PinT 2020: 9th Workshop on Parallel-in-Time Integration

Parallel-in-Time Scientific Committee:

Stephanie Friedhoff*
Laurence Halpern
Benjamin Ong*
Kenji Ono
Debasmita Samaddar
Jacob Schroder*
Jemma Shipton*

June 8 – 12, 2020

*PinT 2020 co-organizers

Welcome Address

The PinT 2020 workshop is being held virtually, June 8–12, 2020, due to COVID-19. We thank the speakers and the participants for their flexibility. Each day will feature a series talks scheduled within a 2-hour block, 3–5pm GMT. Attendees and presenters need to pre-register at
`https://michigantech.zoom.us/meeting/register/upUsd-6hqj4iGNa05WE3Ey2RkX26G6K7-Q`

Some meeting logistics:

- We recommend that you download and install the latest version of the zoom app, `https://zoom.us/support/download`. All attendees and presenters will require version 5.x.
- Only the presenter and host will be able to share screens.
- Participants are automatically muted upon entry to the zoom meeting. During the question period, the host will allow participants to unmute themselves to ask questions. If you prefer, you may also post a question to the chat, and a host will share your question with the presenter.
- If you have sufficient bandwidth, please enable your video feed – it makes the session feel more interactive.

Related PinT news:

- Following the conclusion of the virtual PinT 2020 workshop, we will release a call for papers, to be published within a proceedings issue of Springer's Proceedings in Mathematics and Statistics. Registered attendees and members of the parallelintime@googlegroups.com mailing list will be notified by email when the call is open. To subscribe to the parallelintime mailing list, please send an email to parallelintime+subscribe@googlegroups.com.
- An NSF-sponsored PinT training conference will be held June 7–11, 2021. Please visit that conference website at: `http://conferences.math.mtu.edu/cbms2020/`
- PinT 2021 will be held in Houghton, MI, Jun 14–18, 2021. Please visit that conference website at: `http://conferences.math.mtu.edu/pint2021/`

Contents

Monday June 8, 2020

A universal parallel predictor algorithm of arbitrarily high order for implicit methods

Jun 8
3:00pm
GMT

Laurent Jay

laurent-jay@uiowa.edu

University of Iowa

We consider the application of implicit methods to any type of initial value problems (IVPs). We present a new predictor algorithm for the iterates of modified Newton type/fixed point iterations for the solution to the nonlinear equations. The new predictor algorithm is universal since it applies to any implicit method applied to any type of IVPs for ODEs, DAEs, and time-varying PDEs. It generalizes, unifies, and improves previous approaches. We illustrate our results for the internal stages of implicit Runge-Kutta (IRK) type methods applied to ODEs and the iterated corrections of starting approximation algorithms based for example on continuous/dense output. Our methodology relies on the existence of an asymptotic expansion in the stepsize h of the error between the exact discrete values and an initial starting approximation. Arbitrarily high order approximations can be obtained assuming sufficient smoothness of the solution. Not only starting approximations (i.e., the first iterate) can be predicted, but also the inner iterates of modified Newton type/fixed point iterations can also be predicted leading to a new type of iterations. These improved prediction algorithms require some extra memory to store previous exact errors, but they require minimal computational effort since no new evaluation of the functions to describe the vector field and the constraints is needed. Moreover, for a given iteration each component and each correction can be obtained completely independently from the others allowing for an embarrassingly parallel implementation. This methodology allows a drastic reduction of the number of Newton-type iterations in implicit methods at the cost of some extra memory storage illustrating the well-known computer science principle of time-memory trade-off.

AIR for a space-time hybridizable discontinuous Galerkin method

Jun 8
3:30pm
GMT

Abdullah Ali Sivas[a,o], Ben Southworth[b] and Sander Rhebergen[a]

[o]aasivas@uwaterloo.ca

[a]University of Waterloo, [b]University of Colorado Boulder

Space-time finite element methods are excellent for the discretization of partial differential equations (PDEs), including on time-dependent domains. Unlike classical time-stepping methods, such as Runge-Kutta or multistep methods, space-time methods make no distinction between spatial and temporal variables. Instead, the PDE is discretized directly in $d+1$-dimensional space-time, where d is the spatial dimension. Consider, for example, the time-dependent advection equation in d spatial dimensions, $\partial_t u + \mathbf{a} \cdot \nabla u = f$. To apply the space-time finite element method, we introduce first the space-time gradient $\tilde{\nabla} = (\partial_t, \nabla)$ and space-time advective velocity $\tilde{\mathbf{a}} = (1, \mathbf{a})$. We, then, write the time-dependent advection equation as a 'steady' advection equation in space-time: $\tilde{\mathbf{a}} \cdot \tilde{\nabla} u = f$. We discretize this equation by the space-time HDG method of [1,3].

In this talk, we discuss the solution of the space-time HDG discretization of the advection and advection-diffusion equation on time-dependent domains by ℓAIR algebraic multigrid [2]. ℓAIR was shown in [2] to be an optimal solver for hyperbolic and advection-dominated problems. This makes ℓAIR ideal also as a solver for space-time discretizations of advection dominated flows. We will also discuss and compare the solution of space-time HDG discretizations resulting from an all-at-once discretization, in which the $d+1$-dimensional space-time domain has been discretized into a $d+1$-dimensional unstructured mesh, and a slab approach, in which the space-time domain is first partitioned into time-slabs and the problem is solved one slab at a time. We investigate the efficiency of ℓAIR for purely hyperbolic and strongly advection-dominated problems, which are difficult or intractable for many parallel-in-time methods, and also consider the weakly advection-dominated case. We furthermore investigate ℓAIR in combination with space-time adaptive mesh refinement, a unique advantage of space-time finite elements over a traditional separation of space and time.

[1] K.L.A. Kirk, et al., Analysis of a space-time hybridizable discontinuous Galerkin method for the advection-diffusion problem on time-dependent do- mains, SIAM J. Numer. Anal., 57/4 (2019).

[2] T. A. Manteuffel, et al., Nonsymmetric algebraic multigrid based on local approximate ideal restriction (ℓAIR), SIAM J. Sci. Comput., 40/6 (2018).

[3] S. Rhebergen and B. Cockburn, Space-time hybridizable discontinuous Galerkin method for the advection-diffusion equation on moving and deforming meshes, in The Courant-Friedrichs-Lewy (CFL) condition, 80 years after its discovery, Birkhauser Science, 2013.

Solving PDEs in Space-Time: 4D Tree-Based Adaptivity, Mesh-Free and Matrix-Free Approaches

Jun 8
4:00pm
GMT

Milinda Fernando

milinda@cs.utah.edu

University of Utah

Numerically solving partial differential equations (PDEs) remains a compelling application of supercomputing resources. The next generation of computing resources – exhibiting increased parallelism and deep memory hierarchies– provides an opportunity to rethink how to solve PDEs, especially time-dependent PDEs. Here, we consider time as an additional dimension and simultaneously solve for the unknown in large blocks of time (i.e. in 4D space-time), instead of the standard approach of sequential time-stepping. We discretize the 4D space-time domain using a mesh-free kD-tree construction that enables good parallel performance as well as on-the-fly construction of adaptive 4D meshes. To best use the 4D space-time mesh adaptivity, we invoke concepts from PDE analysis to establish rigorous posterior error estimates for a general class of PDEs. We solve canonical linear as well as non-linear PDEs (heat diffusion, advection-diffusion, and Allen-Cahn) in space-time, and illustrate the following advantages: (a) sustained scaling behavior across a larger processor count compared to sequential time-stepping approaches, (b) the ability to capture"localized" behavior in space and time using the adaptive space-time mesh, and (c) removal of anytime-stepping constraints like the Courant-Friedrichs-Lewy (CFL) condition, as well as the ability to utilize spatially varying time-steps. We believe that the algorithmic and mathematical developments along with efficient deployment on modern architectures shown in this work constitute an important step towards improving the scalability of PDE solvers on the next generation of supercomputers.

Parallel space - time solution strategy for the Navier-Stokes equation

Jun 8
4:30pm
GMT

Biswajit Khara[a,o], Robert Dyja [b], Kumar Saurabh[a], Baskar Ganapathysubramanian[a]

[o]bkhara@iastate.edu

[a]Iowa State University, [b] Czestochowa University of Technology

We study solution strategies for solving time-dependent flow problems through a fully coupled space-time formulation. Such a method could be considered a member of the class of methods that attempts to exploit parallelism in both space as well as time when solving partial differential equations numerically. When developing such a methodology, the discrete problem presents an array of challenges for both linear as well as nonlinear equations. For example, the existence of an "advection" in time direction and the nonexistence of diffusion in the time direction renders the global Peclet number infinite. Thus proper stabilized methods need to devised for these problems. We formulate and implement a stabilized method based on the variational multiscale method (VMS). Furthermore, when solving nonlinear problems, in the absence of any "marching", the lack of a good guess makes any quasi-Newton solve difficult to converge. We find a way around this issue by using an adaptive refinement strategy in space-time and seek coarse scale solution in earlier iterations and resolve smaller features progressively. To this end, we develop an aposteriori error indicator for the space-time stabilized variational problem. For specific application to flow problems using this method, we demonstrate examples using two benchmark problems in computational fluid dynamics: (i) the lid driven cavity and (ii) flow past a cylinder.

Tuesday June 9, 2020

ParaDIAG: Parallel-in-Time Algorithms Based on the Diagonalization Technique

Jun 9
3:00pm
GMT

Shulin Wu

wushulin84@hotmail.com

Northeast Normal University

n 2008, Maday and Ronquist introduced an interesting new approach for the direct parallel-in-time (PinT) solution of time-dependent PDEs. The idea is to diagonalize the time stepping matrix, keeping the matrices for the space discretization unchanged, and then to solve all time steps in parallel. Since then, several variants appeared, and we call these closely related algorithms ParaDIAG algorithms. ParaDIAG algorithms in the literature can be classified into two groups: ParaDIAG-I: direct standalone solvers, and ParaDIAG-II: iterative solvers,

We will explain the basic features of each group in this note. To have concrete examples, we will introduce ParaDIAG-I and ParaDIAG-II for the advection-diffusion equation. We will also introduce ParaDIAG-II for the wave equation and an optimal control problem for the wave equation. We could have used the advection-diffusion equation as well to illustrate ParaDIAG-II, but wave equations are known to cause problems for certain PinT algorithms and thus constitute an especially interesting example for which ParaDIAG algorithms were tested. In this talk, we try to explain the main idea and the main known theoretical results in each case together with some numerical results.

A linear-algebra perspective on convergence of Parareal and MGRiT

Jun 9
3:30pm
GMT

Ben Southworth

ben.s.southworth@gmail.com

University of Colorado

Parareal and multigrid reduction in time (MGRiT) are two of the most popular parallel-in-time methods. Both can be posed as preconditioned fixed-point iterations, for which convergence in an appropriate norm is defined by the (discrete) error- and residual-propagation operators. Here, we introduce a linear-algebra analysis for convergence of Parareal and two-level MGRiT, measuring the norm of error- and residual-propagation. The analysis is based on a ”temporal approximation property” (TAP), which provides a measurement of how the coarse-grid time stepper approximates the fine-grid time stepper. For linear problems independent of time, satisfying the TAP provides necessary and sufficient conditions for two-grid convergence. Furthermore, these conditions are asymptotically exact, that is, the accuracy of the TAP defines the norm of error- and residual-propagation in the limit of a large number of time steps, N. For diagonalizable operators, explicit upper and lower bounds on convergence based on N can also be derived, confirming the asymptotic results are tight in practice.

We emphasize that for linear problems, the TAP defines exactly what is necessary of the coarse- and fine-grid time steppers for convergence of Paraeal and two-level MGRiT. Theory is demonstrated on two hyperbolic examples, the wave equation and advection-reaction equation, and is also used to prove which Runge-Kutta schemes yield convergence independent of spatial and temporal grid sizes for SPD spatial operators. We conclude by commenting on ongoing extensions of the theory to the linear time-dependent setting, which appears to follow a natural generalization by appealing to recent developments in generalized locally Toeplitz sequences.

TriMGRIT: An Extension of Multigrid Reduction in Time for Constrained Optimization

Jun 9
4:00pm
GMT

Robert Falgout

rfalgout@llnl.gov

Lawrence Livermore National Laboratory

Since clock speeds are no longer increasing, time integration is becoming a sequential bottleneck. The multigrid reduction in time (MGRIT) algorithm is an approach for creating concurrency in the time dimension that can be exploited to overcome this bottleneck and is designed to build on existing codes and time integration techniques in a non-intrusive manner. In this talk, we will discuss an extension of MGRIT for solving time-dependent constrained optimization problems. In the linear case, the MGRIT algorithm can be viewed as an approximate block cyclic reduction algorithm applied to a block lower bi-diagonal system. TriMGRIT extends this idea to block tri-diagonal systems such as those that arise in time-dependent constrained optimization. We will present a linear and a nonlinear algorithm applied to several model problems.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

PyMGRIT - a Python Package for the multigrid-reduction-in-time algorithm

Jun 9
4:30pm
GMT

Jens Hahne

jens.hahne@math.uni-wuppertal.de

Bergische Universität Wuppertal, Germany

The multigrid-reduction-in-time (MGRIT) algorithm is a reduction-based time-multigrid method for solving time-dependent problems. The MGRIT method is a non-intrusive approach that essentially uses the same time integrator as a traditional time-stepping algorithm. Therefore, it is particularly well suited for introducing time parallelism in simulations using existing application codes. In this talk, we introduce the Python framework PyMGRIT, which implements the MGRIT algorithm. The PyMGRIT framework features many different variants of the MGRIT algorithm, from different cycle types and relaxation schemes, as well as various coarsening strategies, including time-only and space-time coarsening, to different time integrators on different levels in the multigrid hierarchy. Thereby, PyMGRIT allows serial runs for prototyping and testing of new approaches, as well as parallel runs using the Message Passing Interface (MPI). Examples illustrate different aspects of the package, including pure time parallelism as well as space-time parallelism by coupling PyMGRIT with PETSc or Firedrake, which enable spatial parallelism through MPI.

Wednesday June 10, 2020

Parareal — RBF algorithms for solving time-dependent PDEs

Jun 10
3:00pm
GMT

Nadun Dissanayake[o], Cecile Piret and Benjamin Ong

[o]nldissan@mtu.edu

Michigan Technological University

Radial basis functions (RBF) are a mesh-less approach to discretize differential operators in space. Over the past two decades, the RBF method has gained attention from numerical analysts for its ability to achieve spectral/high-order accuracy. When solving time-dependent PDEs with RBFs, choosing a time integrator that couples well with the RBF discretizations has been an important research topic within the RBF community. In this talk, we explore how the parareal framework can be used to provide a time-parallel approach to solving time-dependent PDEs discretized spatially using RBFs, focusing on how the RBF discretizations can be modified (enriched) to generate desirable coarse parareal solvers.

Parallel-in-Time Solution of Time-Periodic Problems with Unknown Period

Jun 10
3:30pm
GMT

Iryna Kulchytska-Ruchka[1] and Sebastian Schöps

[1]kulchytska@temf.tu-darmstadt.de

Technical University of Darmstadt

In this talk we consider a novel parallel-in-time algorithm for time-periodic problems where the period is not given. Inheriting the idea of the periodic Parareal approach PP-PC [1], the proposed method uses discretization on a two-level grid and calculates not only the initial values at each subinterval but also the corresponding period iteratively. This approach extends the multiple shooting method with unknown period [2] by the Parareal-based approximation of the Jacobian. A particular diagonalization of the resulting nonlinear time-periodic coarse-grid system [3] is introduced, thereby permitting further parallelization on the coarse level. Performance of the introduced algorithm is illustrated for a Colpitt oscillator model.

[1] M. J. Gander, Y.-L. Jiang, B. Song, and H. Zhang. Analysis of two parareal algorithms for time-periodic problems. SIAM J. Sci. Comput., 35(5):A2393–A2415, 2013.

[2] P. Deuflhard. Computation of periodic solutions of nonlinear ODEs. BIT, (24):456–466, 1984.

[3] I. Kulchytska-Ruchka and S. Schöps. Efficient parallel-in-time solution of time-periodic problems using a multi-harmonic coarse grid correction, 2019. ArXiv: 1908.05245.

This work is supported by the 'Excellence Initiative' of the German Federal and State Governments, the Graduate School of Computational Engineering at Technische Universität Darmstadt, the BMBF grant No. 05M2018RDA (PASIROM).

Performance Analysis and Benchmarking for pySDC

Jun 10
4:00pm
GMT

Robert Speck

r.speck@fz-juelich.de

Juelich Supercomputing Centre, Forschungszentrum Juelich GmbH

The parallel full approximation scheme in space and time (PFASST) allows to integrate multiple time-steps simultaneously. Based on iterative spectral deferred correction (SDC) methods, PFASST uses a space-time hierarchy with various coarsening strategies to maximize parallel efficiency. In numerous studies, this approach has been used on up to 448K cores and coupled to space-parallel solvers which use finite differences, spectral methods or even particles for discretization in space. However, since the integration of SDC or PFASST into an existing application code is not straightforward and the potential gain is typically uncertain, we have developed the Python prototyping framework pySDC. While it allows to rapidly test new ideas and to implement first toy problems more easily, it can also be used to run space-time parallel tests and applications using mpi4py. In this talk, we examine pySDC's performance on an HPC cluster and demonstrate the application of the "Scalable Performance Measurement Infrastructure for Parallel Codes" (Score-P) for analyzing the performance of our code. We highlight Python-, MPI- and PinT-specific aspects of our results and show the benefits of a structured benchmarking workflow.

N Ways to Fool the Masses (or Yourself) When Presenting PinT Results

Jun 10
4:30pm
GMT

Michael Minion[a], Robert Speck[b], Sebastian Götschel[c] and Daniel Ruprecht[d]

[a]mlminion@lbl.gov, [b]r.speck@fz-juelich.de, [c]sebastian.goetschel@tuhh.de and [d]ruprecht@tuhh.de

[a]Lawrence Berkeley National Laboratory,
[b]Juelich Supercomputing Centre,
[c,d]Hamburg University of Technology

Designing a good test case for a parallel in time method is not always straightforward. We present parallel in time examples where the results can fool you or fool an audience in a talk. Parts of the presentation are meant to be lighthearted, but the examples are inspired by real events.

Thursday June 11, 2020

Space-time adaptivity for parabolic evolution equations

Jun 11
3:00pm
GMT

Jan Westerdiep

j.h.westerdiep@uva.nl

Korteweg-de Vries Institute for Mathematics, University of Amsterdam

Taking the well-posed mixed simultaneous space-time variational formulation introduced in [And13], we use methods previously developed in [SW20] to construct an adaptive loop that produces space-time approximations as linear combinations of tensor-products of wavelets in time and finite elements in space.

Using an efficient and reliable 'hierarchical basis' error estimator, we apply bulk chasing to show convergence of the iterands. Moreover, we provide an algorithm for linear-complexity application of the system matrix and an optimal preconditioner. Lastly, we include an extensive numerical study to show that this method is competitive in terms of speed and moreover exhibits optimal convergence rate in the number of degrees of freedom.

References:

[And13] R. Andreev. Stability of sparse space-time finite element discretizations of linear parabolic evolution equations. IMA J. Numer. Anal., 33(1):242–260, 2013.

[SW20] Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations. IMA J. Numer. Anal., 2020.

Exploring space-time adaptivity and parallel-in-time convergence for hyperbolic PDEs

Jun 11
3:30pm
GMT

Hans Johansen

hjohansen@lbl.gov

Lawrence Berkeley National Laboratory

We present an analysis of an adaptive space-time algorithm for hyperbolic partial differential equations. The spatial discretization we use is either explicit or based on implicitly-defined compact stencils, which require solving a linear system but can also have improved spectral properties. The space-time refinement uses nested regions with finer grid spacing/time step, that can be used to improve the error of solutions near steep gradients or material features. We explore several time integration options, including explicit, ADI, spectral deferred corrections, extrapolation, and exponential methods, and analyze the results in terms of spectral accuracy. Finally, we demonstrate a combined parallel-refinement-in-time algorithm and demonstrate why the spectral properties of operators in space and time must be considered together for optimal efficiency and convergence.

Simulating interface evolution using a parallel space-time approach: Space-time adaptivity for the phase field equations

Jun 11
4:00pm
GMT

Kumar Saurabh[a,o], Biswajit Khara[a], Milinda Fernando[b], Hari Sundar[b], Baskar Ganapathysubramanian[a]

[o]maksbh@iastate.edu

[a]Iowa State University, [b]University of Utah

Interface evolution (solidification, melting, phase-separation) phenomena exhibit spatially and temporally localized regions of steep gradients. Conventional approaches (i.e. time marching approaches) utilize localized adaptivity in space, but generally utilize global adaptivity in time. Here, we consider time as an additional dimension and simultaneously solve for the unknown in large blocks of time (i.e. in space-time). We focus on space-time solutions of a generalized class of equations called the phase-field equations. We formulate a variational multiscale (VMS) based space-time strategy that allows us to (a) exploit parallelism not only in space but also in time, (b) gain high order accuracy in time, and (c) exploit adaptive refinement approaches to locally refine the region of interest in both space and time. We illustrate this approach with several canonical problems including melting and solidification of complex dendritic/snowflake structures and phase separation simulation.

A parallel-in-time, implicit/explicit multiderivative solver

Jun 11
4:30pm
GMT

Jochen Schuetz[a,o] and David Seal[b]

[o]jochen.schuetz@uhasselt.be

[a]Hasselt University, [b]US Naval Academy

In this talk, we present a novel scheme for a class of extremely stiff ordinary differential equations.

Frequently, singularly perturbed differential equations can be split into stiff and non-stiff parts. If so, a flux splitting can typically be constructed, and stiff parts are treated implicitly-in-time for stability, while the other parts are treated explicitly for efficiency. This treatment has been termed IMEX. It is, amongst others, very successful for the computation of solutions to relaxation problems and low-Mach fluid flow equations.

To the best of our knowledge, the algorithm that we present in this talk is the first attempt ever to combine IMEX methods with so-called multiderivative methods. For multiderivative methods and an ODE of form y'(t) = f(y(t)), say, the algorithm does not only take y'(t) = f(y) into account, but also y"(t) = f'(y) * f(y); this typically results in high-order methods with very few storage requirements. The resulting method that we present here is provably stable for prototypical equations and can be extended to partial differential equations. The scheme is of predictor-corrector type which makes it easily amenable to temporal parallelisation.

In this talk, we will present both analytical and numerical results for the use of the method with ordinary differential equations, including its use parallel-in-time. Subsequently, we will show how to extend the method to the low-Mach Euler equations.

[1] D. Seal and J. Schütz. An asymptotic preserving semi-implicit multiderivative solver. CMAT Preprint UP-19-09, http://www.uhasselt.be/Documents/CMAT/Preprints/2019/UP1909.pdf , 2019.

Friday June 12, 2020

A Comparison of Space-Time Multigrid and PFASST with applications to Cardiac Electrophysiology

Jun 12
3:00pm
GMT

Pietro Benedusi

benedp@usi.ch

Institute of Computational Science, Universitàella Svizzera Italiana

We present a space-time multilevel method that uses a hierarchy of non-nested meshes, created by semi-geometric coarsening. The "grey box" multigrid starts from a single fine spatial mesh and automatically generates space-time coarse meshes of any dimension over complex geometries.

Two model problems are considered: the heat equation with anisotropy and jumping coefficients; the monodomain equation, a non-linear reaction-diffusion model arising from the study of excitable media such as the myocardium.

We analyze the convergence and scaling properties of the proposed solution strategies, focusing on the spectral properties and conditioning of the underlying discrete operators that arise from the tensor space-time finite element discretization.

Strong and weak scaling of the multilevel space-time approach is compared to PFASST (Parallel Full Approximation Scheme in Space and Time), highlighting properties and conceptual and quantitative differences of both approaches.

Optimal Relaxation Weights for Multigrid Reduction In Time (MGRIT)

Jun 12
3:30pm
GMT

Masumi Sugiyama

msugiyama@unm.edu

University of New Mexico

Based on current trends in computer architectures, faster compute speeds must come from increased parallelism rather than increased clock speeds, which are stagnate. This situation has created the well-known bottleneck for sequential time-integration, where each individual time-value (i.e., time-step) is computed sequentially. One approach to alleviate this and achieve parallelism in time is with multigrid. In this work, we consider the scheme known as multigrid-reduction-in-time (MGRIT), but note that there exist other parallel-in-time methods such as parareal and the parallel full approximation scheme in space and time (PFASST). MGRIT is a full multi-level method applied to the time dimension and computes multiple time-steps in parallel. Like all multigrid methods, MGRIT relies on the complementary relationship between relaxation on a fine-grid and a correction from the coarse grid to solve the problem. In this work, we analyze and select relaxation weights for MGRIT using a convergence analysis and find that this is beneficial since it improves the convergence rate and consequently improves the efficiency of computation. We note that choosing appropriate weights for relaxation (here weighted-Jacobi) has a long history for improving the convergence of spatial multigrid methods, and thus it is no surprise that such weight selection can be beneficial for MGRIT, too. Our numerical results demonstrate an improved convergence rate and lower iteration count for MGRIT when non-unitary weights are used for weighted-Jacobi.

Parallel-in-Time with SUNDIALS and XBraid

Jun 12
4:00pm
GMT

David Gardner

gardner48@llnl.gov

Lawrence Livermore National Laboratory

SUNDIALS is a suite of robust and scalable integrators and solvers for systems of ordinary differential equations, differential-algebraic equations, and nonlinear equations used in numerous applications codes for research and industry. The suite consists of six packages, CVODE(S), ARKode, IDA(S), and KINSOL, all built on shared vector, matrix, linear solver, and nonlinear solver APIs allowing for user-defined/application-specific data structures and solvers, encapsulated parallelism, and algorithmic flexibility.

Presently the numerical methods in SUNDIALS utilize sequential time marching schemes with parallelization only in the spatial dimension. Parallel-in-time methods introduce an additional dimension of parallelism to better leverage the increased concurrency available on massively parallel systems. In particular recent work utilizing multigrid-reduction-in-time (MGRIT) has shown significant speedups over sequential time stepping and can be implemented in a non-intrusive manner. In this talk we will present results from recent efforts to combine the adaptive-step explicit, implicit, and IMEX time integration methods from the SUNDIALS ARKode library with the XBraid MGRIT library to provide parallel-in-time integration with SUNDIALS.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-ABS-810233

Parallel-in-Time Training for Deep Residual Neural Networks

Jun 12
4:30pm
GMT

Jacob Schroder

jbschroder@unm.edu

University of New Mexico

Residual neural networks (ResNets) are a type of deep neural network and they exhibit excellent performance for many learning tasks, e.g., image classification and recognition. A ResNet architecture can be interpreted as a discretization of a time-dependent ordinary differential equation (ODE), with the overall training process being an ODE-constrained optimal control problem. The time-dependent control variables are the network weights and each network layer is associated with a time-step. However, ResNet training often suffers from prohibitively long run-times because of the many sequential sweeps forwards and backwards across layers (i.e., time-steps) to carry out the optimization. This work first investigates one possible remedy (parallel-in-time methods) for the long run-times by demonstrating the multigrid-reduction-in-time method for the efficient and effective training of deep ResNets. The proposed layer-parallel algorithm replaces the classical (sequential) forward and backward propagation through the network layers by a parallel nonlinear multigrid iteration applied to the layer domain. However, the question remains how one initializes networks with hundreds or thousands of layers, which leads to the second part. Here, a multilevel initialization strategy is developed for deep networks, where we apply a refinement strategy across the time domain, that is equivalent to refining in the layer dimension. The resulting refinements create deep networks, with good initializations for the network parameters coming from the coarser trained networks. We investigate this multilevel "nested iteration" initialization strategy for faster training times and for regularization benefits, e.g., reduced sensitivity to hyperparameters and randomness in initial network parameters.

Author Index