COVER PAGE

Project Title:	Combinatorial Algorithms in Scientific Computing
Federal Award Identification Number:	DE-SC0010205
Organization:	Office of Advanced Scientific Computing Research, DOE
Project Period:	07/01/2016 - 06/30/2020
Report Term:	07/01/2016 - 06/30/2020 (Final Report)
Principal Investigator:	Alex Pothen
Position:	Professor of Computer Science
Postal Address:	305 N. University Street
	West Lafayette IN 47907-2107
Telephone Number:	(765) 496-9376
Email Address:	apothen@purdue.edu
Recipient Institution:	Purdue University
Street Address/City/State/Zip:	155 South Grant Street
	West Lafayette IN 47907-2114
DUNS:	072051394
EIN:	356002041

ACCOMPLISHMENTS

1. **What are the major goals of the project?**

   This project addressed three goals:

   (a) The design and implementation of algorithms for computing second and higher order derivatives via the Reverse Mode of Algorithmic Differentiation (Automatic Differentiation, AD);

   (b) Fast updates of solutions to linear systems of equations, when only a small principal submatrix is updated, for applications to power grids and surgery simulations;

   (c) Computing matchings and edge covers on parallel computers, including multithreaded shared memory processors and leadership-class distributed memory machines, and applying these to machine learning and data privacy problems.

2. **What was accomplished under these goals?**

   (a) We have developed new Reverse Mode algorithms for computing Hessians and higher order derivatives using the higher order chain rule. This method applies the high order chain rule directly in the reverse mode, following a forward mode computation of the function values, instead of computing the derivatives by making multiple passes through the computational graph. Sparsity and symmetry are fully exploited. The method does not need to have a prior step to compute the sparsity in the Hessian, unlike previous approaches we also show that the time complexity of the algorithm is an order of magnitude smaller than the earlier forward mode algorithms for this problem. A code implementing this algorithm, ReverseAD, has been made available on Github.

   (b) We have developed an augmented matrix approach to update the solution to a linear system of equations when the matrix is modified by a few elements within a principal submatrix. This problem arose in two contexts: First, the dynamic security analysis of a power grid, where operators need to perform N-k contingency analysis, i.e., determine the state of the system when exactly k links from N fail. Our algorithms use an augmented matrix formulation to account for the changes in the grid, and then compute the solution to the augmented system without refactoring the modified matrix. We have provided a hybrid direct-iterative method for solving the augmented system. We also exploit the sparsity of the matrices and vectors to accelerate the overall computation. The time complexity of the algorithm is bounded by the number of nonzeros in a subset of the columns of the Cholesky factor that are selected by the nonzeros in the sparse right-hand-side vector. We show that our augmented algorithms outperforms a direct solver based method by two orders of magnitude. Our algorithms are capable of computing N-k contingency analysis on grids with millions of buses in seconds on a desktop processor.

   The same methodology applies to surgery simulations except that here as a finite element mesh model of an eye (or other organ) is cut, we need to insert or delete new elements and nodes, and hence the dimension of the matrix changes at each update. The augmented matrix approach was extended to deal with this situation, and we showed that our work could be applied to model astigmatism surgery, where small precise cuts are made to the cornea to relieve pressure in the eye and to reduce distortion. A paper on this work was published in the journal Numerical Linear Algebra with its Applications.

   (c) We have designed and implemented several $1/2$-approximation algorithms for maximum weighted b-matching on serial, shared memory parallel, and distributed-memory parallel computers. It was scaled to obtain good performance on 16K processors on Cori. A $(1-1/k)$-approximation algorithm for maximum vertex-weighted matchings was designed, where we can choose k to make the approximation as close to one as possible. This leads to a $2/3$-approximation algorithm, when $k = 2$, and it was also implemented in parallel on a shared memory machine.

One of the findings from this work is that matching problems with random vertex weights generate hard instances for the maximum edge-weighted matching problem, when edge weights are equal to the sum of the vertex weights at the endpoints. This work also led to the first practical parallel approximation algorithm for a matching problem with approximation ratio better than $1/2$.

We also designed several $3/2$- and 2-approximation algorithms for the b-edge cover problem. These algorithms were based on several algorithmic paradigms: Greedy, Locally Dominant edge-based, Primal-dual, and different reductions to matching. We also showed that the widely used k-nearest neighbor graph construction algorithm is a 2-approximate solution to the minimum weight b-EDGE COVER problem. We have applied the approximate b-EDGE COVER algorithm to solve the adaptive anonymity problem in data privacy, solving problems larger by a factor of 1000 over previously reported approaches.

Our work was summarized in an invited article of 93 pages in *Acta Numerica*; this is the highest ranked journal in Applied Mathematics in terms of impact factors.

3. **What opportunities for training and professional development has the project provided?**

   During the project period, four PhD students completed their theses. Arif Khan and Yu-Hong Yeung have joined Pacific Northwest National Labs, where Dr. Khan is a staff scientist in the Scalable Data Analytics group, and Dr. Yeung is a post-doctoral scientist in the scientific machine learning group. Dr. Khan was awarded the John R. Rice Fellowship in Computational Science and Engineerng for his research. Dr. Mu Wang completed this PhD thesis and joined Google Corporation. Ahmed Al-Herz defended his PhD thesis on vertex-weighted matching, and has joined the King Fahd University in Dhahran, Saudi Arabia as an Assistant Professor.

   Five of the eight current PhD students of the PI have spent summer internships at the DOE Labs, at PNNL and Lawrence Berkeley Lab. Dr. Ariful Azad, a former PhD student who was supported in earlier years on a DOE grant, spent several years as a staff scientist at Lawrence Berkeley Lab, and is now an Assistant Professor at Indiana University. Ahmed Al-Herz defended his PhD thesis on vertex-weighted matching, and has rejoined the King Fahd University in Dhahran, Saudi Arabia.

4. **How have the results been disseminated to communities of interest?** The results have been disseminated mainly through publications in journals such as Acta Numerica, SIAM Journal on Scientific Computing, Numerical Linear Algebra with its Applications, IEEE Transaction on the Smart Grid, IEEE Transactions on Parallel and Distributed Systems, etc. They have also been presented at major SIAM conferences (CSE, Parallel Processing, Combinatorial Scientific Computing) ACM conferences such as Supercomputing and IEEE conferences such as IPDPS. Keynote and invited talks were given on these talks at a number of these conferences and at institutions. Copies of publications are also available from the PI's website. Software implementing the algorithms have also been disseminated via Github (ReverseAD for computing high order derivatives; MatchBox for computing several variant matching problems), from the PI's web page (b-matching), and through direct contact. The codes are being used at Pacific Northwest National Lab, Netflix, and have also been incorporated into other software libraries such as ADOL-C.

PRODUCTS DETAILS

PUBLICATIONS

We have acknowledged DOE support on all of these publications.

(a) **Journal articles**

i. Pothen, Alex and Ferdous, S. M. and Manne, Fredrik. (2019). *Approximation algorithms in combinatorial scientific computing*. Acta Numerica. 28, pp. 541-633.

ii. Al-Herz, Ahmed and Pothen, Alex. (2019). *A 2/3-approximation algorithm for vertex-weighted matching*. Discrete Applied Mathematics, 22pp. Published online October 2019. doi:10.1016/j.dam.2019.09.013

iii. Yeung, Yu-Hong and Pothen, Alex and Crouch, Jessica. (2020). *AMPS: Real-time mesh cutting with augmented matrices for surgical simulations*. Numerical linear algebra with applications, 22 pp. Published online Sep. 2020.

iv. Dobrian, Florin and Halappanavar, Mahantesh and Pothen, Alex and Al-Herz, Ahmed. (2019). *A 2/3-Approximation Algorithm for Vertex Weighted Matching in Bipartite Graphs*. SIAM Journal on Scientific Computing. 41 (1), pp.A566–A591.

v. Yeung, Yu-Hong and Pothen, Alex and Halappanavar, Mahantesh and Huang, Zhenyu. (2017). *AMPS: An Augmented Matrix Formulation for Principal Submatrix Updates with Application to Power Grids*. SIAM Journal on Scientific Computing. 39 (5), pp. S809–S827. doi:10.1137/16M1082755

vi. Wang, Mu and Lin, Guang and Pothen, Alex. (2017). *Using automatic differentiation for compressive sensing in uncertainty quantification*. Optimization Methods and Software. pp.1–14. doi:10.1080/10556788.2017.1359267

vii. Wu, Tianying and Venkatasubramanian, Vaithianathan Mani and Pothen, Alex. (2017). *Fast Parallel Stochastic Subspace Algorithms for Large-Scale Ambient Oscillation Monitoring*. IEEE Transactions on Smart Grid. 8 (3) 1494 to 1503. doi:10.1109/TSG.2016.2608965

viii. Azad, Ariful and Buluc, Aydn and Pothen, Alex. (2017). *Computing Maximum Cardinality Matchings in Parallel on Bipartite Graphs via Tree-Grafting*. IEEE Transactions on Parallel and Distributed Systems. 28 (1), pp. 44–59. doi:10.1109/TPDS.2016.2546258

ix. Azad, Ariful and Rajwa, Bartek and Pothen, Alex. (2016). *flowVS: channel-specific variance stabilization in flow cytometry*. BMC Bioinformatics. 17 (1). doi:10.1186/s12859-016-1083-9

x. Khan, Arif and Pothen, Alex and Mostofa Ali Patwary, Md. and Satish, Nadathur Rajagopalan and Sundaram, Narayanan and Manne, Fredrik and Halappanavar, Mahantesh and Dubey, Pradeep. (2016). *Efficient Approximation Algorithms for Weighted b-Matching*. SIAM Journal on Scientific Computing. 38 (5), S593–S619. doi:10.1137/15M1026304

(b) **Refereed Conference Publications**

i. Al-Herz, Ahmed and Pothen, Alex. (2020). A Parallel *2/3-Approximation Algorithm for Vertex-Weighted Matching*. SIAM 2020 Workshop on Combinatorial Scientific Computing, pp. 12-21.

ii. Cheng, Xin and Maji, Hemanta and Pothen, Alex. (2020). *Graphs with Tunable Chromatic Numbers for Parallel Coloring*. SIAM 2020 Workshop on Combinatorial Scientific Computing, pp. 54-64.

iii. Arif Khan, Krzysztof Choromanski. (2018). *Adaptive Anonymization of Data Using b-Edge Cover*. Proceedings of ACM/IEEE Supercomputing Conference (SC18).

iv. Ye Chen, Ryan D. Calvert,Ariful Azad, Bartek Rajwa, James Fleet, Timothy Ratliff and Alex Pothen, *Phenotyping immune cells in tumor and healthy tissue using flow cytometry data*.

Proceedings of the ACM Conference on Bioinformatics and Computational Biology (BCB), pp. 73–78, Aug. 2018.

v. S M Ferdous, Alex Pothen. (2018). *New Approximation Algorithms for Minimum Weighted Edge Cover*. 2018 Proceedings of the Eighth SIAM Workshop on Combinatorial Scientific Computing. doi:10.1137/1.9781611975215.10

vi. Khan, Arif and Pothen, Alex and Ferdous, SM. (2018). *Parallel Algorithms through Approximation: b-Edge cover*. 2018 IEEE International Parallel and Distributed Processing Symposium. pp. 22-33. doi:10.1109/IPDPS.2018.00013

vii. Khan, Arif and Pothen, Alex and Patwary, Md. Mostofa and Halappanavar, Mahantesh and Satish, Nadathur Rajagopalan and Sundaram, Narayanan and Dubey, Pradeep. (2016). *Designing Scalable b-MATCHING Algorithms on Distributed Memory Multiprocessors by Approximation*. Proceedings of ACM/IEEE Supercomputing Conference (SC16). pp. 773–783. doi:10.1109/SC.2016.65

viii. Khan, Arif H. and Pothen, Alex. (2016). *A new 3/2-approximation algorithm for the b-edge cover problem*. Proceedings of the Seventh SIAM Workshop on Combinatorial Scientific Computing. pp. 52-61. doi:10.1137/1.9781611974690.ch6

ix. Mu Wang, Alex Pothen. (2016). *Edge pushing is equivalent to vertex elimination for computing Hessians*. Proceedings of SIAM Workshop on Combinatorial Scientific Computing. pp. 102-111. doi:10.1137/1.9781611974690.ch11

(c) **Software**

i. ColPack, software for graph coloring problems that enable the efficient computation of sparse Jacobian and Hessian matrices. Available at www.github.com/CSCsw/ColPack

ii. Matchbox, software for matching and edge cover problems, including maximum cardinality, vertex-weighted, edge-weighted, b-matchings; and minimum weight b-edge covers. These programs are used in sparse matrix solvers, classification problems in machine learning, graph construction and sparsification for converting unstructured data into graph-structured data. Available at www.github.com/CSCsw/MatchBox

iii. ReverseAD, software for computing high order derivative tensors via the reverse mode of Algorithmic Differentiation. Available at www.github.com/CSCsw/ReverseAD

(d) **Conference Presentations**: We have given several keynote, invited, and contributed presentations at several international conferences organized by the Society for Industrial and Applied Mathematics (SIAM), ACM, IEEE, etc. We do not list them individually here.

PARTICIPANTS DETAILS

1. Alex Pothen: PI on the project, led research in approximation algorithms for combinatorial problems, algorithmic differentation, and sparse matrix solvers for power grid and surgery simulations. Advised the PhD students, presented the work at conferences, and provided reports to DOE.

2. Arif Khan: PhD student who completed a thesis on scalable, parallel, approximation algorithms for b-matching problems and applications in data privacy, machine learning, etc. He has joined PNNL as a staff scientist.

3. Yu-Hong Yeung: PhD student who completed a thesis on fast Augmented System Solvers to update solutions of linear systems of equations arising in power grid analyses and surgery simulations. He has produced code for these problems, evaluated its performance on large power grid models as well as eye surgery models. He has presented his work at SIGGRAPH and published in top-ranked journals. He has joined PNNL as a post-doctoral fellow in physics-based machine learning.

4. Mu Wang: wrote a PhD thesis on Reverse Mode algorithms for computing high order derivatives in Automatic Differentiation, using the high order chain rule. He has produced a code, ReverseAD, and applied his work to uncertainty quantification problems. He has joined Google Corporation.

5. Siddhartha Das: is completing a thesis on scalable machine learning algorithms for problems in software security, in collaboration with colleagues at PNNL.

6. S M Ferdous: is completing a thesis on edge cover problems and applications in machine learning, load balancing in quantum chemistry on parallel computers (collaboration with the NWChem group at PNNL), and sparse matrix solvers at the extreme scale.

IMPACT

1. **What is the impact on the development of the principal discipline(s) of the project?**

   The computation of gradients, Jacobians and Hessians is vital to the solution of nonlinear problems in optimization and differential equations. The Reverse mode of Algorithmic Differentiation (AD) is known to have optimal time complexity for computing gradients. In our work, we have shown that the Reverse mode can be employed with the chain rule for higher order derivatives to obtain efficient algorithms. This approach makes it feasible to compute large-scale Hessians exploiting sparsity and symmetry, and we can compute the sparsity pattern as the derivative tensors are being computed. The computation of the sparsity pattern of the Hessian is typically the most expensive step in the Hessian evaluation for the most widely used approach that employs coloring, and the Reverse mode does not require it as a separate step. Our work also shows that the Hessians of large scale mesh optimization problems can be computed about fifteen times faster with the Reverse mode algorithm when compared with the best previous algorithms. The new algorithm achieves this while using less memory than earlier algorithms. Our code is available on Github. We are evaluating this work on test problems from several application areas such as quantum chemistry, machine learning, optimization of meshes, etc. We have developed the approximation paradigm for designing parallel algorithms. Our work introduced the first practical parallel algorithms for computing matchings in massive graphs with billions of edges. We show that by computing maximum weighted matchings approximately rather than exactly, we can obtain parallel algorithms that run in a few seconds on a parallel processor, whereas exact algorithms fail to complete in hundreds of hours. We also show that existing approximation algorithms do not possess sufficient concurrency, but that they can be modified to increase their concurrency. In the b-Matching problem, we transformed a Greedy algorithm that processes edges in decreasing order of weights (no concurrency) to a Locally Dominant edge algorithm (with more concurrency), and finally to a b-Suitor algorithm (where vertices make bids to match to each other with high concurrency). All three algorithms compute the same matching inspite of these transformations! Currently the bSuitor algorithm is the fastest bMatching algorithm on serial, shared memory, and distributed memory multiprocessors. We have scaled the b-Suitor algorithm to sixteen thousand f cores on the Cori supercomputer at NERSC. The augmented system solver, AMPS, for updating solutions to linear systems of equations in the power grid and computational surgery, is capable of providing real time updates (10-100 updates per second). This algorithm makes use of sparsity in the matrices, the right-hand-side vector and the solution.

2. **What is the impact on other disciplines?**

   The Algorithmic Differentiation work for computing Hessians would be useful in optimization problems in many areas: we have investigated quantum chemistry problems (density functional theory) where exchange correlation functions are being computed. We have also applied it to a problem in uncertainty quantification (UQ), where with AD the gradients can be computed accurately and fast, since an iterative procedure to compute the gradient is not needed. We have shown that the AD-based UQ method is ten to fifteen times faster, and also leads to sparser generalized polynomial chaos expansions on some test problems. The Augmented System solver we have developed can be used to perform contingency analyses of power grids with nearly a million nodes in a matter of seconds. The student who did this work has spent two summers working with colleagues at the Future Power Grid Initiative at PNNL, and we have collaborated with them on this work. Our bSuitor algorithm for computing b-Matchings has been shared with colleagues at Intel, Google, Netflix, Columbia, PNNL, SUNY Stony Brook, etc. They are using it to solve problems in data anonymization, recommender systems, image processing, etc.

3. **What is the impact on the development of human resources?**

   Three PhD students have completed their PhD theses with support from this project. Two of them joined Pacific Northwest National Lab, where one is staff scientist and the second is a post-doctoral fellow. A third student has joined Google. Two of these students were awarded the John R. Rice Fellowship in computational science and engineering at Purdue University in recognition of the research contributions that they have made. Three other PhD students who have been partially supported by this grant are working towards their theses at this time.

4. **What is the impact on physical, institutional, and information resources that form infrastructure?**

   We have received a grant from the Purdue Provost and have secured a 8 x 24 core Xeon machine with 6 TB memory, to study large-scale graph problems. This award was received due to the research accomplishments reported here as well as that of my colleagues in the computational science and engineering group at Purdue.

5. **What is the impact on technology transfer?**

   The codes that we are developing are made available on Github, our website for software, and through integration with other software libraries for Algorithmic Differentiation, Control, etc. such as ADOL-C, CasADI. These codes are being used by companies such as Netflix, Google, and Intel, by DOE Labs (PNNL, Berkeley, Sandia), and academic institutions (Columbia, Michigan, etc.).