2022-2023 Grand Challenge Award Final Report

Awardee: **Gunnar Martinsson, Professor**
Mathematics

Research Award Title: **Randomized Algorithms for Solving Linear Systems**

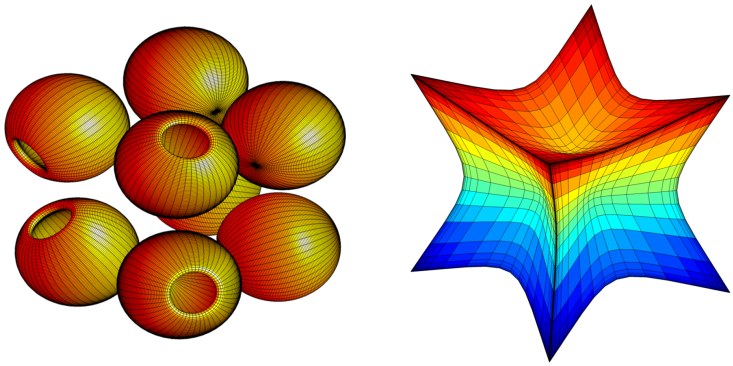
Research Summary A core challenge in scientific computing is that many of the operators we need to work with are global in nature. In other words, every point of the computational domain directly communicates with every other point. Examples include: (1) The solution operators to the partial differential equations that govern heat conduction, electrostatics, time harmonic wave propagation, and the deformation of solid bodies. (2) Scattering matrices in the modeling of radar, sonar, and medical imaging. (3) Surface-to-surface operators such as the map that converts a set of applied loads to a solid body to the resulting deformation of its surface.

When such global operators are discretized, they would typically take the form of a dense matrix, which greatly limits the size of problems that can be handled. For instance, a problem with just N=1E5 degrees of freedom would require 75GB of memory merely to store the corresponding dense matrix. Going to N=1E6 would require 7.5TB, at which point almost nothing can be done in terms of computational modeling. In order to model complex three dimensional geometries, far larger numbers of degrees of freedom are required.

The prevailing paradigm for dealing with global operators in large scale simulations is to never form them explicitly. Instead, indirect techniques for applying them to vectors are combined with iterative methods to solve equations that arise, to compute approximate eigenvectors, etc. In contrast, the Grand Challenge project led to the discovery of new techniques that explicitly compute data sparse representations of the operators, and to new algorithms for directly inverting and factorizing them.

This work drew on existing methodologies such as the Fast Multipole Method of Rokhlin and Greengard, and the H-matrix arithmetic of Hackbusch. However, we significantly expanded the range of problems that can be handled, accelerated the existing methods, and improved numerical accuracy and stability.

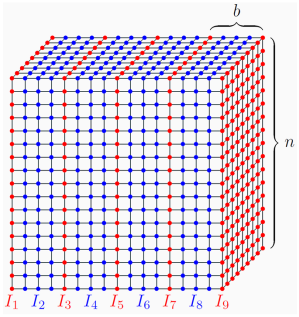
O(N) direct solvers for BIEs. Major progress was made on building scalable *O(N)* direct (as opposed to iterative) solvers for the dense linear systems arising from the discretization of boundary integral equations for problems in three dimensions. These new solvers were coupled with high order discretizations and were designed to be capable of solving problems involving challenging geometries.

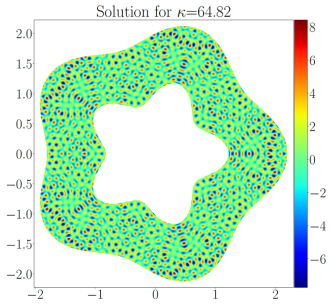


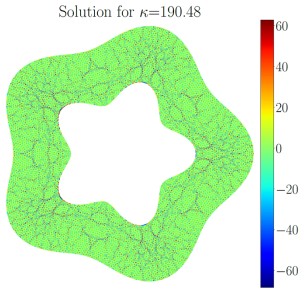
The novelty in the work is in the high performance computing aspect, where we developed new GPU accelerated randomized algorithms for constructing the data sparse representations, and a new distributed memory version of the "strong recursive skeletonization solver" introduced by Minden, Ho, Damle, & Ying in 2017, and later improved by a group involving D. Sushnikova, L. Greengard, M. O'Neil, & M. Rachh at the Flatirons Institute in New York City. A version of the solver that works for simple geometries is under review [9], and the full 3D code that involves high order quadratures will be submitted shortly. To the best of our knowledge, this is the first solver with true linear scaling that has ever been published.

Block nullification: An O(N) completely black box algorithm. We have discovered a linear complexity algorithm for solving the black box compression problem for *hierarchically block separable (HBS)* matrices (a.k.a. H^2-matrices with weak admissibility). The method is entirely black box and is ideal for the acceleration of sparse direct solvers. High practical speed is attained even at very high compression precision (five or ten correct digits). It relies on two new ideas, *block nullification and block extraction* [7, 8]. *Randomized strong recursive skeletonization (RSRS) – Linear complexity simultaneous compression and inversion.* For truly large scale problems in three dimensions, data structures similar to those used in the Fast Multipole Methods (FMM) must be deployed (in technical language, "H^2 matrices with strong admissibility"). A key challenge to using this format is that the only known linear complexity inversion algorithm (*strong recursive skeletonization (SRS)* by Minden, Ho, Damle, and Ying) requires the explicit construction and storage of a large part of the matrix. This slows the method down, and makes it very memory intense. Our new method RSRS overcomes this challenge by working exclusively with *random sketches* of the matrix, which obviates the need to explicitly build the modified blocks. RSRS attains high practical speed, and is very well suited to parallelization. Ongoing work incorporates RSRS into sparse direct solvers for elliptic PDEs in 3D that are intractable for iterative methods.

SlabLU: High order domain decomposition. Our new black box compression scheme "RSRS" for H^2-matrices has been exploited to build an *O(N)* direct solver for elliptic PDEs with variable coefficients on rectangular domains (or domains that can be parameterized over a union of rectangles).







Our solver SlabLU has the following characteristics:

- A very high order ($p = 10$ or $p = 20$ are typical) local spectral collocation discretization is used (the "hierarchical Poincaré-Steklov (HPS scheme). Ten correct digits can often be attained.
- The domain is partitioned into thin "slabs".
- The thinness of each slab is exploited by a local sparse direct solver to eliminate the interior nodes in each slab (blue in the figure).
- RSRS is used to build and factorize the Schur complements living on the slab interfaces (red).
- RSRS is then again used to solve the resulting block tridiagonal system.

The 2D version can handle a domain of size 1 000λ×1 000λ (using 100M dofs) on a desktop [11]. Time to factorize is 10 minutes, and time to solve is 40 seconds. We obtain three digits of relative accuracy *in the solution.* A 3D version distributed memory version of the method is under development.

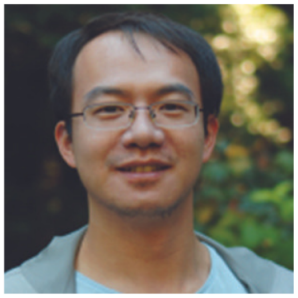
À posteriori error bounds for randomized algorithms. A key outcome of the Grand Challenge award is a new theoretical analysis of a class of randomized algorithms in linear algebra that provides estimates of the error incurred in a specific instantiation of a randomized algorithm [2]. Our new error bounds rely only on information that is available to the user at the time of computation. (In contrast to previously existing theorems that involve unknown quantities such as the exact singular values.)

New funded project. During the project year, the PI visited the University of Oxford several times to work with collaborator Yuji Nakatsukasa. This work led to the new à posteriori error estimates described in the previous paragraph, and also the successful application of a joint grant from the NSF and the EPSRC (the UK equivalent of the NSF) that will support our continued work over the next three years.

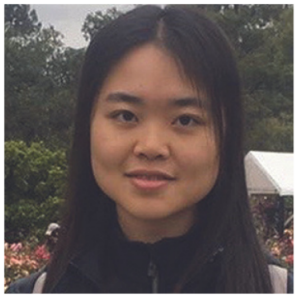
Anna Yesypenko
PhD '23

Yunhui Cai
PhD '25

Heather Wilber
PDR



Chao Chen
PDR



Yijun Dong
PhD '23

Kate Pearce
PDR

Research group. The research conducted during the course of project benefited the members of the Martinsson research group. The successful work resulting from the Grand Challenge award during the 2022/23 years greatly helped postdocs Heather Wilber and Chao Chen in securing tenure track assistant professorships at the University of Washington, and at North Carolina State University, respectively. Dr. Pearce used the work as a foundation for her successful application for an NSF Ascend postdoctoral fellowship, which will support her work through August 2027. Yesypenko and Dong both defended their dissertations shortly after the project concluded, drawing in an essential way on the joint research. Dong is currently a Courant Instructor at NYU, and Yesypenko is a postdoctoral scholar at Oden working with Drs. Biros and Moser.

Presentations

The PI presented the research at national and international venues:

- Mini-symposium talk: SIAM MDS Conference in San Diego, September 2022.
- Plenary presentation: 4^{th} Workshop on Scientific Computing in Gothenburg, Sweden, October 2022.
- Talk: Workshop at the Oden Institute and the Alan Turing Institute in London, UK, January 2023.

- Mini-symposium talk: SIAM CSE Conference in Amsterdam, Netherlands, February 2023.
- Numerical Analysis Seminar at the University of Oxford, UK, May 2023.
- Semi-plenary lecture: Foundations of Computational Mathematics conference, Paris, France, June 2023.
- Workshop lecture: Modern Applied and Computational Analysis, Brown University, June 2023.
- Invited plenary lecture: Numerical Analysis in the 21st century, University of Oxford, UK, August 2023.
- Mini-symposium talk: International Congress on Industrial and Applied Mathematics, August 2023.

Publications

- In addition to works previously referenced, manuscripts 10, 5, 6, 4, 3, and 1 were completed.

References

- C. Chen AND P.-G. Martinsson, Solving linear systems on a gpu with hierarchically off-diagonal low-rank appr in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, SC '22, IEEE Press, 2022.
- Y. Dong, P.-G. Martinsson, AND Y. Nakatsukasa, Efficient bounds and estimates for canonical angles in rando 2022.
- A. Gopal AND P.-G. Martinsson, An accelerated, high-order accurate direct solver for the lippmann–schwinger Advances in Computational Mathematics, 48 (2022), pp. 1–31.
- Broadband recursive skeletonization, in Spectral and High Order Methods for Partial Differential Equations ICOSA-HOM 2020+1, J. M. Melenk, I. Perugia, J. Schoberl, and C. Schwab, eds., Cham, 2023, Springer International Publishing, pp. 31–66
- N. Heavner, C. Chen, A. Gopal, AND P.-G. Martinsson, Efficient algorithms for computing rank-revealing facto Numerical Linear Algebra with Applications, n/a, p. e2515.
- N. Heavner, P. G. Martinsson, AND G. Quintana-Ortì, Computing rank-revealing factorizations of matrices sto Concurrency and Computation: Practice and Experience, n/a, p. e7726.
- J. Levitt AND P.-G. Martinsson, Linear-complexity black-box randomized compression of hierarchically block s 2022.
- J. Levitt AND P.-G. Martinsson, Randomized compression of rank-structured matrices accelerated with graph 2022.
- T. Liang, C. Chen, P.-G. Martinsson, AND G. Biros, A distributed-memory parallel algorithm for discretized in 2023.

- B. Wu AND P.-G. Martinsson, A unified trapezoidal quadrature method for singular and hypersingular bounda SIAM Journal on Numerical Analysis, 61 (2023), pp. 2182–2208.
- A. Yesypenko AND P.-G. Martinsson, Slablu: A sparse direct solver for elliptic pdes on rectangular domains, 2022.