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

siam news

Volume 52/ Issue 4
May 2019

Special Issue on Computational Science and Engineering

With almost 1,900 attendees, the 2019 SIAM Conference on Computational Science and Engineering was SIAM's largest meeting to date! This ***special issue*** *offers coverage of conference minisymposia, invited talks, prizes, and panel discussions, as well as articles on other CSE-related topics.*

Attendees of the 2019 SIAM Conference on Computational Science and Engineering, which took place earlier this year in Spokane, Wash., network and mingle between sessions. SIAM photo.

Scientific Computing, Machine Learning, and Data Science: Recurring Themes at CSE19

By Paul Davis

To its nearly 2,000 attendees, the 2019 SIAM Conference on Computational Science and Engineering (CSE19), which took place from February 25-March 1 in Spokane, Wash., may have seemed more like a lively music festival than a specialized scientific meeting. On the central stage, prize lecturers reprised beloved favorites while up-and-comers debuted newer hits. A host of minisymposia on the numerous smaller platforms drew their own devoted audiences. The overall effect was an engaging mix of robust scientific computing, innovative machine learning, and insightful data science. Old barriers fell while new ones were scaled.

Jack Dongarra, recipient of the SIAM/ACM Prize in Computational Science and Engineering, delivered a prize lecture that offered a numerical nostalgia (if not magical mystery!) tour of the singular value decomposition (SVD), which has been the "working horse" of linear algebra for nearly half a century. The importance of efficient SVD algorithms is difficult to overstate. For example, SVD is a key tool for constructing reduced-order models — the subject of more than 20 minisymposia at this meeting alone.

Dongarra's tour began with EISPACK and its circa 1970 element-wise Fortran operations. It then proceeded through various versions of LAPACK and increasingly complex arrangements of Basic Linear Algebra Subprograms (BLAS) before concluding with the latest and greatest multiprocessors. Along the way, many attendees likely heard at least one tune that they knew. Most could probably recall the first time they had "played" that tune, as well as the machine they used then.

Dongarra, who holds appointments at the University of Tennessee and Oak Ridge National Laboratory, made his point convincingly: a steady stream of software hits arises from an artful coupling of the right mathematical formulation with algorithms and software shaped to exploit evolving computational architectures [1].

A segue from Dongarra's computational retrospective could have led to the two-part computational environment spawned partly by his and his colleagues' work: one class

See **Themes at CSE19** *on page 2*

Training Quantized Deep Neural Networks and Applications with Blended Coarse Gradient Descent

By Jack Xin

In recent years, deep neural networks (DNNs) have seen enormous success in big data-driven applications such as image and speech classification, natural language processing, and health sciences [5, 11]. However, DNNs typically require hundreds of megabytes of memory and billions of flops per inference, making them challenging to deploy in resource-limited environments. Researchers have lately been devoting considerable efforts to the development of low-precision networks for memory and computational savings that nearly maintain full-precision network performance. *Quantization*—an efficient means of producing low-precision DNNs—gives rise to interesting mathematical issues.

DNNs act on an input vector v by repeated composition, written as $F(v)=A_L(F_{L-1}(\cdots F_2(F_1(v))))+b_L$, where $F_l(\cdot)=\max(A_l\cdot+b_l,0)$ [7]; the nonlinearity is the so-called ReLU function, A_l represents the weight matrices, and b_l refers to the bias vectors. For image and sound data, A_l filters and extracts multiscale features from low to high levels across hidden layers ($l\le L-1$). The output $F(v)$ is normalized into class probabilities for a classification task [11]. These DNNs are termed convolutional neural networks (CNNs) because A_l performs convolutions [5, 11]. Given labeled samples $(v^{(n)},u^{(n)})$, network training minimizes the empirical risk: $N^{-1}\sum_{n=1}^{N}\ell(F(v^{(n)});u^{(n)})$. Here $\ell(\cdot;u)$ is a smooth function that measures the discrepancy between the network-predicted label from input $v^{(n)}$ and the true label $u^{(n)}$. Use of stochastic gradient descent (SGD)—a variant of the gradient descent method based on small batches of a large training data set—achieves the minimization [5, 11]. For simplicity, we ignore the forthcoming bias vectors.

Quantization refers to replacement of the ReLU function $\max(\cdot,0)$ with a piecewise constant function and restriction of weight values to a discrete set. In activation quantization, the quantized network function is $F_Q(v):=A_L\sigma(A_{L-1}\cdots A_2\sigma(A_1v,\alpha_1),\cdots,\alpha_{L-1})$. The lth quantized ReLU $\sigma(x_l,\alpha_l)$ acts element-wise on vector x_l from a previous layer and is parameterized by trainable scalar $\alpha_l>0$. In uniform quantization,

$$\sigma(x,\alpha)=\begin{cases}0, & \text{if}\quad x\le 0,\\ k\alpha, & \text{if}\quad (k-1)\alpha<x\\ & \le k\alpha,\ k=1,2,\ldots,\\ & 2^{b_a}-1,\\ (2^{b_a}-1)\alpha, & \text{if}\quad x>(2^{b_a}-1)\alpha,\end{cases}\tag{1}$$

where x is the scalar input, b_a is the bit-width, and k is the quantization level. For a 4-bit quantization, $b_a=4$ and $2^{b_a}=16$ levels exist, including zero.

The empirical risk minimization problem is non-convex, piecewise constant, and high dimensional for σ in (1). In SGD, the gradients of empirical risk function in $A_l(l\le L-1)$ and $\alpha_l(l\le L-2)$ are almost always zero, or simply zero under

See **Deep Neural Networks** *on page 3*

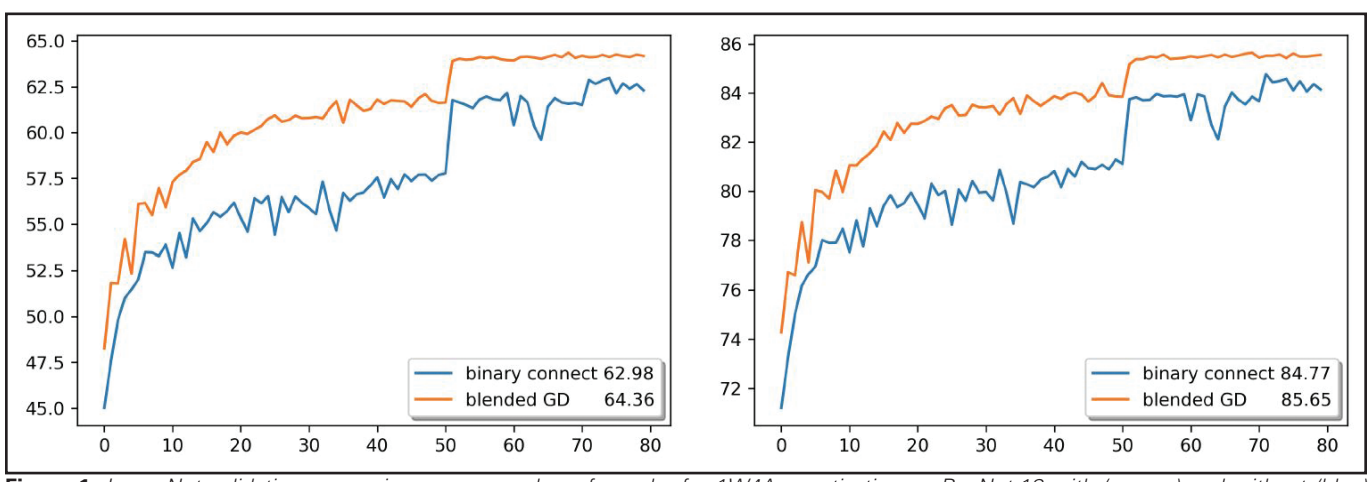


Figure 1. *ImageNet validation accuracies versus number of epochs for 1W4A quantization on ResNet-18 with (orange) and without (blue) blending for 3-valued σ_α. Image courtesy of [8].*

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

siam news
Volume 52/ Issue 4/ May 2019

Themes at CSE19

Continued from page 1

of languages for algorithmic prototypes and another for highly efficient production software. Jeffrey Bezanson, Stefan Karpinski, and Viral Shah of Julia Computing, Inc. received the James H. Wilkinson Prize for Numerical Software for their success in harmonizing the two. Their corresponding talk surveyed some of the ways in which Julia solves that two-language problem—one for prototyping, one for production—in scientific computing and machine learning [2].[1]

Another player might have turned from Dongarra's study of matrices to the potential role of tensors—matrices' higher-order cousins—in machine learning. During her invited lecture, Anima Anandkumar of NVIDIA and the California Institute of Technology did exactly that. Instead of the pairwise correlations that matrices capture, tensors can encode third-order correlations — to detect topics in texts through the co-occurrence of word triplets, for instance. Anandkumar observed that BLAS has successively evolved from Level I vector-vector computations to Level II matrix-vector and finally Level III matrix-matrix computations. Might we now, she conjectured, expect BLAS Level IV tensor-tensor computations?

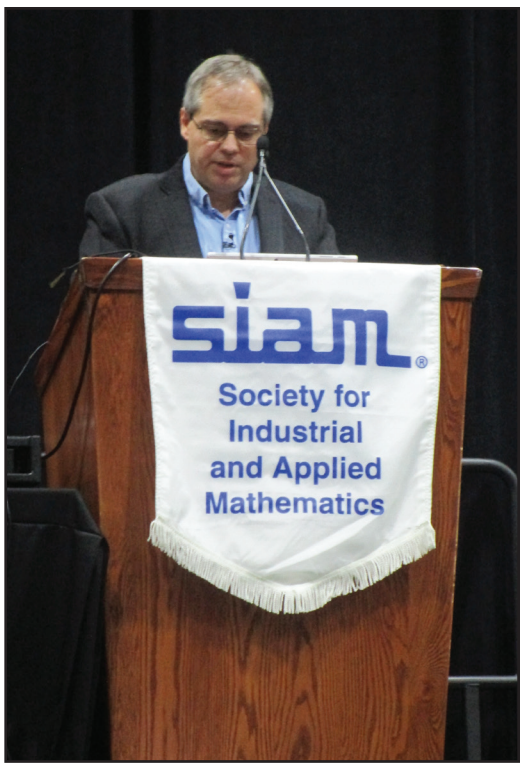
Nitty-gritty computational modeling was well-represented on the main stage. Michael Ferris of the University of Wisconsin, Madison described how stochastic optimization illuminates the policy decisions—and sometimes the unexpected consequences—of New Zealand's bold commitment to 100 percent renewable electrical energy. As another example, Boyce Griffith of the University of North Carolina at Chapel Hill explored methods, models, and applications for fluid-structure interactions in medicine and biology.

On a vastly different scale, Alistair Adcroft of Princeton University and the U.S. National Oceanic and Atmospheric Administration outlined a critical need to improve studies of the role of the world's oceans in climate warming. They absorb and circulate significant amounts of excess solar heat, but current practical ocean models cannot resolve the dynamic details of convecting eddies with sufficient resolution.

Steven Brunton of the University of Washington addressed data-driven discovery of underlying physical laws in both a minisymposium presentation and his SIAG/CSE Early Career Prize lecture. He suggested that the challenge of learning physics from data—a recurring theme at CSE19—draws upon many aspects of CSE, including data science, machine learning, computational modeling, high-performance computing, and optimization. Brunton and his colleagues judiciously choose coordinates and measurements to derive interpretable and generalizable models by simultaneously identifying a model's structure and its parameters' values.

Opening a minisymposium on scientific machine learning, Nathan Baker of Pacific Northwest National Laboratory summarized a workshop on physics-based machine learning that was conducted through the Department of Energy's Advanced Scientific Computing Research program. The workshop aimed to extract from the larger mass of machine learning challenges those that are specific to science and engineering — or as Baker aptly put it, to distinguish the problems of managing "the electric grid from targeting ads for diapers." For instance, many machine learning methods lack the mathematical foundation necessary for understanding their robustness and sensitivity. Those matters become more critical when they guide higher-stakes decisions.

Risk and uncertainty quantification constitute yet another technically difficult but vitally important genre within CSE, especially when it comes to choosing models and methods. Tobin Isaac of the Georgia Institute of Technology, a rising young performer speaking from the main stage, addressed these problems in his SIAG/CSE Best Paper Prize lecture. He described his and his colleagues' careful identification of the methodological components needed to model uncertainty propagation in problems like predicting the flow of the Antarctic ice sheet [3].



Alistair Adcroft of Princeton University and the U.S. National Oceanic and Atmospheric Administration speaks about ocean modeling during the 2019 SIAM Conference on Computational Science and Engineering, held in Spokane, Wash., earlier this year. SIAM photo.

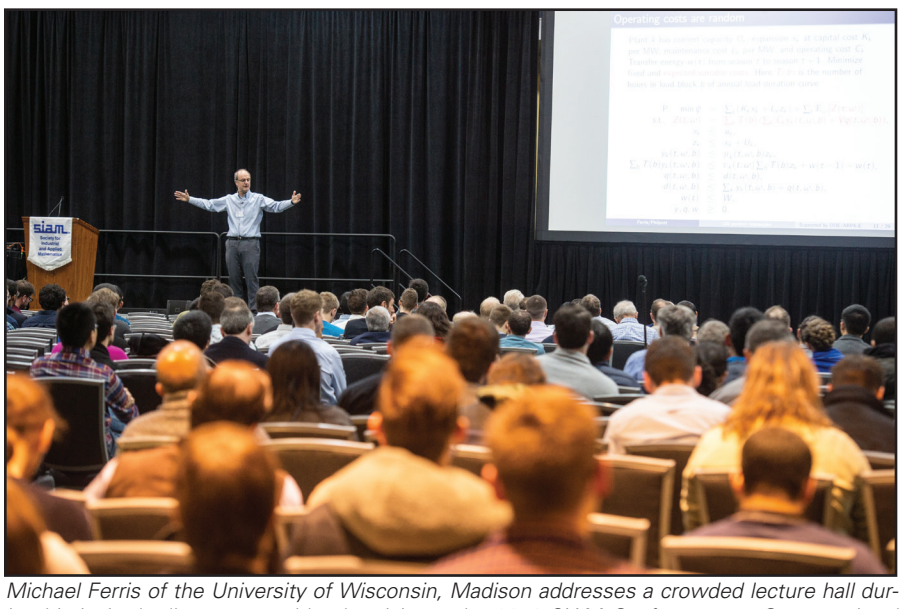
The data sciences were no less ubiquitous. For example, Deanna Needell of the University of California, Los Angeles and Giseon Heo of the University of Alberta worked with the Association for Women in Mathematics to organize a two-part minisymposium on data science. The research reported during these sessions ranged from detecting data anomalies to automating the diagnosis of sleep apnea. The eight papers in the minisymposium were coauthored by a variety of teams that involved 16 distinct individuals in total. The teams began their collaborations in 2017 during a one-week summer research workshop at Brown University's Institute for Computational and Experimental Research in Mathematics. That week was certainly effective, as those 16 collaborators represented almost one-third of workshop participants!

Other minisymposia talks addressed some wide-ranging data science problems. Two examples are Microsoft's efforts to automate farmers' management of their fields and Vanguard's studies of online financial advice driven by artificial intelligence.

No SIAM meeting would be complete without formal and informal attention to the experiences of individual members. SIAM President Lisa Fauci organized and moderated a panel on recruiting strategies for diversity and inclusion.[2] A minisymposium highlighted the exciting work of younger underrepresented mathematicians.[3] And multiple panels explored workplace issues and experiences in the data sciences.

Be sure to check out future issues of *SIAM News* for more detailed coverage of these and other hits from CSE19.

[1] Bezanson, Karpinski, and Shah acknowledged their fourth collaborator, Alan Edelman of the Massachusetts Institute of Technology.



Michael Ferris of the University of Wisconsin, Madison addresses a crowded lecture hall during his invited talk on renewable electricity at the 2019 SIAM Conference on Computational Science and Engineering, which took place earlier this year in Spokane, Wash. SIAM photo.

References

[1] Dongarra, J., Gates, M., Haidar, A., Kurzak, J., Luszczek, P., Tomov, S., & Yamazaki, I. (2018). The Singular Value Decomposition: Anatomy of Optimizing an Algorithm for Extreme Scale. *SIAM Rev., 60*(4), 808-865.

[2] Edelman, A. (2016, March). Julia: A Fast Language for Numerical Computing. *SIAM News, 49*(2), p. 5.

[3] Isaac, T., Petra, N., Stadler, G., & Ghattas, O. (2015). Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet. *J. Comput. Phys., 296*, 348-368.

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

[2] See "CSE19 Panel Promotes Strategies to Increase Diversity and Inclusion in Academia" on page 6.

[3] See "Oscillating Wait Times and Queues with Information Updates" on page 8.

siam news

ISSN 1557–9573. To be published 10 times in 2019: January/February, March, April, May, June, July/August, September, October, November, and December.

Deep Neural Networks

Continued from page 1

auto-differentiation for a deep learning platform; this presents an immediate difficulty. The SGD is stuck with *vanishing gradients*, though descent directions may still exist. A familiar landscape of this kind is a piecewise-constant spiral staircase that contains a descent direction along its envelope. This large-scale view motivates a descent direction via the notion of coarse gradient [8]. The modified chain rule yields the course gradient *wherein the partial derivative σ_α is replaced by $\tilde{\sigma}_\alpha$, with $\tilde{\sigma}$ as the ramp function or clipped ReLU:*

$$\tilde{\sigma}(x,\alpha)=\begin{cases}0, & \text{if} \quad x\le 0,\\ x, & \text{if} \quad 0<x\\ & \le (2^{b_a}-1)\alpha,\\ \left(2^{b_a}-1\right)\alpha, & \text{if} \quad x>\left(2^{b_a}-1\right)\alpha.\end{cases} \tag{2}$$

The partial derivative σ_α is piecewise constant with 2^{b_a} values. Numerical experiments on ImageNet [8] have confirmed the success of coarse gradient descent and revealed that an averaged version of σ_α with 3 or 2 values is a better choice. The 3-valued σ_α is

$$\frac{\partial\sigma}{\partial\alpha}(x,\alpha)\approx\begin{cases}0, & \text{if} \quad x\le 0,\\ 2^{(b_a-1)}, & \text{if} \quad 0<x\\ & \le \left(2^{b_a}-1\right)\alpha,\\ 2^{b_a}-1, & \text{if} \quad x>\left(2^{b_a}-1\right)\alpha.\end{cases} \tag{3}$$

The middle value 2^{b_a-1} in (3) is the arithmetic mean of the intermediate k values in (1). The 2-valued σ_α is equivalent to replacing σ_α with $\tilde{\sigma}_\alpha$, or zeroing out the middle value of (3).

The b_w-bit weight quantization restricts entries of A_l (total M in all layers) to the set $\mathbb{Q}=\mathbb{R}_+\times\{\pm1\}^M$ for $b_w=1$ and $\mathbb{Q}=\mathbb{R}_+\times\left\{0,\pm1,\ldots,\pm(2^{b_w-1}-1)\right\}^M$ for $b_w\ge 2$. Training of a fully-quantized network seeks A_ls and α_ls to minimize the empirical risk, subject to the $\mathbb{Q}$ constraint. The classical approach is projected gradient descent (PGD): $\mathbf{w}^{k+1}=\text{proj}_{\mathbb{Q}}(\mathbf{w}^k-\eta\tilde{\nabla}_{\mathbf{w}}f(\mathbf{w}^k))$, where $\eta>0$ is small, f is the objective function, and $\tilde{\nabla}_{\mathbf{w}}$ is a proper (coarse) gradient in weight $\mathbf{w}$. The $\text{proj}_{\mathbb{Q}}(\cdot)$ has affordable closed-form solutions at $b_w=1,2$ and approximate solutions at $b_w\ge 3$ [10]. However, a small variation of $\mathbf{w}^k$ may vanish under $\text{proj}_{\mathbb{Q}}$ due to $\mathbb{Q}$'s discreteness; this can cause stagnation. The BinaryConnect (BC) method [2] provides a cure with a modified update: $\mathbf{w}_f^{k+1}=\mathbf{w}_f^k-\eta\tilde{\nabla}_{\mathbf{w}}f(\mathbf{w}^k)$, $\mathbf{w}^{k+1}=\text{proj}_{\mathbb{Q}}(\mathbf{w}_f^{k+1})$, where the full-precision weight $\mathbf{w}_f^k$ continues to evolve and eventually moves the quantized weight $\mathbf{w}^k\in\mathbb{Q}$. However, the sufficient descent inequality might not hold for BC even when f has a Lipschitz gradient, i.e., $f(\mathbf{w}^{k+1})-f(\mathbf{w}^k)\le -c\left\|\mathbf{w}^{k+1}-\mathbf{w}^k\right\|^2$, for constant $c>0$ and small learning rate η. Without sufficient descent, $\{\mathbf{w}^k\}$ may not converge to a critical point even if $\{f(\mathbf{w}^k)\}$ does. Blending PGD and BC regains the sufficient descent inequality and generates the following blended coarse gradient decent (BCGD) [8]:

$$\mathbf{w}_f^{k+1}=(1-\rho)\,\mathbf{w}_f^k+\rho\,\mathbf{w}^k-\eta\tilde{\nabla}_{\mathbf{w}}f(\mathbf{w}^k),$$
$$\mathbf{w}^{k+1}=\text{proj}_{\mathbb{Q}}(\mathbf{w}_f^{k+1}), \tag{4}$$

with a blending parameter $0<\rho\ll 1$. At $\rho=10^{-5}$, Figure 1 (on page 1) depicts validation accuracies of ResNet-18 on ImageNet at $(b_w,b_a)=(1,4)$, or 1W4A versus training epochs with and without blending via a 3-valued σ_α. The accuracies noticeably improve with blending. Table 1 lists top-1 and top-5 validation accuracies of full-precision ResNet-18 (32W32A) versus low-precision models trained by BCGD with 3- and 2-valued σ_α. The 3-valued σ_α performs better at a lower precision. At 4W8A, either choice is within one percent of the full-precision model.

Coarse gradients are non-unique. Use of derivatives of a proxy activation function (straight-through estimator) in gradient-based training dates back to the 1950s and the learning of single-layer networks. Researchers have proposed various estimators for multi-layer networks [1, 3, 4]; depending on the problem, some perform better than others [9]. Consider a neural network regression problem with binary activation $\sigma(x)=1_{\mathbb{R}^+}(x)$ and square loss function $\ell(v,w;Z):=\left(v^\top\sigma(Zw)-(v^*)^\top\sigma(Zw^*)\right)^2$, where $v^*\in\mathbb{R}^m$ and $w^*\in\mathbb{R}^n$ are the underlying (nonzero) teacher parameters, and the entries of data matrix $Z\in\mathbb{R}^{m\times n}$ are identically distributed unit normal.

We show that the coarse gradient (using the derivative of either ReLU or clipped ReLU) on the population loss function $E_Z\left[\ell(v,w;Z)\right]$ forms an acute angle with the underlying true gradient [8, 9]. The negative coarse gradient is a descent direction that minimizes population loss so the corresponding iterates converge to a critical point. If the initialization obeys certain geometric conditions, convergence to the global minimizer (v^*,w^*) holds. There is still much more to understand about coarse gradient descent.

We have applied quantized CNNs to object detection and keyword spotting problems [6, 10]. The keyword CNN classifies an audio clip as either silence, an unknown word, or a word on a short list. The CNN has two convolutional layers — one fully-connected layer followed by the output class probabilities. A binary weight CNN ($b_w=1$) maintains the full-precision network's accuracy (93 percent) within one percent. Both are imported to an app on an Android cellular phone for runtime comparison. In Figure 2, a user says "off" and the app correctly recognizes the word and reports a runtime in milliseconds. The binary model doubles the speed on this particular hardware.

The redundancies of DNNs thus allow the development of complexity reduction methods—such as quantization—with minimal performance loss. Coarse gradient descent is useful to other non-convex, nonsmooth optimization problems as well.

	Full Precision	1W4A		4W4A		4W8A	
		3 valued	2 valued	3 valued	2 valued	3 valued	2 valued
top-1	69.64	64.36	63.37	67.36	66.97	68.85	68.83
top-5	88.98	85.65	84.93	87.76	87.41	88.71	88.84

Table 1. *ImageNet validation accuracies (in percentage) with blended coarse gradient descent on ResNet-18 for 3- and 2-valued σ_α. Table courtesy of [8].*

References

[1] Cai, Z., He, X., Sun, J., & Vasconcelos, N. (2017). Deep Learning with Low Precision by Half-wave Gaussian Quantization. In *2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI.

[2] Courbariaux, M., Bengio Y., & David, J. (2015). BinaryConnect: Training Deep Neural Networks with Binary Weights during Propagations. In *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems — Volume 2* (pp. 3123-3131). Montreal, Canada.

[3] Hinton, G. (2012). Neural Networks for Machine Learning. In *Coursera video lectures*. Retrieved from https://www.youtube.com/playlist?list=PLoRl3Ht4JOcdU872GhiYWf6jwrk_SNhz9.

[4] Hubara, I., Courbariaux, M., Soudry, D., El-Yaniv, R., & Bengio, Y. (2018). Quantized Neural Networks: Training Neural Networks with Low Precision Weights and Activations. *J. Mach. Lear. Res., 18*, 1-30.

[5] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature, 512*, 436-444.

[6] Sheen, S., & Lyu, J. (2019). Median Binary-Connect Method and a Binary Weight Convolutional Neural Network for Word Recognition. In *2019 IEEE Data Compression Conference.* Snowbird, UT.

[7] Strang, G. (2018). The Functions of Deep Learning. *SIAM News, 51*(10), p. 1.

[8] Yin, P., Zhang, S., Lyu, J., Osher, S., Qi, Y., & Xin, J. (2019). Blended Coarse Gradient Descent for Full Quantization of Deep Neural Networks. *Res. Math. Sci., 6*(14). Preprint, *arXiv:1808.05240*.

[9] Yin, P., Zhang, S., Lyu, J., Osher, S., Qi, Y., & Xin, J. (2019). Understanding Straight-Through Estimator in Training Activation Quantized Neural Nets. *To be presented at the 2019 International Conference on Learning Representations.* New Orleans, LA.

[10] Yin, P., Zhang, S., Qi, Y., & Xin, J. (2019). Quantization and Training of Low Bit-Width Convolutional Neural Networks for Object Detection. *J. Comp. Math., 37*(3), 349-360. Preprint, *arXiv:1612.06052v2*.

[11] Yu, D., & Deng, L. (2015). *Automatic Speech Recognition: A Deep Learning Approach*. In *Signals and Communication Technology*. New York, NY: Springer.

Jack Xin is a professor of mathematics at the University of California, Irvine. His research interests are in the analysis and computation of multiscale and data science problems.

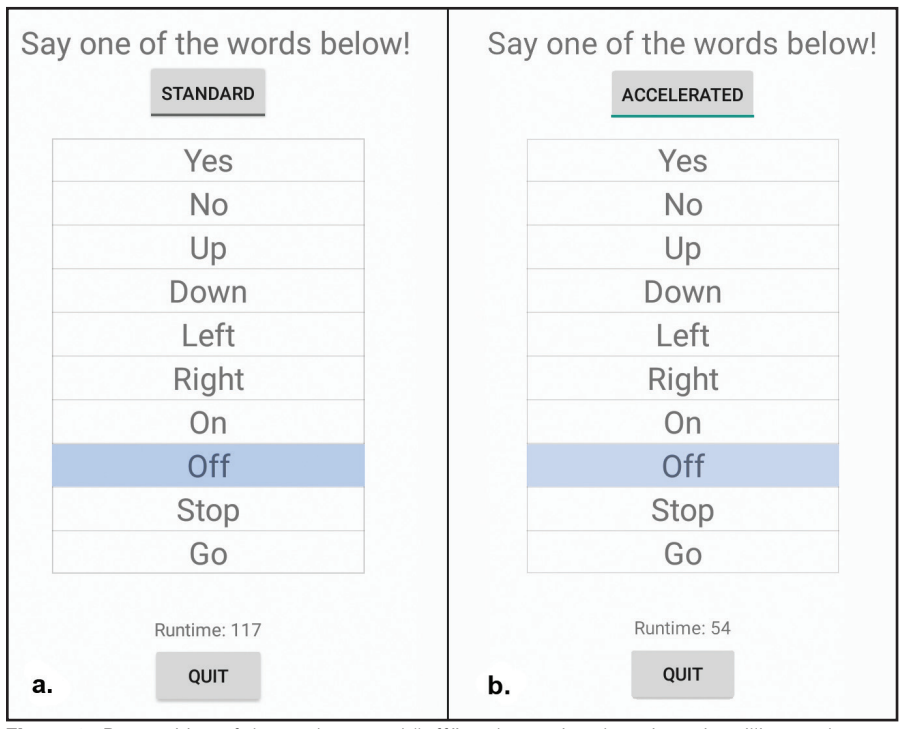


Figure 2. *Recognition of the spoken word "off" and associated runtimes in milliseconds on an Android app.* **2a.** *Full-precision convolutional neural network (CNN) model.* **2b.** *Binary weight CNN model with 2x speedup. Image courtesy of [6].*

Markdown: A Writing Tool for Every Applied Mathematician's Toolbox

By Roy H. Goodman

Much of an applied mathematician's working life is spent writing. Primary tasks of course involve drafting articles and grant proposals, and LaTeX remains the most important tool for this purpose. However, Markdown is now a better option for smaller writing projects, and *SIAM News* readers should be aware of its many capabilities.

Like LaTeX and HTML, Markdown is a markup language. A large ecosystem of applications allows documents written in Markdown to be exported to a variety of formats, including PDF and HTML. Designer John Gruber wanted plain-text Markdown documents to be human readable, without markup instructions overwhelming the text. When working with simple documents devoid of heavy mathematical typesetting, writing in Markdown is faster and easier than using LaTeX. The time investment required to learn the language is also much less than that of LaTeX, especially for experienced TeX users. Indeed, I wrote this article in Markdown.

LaTeX Sample

```
\section{Writing stuff in $\LaTeX$}
\subsection{Smaller Heading}

This is \emph{italics}. This is \textbf{bold}.

\begin{enumerate}
\item Newton
\begin{enumerate}
\item Apples
\item Iteration
\end{enumerate}
\item Gauss
\item Euler
\end{enumerate}

\begin{itemize}
\item Shapes
\begin{itemize}
\item Square
\item Circle
\end{itemize}
\item Colors
\begin{itemize}
\item Purple
\item Green
\end{itemize}
\end{itemize}

\begin{tabular}{|l | c | r|}
\hline
groceries & numbers & plants \\
\hline
eggs & 1 & flowers \\
bread & $e^{i\pi}$ & trees \\
milk & 100 & grass \\
\hline
\end{tabular}
```

Figure 1. *In a LaTeX document, it becomes difficult for users to see the content amid all of the markup text.*

A Brief History

In 2004, Gruber created Markdown with two components: the language itself and a Perl script to convert Markdown documents to valid HTML. His website[1] is one of the best sources of relevant information and features an introduction to the design philosophy, a description of the syntax, and an online sandbox in which users can experiment with the language and preview results.

Many variants or *flavors* of Markdown have since been created; these variations extend the language with additional features—such as tables, footnotes, and citations—at the cost of complexity and some human readability of plain-text documents. *Github Flavored Markdown*, utilized by the software-hosting and version-control site Github to render README files and other documentation, is the most widely-used type. Readers may have used it themselves when posting a question to a Stack Exchange forum or adding formatted text to a Jupyter Notebook.

[1] https://daringfireball.net/projects/markdown

The Markdown Language

Markdown is simple to both read and write, and has much less document overhead than LaTeX. A "Help" page with an overview of the syntax is available through any online editor, and a more complete description is accessible via the *MarkDown Guide*.[2] Figures 1 and 2 depict a basic comparison for headings, itemized and enumerated lists, and tables for the formatted output of Figure 3.

SOFTWARE AND PROGRAMMING

Software

As with LaTeX, Markdown requires a plain-text editor and a program that converts the text to either HTML or another format. While any text editor works sufficiently, several have modes or extensions that perform syntax highlighting and live preview of the formatted text. I am partial to the Atom text editor with the "Markdown Writer" and "Markdown Preview Enhanced" extensions.[3] Users should also know about Pandoc, a document format converter that can convert between multiple markup formats, including PDF, HTML, Markdown, LaTeX, and—to some extent—Microsoft Word. One can use this to output Markdown text to multiple formats.

Trying one of the many online editors/previewers is the easiest way to begin. Most can also export documents to HTML or PDF. Like Google Docs, HackMD[4] allows multiple authors to log in and simultaneously edit the same file, keep track of projects, and share notes containing integrated mathematics with collaborators and students.

My Experience

In the past few years, I have used Markdown for a variety of documents that require some—but not much—mathematical typesetting, especially for online use. For example, I employ it to create syllabi and problem sets for classes that I teach. I also utilize it to respond to referees. These responses typically contain minimal mathematical content, and the live preview makes writing significantly easier. I have previously tried to keep a research diary using Evernote, but was frustrated by its inability to display equations. I now use an inexpensive alternative called Quiver that works with Markdown and LaTeX, which solves this problem.

[2] https://www.markdownguide.org
[3] https://atom.io
[4] https://hackmd.io

Markdown has impacted my workflow most substantially in the management of numerical simulations. When I run simulations, I tend to run the same code repeatedly with different parameters; this necessitates a way to store and—more importantly—retrieve the results when writing a paper. A MATLAB program called cell2md allows me to export a Markdown table that lists the parameters, date, and time of each run, along with links to image files and exported data. I have written a program to create a unique directory for the data and add a new line to the Markdown-based index every time I run the code. The whole index is visible in the open-source viewer/editor MacDown, thus providing an easy interface to my runs (see Figure 4).

Markdown sample

```
# Writing stuff in Markdown

## Smaller Heading
This is *italics*. This is **bold**.
1. Newton
   1. Apples
   2. Iteration
1. Gauss
1. Euler

* Shapes
  * Square
  * Circle
* Colors
  * Purple
  * Green

| groceries | numbers | plants |
|:----------|:-------:|-------:|
| eggs      | 1       | flowers |
| bread     | $e^{i\pi}$ | trees |
| milk      | 100     | grass |
```

Figure 2. *When written in Markdown, the content from Figure 1 is easy to read.*

Limitations

By design, Markdown has a limited feature set. It is currently not powerful enough to replace LaTeX for the purpose of writing articles, though several projects are trying to overcome this deficiency. While extensions exist for equation numbering, bibliography management, and BibTeX integration, such extensions are not standardized; this makes them inconsistent when exporting between different formats. I have found several pages whose authors describe their personal workflow for writing scientific articles in Markdown, but the documentation and tools that would allow an average user to begin adopting Markdown for this purpose simply do not exist.

The Atom editor can generate a live preview of equations produced with a large subset of LaTeX commands. However, its failure to read a document preamble makes it incapable of live-typesetting commands from additional packages or user-defined macros.

I would love to create slides in Markdown and have tried several Markdown-based presentation programs. All seem adequate for very basic presentations, but I have not yet found one that can handle the level of typesetting and structuring necessary for a mathematical presentation.

Parting Words

In closing, do you have a homepage written by hand in Emacs or via Mozilla Composer (which was last updated in 2005, and looks like it)? Use Pandoc to export the page to Markdown and take 20 minutes to rewrite it. If you want a more full-featured homepage, try Hugo,[5] which also uses Markdown. While Markdown cannot replace LaTeX for all of your writing needs, it is quick to learn and will make your life easier and your web pages more readable and professional.

Roy H. Goodman is an associate professor in the Department of Mathematical Sciences at the New Jersey Institute of Technology. He is a member of the SIAM Activity Group on Dynamical Systems and the Activity Group on Nonlinear Waves and Coherent Structures.

[5] https://gohugo.io

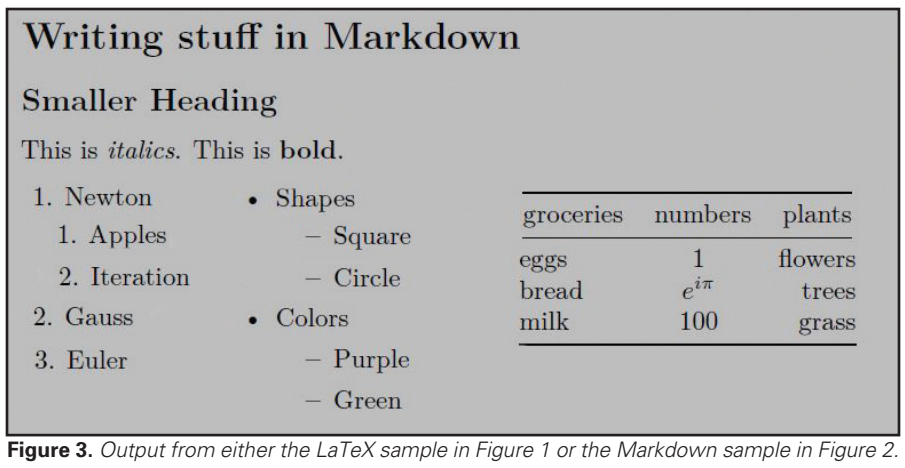


Figure 3. *Output from either the LaTeX sample in Figure 1 or the Markdown sample in Figure 2.*

```
### Stability calculation for QG problem

|Folder | L |nx |$\epsilon$|$t_{final}$|Date |Image |
Comment/*after* |
|:--:|:--:|:--:|:--:|:--:|:--:|:--:|:--:|:--:|:--:|
|[001](001/indexNLStability.html) |10 |100| 0.1 |20 |
08/16/18 19:32|[pcolor](001/indexNLStability_02.png)
[initial/final](001/indexNLStability_04.png) [movie]
(001/animatedStability.mp4)|Trying to demonstrate
dynamics of unstable bound state. *Under-resolved in
space.* |
|[002](002/indexNLStability.html) |10 |200| 0.1 |20 |
08/16/18 22:21|[pcolor](002/indexNLStability_02.png)
[initial/final](002/indexNLStability_04.png) [movie]
(002/animatedStability.mp4)|Trying again with more
points. *Better resolution but slow.* |
|[003](003/indexNLStability.html) |10 |200|0.25 |20 |
08/16/18 22:49|[pcolor](003/indexNLStability_02.png)
[initial/final](003/indexNLStability_04.png) [movie]
(003/animatedStability.mp4)|Note larger $\epsilon$.
*Better. Use this one in the paper!* |
```

Stability calculation for QG problem

Folder	L	nx	ϵ	t_{final}	Date	Image	Comment/*after*
001	10	100	0.1	20	08/16/18 19:32	pcolor initial/final movie	Trying to demonstrate dynamics of unstable bound state. *Under-resolved in space.*
002	10	200	0.1	20	08/16/18 22:21	pcolor initial/final movie	Trying again with more points. *Better resolution but slow.*
003	10	200	0.25	20	08/16/18 22:49	pcolor initial/final movie	Note larger ϵ. *Better. Use this one in the paper!*

Figure 4. *Table of runs, with Markdown on the left and details on the right.*

2019 SIAM Fellows

SIAM is pleased to announce the newly selected Class of SIAM Fellows—a group of distinguished members of SIAM who were nominated by their peers for exceptional contributions to the fields of applied mathematics and computational science. Please join us in congratulating these 28 members of our community.

siam Fellow

Mihai Anitescu
Argonne National Laboratory & University of Chicago

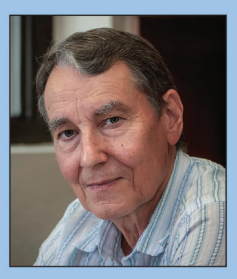
David A. Bader
Georgia Institute of Technology

Francesco Bullo
University of California Santa Barbara

Jose Antonio Carrillo de la Plata
Imperial College London

Stephen Jonathan Chapman
University of Oxford

Pierre Comon
CNRS



Wolfgang A. Dahmen
University of South Carolina Columbia

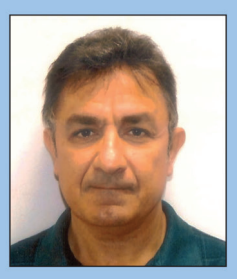
Jesus Antonio De Loera
University of California Davis

Froilan Dopico
Universidad Carlos III de Madrid

Ernesto Estrada
University of Zaragoza & ARAID Foundation

Fariba Fahroo
AFRL/Air Force Office of Scientific Research

Andreas Frommer
Universitat Wuppertal



Roger Ghanem
University of Southern California

Sigal Gottlieb
University of Massachusetts Dartmouth

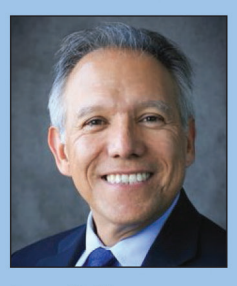
Michael Allen Heroux
Sandia National Laboratories & St. John's University

Misha Kilmer
Tufts University

Ron Kimmel
Technion - Israel Institute of Technology & Intel Corp.

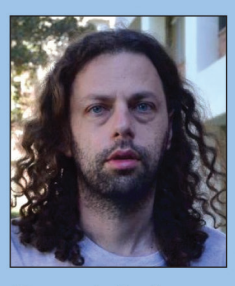
Gitta Kutyniok
Technische Universitat Berlin

Irena Lasiecka
University of Memphis

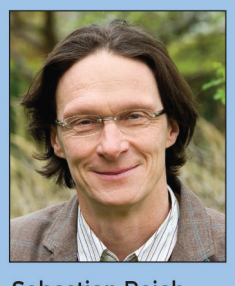


Juan C. Meza
University of California Merced

Jill C. Pipher
Brown University



Mason A. Porter
University of California Los Angeles



Sebastian Reich
Universitaet Potsdam

Carla D. Savage
North Carolina State University

Zuowei Shen
National University of Singapore

Joel A. Tropp
California Institute of Technology

Yin Zhang
Rice University and The Chinese University of Hong Kong, Shenzhen

Jun Zou
The Chinese University of Hong Kong

www.siam.org/Prizes-Recognition/Fellows-Program

Nominate a colleague by October 16, 2019 for the 2020 Class of Fellows.

siam | Society for Industrial and Applied Mathematics

CSE19 Panel Promotes Strategies to Increase Diversity and Inclusion in Academia

By Lina Sorg

Diverse representation within the fields of science, technology, engineering, and mathematics is a highly-charged issue that inspires much debate. Academics regularly advocate in favor of equal opportunities for underrepresented minorities, but total parity remains a work-in-progress. During a panel discussion at the 2019 SIAM Conference on Computational Science and Engineering, which took place earlier this year in Spokane, Wash., Carlos Castillo-Chavez (Arizona State University), Rachel Kuske (Georgia Institute of Technology), Eve Riskin (University of Washington), and Jamol Pender (Cornell University)[1] addressed the topic of diversity from both the student and faculty perspective. The panelists shared firsthand experiences, offered practical suggestions, and proposed various strategies that promote equality and inclusion for minority mathematicians.

While higher institutions continuously work to acknowledge and prevent discrimination, disheartening biases persist. "This problem of making one feel inadequate and incompetent is still very pervasive in many math departments," Castillo-Chavez said. "There's usually a few faculty members that make life very complicated and discourage women or minorities." As a faculty member, educating oneself about such biases—whether involuntary or deliberate—is a valuable first step towards their ultimate eradication.

When hiring new employees or recruiting graduate students, Riskin encouraged existing staff to keep the applicant search broad and open. "Recognize that underrepresented candidates are subject to different expectations," she said. "The bar automatically goes up and people scrutinize more carefully in ways they wouldn't otherwise." Collecting data on the candidate pool is also helpful. For example, acknowledging that females comprise 20 percent of applicants for an open position prevents employers from claiming that no qualified women exist. "We're not saying you should hire someone because they are a woman," Riskin continued, "we're saying you shouldn't *not* hire someone because they are a woman."

Kuske spoke briefly about her time as senior advisor to the provost at the University of British Columbia, during which multiple departments focused on embedding diversity as a strategy for excellence. This objective changed recruitment processes and encouraged committees to design procedures that targeted diversity in a professional manner. Kuske listed active communication, diverse committee composition, procedural consistency, and confidentiality as effective ways to avoid discrimination when evaluating applicants. "You need to take the time, plan, and give yourself the chance to do this purposeful search, collection process, and recruitment," she said. "You also want to avoid constructed criteria, wherein people make up criteria as they go rather than have a predetermined set of standards." For instance, underrepresented students often have less direct career paths; limiting oneself to candidates that adhere to a so-called "normal" trajectory is both an unconscious bias and a disservice to the applicants.

Pender encouraged faculty members to reflect on things they already do, such as delivering invited talks and attending conferences, and consider how they can augment these activities to practice inclusiveness and reach a larger audience. For example, he asks to meet with minority-serving student groups—like the National Society of Black Engineers or the Society of Hispanic Professional Engineers—whenever he speaks at a university. "This is a recruiting tool," Pender said. "You can meet a wide variety of students who you're trying to target for your graduate programs." Additionally, he visits historically black colleges and universities and Hispanic-serving institutions whenever possible to interact with and encourage underrepresented populations. If Pender is invited to give a talk at Georgia Tech, he also goes to Morehouse College. When traveling to Texas A&M University, he stops at Prairie View A&M University. Students see themselves in the speakers, he said, so it is helpful for them to interact with people who look like they do.

Pender also divulged a more casual strategy for connecting with pupils from his classes — he invites them to dinner at the dining hall. "They're already paying to go to the dining hall, so I just show up and eat with some of the students there," he said. "You have to eat, so this is a good way to engage with students around you and understand their perspective on how things are going."

Conversation then shifted from the viewpoint of faculty to that of the underrepresented students themselves. When applying to graduate programs, it is sometimes difficult for minority students to identify universities and educators that will actively help them succeed. "Talk to faculty, visit, talk to grad students," Castillo-Chavez said. "Look at the records, as records are very telling. At the end of the day you have to feel comfortable."

Pender reflected on his own experiences when deciding whether to pursue a Ph.D. at Georgia Tech or Princeton University. He visited both campuses and did not see many people who looked like him at the former — a factor that contributed to his decision to attend Princeton. His future advisor also made it clear that he wanted Pender to succeed, and brought a former Ph.D. student to meet him on his visitation day. "That was a useful thing for me to see, that I could be just like this person and also be successful there," Pender said. Now a professor himself, he has spoken to several students who applied to Cornell because they saw his face on the webpage. "Having a more diverse faculty can attract some applicants that the school wouldn't have gotten otherwise," he added.

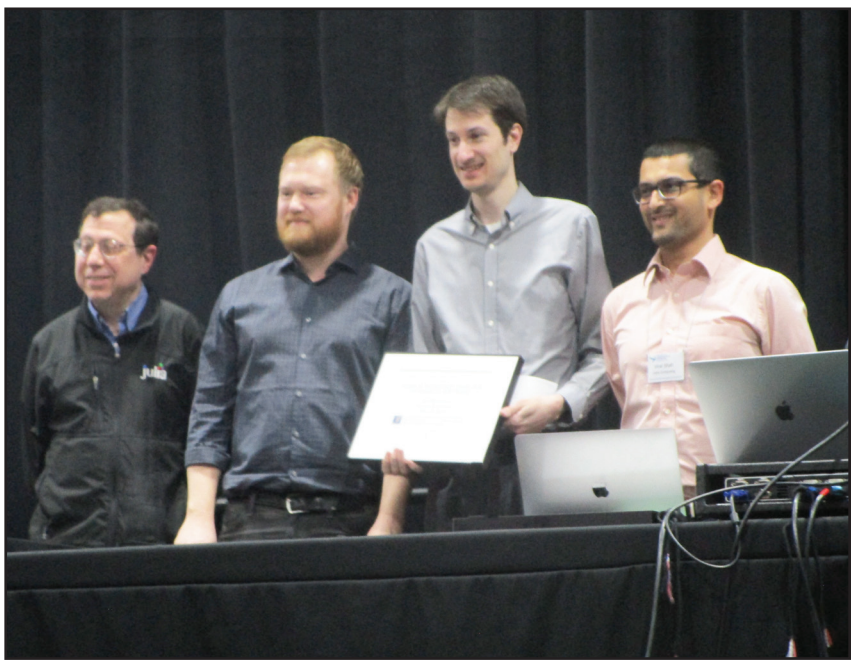
Kuske suggested that prospective students look for programs that include coursework flexibility, easy access to mentors, and opportunities for regular engagement with classmates via groups or organizations that build one's support system. "Those are simple things you can find on the websites or ask when you're visiting," she said. "It will come out fairly clearly whether there's a sense of community, or a sense of trying to find a flexible way that's not just one-size-fits-all so people can make it through."

See **Diversity and Inclusion** *on page 8*

[1] Read about Jamol Pender's minisymposium presentation on page 8.

Photo Highlights from CSE19 and GS19

Jack Dongarra of the University of Tennessee and Oak Ridge National Laboratory receives the SIAM/ACM Prize in Computational Science and Engineering at the 2019 SIAM Conference on Computational Science and Engineering, which took place earlier this year in Spokane, Wash. Dongarra's prize lecture was titled "The Singular Value Decomposition: Anatomy of an Algorithm, Optimizing for Performance." SIAM photo.



From left to right: Julia Computing, Inc. co-founder Alan Edelman of the Massachusetts Institute of Technology joins fellow co-founders Stefan Karpinski, Jeffrey Bezanson, and Viral Shah as the latter three receive the James H. Wilkinson Prize for Numerical Software at the 2019 SIAM Conference on Computational Science and Engineering, held in Spokane, Wash., earlier this year. Later that day, Karpinski, Bezanson, and Shah delivered a prize lecture entitled "Solving the Two Language Problem in Scientific Computing and Machine Learning with Julia." SIAM photo.

Olav Moyner of SINTEF Digital presents the SIAM Activity Group on Geosciences (SIAG/GS) Early Career Prize Lecture, entitled "Multiscale Simulation of Porous Media Flow: Obstacles, Opportunities, and Open-source," during the 2019 SIAM Conference on Mathematical & Computational Issues in the Geosciences, which took place in Houston, Texas, this March. SIAM photo.

SIAM Activity Group on Computational Science and Engineering (SIAG/CSE) chair Karen Devine (right) presents Steven Brunton of the University of Washington with the SIAG/CSE Early Career Prize during the 2019 SIAM Conference on Computational Science and Engineering, which took place in Spokane, Wash., earlier this year. Brunton then delivered a lecture on "Data-Driven Discovery and Control of Complex Systems: Uncovering Interpretable and Generalizable Nonlinear Models." SIAM photo.

Omar Ghattas (left) of the University of Texas at Austin accepts the SIAM Activity Group on Geosciences (SIAG/GS) Career Prize from SIAG/GS chair Béatrice Rivière at the 2019 SIAM Conference on Mathematical & Computational Issues in the Geosciences, held this March in Houston, Texas. Ghattas presented a lecture about "Large-scale Bayesian Inversion for Geoscience Problems." SIAM photo.

Professional Opportunities and Announcements

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

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

Exceptional Mathematician Sought

Exceptional mathematician with background in number theory and/or recursive function theory sought to help prepare paper titled "The Remarkably Simple Structure of the $3x+1$ Function" for submission to journals. Not one claim of an error in the paper has been received in over two years; last year, a mathematician wrote stating his belief that the paper is correct (he did not have time to help prepare it for submission to journals). The paper is accessible on **occampress.com**.

Mathematician must be unafraid of new, unorthodox ideas concerning very difficult problems (in this case, the $3x+1$ Problem). He or she must be willing to write to journal editors stating his/her belief that the paper is correct. At present, journal editors decline to consider the paper because they cannot believe that a non-academic mathematician can have made progress on such a difficult problem (the author's degree is in computer science and he has spent most of his career doing research in the computer industry).

Author will pay any reasonable consulting fee and give generous credit in the "Acknowledgments" section (but only with the mathematician's prior written approval). Author will also offer shared authorship for significant contribution to content.

— *Peter Schorer*, peteschorer@gmail.com

Institute for Pure and Applied Mathematics

Call for Proposals

The Institute for Pure and Applied Mathematics (IPAM) seeks program proposals from the mathematical, statistical, and scientific communities for long programs and workshops, to be reviewed at IPAM's Science Advisory Board meeting in November. Long programs (three months) bring together researchers from mathematics and other disciplines—or multiple areas of mathematics—with the goal of facilitating collaborative, cross-disciplinary research. Winter workshops are typically five days in length. Exploratory workshops, which address an emerging problem or new application of math, typically last three days. Proposals for workshops on multiscale physics will be considered for inclusion in a series of workshops made possible by the Julian Schwinger Foundation for Physics Research. For more information, go to **www.ipam.ucla.edu/propose-a-program** or contact the IPAM director at **director@ipam.ucla.edu**. Proposals should also address the inclusion of women and members of underrepresented minorities as speakers, organizers, and participants.

Oscillating Wait Times and Queues with Information Updates

By Paul Davis

Murphy's Law—the idea that anything that can go wrong will go wrong—seems especially relevant when picking a waiting line. No matter how reliable your information, you inevitably end up in the longer queue. The supposedly quicker route recommended at the highway rest area resembles a parking lot by the time you arrive. The emergency room wait time that you verified on your phone before driving across town to the allegedly deserted location? Standing room only upon arrival (see Figure 1). And the phone app at Disneyland promising a much shorter queue for a ride just a few steps away? It outsmarts you every time (see Figure 2).

Jamol Pender of Cornell University blamed Hopf bifurcation, not Murphy's Law, for fluctuating wait times when he spoke during a minisymposium at the 2019 SIAM Conference on Computational Science and Engineering (CSE19), held in Spokane, Wash., earlier this year. His model shows that the technological blessing of posting nearly up-to-date queue lengths can incur the curse of large variations in those lengths if the associated information is a bit too old.

Oscillations induced by dated information are much more than simple irritations when they affect access to critical services like an emergency room. Pender's conclusion that the oscillations arise as a Hopf bifurcation points to a clear fix — estimates of queue lengths have a predictable shelf life. Exceed that shelf life and queue lengths begin to oscillate.

Pender formulated a stochastic queuing model that updates at fixed time intervals the wait times for each of two queues. The simplest version updates just once at each time step. His key analytic move is a clever continuum approximation of the initial time series formulation. Pender asked his audience to imagine the arriving customers as "droplets merging into a stream flowing out of a bathtub spout." He then derived the limiting continuum model by scaling both the queuing process and the arrival rate by the number of droplets.

Figure 1. *Public posting of the wait time at an emergency room. Is this information old enough to trigger oscillations via Hopf bifurcation? Image courtesy of [1].*

The resulting functional differential equation does indeed display a Hopf bifurcation if the information delay is too long, i.e., if the scaled information delay parameter is too large. One can explicitly calculate the critical value. Comparing discrete computational simulations of the oscillations in queue length with the continuum model's predicted oscillations indicates good agreement, even with a modest population of 100.

Having constructed a breakthrough model that reliably captures the effects of information delay on queue length, Pender and his team are now exploring that rich lode for additional insights into these behaviors. Does the oscillatory performance change significantly if queue lengths are updated several times per time step, rather than just once? Roughly, no. Hopf bifurcation still yields disruptive oscillations in queue lengths. The bifurcation parameter's critical value—now the root of a higher-order polynomial—is simply harder to calculate. But what if arrival rates vary with time? Or information about a queue's length and velocity is accessible? Answers and yet more questions are available in Pender's original paper [1] and his subsequent publications.[1]

Pender's presentation took place during a minisymposium entitled "Exciting Work by Early Career Underrepresented Minority Researchers," part of the SIAM Workshop Celebrating Diversity. He also contributed to a fruitful panel discussion at CSE19 that addressed "Strategies for Promoting Diversity and Inclusion within our Profession,"[2] organized and moderated by SIAM President Lisa Fauci.

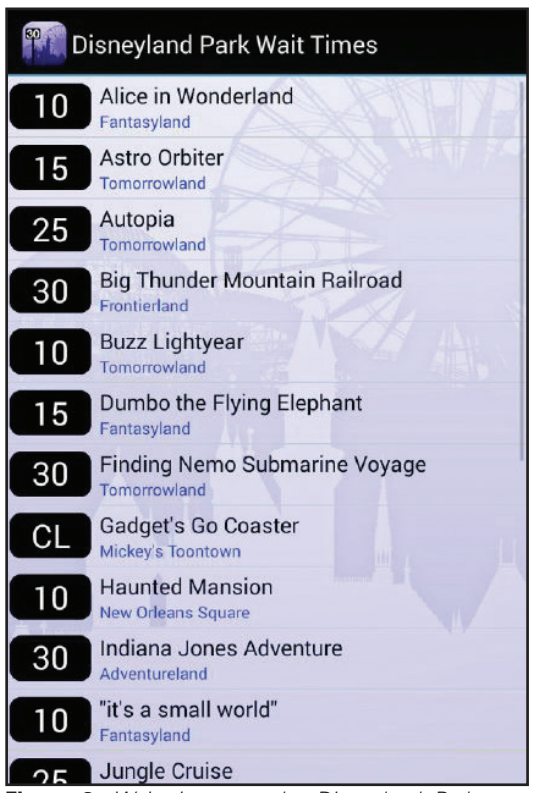


Figure 2. *Wait times on the Disneyland Park app. Image courtesy of [1].*

References

[1] Pender, J., Rand, R.H., & Wesson, E. (2017). Queues with Choice via Delay Differential Equations. *Int. J. Bifur. Chaos, 27*(4), 1730016.

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

[1] https://blogs.cornell.edu/jamolpender/publications

[2] See "CSE19 Panel Promotes Strategies to Increase Diversity and Inclusion in Academia" on page 6.

Diversity and Inclusion

Continued from page 6

Riskin touted the Washington State Academic RedShirt program as an example of a beneficial flexibility. The two-year program helps engineering and computer science students from low-income, first-generation, and underserved backgrounds transition to college-level engineering courses. It involves a specialized curriculum that prepares participants for core math and science prerequisites, and ultimately guarantees them placement in an engineering or computer science major. Riskin added that students should pitch their unique situations as strengths rather than liabilities. She recalled one scholar who used her first-generation status as a testament to her work ethic; that same woman is now applying to graduate programs. "What might seem like a bug can actually be an asset because we faculty want students who are going to work very hard," she said.

When an audience member wondered how students can work with faculty to increase diversity within their departments, Pender suggested having a specific goal in mind and assessing the different courses of action to best achieve it. When he was a graduate student, his department sought to increase the number of underrepresented minorities. Pender proposed a conference that invited undergraduate students to present their research at Princeton. The administration was extremely supportive and even brought in mathematicians from around the country to speak to the students in attendance. The conference became a recurring event, familiarized attendees with Princeton, and ultimately increased diversity among graduate-level applications.

In situations where the faculty is less open to direct input from students, Riskin suggested that graduate students sign a collective letter communicating their concerns and requests. However, she cautioned that they should do their research beforehand, because sometimes a department might be addressing relevant issues behind the scenes but not communicating its efforts. "Certainly those letters do get attention," she said. "They can cause a lot of heartbreak for the department chair, but sometimes the heartbreak is needed."

Kuske noted that students who ask for something that already exists—but to which they lack access—are likely indicative of a need for departmental restructuring. She recalled having witnessed a few departments whose hiring processes were stuck in a rut. As a result, students voiced their collective opinion that the departments were not offering courses that appealed to their professional interests and goals. These candid discussions gradually helped shape diversity in the hiring process and encouraged department staff to be more forward-looking. Pender shared that Cornell's School of Operations Research and Information Engineering allows graduate students to meet with and offer feedback on prospective hires.

Castillo-Chavez acknowledged that while interactions between students and staff can be instrumental, the hiring process is ultimately up to the faculty. He thus recommended that hiring committees incorporate colleagues from different departments to minimize bias and work towards making open-minded thinking the standard rather than the exception. "Often faculty members want to hire students from schools they know or colleagues they respect," he said. "That's where the problem lies. I think that's the biggest roadblock to increasing diversity."

Lina Sorg is the associate editor of SIAM News.

Carlos Castillo-Chavez (Arizona State University), Rachel Kuske (Georgia Institute of Technology), Eve Riskin (University of Washington), and Jamol Pender (Cornell University) share personal experiences and strategies for promoting inclusion in the field of mathematics during the 2019 SIAM Conference on Computational Science and Engineering, held in Spokane, Wash., earlier this year. SIAM photo.

Jamol Pender (Cornell University) addresses attendees of a panel discussion about diversity at the 2019 SIAM Conference on Computational Science and Engineering, which took place earlier this year in Spokane, Wash. SIAM photo.