

Advancing Nuclear Physics Through TOPS Solvers and Tools

E Ng¹, J Sarich², S M Wild², T Munson², H Aktulga¹, C Yang¹,
P Maris³, J P Vary³, N Schunck⁴, M G Bertolli⁵, M Kortelainen^{5,6},
W Nazarewicz^{5,6}, T Papenbrock^{5,6}, M V Stoitsov^{5,6}

¹ Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

² Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA

³ Department of Physics, Iowa State University, Ames, IA 50011, USA

⁴ Physics Division, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

⁵ Department of Physics and Astronomy, University of Tennessee, Knoxville, TN 37996, USA

⁶ Physics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

E-mail: egng@lbl.gov

Abstract. At the heart of many scientific applications is the solution of algebraic systems, such as linear systems of equations, eigenvalue problems, and optimization problems. TOPS, which stands for Towards Optimal Petascale Simulations, is a SciDAC applied math center focused on the development of solvers for tackling these algebraic systems, as well as the deployment of such technologies in large-scale scientific applications of interest to the U.S. Department of Energy. In this paper, we highlight some of the solver technologies we have developed in optimization and matrix computations. We also describe some accomplishments achieved using these technologies in UNEDF, a SciDAC application project on nuclear physics.

1. Introduction

Over the past couple of decades, simulation science has become as important as theoretical and experimental science. The success of simulation sciences hinges on the ability to perform the calculations efficiently. The innermost kernel in these calculations is often the solution of algebraic systems, including systems of linear and nonlinear equations, eigenvalue problems, optimization problems, and sensitive analysis. TOPS, which stands for *Towards Optimal Petascale Simulations*, is a multi-institutional SciDAC applied math center that focuses on the development of solvers for tackling these algebraic systems, as well as the deployment of such technologies in large-scale scientific applications, particularly those of interest to the U.S. Department of Energy.

In this paper, we highlight two areas of TOPS: eigenvalue calculations and optimization. In particular, we highlight some accomplishments we have achieved in collaboration with computational physicists in UNEDF. The goal of the UNEDF SciDAC application project is to obtain a comprehensive understanding of nuclei and their reactions based on the most accurate knowledge of the strong nuclear interaction [1]. Eigenvalue calculations come up in the solution of the nuclear Schrödinger equation [2, 3]. The eigenvalues and the eigenvectors correspond to the energy states and wave functions. Numerical optimization techniques are needed in building the next generation of nuclear energy functionals, which will provide nuclear physicists better tools for predicting the properties and behavior of atomic nuclei over the entire nuclear table.

2. Eigenvalue Calculations

One type of nuclear configuration interaction calculation, which we refer to in this paper as a *total- J calculation*, is particularly useful for investigating nuclear-level densities as a function of the total angular momentum J and excitation energy and for evaluating scattering amplitudes as a function of channel J [4]. In this type of calculation, we are interested in computing a relatively large number of low energy states with a prescribed J value.

One brute-force approach to a total- J calculation is to simply compute a large number of eigenvalues and wave functions of a nuclear many-body Hamiltonian, for example in an M-scheme basis (good angular momentum projection along the z-axis), and select, among these wave functions, the ones that have a prescribed J value. This approach is appropriate when the number of desired energy states and wave functions is small (e.g., ten to twenty states). When that is not true, or when certain properties of a nucleus pertaining to a fixed J are to be calculated, the brute-force approach may require computing a very large number of wave functions, and the computational cost for performing this type of calculation may be prohibitively high. Furthermore, even if we can afford to perform this type of calculation, this may not be an efficient use of resources because we compute a large number of wave functions only to throw away most of them because they do not have the desired J value.

We have developed an alternative approach where we construct an invariant subspace Z that contains all wave functions associated with a fixed J value in advance and project the nuclear many-body Hamiltonian into this subspace to produce a projected Hamiltonian with the minimum dimension consistent with that chosen J . A sparse matrix diagonalization procedure [5, 6, 7] is then applied to this projected Hamiltonian to obtain the desired energy states and their corresponding wave functions.

To construct Z , we need to work with the total angular momentum square operator \hat{J}^2 and compute the null space of $\hat{J}^2 - \lambda I$, where $\lambda = J(J + 1)$ is a known eigenvalue of \hat{J}^2 .

When the many-body basis states associated with the configuration space are properly ordered and grouped, \hat{J}^2 becomes block diagonal: $\hat{J}^2 = \text{diag}(\hat{J}_1^2, \hat{J}_2^2, \dots, \hat{J}_{n_g}^2)$. Therefore, the task of computing the desired null space of $\hat{J}^2 - \lambda I$ reduces to that of computing the desired null spaces of $\hat{J}_i^2 - \lambda I$, for $i = 1, 2, \dots, n_g$.

However, because the dimensions of \hat{J}_i^2 's vary in a wide range (e.g., from 1 to more than 36,000 for ^{12}C , $N_{\text{max}}=6$).¹ it is difficult to maintain a good load balance in the null space calculation.

To achieve optimal parallel performance in the null space calculation, we developed a multilevel task and data distribution scheme with the following goals:

- (i) Limiting the granularity of the parallelism; that is, we try to divide the overall task into many small tasks of limited sizes so that good load balance arises from distributing these small tasks evenly among different processors.
- (ii) Limiting the communication overhead incurred in the null space calculation so that the overall time of the computation can be minimized.

To achieve these inherently conflicting goals, we classified \hat{J}_i^2 blocks into small, medium, and large groups based on the estimated computational loads associated with computing the desired null space of $\hat{J}_i^2 - \lambda I$ and the estimated ratio of communication volume to floating-point operations count.

The small \hat{J}_i^2 blocks are distributed by a greedy algorithm among all processors based on their computational load. The null spaces of these matrices are computed by a sequential LAPACK rank-revealing QR subroutine. No communication is involved in these calculations. Each of the medium-sized \hat{J}_i^2 blocks is assigned to a subgroup of processors by the same greedy

¹ N_{max} is a parameter limiting the total number of oscillator quanta allowed in the many-body states.

algorithm. The null space calculation for such a block is parallelized among processors within the same subgroup, which will incur some communication overhead. Finally, the desired null space calculation for a large \hat{J}_i^2 block is carried out in parallel on all processors.

We implemented three algorithms for computing the null space of $\hat{J}_i^2 - \lambda I$ for medium and large blocks.

- (i) Randomized rank-revealing (RQR). The algorithm performs two standard QR factorizations of dense matrices without pivoting. Although we do not take advantage of the sparsity of \hat{J}_i^2 in this approach, it is more efficient than other approaches when the dimension of the desired null space is relative large (e.g., larger than 10% of the dimension of \hat{J}_i^2).
- (ii) Shift-invert Lanczos (SIL), which requires solutions of sparse linear systems.
- (iii) Polynomial accelerated subspace iteration (PASI). We apply a standard subspace iteration [8] to the matrix $p(\hat{J}_i^2)$, where $p(\omega)$ is a polynomial that assumes the value of 1 at $\omega = \lambda$, and has a much smaller magnitude (than 1) in other parts of the spectrum of \hat{J}_i^2 .

Table 1 shows that our load-balance scheme is much better than a brute-force approach of distributing \hat{J}_i^2 in a cyclic fashion to different processors. Table 2 shows that PASI is more efficient when $J=0$. The randomized QR algorithm appears to be more efficient for larger J values. However, when J becomes very large, which typically results in smaller dimension of the null space, PASI becomes more efficient again.

Table 1. Comparison between the greedy load-balancing algorithm with a parallel algorithm based a cyclic distribution of \hat{J}_i^2 blocks in terms of wall clock time (in secs). (n_p is the number of processors.)

Core	N_{\max}	Alg.	n_p	Cyclic	Greedy
${}^6\text{Li}$	12	PASI	120	131	132
${}^{12}\text{C}$	4	PASI	120	6.1	5.2
${}^{12}\text{C}$	6	PASI	496	608	295
${}^6\text{Li}$	12	RQR	120	233	193
${}^{12}\text{C}$	4	RQR	120	18.7	17.0
${}^{12}\text{C}$	6	RQR	496	1220	900

Table 2. RQR decomposition vs PASI for different J values. Both methods use the greedy load-balancing technique. Times are in secs.

Core	(N_{\max}, J)	n_p	RQR	PASI
${}^6\text{Li}$	(12, 0)	120	193	132
${}^6\text{Li}$	(12, 1)	120	195	464
${}^6\text{Li}$	(12, 12)	496	140	95
${}^{12}\text{C}$	(6, 0)	496	900	295
${}^{12}\text{C}$	(6, 1)	496	890	> 1,800
${}^{12}\text{C}$	(6, 12)	496	840	105

3. Optimization

Optimization plays a central role in the building of next-generation nuclear energy functionals, including functionals based on density functional theory (DFT) and/or ab initio calculations. In the case of DFT-based functionals, for example, a primary computational bottleneck is determining parameter values so that the functional agrees with data on a set of observables such as binding energies, radii, and odd-even staggering [9]. Mathematically, we need to solve the optimization problem

$$\min_x \left\{ f(x) = \sum_{i=1}^o \left(\frac{d_i - s(\theta_i; x)}{\sigma_i} \right)^2 : l_j \leq x_j \leq u_j, j = 1, \dots, n \right\}, \quad (1)$$

where n parameter values must be determined from a set of data of o observables. Challenges in solving this problem include the computational expense of, and the noise resulting from, the

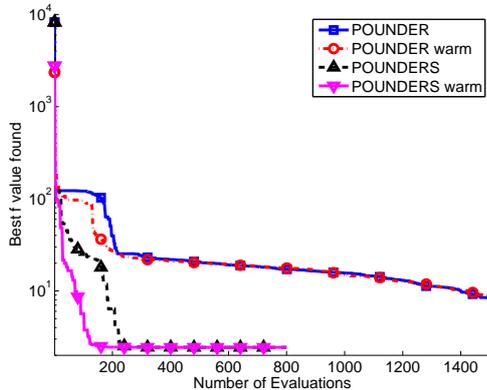


Figure 1. Exploiting structure in parameter estimation problems substantially reduces the number of required simulation evaluations.

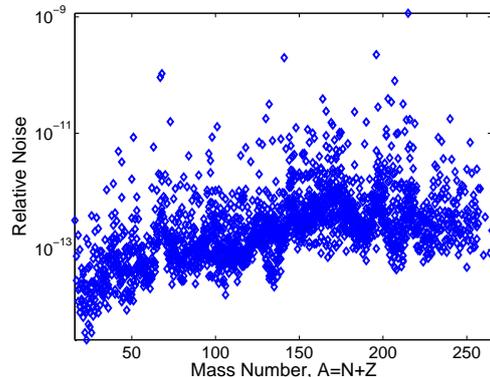


Figure 2. Quantifying the computational noise for a deterministic ab initio simulator across 2,049 nuclei.

iterative calculations performed when simulating the theoretical observable $s(\theta_i; x)$, and the fact that derivatives of some simulated observables with respect to certain parameters x_j may not be available (or even exist) for use by an optimization algorithm.

As part of SciDAC efforts, we have developed POUNDERS (Practical Optimization Using No DERIVATIVES for Sums of squares), an algorithm for derivative-free optimization of nonlinear least-squares problems such as (1). A key benefit of POUNDERS is that it works with the individual residuals $d_i - s(\theta_i; x)/\sigma_i$ rather than the aggregated fit function $f(x)$. As a result, POUNDERS can take advantage of the availability of the derivatives of some observables (e.g., binding energies) and can approximate nonlinearities in f using simulations at fewer x values. As part of the TOPS collaboration, POUNDERS is now available through the open-source Toolkit for Advanced Optimization (TAO) [12].

Figure 1 quantifies the computational savings in this ability to exploit the sums of squares structure in (1) for a fit to 2,049 binding energies. By working with the residuals, the POUNDERS variants obtain far better fits in far fewer evaluations than the analogous variants of POUNDER, a similar algorithm that does not have access to the residuals. The *warm* variants illustrate the benefit of using external simulations, done as part of an initial experimental design, to warm start the optimization.

The savings in Figure 1 can be substantial. For the more complex functional optimized in [10], each evaluation of f requires 14.4 CPU-hours. The resulting parameterization is then used to perform a simulation of nuclei across the mass-table in a calculation requiring 9,000 processors for more than half a day [9].

Mathematical work has also contributed to the sensitivity analysis of nuclear energy functionals. Though the simulations are typically deterministic, the aforementioned computational noise can obfuscate the number of reliable digits in computed functional properties. The ECNoise algorithm described in [11] estimates a standard deviation-like quantity using only a few simulations. Figure 2 illustrates the relative noise in the computed binding energies for 2,049 nuclei with the parameterization obtained from the POUNDERS optimization in Figure 1. Estimates of the noise can, for example, reveal limitations on the predictability of computed functional observables and can enable stable approximations of the noisy derivatives needed for sensitivity analysis.

4. Conclusions

Our work on eigenvalue calculations has had several benefits for nuclear structure calculations. Earlier collaborations with nuclear physicists led to significant improvements to an eigensolver for configuration interaction calculation, which was subsequently used in predicting the properties of ^{14}F before the isotope was observed experimentally. The work described in this paper takes configuration interaction calculation one step further. It enables our physics collaborators to efficiently compute energy states of a nuclei with a prescribed total angular momentum instead of computing many energy states and identifying those corresponding to the prescribed total angular momentum.

The optimal parameters we have delivered to our physics collaborators have resulted in realistic functionals that are now being explored by a variety of groups outside of the UNEDF collaboration (as evidenced in the most recent JUSTIPEN conference). For example, our current results show remarkable power for predicting fission barrier heights, which is a first step toward a microscopic understanding of fission. These results are a consequence of including a richer set of experimental data and more free parameters, resulting in problems that can be solved only by an efficient, state-of-the-art optimization algorithm.

Acknowledgments

The work under TOPS at ANL and LBNL was supported by the Office of Advanced Scientific Computing Research of the U.S. Department of Energy under contracts DE-AC02-06CH11357 (ANL) and DE-AC02-05CH11231 (LBNL). The UNEDF SciDAC collaboration was supported by the U.S. Department of Energy under grant numbers DE-FC02-09ER41582. This work was also supported by the Office of Nuclear Physics of the U.S. Department of Energy under grant numbers DE-FG02-87ER40371 (Iowa State), DE-FG02-07ER41529 (Univ. of Tennessee) and DE-FG02-96ER40963 (Univ. of Tennessee).

Computational resources were provided through an INCITE award (James Vary, PI) at ORNL and ANL, and by the Laboratory Computing Resource Center (LCRC) at ANL, the National Energy Research Scientific Computing Center (NERSC) at LBNL, and the National Center for Computational Sciences at ORNL.

References

- [1] See <http://www.unedf.org>.
- [2] J. P. Vary, P. Maris, E. Ng, C. Yang and M. Sosonkina. Ab initio nuclear structure – the large sparse matrix eigenvalue problem. *Journal of Physics: Conference Series*, 180:012083, 2009.
- [3] P. Sternberg, E. G. Ng, C. Yang, P. Maris, J. P. Vary, M. Sosonkina, and H. V. Le. Accelerating configuration interaction calculations for nuclear structure. In *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing* (Austin, Texas, November 15 - 21, 2008). Conference on High Performance Networking and Computing. IEEE Press, Piscataway, NJ, 1-12.
- [4] A. M. Shirokov, A. I. Mazur, S. A. Zaytsev, J. P. Vary and T. A. Weber. Nucleon-nucleon interaction in the J -matrix inverse scattering approach and few-nucleon systems. *Phys. Rev. C*, 70:044005-1–044005-23, 2004.
- [5] R. B. Lehoucq, D. C. Sorensen, P. Vu, and C. Yang. *ARPACK: An implementation of the Implicitly Re-started Arnoldi Iteration that computes some of the eigenvalues and eigenvectors of a large sparse matrix*, 1995. Available from <ftp.caam.rice.edu> under the directory `pub/software/ARPACK`.
- [6] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *ARPACK Users Guide: Solution of Large Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods*. SIAM, Philadelphia, PA, 1998.
- [7] D. C. Sorensen. Implicit application of polynomial filters in a k -step Arnoldi method. *SIAM Journal on Matrix Analysis and Applications*, 13(1):357–385, January 1992.
- [8] B. N. Parlett. *The Symmetric Eigenvalue Problem*. Prentice-Hall, 1980.
- [9] M. Stoitsov, J. Moré, W. Nazarewicz, J.C. Pei, J. Sarich, N. Schunck, A. Staszczak, and S. Wild. *J. Phys.: Conf. Series* **180** 012082, 2009.
- [10] M. Kortelainen, T. Lesinski, J. Moré, W. Nazarewicz, J. Sarich, N. Schunck, M. V. Stoitsov, and S. Wild. *Phys. Rev. C* **82** 024313, 2010.
- [11] J. Moré and S. Wild. Estimating computational noise *SIAM J. Sci. Comput.*, 2011, to appear.
- [12] S. Benson, L. Curfinan McInnes, J. Moré, T. Munson, and J. Sarich. Toolkit for Advanced Optimization, <http://www.mcs.anl.gov/tao>.