows of opportunity have a pronounced tendency to close quickly.

The book's final chapters describe the condition of the leading figures at RT as they enter their later years. There is particular focus on the role that Mercer and his daughter Rebekah have come to play in politics; the outcry surrounding this involvement led Mercer to step down from his role as co-CEO at RT in 2017. Though none of the other major figures in RT have chosen to become so prominent, many have deployed their wealth in support of various causes, both political and otherwise.

Ultimately, *The Man Who Solved the Market* offers a fascinating glimpse into the life of elusive mathematician Jim Simons and his role as founder of one of the world's most successful investment firms.

James Case writes from Baltimore, Maryland.

A Deeper Way to Practice Deep Learning

Lessons from the 2020 SIAM Conference on Imaging Science

By Stacey Levine and Michael Elad

Deep learning (DL) is a revolution. The performance of deep learning solutions in the arena of image processing and computational imaging has taken a clear lead, pushing aside a wealth of classical knowledge that has accumulated over many decades of extensive research. While theoreticians acknowledge this performance boost, many find the seeds of DL unsettling due to their purely empirical nature, which does not have a strong theoretical backbone. DL also lacks model predictability and explainability, the absence of which could complicate real-world applications in fields like medical imaging and autonomous vehicles. Yet despite these shortcomings, increasingly more researchers from the mathematical imaging community are joining this new line of study.

But what about the classics in imaging science? Decades of powerful, theoretically sound, and successful methods have been built from different branches of mathematics, including variational approaches, partial differential equations (PDEs), harmonic analysis, sparsity-based models, and integral operators. Scientists have applied and intertwined these branches in various ways, resulting in powerful imaging techniques. Should we simply resign ourselves to the idea that these approaches might become obsolete?

Luckily, this DL fever has spread throughout the field of mathematical imaging in a more controlled, thoughtful, and "deep" manner than originally anticipated. Recall that DL's initial framework involves choosing an arbitrary network architecture and training it end-to-end to match inputs to outputs in a supervised fashion. This strategy employs black box solutions to increase performance by leveraging massive amounts of data and computational power while neglecting physical connections and data models. In contrast, while the imaging research community has embraced DL, many individuals have been inclined to pursue work that also remains harmonious with the classics. Recent research in the field includes novel DL architectures that are based on well-posed traditional imaging tools. In this way, DL provides new vantage points for understanding conventional models while simultaneously presenting fresh opportunities for constructing a comprehensive general theory for DL. The cooperation of these two worlds is inspiring the discovery of important connections, questions, and complementary approaches (see Figure 1).

The 2020 SIAM Conference on Imaging Science[1] repeatedly and extensively reinforced these themes. The plenary talks[2] offered snapshots of DL's impact across imaging domains, as well as the thoughtfulness with which leading researchers are effectively merging DL with the "classics."

DL architectures that are motivated by variational and PDE-based models are generating impressive results for image synthesis, restoration, and reconstruction. Gabriel Peyré's (CNRS and École Normale Supérieure) address kicked off this recurring theme.[3] His talk connected the field of optimal transport with Ian Goodfellow's generative adversarial networks, wherein fitting densities that are parametrized by deepnets become powerful frameworks for both image generation and discrimination. Thomas Pock's (Graz University of Technology) lecture on variational networks, which was inspired by the successful total variation functional, continued the same theme. Pock discussed image restoration architectures based on energy functionals that surpass their variational counterparts with respect to performance, while still remaining well-positioned to establish stability and generalization results and afford a much-desired Bayesian interpretation. Maarten V. de Hoop (Rice University) linked the Fourier integral operator and wave equation to DL architectures that researchers use for image reconstruction. He explained how these relationships give way to important generalizability guarantees — a critical challenge when one employs supervised data-driven DL models in practice.

Several plenaries demonstrated the power of fusing classical and DL approaches, utilizing DL only where classical or physics-based models are lacking. Laura Waller (University of California, Berkeley) presented both the state of the art and current challenges in physics-based computational microscopy. She also offered keen insight into specific parts of the problem that can benefit from data-driven DL models and spoke about how the fusion of these approaches is pushing boundaries. The efficacy of using data priors to more cleverly train models via smaller datasets was central to Michal Irani's (Weizmann Institute of Science) talk, which described the ability of patch-based methods to improve a degraded image by learning intrinsically (without requiring training data that is external to the image). Irani's presentation also illustrated how combining intrinsic learning with external data-driven DL models can supply users with the best of both worlds.

The invited presentations likewise provided key insights into foundational questions that lie at the intersection of learning-based and classical approaches. William T. Freeman (Massachusetts Institute of Technology) identified the features of the human visual system that are most critical for replication in an artificial neural vision system; he connected these features with a range of examples and applications. On the computational side, Yuejie Chi (Carnegie Mellon University) bridged the gap between theory and practice in nonconvex approaches for the solution of low-rank matrix estimation problems, which are foundational in many machine learning and classical scenarios. Her talk addressed gradient descent-type algorithms with guarantees for computational complexity, statistical performance, and robustness properties while also emphasizing the need for more unified theory.

A collection of minitutorials by Daniel Cremers (Technical University of Munich), Michael Moeller (University of Siegen), Jeffrey Fessler (University of Michigan), and Peyman Milanfar (Google Research) all continued to build the bridge between classical and data-driven approaches by tackling applications in image restoration, medical image reconstruction, and computational photography. Minisymposia talks reiterated these themes and intertwined the classics with this new DL paradigm. The aforementioned examples provide just a snapshot of DL's well-represented influence within the field of imaging science.

While many theoreticians initially doubted DL, this new paradigm no longer seems offensive — so long as scientists handle it thoughtfully. Analogues and connections to the classical body of imaging literature—ranging from vision modeling to informed DL architectures—are rich and growing. Such relationships lead to provable guarantees, as well as efficient and well-motivated optimization tools that are critical to network training. They are also unveiling connections that allow seemingly "black box solutions" to become more akin to "illuminating approaches."

Our community's perspective seems less like "build it deeper and see what happens" and more like "build it carefully and seek a balance between performance, mathematical foundations, and insight." It is impossible to ignore DL's potential, nor should we. But we are realizing that the classical knowledge and know-how in image processing and computer vision will play a central role in paving the way towards next-generation practice and understanding of DL solutions.

In light of these realizations, one might wonder whether DL has to be involved in every imaging science advancement. We do not believe that this is the case. Indeed, our community is currently making important advancements in various directions with purely classical approaches. New theoretical results in optimization, optimal transport, wave equations, harmonic analysis, variational methods, PDEs, patch-based methods, sparse representations, and other areas continue to impact the field in important ways. In fact, the imaging science community's commitment to the classics—both within and outside the DL regime—is allowing DL to take its proper place in a productive and contextualized manner and is foundational to the field overall.

In summary, the imaging science research community is pursuing its own take on DL. It is a new playground, but we are utilizing our vast arsenal of classical skills so that we do not tackle each piece of equipment as if we have never seen it before. We are treating these new tools like playdough—molding architectures to complement our wealth of knowledge—while using DL to shape and evolve the classics, ultimately enabling the creation of things we never thought possible.

Stacey Levine and Michael Elad were the organizing committee co-chairs for the 2020 SIAM Conference on Imaging Science. Stacey Levine is a professor in the Department of Mathematics and Computer Science at Duquesne University. Michael Elad is a professor in the Computer Science Department at the Technion – Israel Institute of Technology. He is editor-in-chief of the SIAM Journal on Imaging Sciences.

[1] https://www.siam.org/conferences/cm/conference/is20
[2] https://go.siam.org/kAl5sa
[3] See page 7 for an article by Gabriel Peyré that is based on his invited talk.

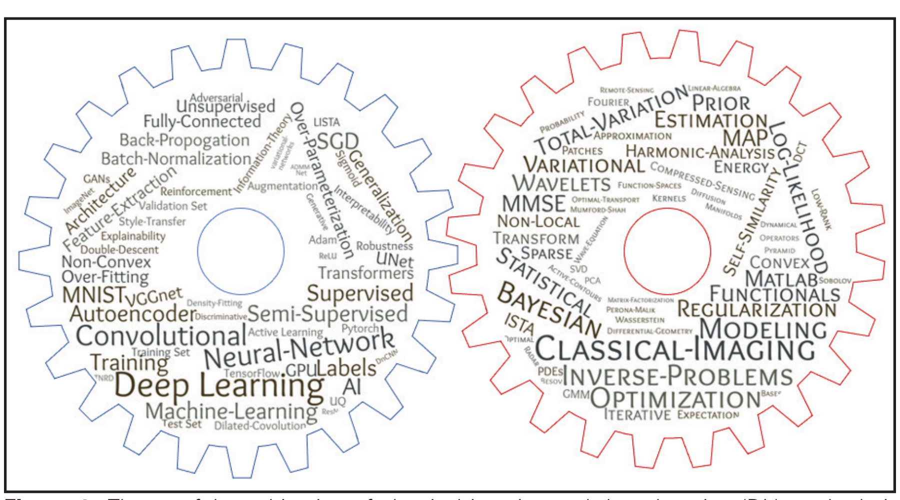


Figure 1. *The careful combination of classical imaging and deep learning (DL) methods is inspiring important connections, questions, and complementary approaches. Image courtesy of Stacey Levine and Michael Elad.*

Healthcare Industry

Continued from page 9

write reports — all of which help us explain findings to our clients and incorporate feedback. Directors normally give the project teams full responsibility for developing recommendations and executing projects on a day-to-day basis, which makes my job both challenging and rewarding. Some of my responsibilities include preparing the necessary coding for a program to run an analysis and meeting with project leaders to review insights from the analysis and assess client suggestions. We then conduct a more detailed review once the analysis is complete and the final document is drafted.

It is common for someone with a background in academia to need some time to adjust to the particular demands of industry work, especially in terms of client communication. It can be difficult to adapt to the comprehension level of clients; applied mathematicians tend to have a very strong analytical and theoretical knowledge base, whereas our clients have a much broader, practical background. When things go as planned, we clock 40 to 50 hours a week, but the occasional time crunch from a client can easily result in 12-hour days. PRECISIONheor is strongly committed to avoiding long hours for its employees, though compliance is largely the responsibility of the employees themselves. This type of time management requires careful planning and occasionally standing our ground with clients to refrain from committing to unreasonable deliverables.

Although I have transitioned to industry and greatly enjoy my work with PRECISIONheor, I continue to remain active in professional academic networks.[5] I am widely involved in many communities, such as journal editorial boards, webinar committees,[6] and research organizations.[7] Maintaining a connection to my academic networks has created a support system that allows me to continually grow and learn while simultaneously providing me with opportunities to teach, mentor, and help those who are just starting their careers. Retaining my excitement for learning new concepts and methods likely contributed to my smooth transition, which has been an absolute pleasure.

Anuj Mubayi[8,9] is an associate director in PRECISIONheor's Advanced Modeling Group. He is also an instructional assistant professor in the Department of Mathematics at Illinois State University (ISU), an adjunct faculty member in the College of Health Solutions at Arizona State University, and an associate research scientist at the Prevention Research Center in Berkeley, Calif. In addition, Mubayi serves as a research mentor at the Center for Collaborative Studies in Mathematical Biology, which is part of ISU's Intercollegiate Biomathematics Alliance.[10] He is an applied and computational mathematical scientist with more than 10 years of experience working on modeling problems that are of interest to public health communities, such as the design and evaluation of cost-effective intervention programs in the healthcare sector and the study of ecology and population dynamics of disease transmission.

[5] https://www.youtube.com/channel/UCuhaV3SmDLMmWsGFJJHtXiA
[6] https://sites.google.com/view/bereedms-series/home
[7] https://www.cs.fsu.edu/vipra
[8] https://www.anujmubayi.com
[9] https://www.linkedin.com/in/anujmubayi
[10] https://about.illinoisstate.edu/iba