Exascale Data Analytics for the DOE

Archibald[1], Bao[4], Bertozzi[3], Gelb[2], Hauck[1], Law[1], Osher[3], Tran[1], Webster[1], Wise[5], and Zhang[1]

[1]Oak Ridge National Laboratory, TN 37831 ({archibaldrk,hauckc,lawkj,tranha,webstercg,zhangg}@ornl.gov)
[2]Dartmouth College, NH 03755 (annegelb@math.dartmouth.edu)
[3]University of California Los Angeles, CA 90095 ({bertozzi,sjo}@math.ucla.edu)
[4]University of Tennessee Chattanooga, TN 37403 (feng-boa@utc.edu)
[5]University of Tennessee Knoxville, TN 37996 (swise@math.utk.edu)

Summary of Position

The next decade will require rigorously analyzed mathematical developments, customized to the interests of the Department of Energy (DOE), that are able to completely and accurately capture and store critical information from large amounts of scientific data. This goal may be accomplished through advances in (1) model and functional driven representation of data; (2) graphical and machine learning; and (3) Bayesian methods. Moreover, such developments must keep pace with the ever evolving computational capability and data generated across the DOE complex in order to ensure fast acceleration towards scientific discovery at the leadership level.

Next Generation Mathematics for Data at the DOE

The DOE generates massive amounts of scientific data through many distinct modalities, most notably, data measured from experimental and observational facilities and data generated by supercomputer simulations. The DOE has witnessed exponential growth of data production, a trend that is anticipated to continue into the future [70, 72]. The life-cycle of scientific data flows through a hierarchy of storage devices in which both the storage capacity and transmission latency generally grows as the data moves from the point of generation to a final archival location [74]. This very broad characterization of scientific data at the DOE describes the salient features of the framework that is expected to continue into the near future [71], and shapes the challenges and restrictions faced by algorithmic development in data analytics at the DOE.

This explosion of data and framework dynamics will force changes in how data is stored, analyzed, and moved. One key shift that will need to occur is the functional representation of data using either approximation or model driven approaches. At the forefront of model [10, 11, 15, 20, 49, 55, 81, 88] and functional [2, 3, 34, 60, 73, 75–80, 85–87] driven representation of data is sparsity, regularization, and fast optimization and reconstruction. Sparsity is inherent in a wide range of data, and many analytics tasks for big data become possible only through detecting sparsity patterns. Near-optimal sparse representations of data are facilitated by methods in compressed sensing [13, 24] and statistical learning [41], both of which have been developed in the last decades. Such methods are characterized by minimal amounts of data acquisition or measurements, as well as sparsity-enforcing convex regularizations. Data growth at the DOE is driving the need for innovative techniques that allow more efficient compression and reconstruction of data. Several methodologies that have recently been explored show great potential, specifically those that include exploiting the structured sparsity of the data [1, 17, 23, 87, 89], offer new strategies of data acquisition [84], and employ nonconvex regularizations [83]. These types of methods must be further developed to ensure fast and optimal compression of data, in particular to minimize what is needed to reduce computational cost, storage and transport.

Graphical and machine learning methods have risen to prominence in the last few decades, and will be central to data analytics in the future [44]. Mathematics has played a unique role in graphical and machine learning methods [4, 5, 26, 45, 48, 61, 63–66], from building algorithms that take advantage of tools built for partial differential equations (PDEs) to accelerating learning on high performance computing (HPC) platforms, as well as increasing the accuracy of classifying

complex features. It is fair to say that PDE based methods have revolutionized image processing, beginning with TV restoration introduced in the later part of last century [68, 69, 75], and continuing on today with missing data restoration and super resolution imaging techniques [12]. The same thing will be true, on an even larger scale, for machine learning, where PDE based methods are improving the architecture and non-convex optimization in deep neural nets [14], and have been demonstrated on the exascale-ready machine Cori at NERSC to efficiently compute low rank representations of fully connected similarity graph, the backbone calculation of spectral clustering or semi-supervised/unsupervised machine learning methods [63]. Over the next ten years, mathematicians must develop new methods capable of learning physical relationships and models directly from data, determine confidence measures in predictions and classifications made by machine learning methods, and take full advantage of the DOE computational facilities with architecture aware algorithms that are memory and communication reducing, thereby enabling efficient utilization of all compute nodes. The proper incorporation of PDE based methods, such as the HPC methods developed by the FASTMATH [50] SciDAC institute [51], will most certainly play an important role in bringing machine learning to the exascale.

Understanding the role uncertainty plays is paramount for the methods described above to be fully understood. Bayesian methods that yield a rigorous framework for describing statistical models of data are well studied, and provide confidence levels for data assimilation and model calibration. Such inverse uncertainty quantification (UQ) allows one to compute, given all available information, posterior expectations of arbitrary quantities of interest. However, capturing full information using Bayesian methods is much too costly compared to deterministic solutions without UQ, and therefore improving computational efficiency as well as larger leadership computers are required. Fully exploiting emerging massively parallel computing architectures, with the explosion of data along with the growing appreciation in the scientific community of the requirement of UQ, generates new challenges across computer science, mathematics, and statistics. A particularly fruitful area of emerging research is the joint numerical and statistical analysis of Bayesian methods, where enormous gains in computational efficiency can be found in comparison to disjoint independent analysis of the respective components. The workhorse of Bayesian computation is indisputably the Markov chain Monte Carlo (MCMC) [28, 31], and within MCMC the most ubiquitous classes of algorithms are Gibbs [27] and Metropolis Hastings (MH) [42, 67]. Many of the more sophisticated algorithms at the forefront of research in computationally intensive statistical inference rely on MH kernels as a crucial component. Algorithms are emerging which are capable of probing posteriors with features such as strong correlation [37, 38], functional (infinite-dimensional) parameters [9, 19], sharp manifolds [32, 33, 35, 62], multi-modality [21, 25, 40, 57, 58, 82], or combinations thereof [7, 15, 20, 59]. In certain scenarios it is natural to emply GPU-acceleration of the linear algebra appearing inside the method [15]. To apply Bayesian methods to computationally expensive forward simulations, or to large multivariate data, or to models with large parameter spaces, new methods must be developed which are carefully designed according to consideration of the joint analysis of the numerical and statistical aspects. One promising emerging avenue is the application of multilevel [18, 29, 43] and multi-index Monte Carlo [30, 39] to problems of statistical inference [7, 8, 16, 22, 36, 46, 47, 52–54, 56]. Some other very promising avenues is the incorporation of Bayesian methods into graphical machine learning models developing a metric for uncertainty of the classification – a very important quantifier in the absence of ground truth or other suitable metrics [6].

Convergence of data- and model-driven discovery is clearly a cross-disciplinary problem that will need dedicated researchers and a commitment to seamless integration of computational facilities and data generated at the DOE . In order to achieve this goal, mathematical development in algorithms and theory is needed to effectively connect the resources at the DOE.

[1] B. Adcock. Infinite-dimensional compressed sensing and function interpolation. *Found Comput Math*, 2017.

[2] R. Archibald, A. Gelb, and R. Platte. Image reconstruction from undersampled Fourier data using the polynomial annihilation transform. *Journal of Scientific Computing*, 67(2):432–452, May 2016.

[3] R. Archibald, A. Gelb, and J. Yoon. Polynomial fitting for edge detection in irregularly sampled signals and images. *SIAM J. Numer. Anal.*, 43(1):259–279, Jan 2005.

[4] A. Bertozzi and A. Flenner. Diffuse interface models on graphs for classification of high dimensional data. *Multiscale Modeling & Simulation*, 10(3):1090–1118, 2012.

[5] A. Bertozzi and A. Flenner. Diffuse interface models on graphs for classification of high dimensional data. *SIAM Review*, 58(2):293–328, 2016.

[6] A. Bertozzi, X. Luo, A. Stuart, and K. Zygalakis. Uncertainty quantification in the classification of high dimensional data. *CoRR*, abs/1703.08816, 2017.

[7] A. Beskos, A. Jasra, K. Law, Y. Marzouk, and Y. Zhou. Multilevel sequential Monte Carlo with dimension-independent likelihood-informed proposals. *arXiv preprint arXiv:1703.04866*, 2017.

[8] A. Beskos, A. Jasra, K. Law, R. Tempone, and Y. Zhou. Multilevel sequential Monte Carlo samplers. *Stochastic Processes and their Applications*, 127(5):1417–1440, 2017.

[9] A. Beskos, G. Roberts, A. Stuart, and J. Voss. MCMC methods for diffusion bridges. *Stochastics and Dynamics*, 8(03):319–350, 2008.

[10] D. Blömker, K. Law, A. Stuart, and K. Zygalakis. Accuracy and stability of the continuous-time 3DVAR filter for the Navier–Stokes equation. *Nonlinearity*, 26(8):2193, 2013.

[11] C. Brett, K-F. Lam, K. Law, D. McCormick, M. Scott, and A. Stuart. Accuracy and stability of filters for dissipative PDEs. *Physica D: Nonlinear Phenomena*, 245(1):34–45, 2013.

[12] M. Burger and S. Osher. *A Guide to the TV Zoo*, pages 1–70. Springer International Publishing, Cham, 2013.

[13] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(1):489–509, 2006.

[14] P. Chaudhari, A. Oberman, S. Osher, S. Soatto, and G. Carlier. Deep relaxation: partial differential equations for optimizing deep neural networks. *CoRR*, abs/1704.04932, 2017.

[15] Y. Chen, D. Keyes, K. Law, and H. Ltaief. Accelerated dimension-independent adaptive Metropolis. *To appear in SISC. arXiv preprint arXiv:1506.05741.*, 2016.

[16] A. Chernov, H. Hoel, K. Law, F. Nobile, and R. Tempone. Multilevel ensemble Kalman filtering for spatially extended models. *arXiv preprint arXiv:1608.08558*, 2016.

[17] A. Chkifa, N. Dexter, H. Tran, and C. Webster. Polynomial approximation via compressed sensing of high-dimensional functions on lower sets. *Mathematics of Computation*, 2016. To appear (arXiv:1602.05823).

[18] K. Cliffe, M. Giles, R. Scheichl, and A. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011.

[19] S. Cotter, G. Roberts, A. Stuart, D. White, et al. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, 28(3):424–446, 2013.

[20] T. Cui, K. Law, and Y. Marzouk. Dimension-independent likelihood-informed MCMC. *Journal of Computational Physics*, 304:109 – 137, 2016.

[21] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(3):411–436, 2006.

[22] P. Del Moral, A. Jasra, K. Law, and Y. Zhou. Multilevel sequential Monte Carlo samplers for normalizing constants. *To appear in TOMACS. arXiv preprint arXiv:1603.01136*, 2016.
[23] N. Dexter, Tran, and C. Webster. Reconstruction of joint sparse signals using forward-backward splitting. *ORNL Technical Report*, 2017.
[24] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
[25] D. Earl and M. Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005.
[26] C. Garcia-Cardona, E. Merkurjev, A. Bertozzi, A. Flenner, and A. Percus. Multiclass data segmentation using diffuse interface methods on graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1600–1613, Aug 2014.
[27] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov 1984.
[28] C. Geyer. Practical Markov chain Monte Carlo. *Statistical Science*, pages 473–483, 1992.
[29] M. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.
[30] M. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 5 2015.
[31] W. Gilks. *Markov Chain Monte Carlo*. Wiley Online Library, 2005.
[32] W. Gilks, G. Roberts, and E. George. Adaptive direction sampling. *The Statistician*, pages 179–189, 1994.
[33] M. Girolami and B. Calderhead. Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.
[34] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–343, 2009.
[35] J. Goodman and J. Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, 2010.
[36] A. Gregory, C. Cotter, and S. Reich. Multilevel ensemble transform particle filtering. *SIAM Journal on Scientific Computing*, 38(3):A1317–A1338, 2016.
[37] H. Haario, M. Laine, A. Mira, and E. Saksman. DRAM: efficient adaptive MCMC. *Statistics and Computing*, 16(4):339–354, 2006.
[38] H. Haario, E. Saksman, and J. Tamminen. An Adaptive Metropolis Algorithm. *Bernoulli*, pages 223–242, 2001.
[39] A.-L. Haji-Ali, F. Nobile, and R. Tempone. Multi-index Monte Carlo: when sparsity meets sampling. *Numerische Mathematik*, pages 1–40, 2015.
[40] U. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters*, 281(1):140–150, 1997.
[41] T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical Learning with Sparsity*, volume 143 of *Monographs on Statistics and Applied Probability*. Taylor & Francis Group, LLC, 2015.
[42] W. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
[43] S. Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer, 2001.
[44] A. Hey, S. Tansley, K. Tolle, et al. *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft research Redmond, WA, 2009.
[45] H.Hu, J. Sunu, and A. Bertozzi. Multi-class graph Mumford-Shah model for plume detection using the MBO scheme. In *Proceedings of the EMMCVPR conference in Hong Kong 2015*, volume 8932, pages 209–222. Springer, 2015. Lecture Notes in Computer Science.

[46] V. Hoang, C. Schwab, and A. Stuart. Complexity analysis of accelerated MCMC methods for Bayesian inversion. *Inverse Problems*, 29(8):085010, 2013.

[47] H. Hoel, K. Law, and R. Tempone. Multilevel ensemble Kalman filtering. *SIAM Journal of Numerical Analysis*, 54(3):1813–1839, 2015.

[48] H.i Hu, T. Laurent, M. Porter, and A. Bertozzi. A method based on total variation for network modularity optimization using the MBO scheme. *SIAM Journal on Applied Mathematics*, 73(6):2224–2246, 2013.

[49] M. Iglesias, K. Law, and A. Stuart. Ensemble Kalman methods for inverse problems. *Inverse Problems*, 29(4):045001, April 2013.

[50] DOE SciDAC Institute. Fastmath. 2017-08-16.

[51] DOE SciDAC Institutes. Homepage. 2017-08-16.

[52] A. Jasra, K. Kamatani, K. Law, and Y. Zhou. A Multi-Index Markov Chain Monte Carlo Method. *arXiv preprint arXiv:1704.00117*, 2017.

[53] A. Jasra, K. Kamatani, K. Law, and Y. Zhou. Bayesian static parameter estimation for partially observed diffusions via multilevel Monte Carlo. *arXiv preprint arXiv:1701.05892*, 2017.

[54] A. Jasra, K. Law, and C. Suciu. Advanced multilevel Monte Carlo methods. *arXiv preprint arXiv:1704.07272*, 2017.

[55] D. Kelly, K. Law, and A. Stuart. Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. *Nonlinearity*, 27(10):2579, 2014.

[56] C. Ketelsen, R. Scheichl, and A. Teckentrup. A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. *arXiv preprint arXiv:1303.7343*, 2013.

[57] S. Kou, Q. Zhou, and W. Wong. Discussion paper equi-energy sampler with applications in statistical inference and statistical mechanics. *The annals of Statistics*, pages 1581–1619, 2006.

[58] D. Landau and F. Wang. Determining the density of states for classical statistical models by a flat-histogram random walk. *Computer physics communications*, 147(1-2):674–677, 2002.

[59] K. Law. Proposals which speed up function-space MCMC. *Journal of Computational and Applied Mathematics*, 262:127–138, 2014.

[60] W. Li, E. Ryu, S. Osher, W. Yin, and W. Gangbo. A parallel method for earth mover's distance. *UCLA Comput. Appl. Math. Pub.(CAM) Rep*, pages 17–12, 2017.

[61] X. Luo and A. Bertozzi. Convergence of the graph Allen–Cahn scheme. *Journal of Statistical Physics*, 167(3):934–958, 2017.

[62] J. Martin, L. Wilcox, C. Burstedde, and O. Ghattas. A stochastic newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487, 2012.

[63] Z. Meng, A. Koniges, Y. He, S. Williams, T. Kurth, J. Deslippe B. Cook, and A. Bertozzi. OpenMP parallelization and optimization of graph-based machine learning algorithms. In N. Maruyama, B. de Supinksi, and M. Wahib, editors, *IWOMP 2016. Lecture Notes in Computer Science, vol 9930*. Springer, 2016.

[64] E. Merkurjev, E. Bae, A. Bertozzi, and X.-C. Tai. Global binary optimization on graphs for classification of high-dimensional data. *Journal of Mathematical Imaging and Vision*, 52(3):414–435, 2015.

[65] E. Merkurjev, A. Bertozzi, X. Yan, and K. Lerman. Modified cheeger and ratio cut methods using the Ginzburg–Landau functional for classification of high-dimensional data. *Inverse Problems*, 33(7):074003, 2017.

[66] E. Merkurjev, T. Kostić, and A. Bertozzi. An MBO scheme on graphs for classification and image processing. *SIAM Journal on Imaging Sciences*, 6(4):1903–1930, 2013.

[67] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[68] J. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhauser Boston Inc., Cambridge, MA, USA, 1995.

[69] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[70] Department of Energy Scientific Grand Challenges Workshop Series. Architectures and technology for extreme scale computing, 2009.

[71] Dept of Energy Scientific Grand Challenges Workshop Series. Cross-cutting technologies for computing at the exascale, 2009.

[72] Dept. of Energy Scientific Grand Challenges Workshop Series. Discovery in basic energy sciences: The role of computing at the extreme scale, 2009.

[73] S. Osher, Z. Shi, and W. Zhu. Low dimensional manifold model for image processing. *UCLA CAM report 16-04*, 2016.

[74] DOE ASCAC Data Subcommittee Report. Synergistic challenges in data-intensive science and exascale computing, 2013.

[75] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, November 1992.

[76] T. Sanders, A. Gelb, and R. Platte. Composite SAR imaging using sequential joint sparsity. *Journal of Computational Physics*, 338:357 – 370, 2017.

[77] T. Sanders, A. Gelb, R. Platte, I. Arslan, and K. Landskron. Recovering fine details from under-resolved electron tomography data using higher order total variation regularization. *Ultramicroscopy*, 174:97 – 105, 2017.

[78] Z. Shi, S. Osher, and W. Zhu. Weighted nonlocal laplacian on interpolation from sparse data. *Journal of Scientific Computing*, Apr 2017.

[79] Z. Shi, W. Zhu, and S. Osher. Low dimensional manifold model in hyperspectral image reconstruction. *CoRR*, abs/1605.05652, 2016.

[80] Z. Shi, W. Zhu, and S. Osher. Low dimensional manifold model with semi-local patches. *UCLA CAM report 16-63*, 2016.

[81] A. Stuart, A. Shukla, and K. Law. Analysis of the 3DVAR filter for the partially observed Lorenz'63 model. *Discrete and Continuous Dynamical Systems*, 34(3):1061–1078, 2013.

[82] R. Swendsen and J.-S. Wang. Replica Monte Carlo simulation of spin-glasses. *Physical Review Letters*, 57(21):2607, 1986.

[83] H. Tran and C. Webster. Unified sufficient conditions for uniform recovery of sparse signals via nonconvex minimizations. *submitted to Constructive Approximation*, 2017.

[84] H. Tran and C. Webster. A uniform bound of sample complexity for sparse Legendre approximation via ℓ_1 minimization in arbitrary polynomial subspace. *submitted to Journal of Complexity*, 2017.

[85] H. Tran, C. Webster, and G. Zhang. Analysis of quasi-optimal polynomial approximations for parameterized PDEs with deterministic and stochastic coefficients. *Numerische Mathematik*, 2017.

[86] G. Wasserman, R. Archibald, and A. Gelb. Image reconstruction from Fourier data using sparsity of edges. *Journal of Scientific Computing*, 65(2):533–552, Nov 2015.

[87] W. Zhu, B. Wang, R. Barnard, C. Hauck, F. Jenko, and S. Osher. Scientific data interpolation with low dimensional manifold model. *UCLA CAM report 17-22*, 2017.

[88] D. Zosso, J. An, J. Stevick, N. Takaki, M. Weiss, L. Slaughter, H. Cao, P. Weiss, and A. Bertozzi. Image segmentation with dynamic artifacts detection and bias correction. *Inverse Problems and Imaging*, 11(3):577–600, 2017.

[89] D. Zosso, K. Dragomiretskiy, A. Bertozzi, and P. Weiss. Two-dimensional compact variational mode decomposition. *Journal of Mathematical Imaging and Vision*, 58(2):294–320, 2017.