High End Data Science and HPC for the Electrical Power Grid

Alex Pothen and Ariful Azad
Purdue University and Indiana University

The U.S. electrical power grid is the largest machine ever built, and the National Science Foundation, the Department of Energy and other federal agencies are calling for research to improve the power system reliability through wide area measurement and control by employing syncrophasor technology (Phasor Measurement Unit sensors, PMUs). Over 2500 PMU sensor units are deployed in the North American power grid. PMUs obtain 30-60 samples per second, generating Petabytes of data per day, and providing real-time situational awareness leading to early warning of grid events and dynamic behavior of the grid. It can help with strategies to recover from natural disasters, large variations in the power generation from renewable sources, cyberattacks, etc. These include planning ahead for events and shocks, decision and control as the event unfolds and for post-shock recovery to a steady state. The North American Power Grid Initiative (NASPI) promotes the use of syncrophasor data to estimate and control the state of the grid.

An important use of PMUs is in oscillation detection in the power grid, and the predominant computation here is computing the dominant singular values and singular vectors of a covariance matrix computed from the time varying data matrices from a formulation for controlling the grid. We have recently deployed randomized SVD algorithms and Lanczos algorithms on parallel computers to enable this computation without forming the covariance matrix itself, employing only matrix-vector products for this purpose [1]. We are able to compute SVDs for data from hundreds of PMU sensors in a matter of seconds.

A second application of high-end data analytics in the power grid that we have considered is in contingency analysis, where power grid operators have to determine how to stabilize the grid when a few transmission lines or generators should go down. We have modeled this scenario in the case of DC power flow and have shown that an augmented matrix formulation can be used to update solutions to Kirchoff's equations three orders of magnitude faster than a method that would solve the modified system of equations directly [2].

A third application of high-end data analytics in the power grid is in vulnerability analysis of power networks. Vulnerability analysis is aimed at detecting vulnerable sections of a power grid subject to cyberattack, natural disasters or mechanical failures. Traditionally, power grids and flows are analyzed using non-linear, numerical methods. However, under certain assumptions, graph-theoretical methods can be useful to study optimal power flow and vulnerability of power networks. It is widely believed that power grids exhibit the properties of scale-free networks, especially when a power network is described as a function of its “electrical topology” [3]. Consequently, power networks are robust to random attacks and failures, but they can be vulnerable to targeted attacks. To model power flow and vulnerability, prior work considered the maximum flow problem from graph theory [4]. However, as electricity flows differently than fluids, specialized flow models are required for power networks. To this end, an exciting avenue of research is in designing realistic flow models for power networks and using them to analyze

dynamic power network with random failure of equipment. The problem is even more exciting when power flow can be controlled such as with Flexible Alternating Current Transmission System (FACTS) [5].

In developing algorithms and libraries for power grids, computer scientists are often challenged by lack of reliable data due to privacy and security concerns. One approach to tackle this challenge is to generate synthetic data based on public information about power stations and demographic data. For example, ARPA-E Grid Data program supports several projects to generate synthetic data with realistic generation and load profiles (https://electricgrids.engr.tamu.edu/). These datasets can be augmented further to make them reliable replicas of real grids without exposing secure information.

Today's dynamic and massive power grids require easy-to-use algorithms and libraries deployed on HPC infrastructures. A joint effort from power grid analysts, data scientists and HPC experts can play pivotal roles in rapid analysis of power networks for potential cyber threats, vulnerability and swift recovery schemes after an attack or mechanical failure. The HPC community can provide high-performance software that can be easily deployed to the cloud for out-of-the-box analysis.

References

[1] Tianying Wu, Vaithianathan ``Mani" Venkatasubramanian, and Alex Pothen, Fast parallel stochastic subspace algorithms for large-scale ambient oscillation monitoring, IEEE Transactions on the Smart Grid, 8(3), pp. 1494-1503, 2017.

[2] Yu-Hong Yeung, Alex Pothen, Mahantesh Halappanavar and Zhenyu Huang, AMPS: An augmented matrix formulation for principal submatrix updates with application to power grids, SIAM J. Scientific Computing, 39 (3), S809-S827, 2017.

[3] Paul Hines, and Seth Blumsack, A centrality measure for electrical networks, Proceedings of the 41st Annual Hawaii International Conference on System Sciences, 2008.

[4] Ajendra Dwivedi, Xinghuo Yu, and Peter Sokolowski, Analyzing power network vulnerability with maximum flow based centrality approach, 8th IEEE International Conference on Industrial Informatics, 2010.

[5] Franziska Wegner, Network Flow Models for Power Grids. Diss. Karlsruhe Institute of Technology, 2014.