Efficient Parallel Numerical Integration Algorithms for the Atmospheric Sciences

Matthew R. Norman

SIAM GS11

July 20, 2011

Constraints from Atmospheric Science Needs

- Climate Runs (e.g., IPCC) Require High Throughput
  - Generally held at 5 Sim. Years Per Day (SYPD)
  - 1,825 times faster than realtime
  - Implicit methods need enough work per node for efficiency
  - Sufficient problem sizes fall below 5 SYPD (currently)
- Thus, we resort to time-explicit methods
  - Good for parallel communication
  - Not so good for the time step
- Graduate work
  - Increase the time step
  - Decrease parallel communication

Algorithm Design Choices In This Direction

- Fully discrete: Only 1 stage of comm. per time step
  - Must translate spatial information into temporal
  - Characteristics-based methods do this via trajectories
    - Can have large CFL time steps
    - Majority of my graduate work
  - ADER uses PDE itself to obtain time derivs from space derivs
- Multiple moments per cell
  - Reconstruction stencil is smaller, sometimes entirely local
  - Generally more accurate that grid refinement
  - However, usually restricts time step
- Increasing FLOPS is “good” if:
  - Increases the accuracy of your solution
  - Decreases data movement on the machine

Current State Of The Art: Spectral Element

- Spectral Element Method (Finite Elem. / Galerkin)
  - Spatially local (no reconstruction halo)
  - Cubed-sphere grid (quasi-uniform)
  - Runge-Kutta time integrator
  - 5 SYPD for $\Delta x \approx 14$km atmosphere with ≈ 90,000 cores
    - 93 Million unique model points
    - $\Delta t_{max} \approx 21$ seconds
    - We typically use $\Delta t = 10$ seconds
    - Average of 5.5 milliseconds per time step
  - Time step scales with $\Delta x^{-1}\ p^{-1.7}$
    - p is number of nodes per element
    - Δx is 1-D length of an element

Reconsidering The Use Of Existing Machinery

- A study limited DG modes with HWENO interp
- Implemented two methods
  - 4^{th}-Order Multi-Moment FV HWENO Method
  - 3^{rd}-Order DG Method Limited by HWENO
- FV-HWENO used CFL over 3x larger than DG
- FV-HWENO accuracy similar or better than 3^{rd}-order DG

Reconsidering The Use Of Existing Machinery

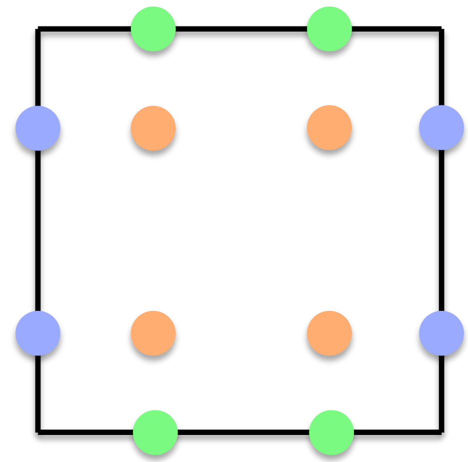
- A study limited DG modes with HWENO interp
- Implemented two methods
  - 4th-Order Multi-Moment FV HWENO Method
  - 3rd-Order DG Method Limited by HWENO
- FV-HWENO used CFL over 3x larger than DG
- FV-HWENO accuracy similar or better than 3rd-order DG

- Is the FV-HWENO scheme competitive in its own right?

ADER-Type Finite Volume Simulation

- Generic Conservation Law: $\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} + \frac{\partial \mathbf{G}(\mathbf{U})}{\partial y} = \mathbf{S}$
- Fully Discrete Finite Volume Framework

$$\overline{\mathbf{U}}_{i,j,n+1} = \overline{\mathbf{U}}_{i,j,n} - \frac{\Delta t}{\Delta x}\left(\hat{\mathbf{F}}_{i+1/2,j} - \hat{\mathbf{F}}_{i-1/2,j}\right) - \frac{\Delta t}{\Delta z}\left(\hat{\mathbf{H}}_{i,j+1/2} - \hat{\mathbf{H}}_{i,j-1/2}\right) + \Delta t \overline{\hat{\mathbf{S}}}_{i,j}$$



- Use time-averaged fluxes and sources
- Cauchy-Kowalewski Procedure at each integration point
  - Form Taylor series in time, use PDE itself to get time derivatives

$$\mathbf{U}\left(x^*, y^*, t\right) = \mathbf{U}\left(x^*, y^*, t_n\right) + \left.\frac{\partial \mathbf{U}}{\partial t}\right|_{x^*, y^*, t_n}\left(t - t_n\right) + \left.\frac{\partial^2 \mathbf{U}}{\partial t^2}\right|_{x^*, y^*, t_n}\frac{\left(t - t_n\right)^2}{2} + \cdots$$

$$\left.\frac{\partial \mathbf{U}}{\partial t}\right|_{x^*, y^*, t_n} = -\frac{\partial \mathbf{F}}{\partial \mathbf{U}}\left.\frac{\partial \mathbf{U}}{\partial x}\right|_{x^*, y^*, t_n} - \frac{\partial \mathbf{G}}{\partial \mathbf{U}}\left.\frac{\partial \mathbf{U}}{\partial y}\right|_{x^*, y^*, t_n} + \mathbf{S}\left(\mathbf{U}\big|_{x^*, y^*, t_n}\right)$$

Local Multi-Moment Finite Volume Simulation

- Need discrete evolution for higher spatial moments
  - Simply differentiate the PDE with respect to space

$$\frac{\partial}{\partial t}\frac{\partial^{m+n}\mathbf{U}}{\partial x^m \partial y^n}+\frac{\partial}{\partial x}\frac{\partial^{m+n}\mathbf{F}(\mathbf{U})}{\partial x^m \partial y^n}+\frac{\partial}{\partial y}\frac{\partial^{m+n}\mathbf{G}(\mathbf{U})}{\partial x^m \partial y^n}=\frac{\partial^{m+n}}{\partial x^m \partial y^n}\mathbf{S}$$

  - Integrate over a space-time domain

$$\overline{\mathbf{U}}_{i,j,n+1}^{(m,n)}=\overline{\mathbf{U}}_{i,j,n}^{(m,n)}-\frac{\Delta t}{\Delta x}\left(\hat{\mathbf{F}}_{i+1/2,j}^{(m,n)}-\hat{\mathbf{F}}_{i-1/2,j}^{(m,n)}\right)-\frac{\Delta t}{\Delta z}\left(\hat{\mathbf{H}}_{i,j+1/2}^{(m,n)}-\hat{\mathbf{H}}_{i,j-1/2}^{(m,n)}\right)+\Delta t\,\overline{\hat{\mathbf{S}}}_{i,j}^{(m,n)}$$

  - Higher moments use the same FV machinery
- Use local moments for local reconstruction
- Local recon & C-K provides local space-time Taylor polyn
- Space-time polyn closes the scheme for all moments
- Use symbolic math software to generate all derivatives

New Proposed Method: Multi-Moment ADER-Taylor

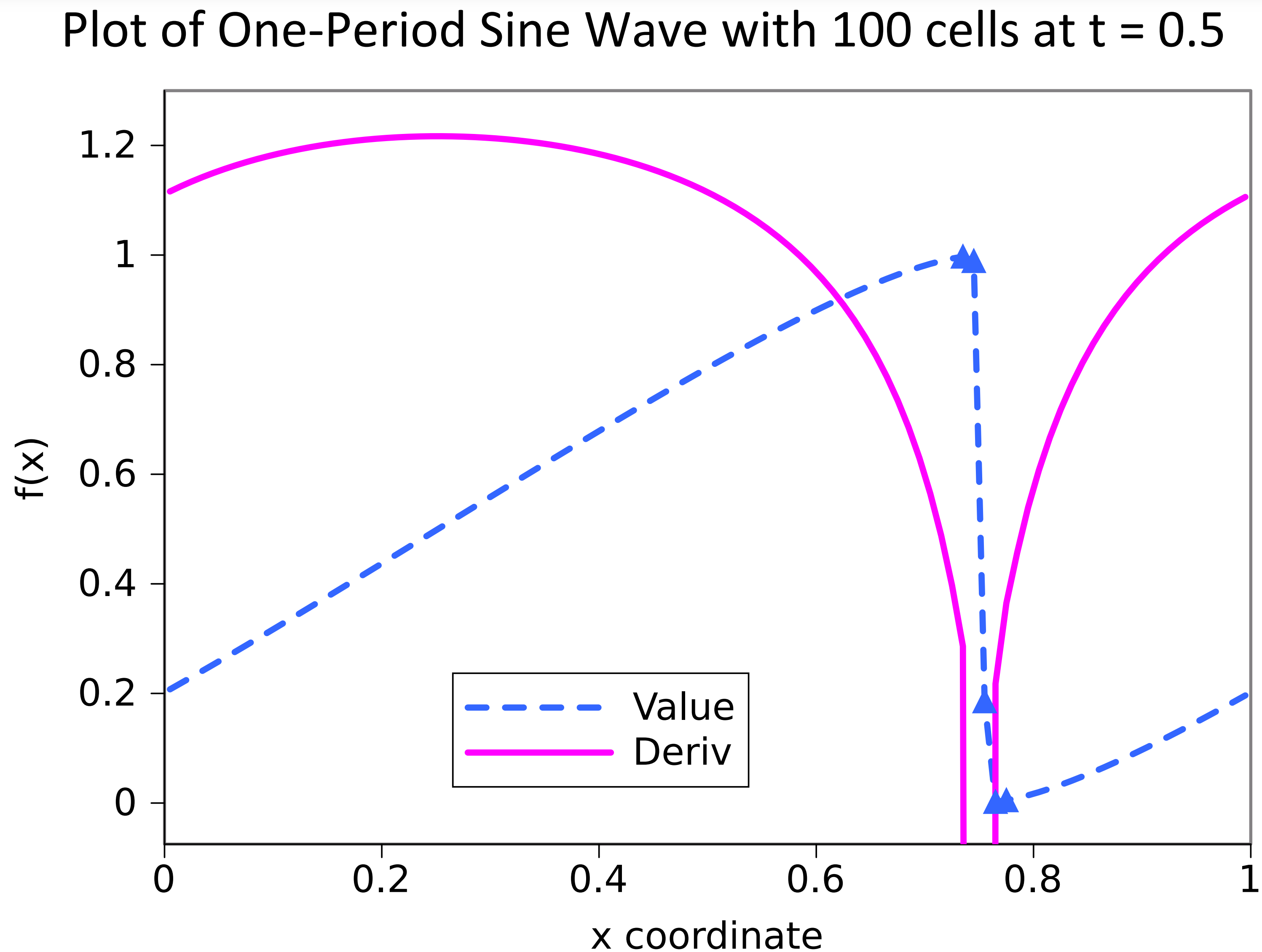
- Fully Discrete Time Stepping (ADER)
  - Cut out Runge-Kutta stage syncs & comms
- Reduce ADER Expense (Taylor expansion of C-K derivs)
- Multiple modes per cell (as with DG or SE)
- Spatially local (as with DG or SE)
- Time step scales Δx^{-1} $\mathbf{p^0}$
  - CFL remains at **unity** for all p-refinement
- Most similar to modal DG with Taylor basis
  - Cell evolves value, 1^{st}-deriv, 2^{nd}-deriv, etc.
- Readily applicable to an arbitrary mesh
- HWENO limiters and hybrid schemes apply easily

Computational Aspects Of Multi-Moment ADER

- Increases computation (Good or Bad?)
  - Significantly decreases communication (Good)
  - Legitimately increases resolution (Good)
  - Does not increase time step (Neutral, better than Galerkin)

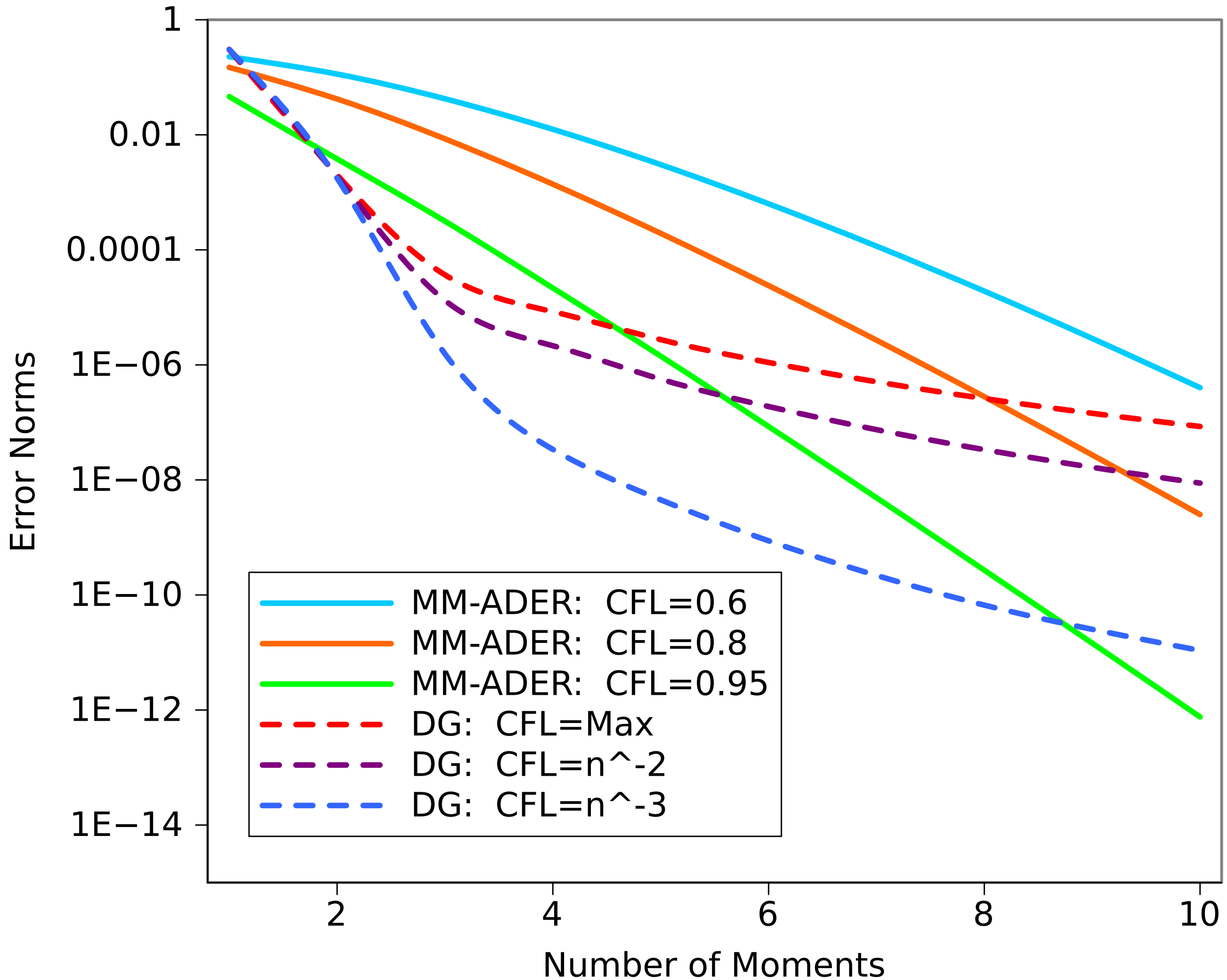
# Modes	RK-SE syncs per MM-ADER Δt	RK-DG syncs per MM-ADER Δt
2	3.5	7.3
3	7.8	14.3
4	12.6	23.1
5	18.2	33.4
6	24.3	45.3
7	31.5	58.8
8	39.7	73.7

Burger’s Equation Shock Simulation: HWENO



Error Comparison between MM-ADER and RKDG

Error for One-Period Sine Wave with 25 Cells after 4 Revolutions



Questions?

ADER-Taylor Modification

- ADER methods notoriously expensive
  - C-K procedure expensive & performed at all points
- ADER-Taylor modification
  - Perform C-K procedure only once at cell center
    - Generates time & mixed space-time derivatives
  - Form space-time Taylor polyn over local space-time domain
  - Sample polyn at points in space and time for fluxes and sources
- Reduces computational constant and complexity
  - In 2-D, C-K needs order n^3 space-time derivatives (n = # modes)
  - Cost of each derivatives grows at up to n^2
  - Number of 2-D space-time quadrature points grows with n^3
  - Taylor polyn removes the growth with quadrature points

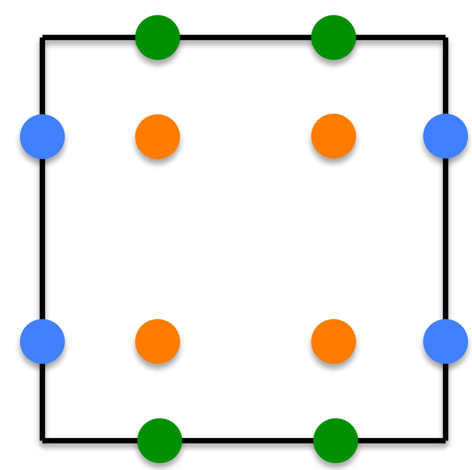
Flux-Based Characteristic Semi-Lagrangian Method

$$\frac{\partial}{\partial t}\begin{bmatrix}\phi\\ \phi u\\ \phi v\end{bmatrix}+\frac{\partial}{\partial x}\begin{bmatrix}\phi u\\ \phi u^2+\phi^2/2\\ \phi uv\end{bmatrix}+\frac{\partial}{\partial y}\begin{bmatrix}\phi v\\ \phi vu\\ \phi v^2+\phi^2/2\end{bmatrix}=\begin{bmatrix}0\\ -\phi\left(\Phi_x-fv\right)\\ -\phi\left(\Phi_y+fu\right)\end{bmatrix} \Rightarrow \frac{D}{Dt}\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\end{bmatrix}=\begin{bmatrix}S_1\\ S_2\\ S_3\end{bmatrix}$$

- Equations reduced to Lagrangian transport
  - Transported quantities are “Characteristic Variables” (CVs)
  - Solve CV transport with semi-Lagrangian method

- Fully-Discrete Finite-Volume Formulation

$$\overline{\mathbf{U}}_{i,j,n+1}=\overline{\mathbf{U}}_{i,j,n}-\frac{\Delta t}{\Delta x}\left(\hat{\mathbf{F}}_{i+1/2,j}-\hat{\mathbf{F}}_{i-1/2,j}\right)-\frac{\Delta t}{\Delta z}\left(\hat{\mathbf{H}}_{i,j+1/2}-\hat{\mathbf{H}}_{i,j-1/2}\right)+\Delta t\,\overline{\hat{\mathbf{S}}}_{i,j}$$



Pros and Cons of the FBCSL Method

- Pros
  - Large time step: Limited by Jacobian gradient only
  - My experience, CFL=3 or larger
  - Communication frequency extremely low
- Cons
  - Genuinely multi-dimensional very complicated
    - Often not a problem on orthogonal meshes
    - Definitely a problem on non-orthogonal (i.e., cubed-sphere)
  - Communication volume higher when performed
  - Source term inclusion can be tricky and expensive

Chronological Survey of Atmos Integration Schemes

- Semi-Implicit, Semi-Lagrangian
  - Linearize & split fast waves, solve implicitly
  - Solve slow dynamics with semi-Lagrangian method
  - Large time step, Very heavy network traffic
- Explicit Eulerian Finite-Volume
  - Sub-cycle fast waves with low stencil, cheap method
  - Solve slow dynamics with Eulerian method
  - Time step roughly CFL=3 or 4, Reduced network traffic
- Explicit, Eulerian Galerkin (i.e., Finite Element)
  - Still sub-cycle fast waves
  - Solve slow dynamics with Galerkin method
  - Time step suffers significantly, Minimal network traffic