Efficient Temporal Integration at The Exascale

Michael L. Minion
Lawrence Berkeley National Lab and Stanford University

Overview: The development of exascale computers will allow researchers to run numerical simulations for new classes of problems that are currently out of reach of the largest machines available. For problems whose mathematical description is given by a partial differential equation (PDE), exascale computing will not only enable larger simulations of existing models, but also enable more physically consistent modeling of applications characterized by multiple space and time scales and/or multiple physical processes. For such applications, increased computational power may not alone be sufficient, rather new efficient parallel algorithms must also be developed both in terms of spatial and temporal discretizations. Iterative temporal integration methods based on deferred corrections have many appealing characteristics for exascale applications and provide ample opportunities for mathematical and computational research in the coming years.

Iterative Temporal Integration Based on Deferred Corrections

Deferred and defect correction methods for ODEs first appeared in the 1960s (e.g. [1, 2]), but for many years did not enjoy the popularity of Runge-Kutta or linear-multistep methods. More recently these methods have seen renewed interest due to several appealing characteristics:

Accuracy: Higher-order spatial discretizations are attractive at the exascale since fewer degrees of freedom are needed to achieve the same error tolerance as lower order methods. For PDEs, higher-order spatial discretizations require higher-order methods in time as well, and spectral deferred correction (SDC) methods [3] based on Gaussian quadrature rules can easily be constructed to have very high formal order of accuracy for both stiff and non-stiff equations with good stability properties.

Flexibility: One of the attractive features of iterative methods like SDC is that higher-order accuracy is attained through repeated use of lower-order methods. This provides great flexibility in terms of how the temporal discretization is designed. In particular, time-marching schemes that treat different terms either explicitly or implicitly depending on type [4, 5], or with different time steps for different terms [6], or even utilizing different numerical codes for different physical terms [7] have been developed for specific cases.

Parallelizability: With core counts on exascale machines expected to be on the order of billions, increasing concurrency in PDE simulations while avoiding low computation to communication ratios is a major challenge. While increasing the number of spatial degrees of freedom in a PDE simulation can increase the concurrency, it also typically requires a reduction in time step, leading to longer run times. Hence the last decade has seen a renewed effort to develop strategies for time parallelization of ODEs and PDEs (e.g. [8, 9]). Time parallel methods based on deferred corrections have recently appeared [10, 11, 12, 13]. Very recent results have shown promise in combining spatial and temporal parallelism on

$O(10^6)$ cores using the parallel full approximation scheme in space and time (PFASST) [12].

Resiliency: Iterative temporal integration methods may also be useful for ensuring resilient temporal integration in the presence of soft and/or hard errors. Since SDC provides explicitly computable residuals in each time step, unexpected convergence behavior could provide one diagnostic for floating point error. Another advantage is that if an error is detected, iterative temporal methods can be restarted from the last known error free iteration rather than starting from scratch.

Mathematical Challenges

Despite the desirable features outlined above, iterative temporal methods present a host of unresolved questions requiring mathematical and computational analysis. Given a particular application, it is not clear how to choose the most efficient combination of formal order, time-step, and number of iterations. Furthermore, there may be multiple options for operator splitting, multirate integration, and multi-level operator coarsening available. In terms of implementation, some procedures to choose the optimal mix of spatial and temporal parallelization need to be developed, and these may change dynamically in a given simulation. In [14] a procedure is introduced that allows the overlap of communication in PFASST with computation, however this procedure is based on a highly synchronous model which is unlikely to be realized on exascale machines. The optimal choices in each instance will clearly be problem dependent, but mathematical analysis can produce broad guidelines to best exploit a given architecture.

For PDE simulations, it is possible to construct SDC methods that use a hierarchy of spatial discretizations within the iterations to reduce the computational cost. Coarsening can be done by reducing the spatial resolution [15, 12], which creates a method similar to parallel space-time multigrid [16, 17]. Hence there is the potential to increase the efficiency of such methods by exploiting the wealth of experience from the multigrid community (see [18] for recent work along these lines). More general coarsening strategies based on reducing the order of spatial operators [15] or floating point precision at coarser levels are also possible but require careful analysis to determine how and when such an approach would provide increased efficiency.

A further intriguing possibility to increase the efficiency of time-parallel methods like PFASST is the use of reduced physical models at coarse levels. This has been investigated already in the context of parareal [19], and preliminary experiments using PFASST for molecular dynamics have shown promise. The benefit of this approach is that it allows one to run a high-fidelity simulation in roughly the same wall clock time as a time-serial low-fidelity simulation. However, success in this approach relies on an understanding of the multi-model analogs of interpolation and restriction in multigrid methods. These questions are hence related to the larger field of multi-scale and multi-physics mathematical models, but require a more tightly coupled hierarchy of models as opposed to techniques developed for equations with well-separated space and time scales.

References

[1] P.E. Zadunaisky. A method for the estimation of errors propagated in the numerical solution of a system of ordinary differential equations. In G. Contopoulos, editor, *The Theory of Orbits in the Solar System and in Stellar Systems. Proceedings of International Astronomical Union, Symposium 25*, pages 281–287, 1964.

[2] V. Pereyra. On improving an approximate solution of a functional equation by deferred corrections. 8:376–391, 1966.

[3] A. Dutt and L. Greengard. Spectral deferred correction methods for ordinary differential equations. *BIT Numerical Mathematics*, 40(2):241–266, 2000.

[4] A. Bourlioux, A. T. Layton, and M. L. Minion. Higher-order multi-implicit spectral deferred correction methods for problems of reacting flow. *J. Comput. Phys.*, 189:651–675, 2003.

[5] M. L. Minion. Semi-implicit spectral deferred correction methods for ordinary differential equations. *Comm. Math. Sci.*, 1:471–500, 2003.

[6] E. L. Bouzarth and M. L. Minion. A multirate time integrator for regularized stokeslets. *JCP*, 229(11):4208–4224, 2010.

[7] A. Nonaka, J. B. Bell, M. S. Day, C. Gilet, A. S. Almgren, and M. L. Minion. A Deferred Correction Coupling Strategy for Low Mach Number Flow with Complex Chemistry. *Combustion Theory and Modelling*, 16(6):1053–1088, 2012.

[8] J.-L. Lions, Y. Maday, and G. Turinici. A parareal in time discretization of PDE's. *C.R. Acad. Sci. Paris, Serie I*, 332:661–668, 2001.

[9] C. Farhat and M. Chandesris. Time-decomposed parallel time-integrators: theory and feasibility studies for fluid, structure, and fluid-structure applications. *Internat. J. Numer. Methods Engrg.*, 58(9):1397–1434, 2003.

[10] M.L. Minion and S.A. Williams. Parareal and spectral deferred corrections. In *AIP Conference Proceedings*, volume 1048, pages 388–391. Citeseer, 2008.

[11] M. L. Minion. A hybrid parareal spectral deferred corrections method. *Communications in Applied Mathematics and Computational Science*, 5(2):265–301, 2010.

[12] M. Emmett and M.L. Minion. Toward an efficient parallel in time method for partial differential equations. *Communications in Applied Mathematics and Computational Science*, 7(1):105–132, 2012.

[13] A. Christlieb, C. B. Macdonald, and B. Ong. Parallel high-order integrators. *SISC*, 32(2):818–835, 2010.

[14] M. Emmett and M. L. Minion. Efficient implementation of a multi-level parallel in time algorithm. In *21st International Conference on Domain Decomposition Methods*, pages 1–8, Rennes, 2012.

[15] A. T. Layton. On the efficiency of spectral deferred correction methods for time-dependent partial differential equations. *Appl. Numer. Math.*, 59(7):1629–1643, 2009.

[16] W. Hackbusch. Parabolic multi-grid methods. In R. Glowinski and Jacques-louis Lions, editors, *Proceedings of the Sixth International Symposium on Computing Methods in Applied Sciences and Engineering*, pages 189–197. North-Holland, June 1984.

[17] G Horton and S Vandewalle. A space-time multigrid method for parabolic partial differential equations. *SIAM Journal on Scientific Computing*, 16(4):848–864, 1995.

[18] S. Friedhoff, R. D. Falgout, T. V. Kolev, S. MacLachlanand, , and J. B. Schroder. A multigrid-in-time algorithm for solving evolution equations in parallel. Technical Report LLNL-CONF-606952, Lawrence Livermore National Lab, 2013.

[19] X. Dai, C. Le Bris, F. Legoll, and Y. Maday. Symmetric parareal algorithms for Hamiltonian systems. November 2010.