LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-PROC-654443

Parallel time integration with multigrid

R. Falgout, T. Kolev, J. B. Schroder, S. Friedhoff, S. MacLachlan

May 15, 2014

85th Annual Meeting of GAMM
Erlangen, Germany
March 10, 2014 through March 14, 2014

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Parallel time integration with multigrid

Robert Falgout[1], **Stephanie Friedhoff**[2,*], **Tzanio Kolev**[1], **Scott MacLachlan**[2], and **Jacob B. Schroder**[1]

[1] Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P. O. Box 808, L-561, Livermore, CA 94551.

[2] Department of Mathematics, Tufts University, 503 Boston Avenue, Medford, MA 02155.

With current trends in computer architectures leading towards systems with more, but not faster, processors, faster time-to-solution must come from greater parallelism. We present a family of truly multilevel approaches to parallel time integration based on multigrid reduction (MGR) principles. The resulting multigrid-reduction-in-time (MGRIT) algorithms are non-intrusive approaches, which directly use an existing time propagator and, thus, can easily exploit substantially more computational resources then standard sequential time-stepping. Furthermore, we demonstrate that MGRIT offers excellent strong and weak parallel scaling up to thousands of processors for solving diffusion equations in two and three space dimensions.

Copyright line will be provided by the publisher

1 Introduction

We consider a system of ordinary differential equations (ODEs) of the form

$$u'(t) = f(t, u(t)), \quad u(0) = \overline{u}_0, \quad t \in [0, T], \tag{1}$$

such as in a method of lines approximation of a parabolic partial differential equation (PDE). Let $t_i = i\delta t,\ i = 0, 1, \ldots, N_t$, be a temporal mesh with constant spacing $\delta t = T/N_t$, and, for $i = 1, \ldots, N_t$, let u_i be an approximation to $u(t_i)$ and $u_0 = u(0)$. In the case that f is a linear function of $u(t)$, the solution to (1) is defined via time-stepping, which can also be represented as a forward solve of the linear system, written in block form as

$$\begin{bmatrix} I & & & & \\ -\Phi_{\delta t} & I & & & \\ & -\Phi_{\delta t} & I & & \\ & & \ddots & \ddots & \\ & & & -\Phi_{\delta t} & I \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ u_{N_t} \end{bmatrix} = \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ \vdots \\ g_{N_t} \end{bmatrix}, \tag{2}$$

where $g_0 = u(0)$ and $\Phi_{\delta t}$ represents the time-stepping operator that takes a solution at time t_i to that at time t_{i+1}, along with a time-dependent forcing term g_i. Hence, in the time dimension, this forward solve is completely sequential.

Alternatively, considering the lower block bidiagonal structure, we could apply cyclic reduction, whereby we first solve the Schur complement system,

$$\begin{bmatrix} I & & & & \\ -\Phi_{\delta t}^m & I & & & \\ & -\Phi_{\delta t}^m & I & & \\ & & \ddots & \ddots & \\ & & & -\Phi_{\delta t}^m & I \end{bmatrix} \begin{bmatrix} u_0 \\ u_m \\ u_{2m} \\ \vdots \\ u_{N_t} \end{bmatrix} = \begin{bmatrix} g_0 \\ \hat{g}_m \\ \hat{g}_{2m} \\ \vdots \\ \hat{g}_{N_t} \end{bmatrix}, \tag{3}$$

for the value of the solution at every m-th temporal point, with consistently restricted forcing terms. Then define the solution at the remaining temporal points by local (and parallel) time-stepping between those points defined from the Schur complement.

2 The multigrid-reduction-in-time (MGRIT) algorithms

Interpreting the above cyclic reduction approach as a two-level MGR algorithm, we can define the coarse temporal mesh, or C-points, to be those points included in the Schur complement system (3), with the remaining temporal points as F-points. We can further define "ideal" interpolation as the map which takes the solution at the C-points and yields a zero residual at the F-points, with a similar definition for "ideal" restriction. The Schur complement then arises as the standard Petrov-Galerkin coarse-grid operator with these definitions of restriction and interpolation. As is typical in the MGR setting, the MGRIT approaches replace the true Schur complement with a simpler operator (typically of the same form as the original bidiagonal system,

* Corresponding author: e-mail stephanie.friedhoff@tufts.edu

Copyright line will be provided by the publisher

but with time-step $m\delta t$), replace ideal restriction with simple injection, and compensate by adding relaxation. Furthermore, the two-level method can be extended to multiple levels in a simple recursive manner. Note that the MGRIT approaches are natural multilevel extensions of the inherently two-level parareal algorithm [1]; thus, they provide techniques that offer parallel scalability for cases where the "coarse-in-time" grid is still too large to be treated sequentially.

3 Numerical results

We apply several MGRIT algorithms to a simple parabolic model problem, the diffusion equation in two space dimensions, $u_t = \Delta u_{xx}$, subject to the initial condition $u(x, y, 0) = \sin(x)\sin(y)$ and homogeneous Dirichlet boundary conditions. We discretize the model problem using central finite differences for the spatial derivatives and backward Euler for the time derivative on the uniform space-time grid given by $(x_j = j\Delta x,\ y_k = k\Delta y,\ t_i = i\delta t),\ j = 0, 1, \ldots, N_x,\ k = 0, 1, \ldots, N_y,\ i = 0, 1, \ldots, N_t$, with spacing $\Delta x = \pi/N_x$, $\Delta y = \pi/N_y$, and $\delta t = T/N_t$, respectively. Hence, in (2), $\Phi_{\delta t} = (I+\delta t M)^{-1}$, where M is the usual central finite-difference discretization of $-\Delta u$ (see [2] for details). The time-step on the finest grid, $l = 0$, is chosen to be $\delta t = (\Delta x)^2 = (\Delta y)^2$, and the time-step on each coarse grid, l, is given by $m^l \delta t,\ l > 0$. Spatial problems are solved using a parallel spatial multigrid method with a heuristic stopping tolerance.

Figure 1 shows weak parallel scaling results for several MGRIT variants applied to the model problem on the space-time domain $[0, \pi]^2 \times [0, \pi^2/64]$. The problem size per processor is fixed at (roughly) 2^7 points in each spatial direction and 2^8 points in the temporal direction. Thus, on one processor, we use a uniform grid of $\Delta x = \Delta y = \pi/128$ and 257 points in time while, on 4096 processors, we use a uniform grid of $\Delta x = \Delta y = \pi/1024$ and 16,385 points in time. Shown are results for three different coarsening schemes: two coarsen uniformly across all grids, with factors $m = 2$ or $m = 16$, while the third, denoted $m = 16/2$ in the figure, coarsens by factors of 16 until fewer than 16 temporal points are left on each processor, then coarsens by factors of 2. For each coarsening scheme, the solid line shows results for standard FCF-relaxation, while the dashed lines correspond to multigrid schemes that use F-relaxation on the finest grid, and FCF-relaxation on all coarse grids. Here, we see excellent weak scaling, with roughly 50% efficiency at 4096 processors.

Strong scaling results for the model problem on the space-time domain $[0, \pi]^2 \times [0, \pi^2]$, discretized on a $129^2 \times 16{,}385$ grid are shown in Figure 2. For the time-stepping approach, we parallelize only in space and use sequential time-stepping. For all three MGRIT variants, we parallelize over 16 processors in the spatial dimensions, with increasing numbers of processors in the temporal dimension. Here, we again see slight improvement from the MGRIT variant using V-cycles with F-relaxation on the finest grid and FCF-relaxation on all coarse grids (denoted F-FCF in the figure) over that with FCF-relaxation on all grids. While an F-cycle variant using F-relaxation on all grids offers some improvement on smaller numbers of processors, it shows somewhat poorer parallel scalability. Overall, we see excellent speedup for the MGRIT results at high processor counts. We note that these results highlight the fact that the MGRIT framework is largely intended for the case where many more processors are available than can be effectively utilized by sequential time-stepping. While there is substantial overhead in MGRIT algorithms over the optimal algorithmic scaling of time-stepping, this extra work can be effectively parallelized at very large scales.

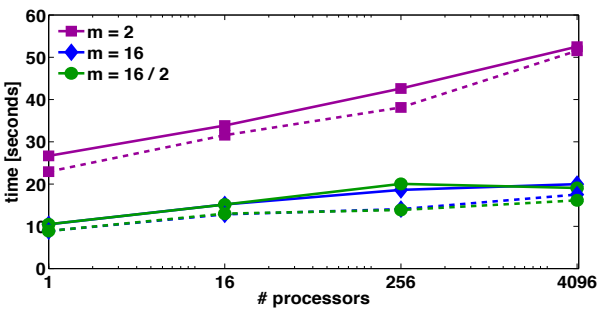

Fig. 1: Weak scaling results for MGRIT variants applied to the diffusion equation in two space dimensions.

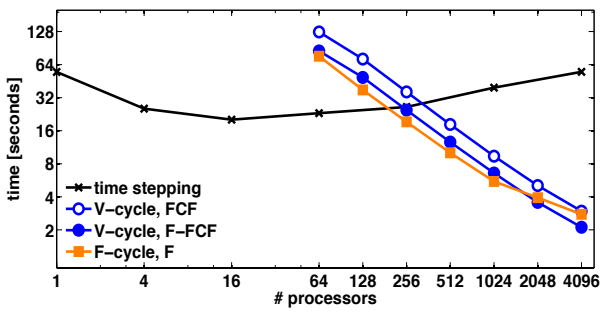

Fig. 2: Strong scaling results for MGRIT variants applied to the diffusion equation in two space dimensions.

Acknowledgements This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-PROC-654443). The work of Stephanie Friedhoff and Scott MacLachlan was partially supported by the National Science Foundation, under grant DMS-1015370.

References

[1] J. L. Lions, Y. Maday, and G. Turinici, C.R. Acad Sci. Paris Sér. I Math **332**, 661–668 (2001).
[2] R. D. Falgout, S. Friedhoff, Tz. V. Kolev, S. P. MacLachlan, and J. B. Schroder, Parallel time integration with multigrid (2013). Submitted.

Copyright line will be provided by the publisher