



Students and Education | January 28, 2019

Students Tackle Bayesian Inverse Problems in the Colorado Rockies

Reflections on the 2018 Gene Golub Summer School

By Omar Ghattas, Youssef Marzouk, Matt Parno, Noemi Petra, Georg Stadler, and Umberto Villa

Recent years have seen rapid growth in the volumes of observational and experimental data acquired from natural or engineered systems. How do we extract knowledge and insight about these systems from all of this data? This learning-from-data problem is at its core a mathematical inverse problem. That is, given (possibly noisy) data and a (possibly uncertain) “forward” model that maps parameters to data, we seek to infer parameters that characterize the model. Inverse problems abound in all areas of science, engineering, medicine, and beyond. Examples include inference of internal defects from scattered ultrasonic wave measurements, anatomical structures from X-ray computed tomography data, coalescence of binary systems from detected gravitational waves, ocean state from surface temperature observations, and subsurface contaminant plume spread from crosswell electromagnetic measurements.

Inverse problems are often ill-posed, i.e., their solutions may not exist or be unique or stable to perturbations in the data. Simply put, the data—even when large-scale—does not provide sufficient information to fully determine the model parameters. Non-uniqueness can stem from noise in the data or model, sparsity or redundancy in the data, or smoothing properties of the map from model parameters to observables. In such cases, uncertainty is a fundamental feature of the inverse problem; we wish to both infer the parameters and quantify the uncertainty associated with this inference. The ability to do the latter opens the door to powerful capabilities for model-based decision-making under uncertainty.

Bayesian inference provides a powerful framework for solution of inverse problems under uncertainty. However, when the forward model is expensive (as for partial differential equations (PDEs)) and the parameter dimension is large (as with discretized fields), Bayesian inversion becomes prohibitive with standard statistical methods. The last few years have seen the development of advanced mathematical and computational methods for Bayesian inverse problems governed by complex, high-dimensional forward models. To introduce graduate students to these recent advances, we organized the 2018 Gene Golub SIAM Summer School (G2S3) last June in Breckenridge, Colo., with the theme of *Inverse Problems: Systematic Integration of Data with Models under Uncertainty*.

The Bayesian inversion theme was very popular, as evidenced by the 257 applications we received from students around the world. The two-week summer school was sponsored by SIAM through a generous endowment from the estate of mathematician Gene Golub. The National Science Foundation's (NSF) Division of Mathematical Sciences provided additional support for U.S.-based students. Together these funds allowed 44 students to attend. However, limitations on available funding



A group of students and instructors hike to the summit of Quandary Peak, one of the 14,000-foot mountains near Breckenridge, Colo., during the 2018 Gene Golub SIAM Summer School. Photo credit: Georg Stadler.

and meeting space meant that many excellent applicants were turned down. Participants formed a highly diverse group with respect to gender, geography, and research interests. They hailed from five continents with a wide range of backgrounds spanning applied math, statistics, engineering, and natural sciences.

The summer school lectures offered an integrated presentation of deterministic and Bayesian inverse theory and algorithms that first introduced ill-posedness and regularization; developed the ideas and tools for deterministic inversion via adjoint-based first- and second-order sensitivity analysis and numerical optimization; acquainted students with the Bayesian statistical framework for finite- and infinite-dimensional inverse problems; and finally explored linearized and sampling-based statistical solution methods, which built on several deterministic tools.

Concepts discussed in the morning lectures were put into practice and examined more thoroughly during hands-on laboratory sessions in the afternoon. These sessions utilized powerful open-source software that implemented state-of-the-art deterministic and Bayesian inversion methods (hIPPYlib, MUQ) and finite element solution of PDEs (FEniCS). The laboratory components featured cloud-based interactive tutorials via Jupyter notebooks, which mixed instruction and theory with editable and runnable code. The notebooks enabled students to run and modify sample codes through their web browsers, permitting rapid experimentation with different inverse problem formulations, discretizations, and solution algorithms. An NSF XSEDE cloud system called Jetstream deployed the software libraries via Docker containers, simplifying their installation and use. To broaden the summer school's impact and availability beyond just attendees, all lab material—including 16 tutorials and associated code—is freely available on



The entire 2018 Gene Golub SIAM Summer School team poses in the shape of a Gaussian curve in Breckenridge, Colo. Photo credit: Georg Stadler.

and wave scattering—on the last day of the program.

The secluded mountain environment of Breckenridge provided a relaxed and inspiring atmosphere highly conducive to learning, collaboration, and idea exchange. Additionally, the Rocky Mountains created many opportunities for recreational activities like hiking, whitewater rafting, and mountain biking.

Seven of Gene Golub's eight most-cited papers treat methods that are important components of inverse problems (least squares optimization, saddle point solvers, singular values, and cross-validation). We think he would have approved of the theme of the 2018 G2S3! We thank Gene Golub's estate, SIAM, and the NSF for making this year's school possible. The students' enthusiasm and positive feedback has encouraged us to offer a similar event in the future.

Omar Ghattas is a professor in the Departments of Geological Sciences and Mechanical Engineering and the Institute for Computational Engineering and Sciences at the University of Texas at Austin. Youssef Marzouk is a professor in the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology. Matt Parno is a researcher at the U.S. Army Cold Regions Research and Engineering Laboratory. Noemi Petra is a professor of applied mathematics at the University of California, Merced. Georg Stadler is a professor at New York University's Courant Institute of Mathematical Sciences. Umberto Villa is a research professor in the Department of Electrical and Systems Engineering at Washington University in St. Louis.

the G2S3 website.

The school culminated with team-based research projects in which participants employed the theory, algorithms, and software they had learned to tackle realistic and challenging inverse problems of their choice. The 11 teams presented their projects—ranging in areas such as photoacoustic tomography, incompressible flows, seismic imaging, elastography, heat conduction,