Solving linear systems is a fundamental task in numerical linear algebra because of the wide range of applications to applied fields such as the sciences, medicine, and economics. Recently, there has been a rapid increase in the amount of data which scientists are able to collect and store. As a result, the linear systems which scientists now seek to solve have also been increasing in size. Iterative methods are often the only tractable way to deal with such large systems, and Krylov subspace methods are among the most successful and widely used iterative methods [1, 2]. However, the standard techniques developed years ago are no longer sufficient for many of today's applications. As such, **new iterative methods, designed explicitly to deal with high-dimensional data, are required to handle the problems scientists now seek to solve.**

Krylov subspace methods for linear systems (henceforth referred to as Krylov methods[1]) iteratively find an approximate solution to the $n \times n$ linear system $Ax = b$ from a sequence of nested subspaces (called Krylov subspaces). In at most n steps the sequence of Krylov subspaces will contain the exact solution. This means that in exact arithmetic, Krylov methods converge in at most n iterations. However, **since in practice finite precision arithmetic is used, the convergence of Krylov methods is usually very different than in exact arithmetic.** When analyzing the numerical convergence of iterative methods important considerations are (i) how long a single iteration takes (ii) how many iterations are required to reach a given level of accuracy and (iii) what is the highest level of accuracy can eventually be attained. These tell us about how long it will take to solve a given problem and how good our solution will be, three critical pieces of information for the scientists applying such methods in the course of their research. **The primary objectives of this proposal are (i) develop parallelized CG methods whose convergence properties are desirable and well understood, and (ii) develop a framework for approximate Krylov methods**.

My current research with Anne Greenbaum focuses on the first objective. Recent work has aimed to reduce the runtime of Krylov methods such as conjugate gradient (CG) and GMRES through parallelization [3, 4, 5]. For a given method, the parallel variants are equivalent to the nonparallel variants in exact arithmetic. However, this is no longer true in finite precision arithmetic where the behavior can be very different. In fact, the parallel CG algorithms can converge much slower and to a lower degree of accuracy than the standard (non-parallel) implementation.

Our current approach is to first understand what criteria will ensure good convergence properties, and then, once these criteria are isolated, develop a method which satisfies them. To determine sufficient criteria for a "good" method we have been running experiments comparing the results of various methods when applied to linear systems arising in structural engineering and fluid dynamics. These tests allow us to look at various indicators such as the residual $b - Ax_k$ for each of the methods. At the moment we have isolated potential criteria based on previous work by Greenbaum [6]. We are working to prove that certain methods satisfy these critera, and that others do not. However, since these critera are quite strong and are independent of the system being solved. This means there is room to find weaker sufficient and problem dependent critera which would allow methods to be chosen on a system by system basis.

The second objective is based on the observation that, while in exact arithmetic Krylov methods minimize some quantity over successive Kyrlov spaces, this is not the case in finite precision arithmetic. Since these quantities are never actually minimized in practice, a reasonable question to ask is if it is necessary to try to exactly minimize them in the first place. Broadly speaking, the class

[1]Krylov subspace methods can be used for the more general problem of computing $f(A)b$. We have limited the scope of the proposal to the special case $f(A) = A^{-1}$ for solving linear systems.

of approximation algorithms I am proposing will only approximately minimize these quantities at each step in order to decrease the computational cost per iteration. For instance certain projections and matrix products could be approximated by using sub-sampling techniques [7]. I hope to decrease the overall compute time without sacrificing the final accuracy by properly balancing how many more iterations are required with how much faster each iteration can be run.

OTHER RESOURCES: The majority of my intended work will be theoretical algorithm design. However, having access to supercomputer time on through XSEDE would provide means of testing the algorithms on data too large for local machines.

INTELLECTUAL MERITS: The proposed work on parallelized Krylov methods for linear systems will provide insights into how to choose the "right tool for the job". **At the moment there is no clear consensus on which, if any, parallel variants of Krylov subspace methods should be chosen for a given system.** Finding critera which would allow researchers to pick a Krylov subspace solver based on the problem they are trying to solve would be an important contribution to numerical analysis. Similarly, creating a framework to develop approximation methods for linear systems based on Krylov subspace methods has the potential to provide insights into how other classes of algorithms can be sped up using randomization.

Past experiences working on interdisciplinary projects has convinced me of the necessity of cross-discipline collaboration. My background in computer science and probability will help me to construct novel new Krylov subspace methods, while my background in physics will allow me to be able to collaborate directly with researchers in the physical sciences to understand how best to apply our developments in algorithm design to their problems.

BROADER IMPACTS: The overarching motivation for this project is the fact that **the less time scientists have to wait for code to run, the more time they can spend thinking about the problems they are tackling.** When working in computational physics, I saw first hand how the faster algorithm I introduced allowed the group to rapidly test new hypotheses, resulting in faster model validation. Faster methods means that the time researchers currently spend waiting for for the results of large computations will be able to be spent analyzing those results of these computations.

In particular, Krylov subspace methods are widely used for solving linear systems too large for direct methods. Among other sources, systems with hundreds of millions or even billions of equations commonly are used in nonlinear solvers (such as Newton type methods), and arise in from the analysis of circuit architecture, and the discretization of partial differential equations. Currently, high precision electrodynamics simulations, or large scale atmospheric simulations can often day days or even weeks to run. **Speeding up such computations has the potential to immediately facilitate tasks such as the development of better weather forecasting, cleaner energy, and more efficient traffic networks.**

References: [1] A. Greenbaum. *Iterative methods for solving linear systems*. Vol. 17. Siam, 1997. [2] C. Musco, C. Musco, and A. Sidford. "Stability of the Lanczos Method for Matrix Function Approximation". In: *CoRR* (2017). [3] E. De Sturler and H. A. van der Vorst. "Reducing the effect of global communication in GMRES (m) and CG on parallel distributed memory computers". In: *Applied Numerical Mathematics* 18.4 (1995). [4] A. T. Chronopoulos and C. W. Gear. "On the efficient implementation of preconditioned s-step conjugate gradient methods on multiprocessors with memory hierarchy." In: *Parallel Computing* 11 (1989). [5] P Ghysels and W. Vanroose. "Hiding global synchronization latency in the preconditioned Conjugate Gradient algorithm". In: *Parallel Computing* (2013). [6] A. Greenbaum. "Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences". In: 113 (Feb. 1989). [7] N. Halko, P. Martinsson, and J. Tropp. "Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions". In: *SIAM Review* 53.2 (2011).