Spectral Methods for Conservation Forms

Agnieszka Lutowska

TU/eindhoven

CASA Seminar, Nov 28 2007

Outline

Convolution Sums

- Nonlinear or variable-coefficient problems.
- Fourier Galerkin for product: $s(x) = u(x)v(x)$.
- An infinite series expansion gives:

$$\hat{s}_k = \sum_{m+n=k} \hat{u}_m \hat{v}_n,$$

where

$$u(x) = \sum_{m=-\infty}^{\infty} \hat{u}_m e^{imx}, \qquad v(x) = \sum_{m=-\infty}^{\infty} \hat{v}_n e^{inx},$$

and

$$\hat{s}_k = \frac{1}{2\pi} \int_0^{2\pi} s(x) e^{-ikx}\, dx.$$

- Our case:
  u and v are finite Fourier series of degree $\leq N/2$, i.e. trigonometric polynomials $u, v \in S_N$, whereas $s \in S_{2N}$ and

$$S_N = \text{span}\{e^{ikx} \mid -N/2 \leq k \leq N/2 - 1\}.$$

- Of interest: values of $\hat{s}_k$ only for $|k| \leq N/2$.
- Truncation of degree $N/2$:

$$\hat{s}_k = \sum_{\substack{m+n=k \\ |m|,|n| \leq N/2}} \hat{u}_m \hat{v}_n, \qquad |k| \leq N/2.$$

- Direct summation: $\qquad O(N^2)$ operations

Transform Methods and Pseudospectral Methods

Goal:

Evaluate

$$\hat{s}_k = \sum_{\substack{m+n=k \\ |m|,|n|\le N/2}} \hat{u}_m \hat{v}_n, \qquad |k| \le N/2.$$

for $u, v \in S_N$.

Idea:

- use inverse discrete Fourier transform (DFT) for $\hat{u}_m$ and $\hat{v}_n$ to transform them to physical space
- perform multiplication similar to $s(x) = u(x)v(x)$
- use DFT to determine $\hat{s}_k$

This must be done carefully!

- Discrete transforms

$$u_j = \sum_{k=-N/2}^{N/2-1} \hat{u}_k e^{ikx_j}, \qquad v_j = \sum_{k=-N/2}^{N/2-1} \hat{v}_k e^{ikx_j},$$

$$j = 0, 1, ..., N-1,$$

- Define

$$s_j = u_j v_j, \qquad j = 0, 1, ..., N-1$$

and

$$\tilde{s}_k = \frac{1}{N} \sum_{j=0}^{N-1} s_j e^{-ikx_j}, \qquad k = -\frac{N}{2}, ..., \frac{N}{2} - 1,$$

where

$$x_j = 2\pi j/N$$

Discrete transforms orthogonality relation gives:

$$\tilde{s}_k = \sum_{m+n=k} \hat{u}_m \hat{v}_n + \sum_{m+n=k\pm N} \hat{u}_m \hat{v}_n = \hat{s}_k + \sum_{m+n=k\pm N} \hat{u}_m \hat{v}_n$$

- The differential equation is not approximated by a true spectral Galerkin method.
- Resulting scheme: Pseudospectral method (Orszag, 1971)
- Operation count: $(15/2)N \log_2 N$

Techniques for removing the aliasing error:

- by padding or truncation
- by phase shifts
- by orthogonal polynomials

Aliasing removal by padding or truncation

Idea: Use a discrete transform with M rather than N points, where $M \geq 3N/2$.

- Let

$$y_j = 2\pi j/M,$$

$$\bar{u}_j = \sum_{k=-M/2}^{M/2-1} \breve{u}_k e^{iky_j}, \quad \bar{v}_j = \sum_{k=-M/2}^{M/2-1} \breve{v}_k e^{iky_j},$$

$$\bar{s}_j = \bar{u}_j \bar{v}_j, \qquad \text{for} \qquad j = 0, 1, \ldots, M-1,$$

where

$$\breve{u}_k = \begin{cases} \hat{u}_k, & |k| \leq N/2 \\ 0 & \text{otherwise} \end{cases}$$

- Similarly, let

$$\breve{s}_k = \frac{1}{M}\sum_{j=0}^{M-1} \bar{s}_j e^{-iky_j}, \qquad k = -\frac{M}{2}, \ldots, \frac{M}{2} - 1$$

- Then

$$\breve{s}_k = \sum_{m+n=k} \breve{u}_m \breve{v}_n + \sum_{m+n=k\pm M} \breve{u}_m \breve{v}_n$$

- Of interest: $\breve{s}_k$ for $|k| \leq N/2$
- Choose $M \geq \frac{3N}{2} - 1$ so that aliasing vanishes. For such M we have:

$$\hat{s}_k = \breve{s}_k, \qquad k = -\frac{N}{2}, \ldots, \frac{N}{2} - 1.$$

- Operation count: $(45/4)N\log_2(\frac{3}{2}N)$

Aliasing removal by phase shifts

Idea: Replace u_j and v_j by

$$u_j^{\Delta} = \sum_{k=-N/2}^{N/2-1} \hat{u}_k e^{ik(x_j+\Delta)}, \quad v_j^{\Delta} = \sum_{k=-N/2}^{N/2-1} \hat{v}_k e^{ik(x_j+\Delta)},$$

j = 0,1,...,N-1,

<u>Compute</u>

$$s_j^{\Delta} = u_j^{\Delta} v_j^{\Delta}, \qquad j = 0, 1, ..., N-1$$

and

$$\hat{s}_j^{\Delta} = \frac{1}{N} \sum_{j=0}^{N-1} s_j^{\Delta} e^{-ik(x_j+\Delta)}, \qquad k = -\frac{N}{2}, ..., \frac{N}{2} - 1$$

which gives

$$\hat{s}_k^{\Delta} = \sum_{m+n=k} \hat{u}_m \hat{v}_n + e^{\pm iN\Delta} \Big(\sum_{m+n=k\pm N} \hat{u}_m \hat{v}_n \Big).$$

For $\Delta = \pi/N$ (shift by half of a grid cell)

$$\hat{s}_k = \frac{1}{2}[\tilde{s}_k + \hat{s}_k^{\Delta}]$$

Aliasing removal by orthogonal polynomials

Convolution sums are also produced in Chebyshev Galerkin and tau methods (by quadratic nonlinearities).

Idea: Examine the nonlinear term from the perspective of quadrature.

- Consider the product:

$$s(x) = u(x)v(x),$$

where u and v are in $\mathbb{P}_N$, i.e.

$$u(x) = \sum_{k=0}^{N} \hat{u}_k T_k(x) \qquad \text{and} \qquad v(x) = \sum_{k=0}^{N} \hat{v}_k T_k(x).$$

- Then

$$\hat{s}_k = \frac{2}{\pi c_k} \int_{-1}^{1} u(x) v(x) T_k(x) w(x) dx, \qquad k = 0, 1, ..., N,$$

  where $w(x)$ is the Chebyshev weight.
  $u(x)v(x)T_k(x)$ - polynomial of degree $\leq 3N$

- Solution: Evaluate $\hat{s}_k$ exactly by a a Chebyshev Gauss-Lobatto quadrature using the points $y_j = \cos(\pi j/M)$, $j = 0, 1, ..., M$ for $M \geq 3M/2 + 1/2$ or...

... use transform methods (for $M \geq 3M/2 + 1/2$):

- padding of $\hat{u}_k$ and $\hat{v}_k$
- inverse discrete Chebyshev transforms
- multiplication: $s_j = u_j v_j,\ j = 0, 1, ..., M$
- discrete Chebyshev transform
- extracting $\hat{s}_k = \check{s}_k,\ k = 0, 1, ..., N$

Set of coefficients not fully de-aliased!
For $M = 2N$ - fully de-aliased set of coefficients (by greater computational cost)

Relation Between Methods

For Burgers' equation:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} - \nu\frac{\partial^2 u}{\partial x^2} = 0,$$

periodic on $(0, 2\pi)$, where ν is a positive constant:

- The Galerkin approximation:

$$\frac{d\hat{u}_k}{dt} + \sum_{m+n=k} \hat{u}_m \hat{v}_n + \nu k^2 \hat{u}_k = 0, \qquad k = -\frac{N}{2}, \ldots, \frac{N}{2} - 1,$$

  where $\hat{v}_k = ik\hat{u}_k$

- Fourier pseudospectral approximation:

$$\frac{d\hat{u}_k}{dt} + \sum_{m+n=k} \hat{u}_m \hat{v}_n + \sum_{m+n=k\pm N} \hat{u}_m \hat{v}_n + \nu k^2 \hat{u}_k = 0,$$

- The collocation approximation:

$$\frac{\partial u^N}{\partial t} + u^N v^N - \nu \frac{\partial^2 u^N}{\partial x^2} \,|_{x=x_j} = 0, \qquad j = 0, ..., N-1,$$

where $v^N = \partial u^N / \partial x$ can be transformed into

$$\frac{d\hat{u}_k}{dt} + \sum_{m+n=k} \hat{u}_m \hat{v}_n + \sum_{m+n=k\pm N} \hat{u}_m \hat{v}_n + \nu k^2 \hat{u}_k = 0,\ k = -\frac{N}{2}, ..., \frac{N}{2} - 1,$$

The Fourier pseudospectral and collocation discretizations of Burgers' equation are equivalent.

The Chebyshev collocation method is not equivalent to the pseudospectral Chebyshev tau method.
The term

$$u^N \frac{\partial u^N}{\partial x}$$

is approximated by:

- Pseudospectral Chebyshev tau method

$$P_{N-2}(I_N(u^N \frac{\partial u^N}{\partial x})) \text{ - trigonometric polynomial}$$

- Chebyshev collocation method

$$\tilde{I}_{N-2}(I_N(u^N \frac{\partial u^N}{\partial x})) = \tilde{I}_{N-2}(u^N \frac{\partial u^N}{\partial x}) \text{ - algebraic polynomial}$$

Conservation Forms

The inviscid, periodic Burgers' equation:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0, \qquad 0 < x < 2\pi, \qquad t > 0$$

satisfies an infinite set of conservation properties (for real-valued solutions):

$$\frac{d}{dt}\int_0^{2\pi} u^k dt = 0, \qquad k = 1, 2, \ldots.$$

- Both the spatial and temporal discretizations affect the conservation properties
- We focus on the spatial discretization and consider the semi-discrete evolution equation

- Assumption: The solution and its approximation are real-valued functions
- Semi-discrete Fourier approximations to the inviscid Burgers equation satisfy only a small number of conservation properties
- Fourier Galerkin approximation for Burgers equation in the weak form with $\nu = 0$ is equivalent to:

$$\int_0^{2\pi} \left(\frac{\partial u^N}{\partial t} + u^N \frac{\partial u^N}{\partial x}\right) v dx = 0 \qquad \text{for all} \qquad v \in S_N$$

- Taking $v \equiv 1$ yields:

$$\frac{d}{dt}\int_0^{2\pi} u^N dx = -\frac{1}{2}\int_0^{2\pi} \frac{\partial}{\partial x}((u^N)^2)dx = -\frac{1}{2}(u^N)^2 \Big|_0^{2\pi} = 0$$

- Taking $v = u^N$ yields:

$$\frac{d}{dt}\int_0^{2\pi} (u^N)^2 dx = -\frac{1}{3}\int_0^{2\pi} \frac{\partial}{\partial x}((u^N)^3)dx = -\frac{1}{3}(u^N)^3 \Big|_0^{2\pi} = 0$$

Fourier Galerkin approximations conserve $\int u^N$ and $\int (u^N)^2$.
They do not necessarily conserve $\int (u^N)^k$ for $k \geq 3$.

- Fourier collocation approximations may conserve one or both of these two quantities, depending on precisely how the nonlinear term is approximated.
- For $\nu = 0$ the approximation is:

$$\frac{\partial u^N}{\partial t} + \frac{1}{2}\mathcal{D}_N((u^N)^2) = 0,$$

where differentiation operator $\mathcal{D}$ is skew-symmetric with respect to the inner product and bilinear form $(u, v)_N$ is an inner product on the space S_N.
It holds (for $v \in S_N$):

$$\frac{d}{dt}(u^N, v)_N = \frac{1}{2}((u^N)^2, \frac{dv}{dx})_N.$$

- taking $v \equiv 1$:

$$\frac{d}{dt}\left(\frac{2\pi}{N}\sum_{j=0}^{N-1} u_j^N\right) = 0,$$

so $\int u^N$ is conserved.

- $\frac{2\pi}{N}\sum(u_j^N)^2$ is not conserved (the inner products are not exact for $v = u^N$.)
- Collocation method applied in the form:

$$\frac{\partial u^N}{\partial t} + \frac{1}{3}\mathcal{D}_N((u^N)^2) + \frac{1}{3}u^N\mathcal{D}_N u^N = 0,$$

conserves both $\frac{2\pi}{N}\sum u_j^N$ and $\frac{2\pi}{N}\sum(u_j^N)^2$.

General form:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathcal{M}(\mathbf{u}) = \mathbf{0} \quad \text{in} \quad \Omega$$

- The independent variables are conserved (except for boundary effects) if the spatial operator is in divergence form, i.e.,

$$\mathcal{M}(\mathbf{u}) = \nabla \cdot \mathcal{F}(\mathbf{u}),$$

  where $\mathcal{F}$ is a flux function.
- From Gauss' theorem:

$$\frac{d}{dt}\int_\Omega \mathbf{u} dV = -\int_{\partial\Omega} \mathcal{F} \cdot \hat{\mathbf{n}} dS$$

  The only integral changes in $\mathbf{u}$ are those due to fluxes through the boundaries.

- If the spatial operator is orthogonal to the solution, i.e.,

$$(\mathcal{M}(\mathbf{u}), \mathbf{u}) = 0,$$

then the quadratic conservation law

$$\frac{d}{dt}(\mathbf{u}, \mathbf{u}) = \frac{d}{dt}||\mathbf{u}||^2 = 0$$

holds.

Approximation to general problem:

- Galerkin:

$$\left(\frac{\partial \mathbf{u}^N}{\partial t} + \mathcal{M}(\mathbf{u}^N), \mathbf{v}\right) = 0, \text{ for all } \mathbf{v} \in \mathbb{P}_N^{0-}(-1,1)$$

- collocation and G-NI approximations:

$$\left(\frac{\partial \mathbf{u}^N}{\partial t} + \mathcal{M}_N(\mathbf{u}^N), \mathbf{v}\right)_N = 0, \text{ for all } \mathbf{v} \in \mathbb{P}_N^{0-}(-1,1),$$

  where $\mathcal{M}_N$ is discrete approximation of $\mathcal{M}$
- Consider Fourier (periodic problems) or Legendre (nonperiodic problems) approximations:
  - for spatial operator in the divergence form $\int \mathbf{u}^N$ is conserved
  - for spatial operator orthogonal to the solution $\int (\mathbf{u}^N)^2$ is conserved

Enforcement of Boundary Conditions

Consider a scalar, one-dimensional, nonperiodic, hyperbolic problem with an explicit time discretization.

Problem: Spectral methods are far more sensitive than finite-difference methods to the boundary treatment.

$$\frac{\partial u}{\partial t} + \beta \frac{\partial u}{\partial x} = 0, \qquad -1 < x < 1, \quad t > 0$$

Assumption: Wave speed β is constant and strictly positive

Hence: $x = -1$ - inflow boundary point.

- Inflow boundary condition:

$$u(-1, t) = u_L(t), \qquad t > 0$$

- Initial condition:

$$u(x, 0) = u_0(x), \qquad -1 < x < 1$$

Strong imposition of the boundary condition.

- Legendre Gauss-Lobatto collocation points: $x_j, j = 0, ..., N$
- Semi-discrete spatial approximation, $u^N(t) \in \mathbb{P}(-1,1)$ for all $t > 0$ is defined:

$$\frac{\partial u^N}{\partial t}(x_j,t) + \beta\frac{\partial u^N}{\partial x}(x_j,t) = 0, \quad j = 1, ..., N \quad t > 0$$

$$u^N(-1,t) = u_L(t), \qquad t > 0$$

$$u^N(x_j,0) = u_0(x_j), \qquad j = 0, ..., N$$

- We obtain G-NI scheme:

$$(u_t^N, v)_N + (\beta u_x^N, v)_N = 0 \qquad \text{for all,} \;\; v \in \mathbb{P}_N^{0-}(-1,1), \;\; t > 0,$$

where $(u,v)_N = \sum_{j=0}^N u(x_j)v(x_j)w_j$ and w_j is the Legendre Gauss-Lobatto weight.

Weak enforcement of boundary conditions

- Useful e.g. in multidomain spectral methods or for systems of equations
- G-NI scheme:
  Find $u^N(t) \in \mathbb{P}_N(-1,1)$ satisfying for all $t > 0$ and all $v \in \mathbb{P}_N(-1,1)$,

$$(u_t^N, v)_N - (\beta u^N, v_x)_N + \beta u^N(1,t)v(1) = \beta u_L(t)v(-1)$$

  as well as initial condition.
- Neither the trial function u^N nor any test function v is required to satisfy a boundary condition

Penalty approach

The spectral approximation u^N is defined as the solution of the polynomial equation

$$(\frac{\partial u^N}{\partial t} + \beta \frac{\partial u^N}{\partial x})(x,t) + \tau \beta Q_N(x)(u^N(-1,t) - u_L(t)) = 0$$

$$-1 \leq x \leq 1, \qquad t > 0,$$

where τ is the penalization parameter and Q_N is a fixed polynomial of degree $\leq N$.

Staggered-grid method

- Two families of interpolation/collocation nodes:
  - Gauss-Lobatto: $x_j, j = 0, ..., N$
  - Gauss: $\bar{x}_j, j = 1, ..., N$

  staggered with respect to each other.
- u is represented as a polynomial of degree $N - 1$ using the Gauss points
- Flux $\mathcal{F}(u) = \beta u$ is represented by a polynomial of degree N using the Gauss-Lobatto points

- Construct the polynomial $\tilde{u}_N \in \mathbb{P}_N$

$$\tilde{u}^N(x_j,t) = \begin{cases} u_L(t) & j=0 \\ \bar{u}^N(x_j,t) & j=1,...,N \end{cases}$$

at the Gauss-Lobatto points.

- Generate the flux $\tilde{F}(x,t) = \mathcal{F}(\tilde{u}^N(x,t))$
- Formulate the collocation conditions at Gauss points

$$\frac{\partial \bar{u}^N}{\partial t}(\bar{x}_j,t) + \frac{\partial F^N}{\partial x}(\bar{x}_j,t) = 0 \qquad j=1,...N, \quad t>0$$

$$u^N(\bar{x}_j,0) = u_0(\bar{x}_j) \qquad j=1,...,N$$

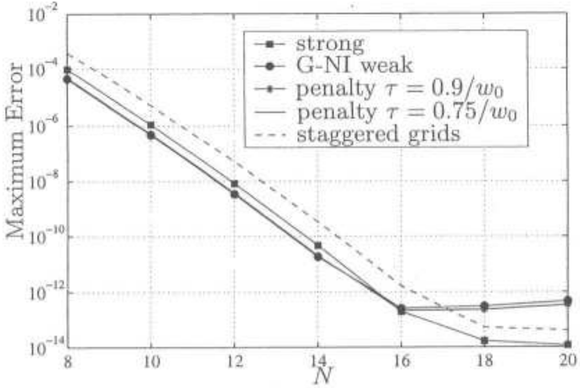
Numerical example

Consider

$$
\begin{aligned}
&\frac{\partial u}{\partial t} + \frac{3}{2}\frac{\partial u}{\partial x} = 0, && -1 < x < 1,\ t > 0,\\
&u(-1,t) = \sin(-2-3t), && t > 0,\\
&u(x,0) = \sin 2x, && -1 < x < 1,
\end{aligned}
$$

Solution: the right-moving wave $u(x,t) = \sin(2x - 3t)$.

- Experiments for Legendre quadrature/collocation points.
- Time discretization conducted with $\Delta t = 10^{-4}$, using the RK4 scheme for all methods.

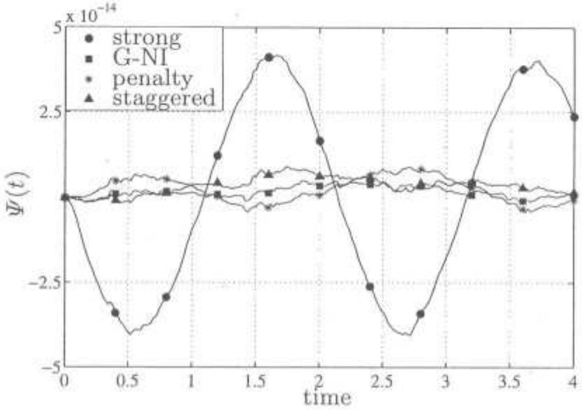


Maximum error at $t = 4$ with different spectral schemes.

Conservation properties.
Evaluate quantity:

$$\Psi(t) = \Big(\int_{-1}^{1} u^N(x,t)dx - \int_{-1}^{1} u^N(x,0)dx\Big) + \Big(\beta \int_0^t u^N(1,s)ds - \beta \int_0^t u_L(s)ds\Big),$$

that is zero for exact solution,
$N = 16$, integrals in time evaluated by Simpson's composite rule (the same accuracy as the RK4 time discretization).



Evolution in time of the quantity $\Psi(t)$ for different spectral schemes.

Conclusions

- Pseudospectral methods for evaluating convolution sums.
- Conservation properties of spectral methods.
- Different types of boundary treatment and their influence on conservation properties.