High-Order, Entropy-Based Models for Linear Transport in Slab Geometries

Cory D. Hauck
Computational Mathematics Group
Computer Science and Mathematics Division
Oak Ridge National Laboratory†
hauckc@ornl.gov

Workshop on Computational Challenges in Hot Dense Plasmas
Institute for Pure and Applied Mathematics
Los Angeles, CA
30 March 2012

†Managed by UT Battelle, LLC under Contract No. De-AC05-00OR22725
for the U.S. Department of Energy.

Collaborators

- Graham Alldrege (Maryland)
- Dianne O'Leary (Maryland)
- André Tits (Maryland)

Outline

Linear Transport

Problem Setting

- We consider the extremely simplified case of linear transport for neutral particles of unit speed
  - Located at **position** $x \in \mathcal{D} \subset \mathbb{R}^3$
  - Traveling with **direction** $\Omega \in \mathbb{S}^2$
- Particles move through a material medium with **scattering**, **absorption**, and **total cross-sections** σ_{s}, σ_{a}, and $\sigma_{\mathrm{t}} = \sigma_{\mathrm{a}} + \sigma_{\mathrm{s}}$
- The material generates a particle source of S.

Kinetic Transport Equation

- The **kinetic density** $\psi = \psi(x, \Omega, t)$ is the density of particles with respect to the measure $d\Omega dx$. It satisfies:

$$\boxed{\partial_t \psi + \Omega \cdot \nabla_x \psi + \sigma_{\mathrm{t}} \psi = \frac{1}{4\pi} \left(\sigma_{\mathrm{s}} \phi + S \right)}$$

  where $\phi = \langle \psi \rangle$, angle brackets denotes (non-normalized) integration over Ω, and S is a source.

- Equation is interesting in its own right, but also a simple prototype for collisional kinetic equations.

Basic Properties

- **Balance Law**. The particle concentration satisfies

$$\boxed{\partial_t \phi + \nabla \cdot \langle \Omega \psi \rangle + \sigma_{\mathrm{a}} \phi = S}$$

- **Diffusion Limit**. On long time-scales and for highly scattering materials, $\psi(x, \Omega) = \phi/4\pi + O(\varepsilon)$ and

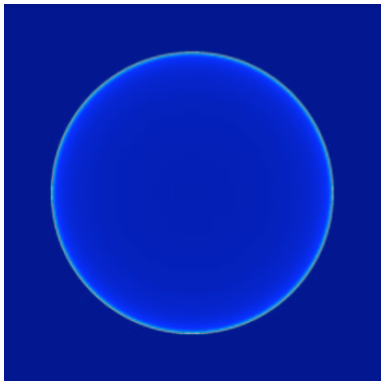
$$\boxed{\partial_t \phi + \sigma_{\mathrm{a}} \phi = \nabla_x \cdot \left(\frac{1}{3\sigma_{\mathrm{s}}} \nabla_x \phi \right) + S + O(\varepsilon^2), \quad \varepsilon \ll 1}$$

- **Streaming Limit**. When $\sigma_{\mathrm{s}} = 0$, the transport equation becomes an infinite set of uncoupled equations.

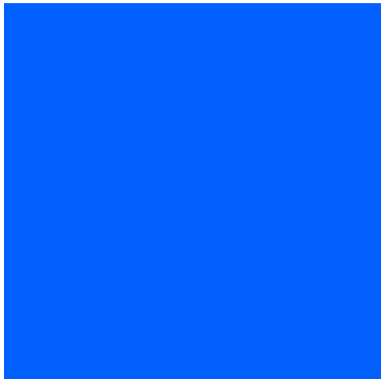
Numerical Simulation

- Challenges:
  - Large phase space.
  - Multiscale behavior.
  - Implicit solutions.
- Common Methods for Angular Discretization:
  - Diffusion
  - Monte Carlo
  - Discrete Ordinates (S_N)
  - Spherical Harmonics Expansion (P_N)

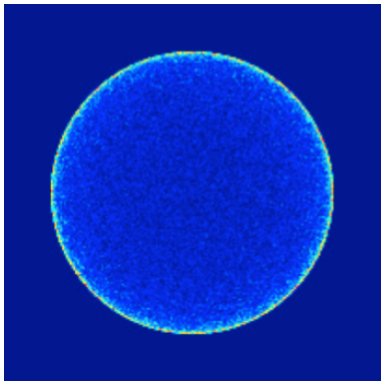
The Line Source Problem: Comparison of Methods[1]



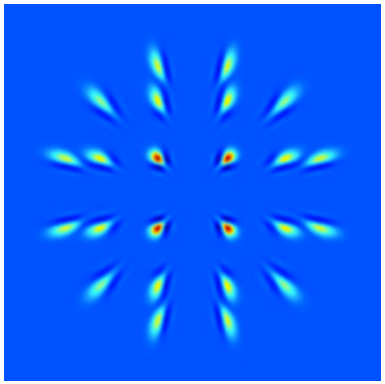
(a) analytic



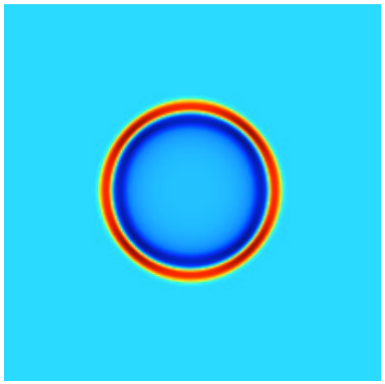
(b) Diffusion



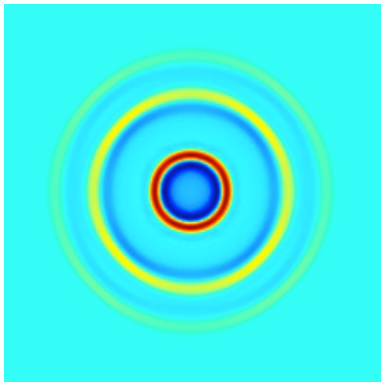
(c) Monte-Carlo



(d) S_6



(e) P_1



(f) P_5

[1] T. A. Brunner. "Forms of Approximate Radiation Transport", Tech. Rep. SAND2002-1778, Sandia National Laboratories, Jul 2002.

Moment Methods

Moment Equations

- Let $\mathbf{m} = \mathbf{m}(\Omega)$ be a vector containing spherical harmonics in Ω.
- Let $\mathbf{u}(x,t) := \langle \mathbf{m}\psi(x,\cdot,t)\rangle$ be moments of ψ with respect to $\mathbf{m}$.
- Multiply the transport equation by $\mathbf{m}$ and integrate over all angles:

$$\boxed{\partial_t \mathbf{u} + \nabla_x \cdot \langle \Omega \mathbf{m}\psi\rangle + \sigma_{\mathrm{t}}\mathbf{u} = \sigma_{\mathrm{s}} Q\mathbf{u} + \mathbf{s}}$$

where $Q = \mathrm{diag}([1, 0\ldots, 0])$ and $\mathbf{s} = [S, 0\ldots, 0]^T$.

Closure

- To close the system, replace ψ by an ansatz $\hat{\psi}(\mathbf{u})$ that satisfies the consistency relation

$$\boxed{\left\langle \mathbf{m}\hat{\psi}(\mathbf{u}(x,t)) \right\rangle = \mathbf{u}(x,t)\ .}$$

- The results gives a closed system of balance laws

$$\boxed{\partial_t \mathbf{u} + \nabla_x \cdot \mathbf{f}(\mathbf{u}) + \sigma_{\mathrm{t}}\mathbf{u} = \sigma_{\mathrm{s}} Q\mathbf{u} + \mathbf{s}}$$

  where $\mathbf{f}(\mathbf{u}) = \langle \Omega \mathbf{m}\hat{\psi}(\mathbf{u}) \rangle$

- The behavior of the moment system depends heavily on the closure.

Entropy-Based Closure

- Let η be a scalar valued, strictly convex function.
- Closures based on the **entropy minimization principle** (or maximization in physics) use the ansatz

$$\boxed{\hat{\psi}(\mathbf{u}) = \arg \min_{g \in L^1(\mathbb{S})} \{\langle \eta(g) \rangle : \langle \mathbf{m} g \rangle = \mathbf{u}\}}$$

  to close the moment system.

Formal Optimality Conditions

- A minimizer, if it exists, is given by $G_{\hat{\boldsymbol{\alpha}}}$, where

$$\boxed{G_{\boldsymbol{\alpha}} \equiv \eta_*'(\boldsymbol{\alpha}^T \mathbf{m})}$$

  where η_* is the Legendre dual of η.

- The dual variable $\hat{\boldsymbol{\alpha}}(\mathbf{u})$ solves

$$\boxed{\min_{\boldsymbol{\alpha} \in \mathbb{R}^n} \left\{ \langle \eta_*(\boldsymbol{\alpha}^T \mathbf{m}) \rangle - \boldsymbol{\alpha}^T \mathbf{u} \right\}}$$

Why Entropy (as a modeling tool) ?

- Entropy minimization is a well-known tool for identifying equilibria, where the moments are conserved quantities (H-Theorem).
- Entropy-based moments systems are symmetric hyperbolic (in the $\hat{\boldsymbol{\alpha}}$ variable) and dissipate the entropy $h(\mathbf{u}) = \langle \eta(G_{\hat{\boldsymbol{\alpha}}(\mathbf{u})}) \rangle$ (in a closed system).
- In nonlinear settings, the collision operator "selects" a specific entropy.
- In linear settings, any strictly convex entropy will be dissipated (in a closed system). Positivity of the ansatz is the practical issue.

Why Entropy (as a practical implementation) ?

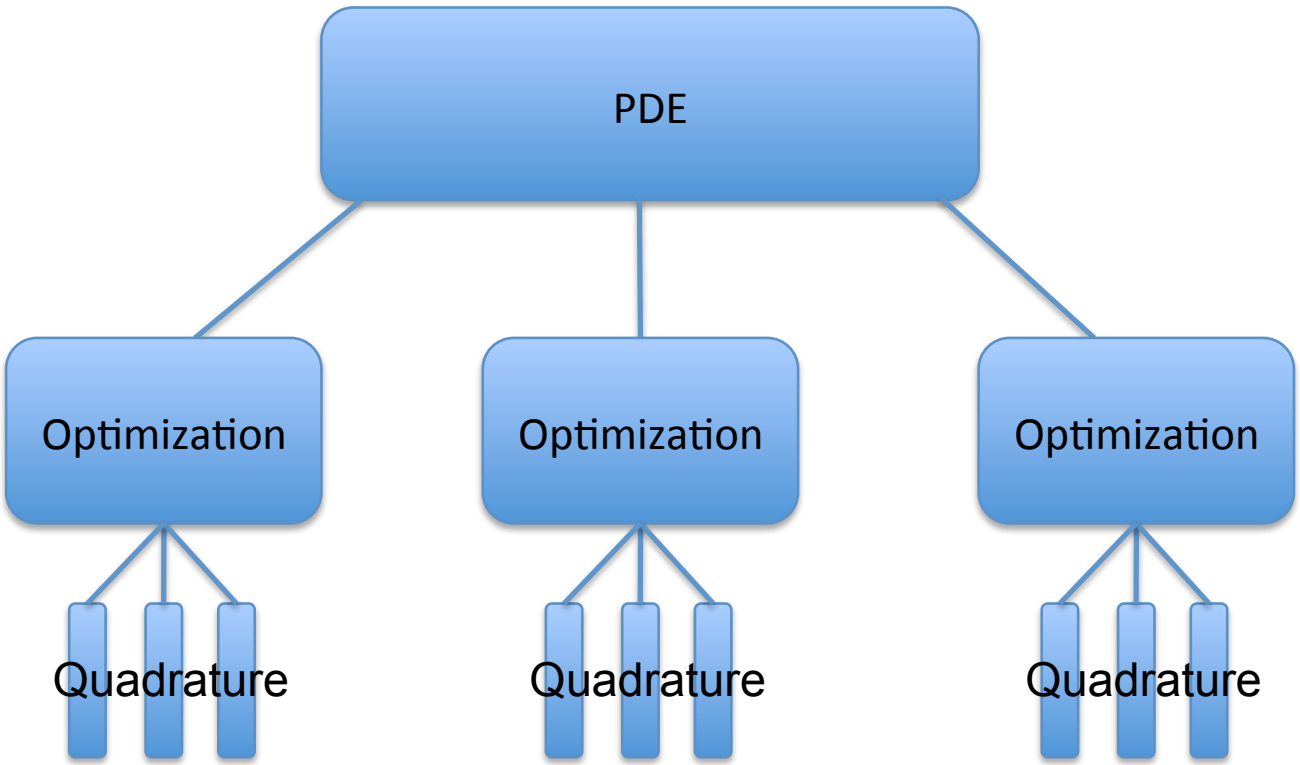
- In many contexts, radiation is a major conduit for energy exchange between materials and high resolution and kinetic simulations are necessary.
- Large scale kinetic simulations are extremely expensive. [Astrophysics supernova group at Oak Ridge estimates $O(10^{16})$ bytes of memory and $O(10^{22}) - O(10^{26})$ flops per run.]
- High performance computers are beginning to show bottlenecks in data management. Problem will get much worse in the future.
- Kinetic simulations will have to go hybrid to limit data: global moments + local kinetics.

Previous Work

- Literally too much to list.
- First use as a non-equilbrium moment closure was by Minerbo in 1978 at Los Alamos: "Maximum Entropy Eddington Factors", *J. Quant. Spectrosc. Radiat. Transfer.*
- Levermore (1996): application to non-equilibrium gas dynamics; formal mathematical properties
- Dubroca and Feugeas (1999): thermal radiative transfer with Bose-Einstein statistics
- Relationship to Extended Thermodyanmics (Müller and Ruggeri)

Solving the Optimization

- In the past, most work has focused on the two moment model M_1.
- Solving the optimization is extremely expensive: iterative methods, quadrature evaluation.
- However ... in large-scale, parallel computing environments, the expense can be mitigated.



Implementation of the Closure

Setup

- We assume a slab geometry with no source, in which case the transport equation is

$$\boxed{\partial_t \psi + \mu \partial_x \psi + \sigma_{\mathrm{t}} \psi = \frac{\sigma_{\mathrm{s}}}{2} \phi}$$

  where $\mu \in [-1, 1]$.
- The vector $\mathbf{m}$ contains the first $N+1$ Legendre polynomials in μ.
- We use the Maxwell-Boltzmann entropy. The entropy ansatz is

$$\boxed{\hat{\psi}(\mathbf{u}) = G_{\hat{\boldsymbol{\alpha}}(\mathbf{u})}\,, \quad \text{where} \quad G_{\boldsymbol{\alpha}} \equiv \exp(\boldsymbol{\alpha}^T \mathbf{m})}$$

  and $\hat{\boldsymbol{\alpha}}(\mathbf{u})$ solves

$$\boxed{\min_{\boldsymbol{\alpha} \in \mathbb{R}^{N+1}} \left\{ \langle G_{\boldsymbol{\alpha}} \rangle - \boldsymbol{\alpha}^T \mathbf{u} \right\}}$$

Realizability

Definition

A vector $\mathbf{v} \in \mathbb{R}^{N+1}$ is *realizable* if $\mathbf{v} = \langle \mathbf{m} g \rangle$ for some non-negative $L^1(d\mu)$ function g with $\langle g \rangle > 0$. The set of all such $\mathbf{v}$ is denoted by $\mathcal{R}_{\mathbf{m}}$.

Realizability is a necessary condition for the optimization problem to have a solution. Unfortunately,

1. A typical numerical algorithm for updating the moments need not preserve realizability.
2. The optimization problem is much harder to solve near the boundary of $\mathcal{R}_{\mathbf{m}}$.

Kinetic Scheme

- To approximate

$$\boxed{\partial_t \mathbf{u} + \partial_x \mathbf{f}(\mathbf{u}) + \sigma_{\mathrm{t}} \mathbf{u} = \sigma_{\mathrm{s}} Q \mathbf{u},}$$

  we use a finite-volume method.

- Let

$$\boxed{\mathbf{u}_j(t) \simeq \frac{1}{\Delta x} \int_{I_j} \mathbf{u}(x,t)\, dx\ .}$$

  approximate the average of $\mathbf{u}$ over spatial cells I_j of width Δx.

- The kinetic scheme has the form

$$\boxed{\partial_t \mathbf{u}_j + \frac{\mathbf{f}_{j+1/2} - \mathbf{f}_{j-1/2}}{\Delta x} + \sigma_{\mathrm{t}} \mathbf{u}_j = \sigma_{\mathrm{s}} Q \mathbf{u}_j\ ,}$$

- For time integration, we use a second-order SSP Runge-Kutta scheme (Heun's method).

Kinetic Scheme (contd.)

- Let $\bar{\boldsymbol{\alpha}}(\mathbf{u})$ denote the optimization algorithm's approximation to the true solution $\hat{\boldsymbol{\alpha}}(\mathbf{u})$
- Let $\bar{G}_j := G_{\bar{\boldsymbol{\alpha}}(\mathbf{u}_j)}$ and $\hat{G}_j := G_{\hat{\boldsymbol{\alpha}}(\mathbf{u}_j)}$
- Edge values for the flux are computing using the entropy ansatz

$$\boxed{\mathbf{f}_{j+1/2} = \left\langle \mu \mathbf{m} \bar{G}_{j+1/2} \right\rangle,}$$

where

$$\boxed{\bar{G}_{j+1/2} := \begin{cases} \bar{G}_j + \frac{\Delta x}{2} s_j & \mu > 0 \\ \bar{G}_{j+1} - \frac{\Delta x}{2} s_{j+1} & \mu < 0 \end{cases}}$$

Kinetic Scheme (contd.)

- The quantities s_j approximate the spatial derivative of $\bar{G}$

$$\boxed{s_j := \text{minmod}\left\{\theta\frac{\bar{G}_j - \bar{G}_{j-1}}{\Delta x}, \frac{\bar{G}_{j+1} - \bar{G}_{j-1}}{2\Delta x}, \theta\frac{\bar{G}_{j+1} - \bar{G}_j}{\Delta x}\right\}.}$$

  for $\theta \in (1, 2)$.

- Note that s_j is computed only by communicating the moments.

Maintaining Realizability

- Realizability depends on the ratio

$$\boxed{\gamma_j(\mu) := \frac{\bar{G}_j(\mu)}{\hat{G}_j(\mu)}.}$$

Theorem

Given the modified CFL condition

$$\max_{\mu,\, j}\{\gamma_j\}\frac{\Delta t}{\Delta x}\frac{\theta+2}{2} + \sigma_{\mathrm{t}}\Delta t < 1,$$

the realizable set is invariant under the kinetic solver.

The Dual Problem

- The dual problem for $\hat{\boldsymbol{\alpha}}(\mathbf{u})$ is

$$\boxed{\min_{\boldsymbol{\alpha}\in\mathbb{R}^n} \left\{ \langle G_{\boldsymbol{\alpha}} \rangle - \boldsymbol{\alpha}^T \mathbf{u} \right\}}$$

  where $G_{\boldsymbol{\alpha}} = \exp(\boldsymbol{\alpha}^T \mathbf{m})$.

- The gradient g and Hessian H are

$$\boxed{g(\boldsymbol{\alpha}) = \langle \mathbf{m} G_{\boldsymbol{\alpha}} \rangle - \mathbf{u} \quad \text{and} \quad H(\boldsymbol{\alpha}) = \langle \mathbf{m}\mathbf{m}^T G_{\boldsymbol{\alpha}} \rangle > 0}$$

Challenges of the Dual: Quadrature

- For moments near the realizable boundary the effective support of the entropy ansatz becomes very narrow.
- The dual problem is very sensitive to the numerical quadrature.

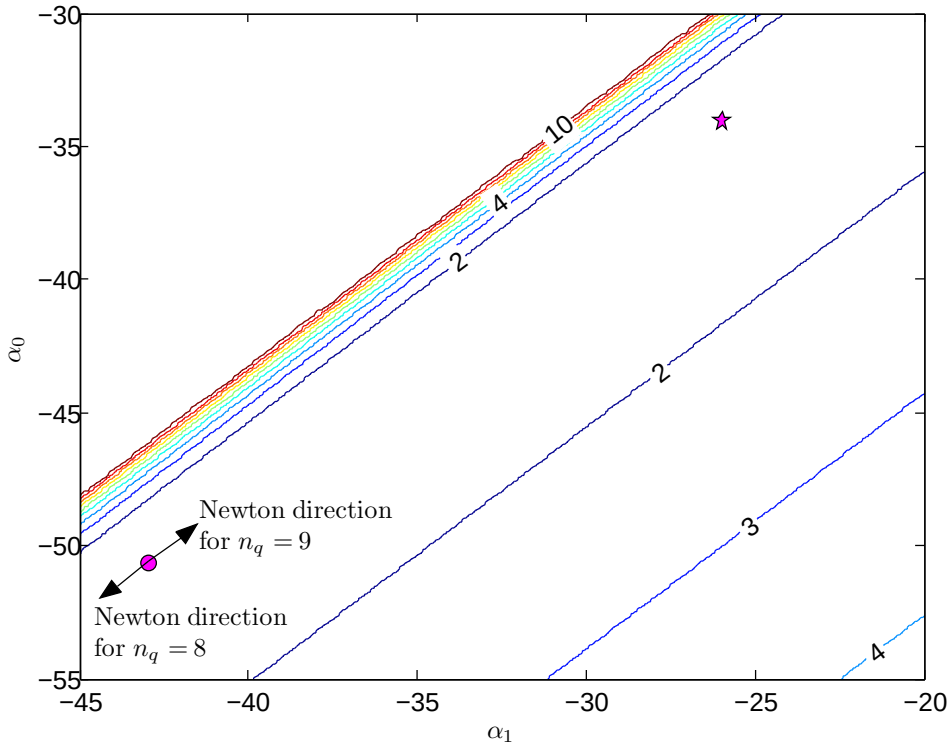


Figure: Level curves of the dual objective in a two-moment problem.

Challenges of the Dual: Sensitivity

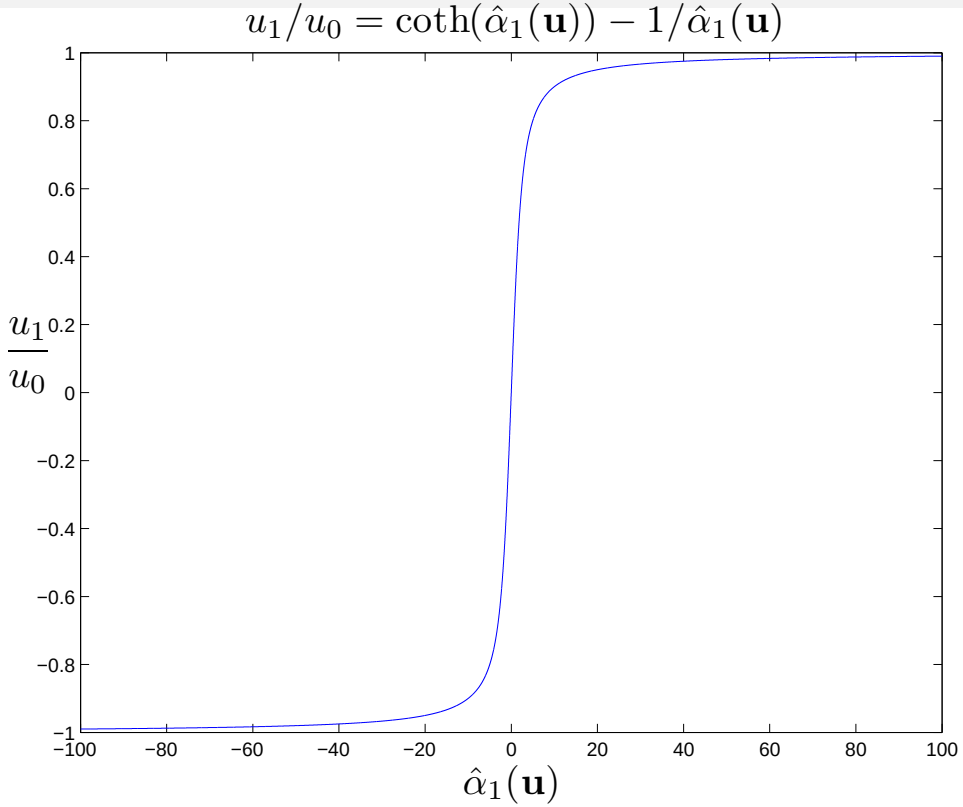
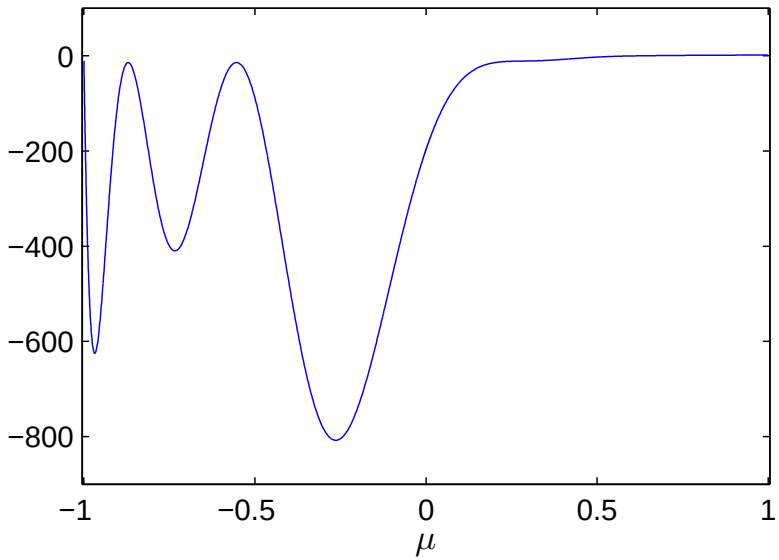


Figure: Relating multipliers to moments in a two moment problem.

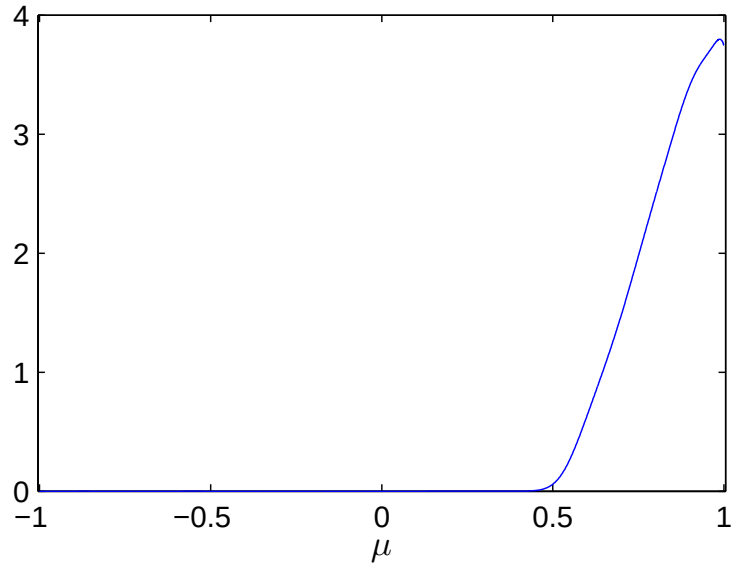
- Near the realizable boundary, the Hessian becomes singular.
- Our numerical scheme maintains realizability under the assumption that the dual problem for $\hat{\boldsymbol{\alpha}}$ is sufficiently converged.

Challenges of the Dual: Poorly Conditioned Hessian

- In the course of solving the dual problem for an M_{15} model we come across a "bad" value of $\boldsymbol{\alpha}$.

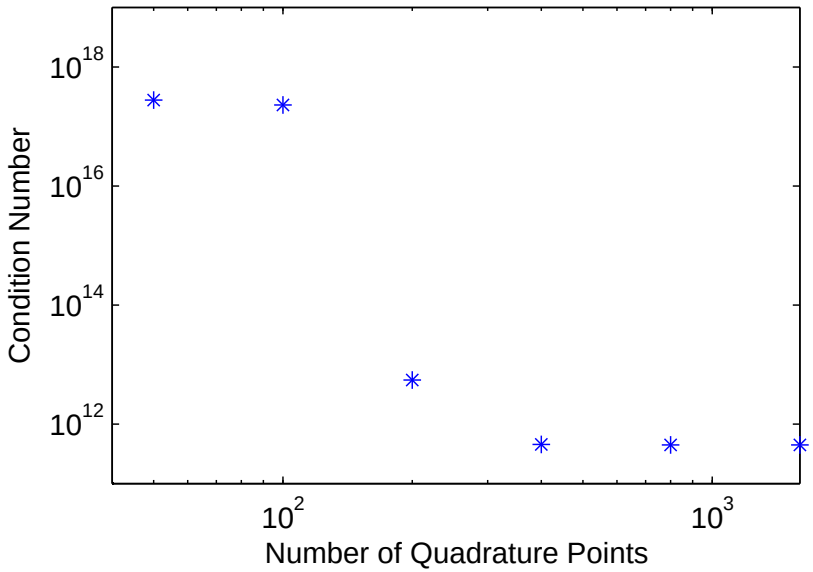


(a) The polynomial $\boldsymbol{\alpha}^T \mathbf{m}$.

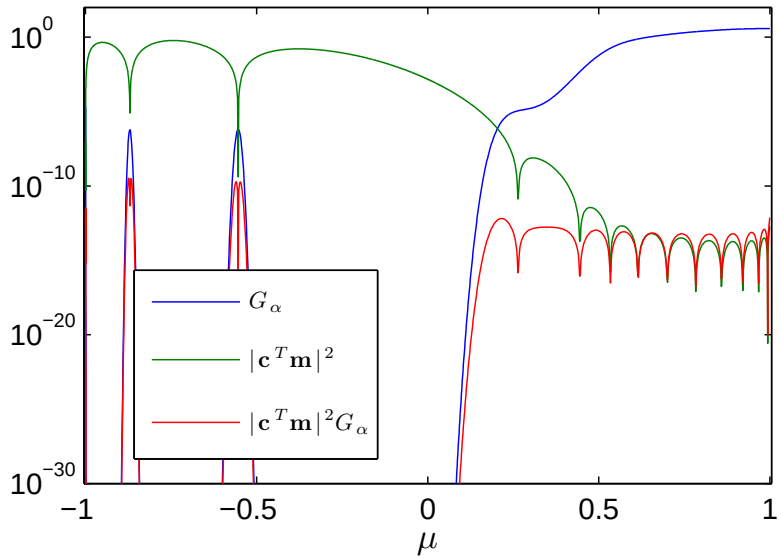


(b) The ansatz $G_{\boldsymbol{\alpha}}$.

Challenges of the Dual: Poorly Conditioned Hessian



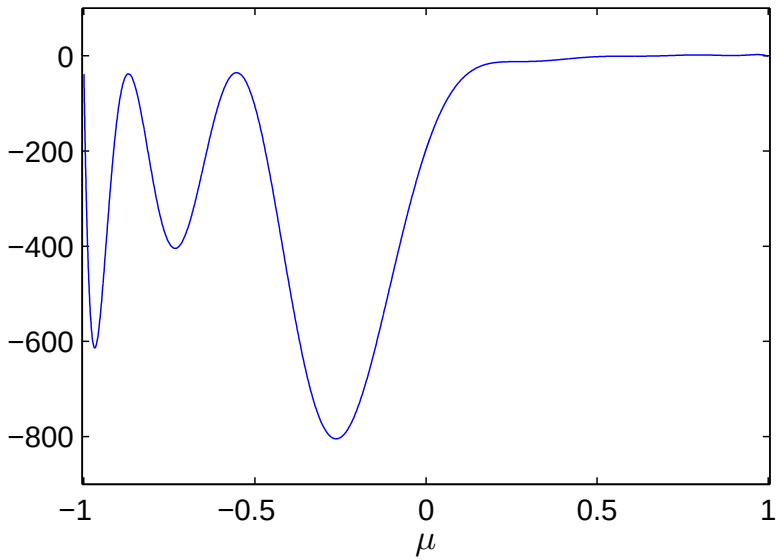
(a) The (approximate) condition number for $H(\boldsymbol{\alpha})$.



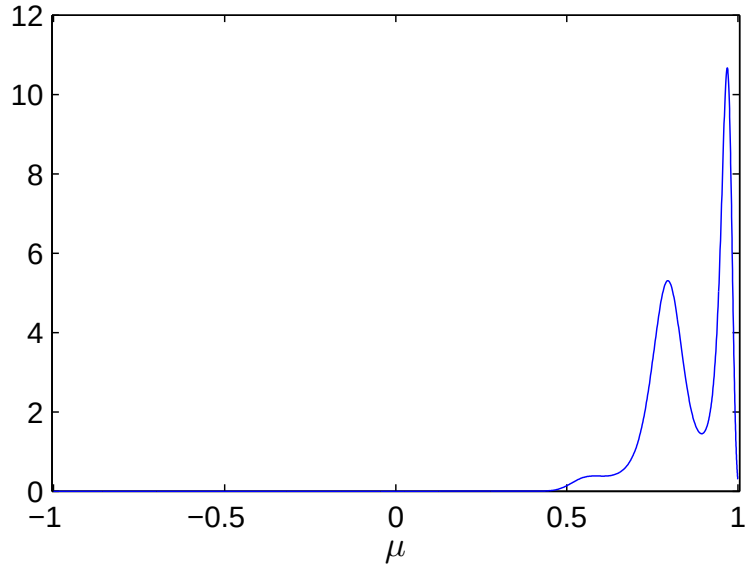
(b) The integrand at the minimum of the quadratic form.

Challenges of the Dual: Poorly Conditioned Hessian

- With a small perturbation, we can generate a "super-bad" value of $\boldsymbol{\alpha}$.

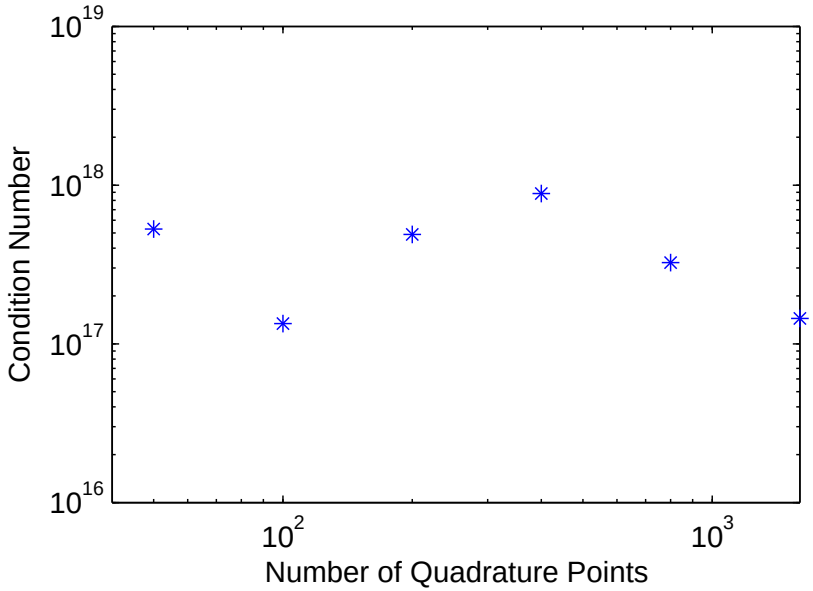


(a) The polynomial $\boldsymbol{\alpha}^T \mathbf{m}$.

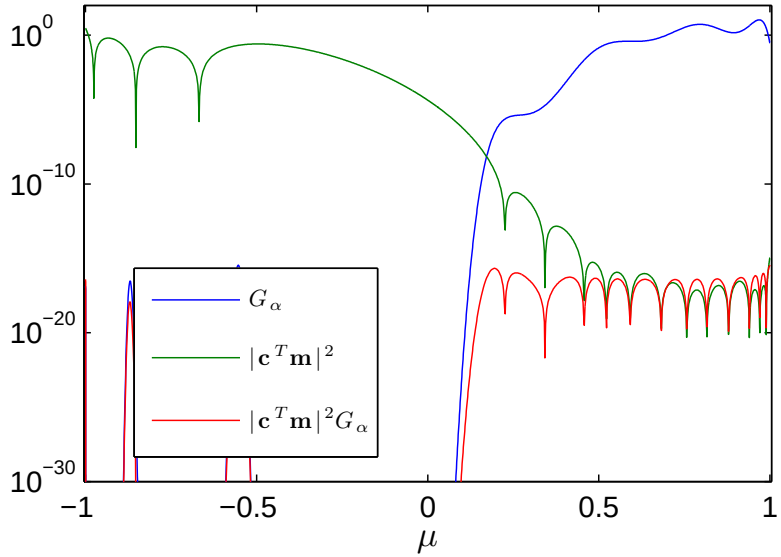


(b) The ansatz $G_{\boldsymbol{\alpha}}$.

Challenges of the Dual: Poorly Conditioned Hessian



(a) The (approximate) condition number for $H(\boldsymbol{\alpha})$.



(b) The integrand at the minimum of the quadratic form.

Regularizing the Moments

- As $\mathbf{u}$ approaches the boundary of realizability, the dual problem becomes impossible to solve in finite-precision arithmetic.

- We find nearby realizable moments by replacing $\mathbf{u}$ by

$$\mathbf{v}(r) = (1 - r)\mathbf{u} + ru_0\mathbf{e},$$

where $\mathbf{e} = (1, 0, \ldots, 0)$ and $0 < r \ll 1$.

Numerical Results: Two-Beam Instability

- Bounded domain: $x \in (x_L, x_R) = (-0.5, 0.5)$
- Boundary conditions:

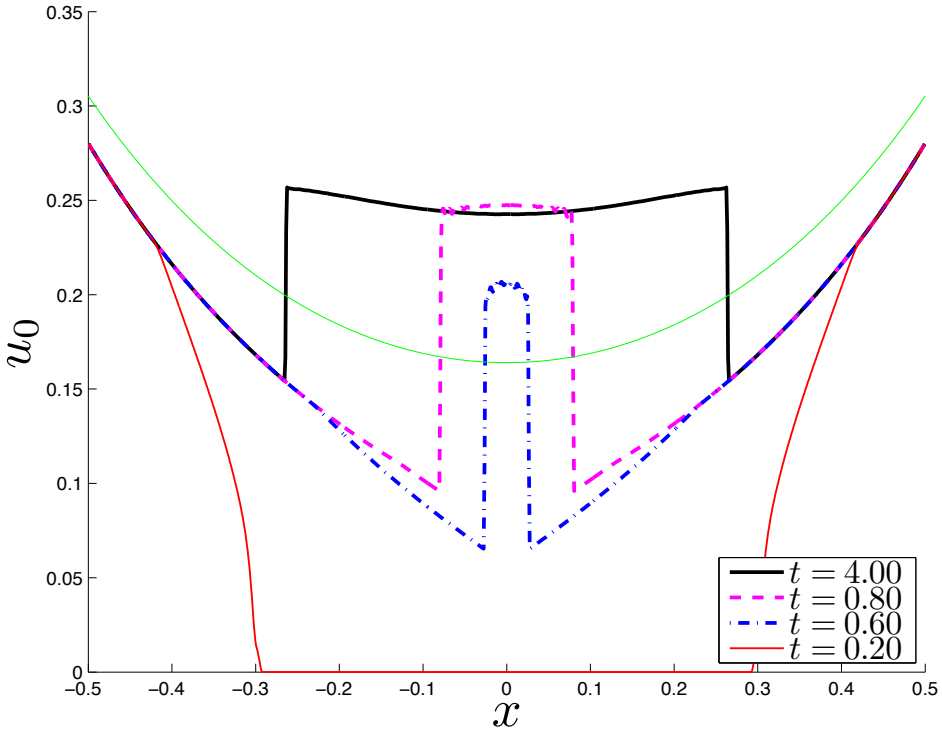
$$\psi(x_L, \mu, t) = \exp(-10(\mu - 1)^2)$$
$$\psi(x_R, \mu, t) = \exp(-10(\mu + 1)^2)$$

- Initially, (almost) a void:

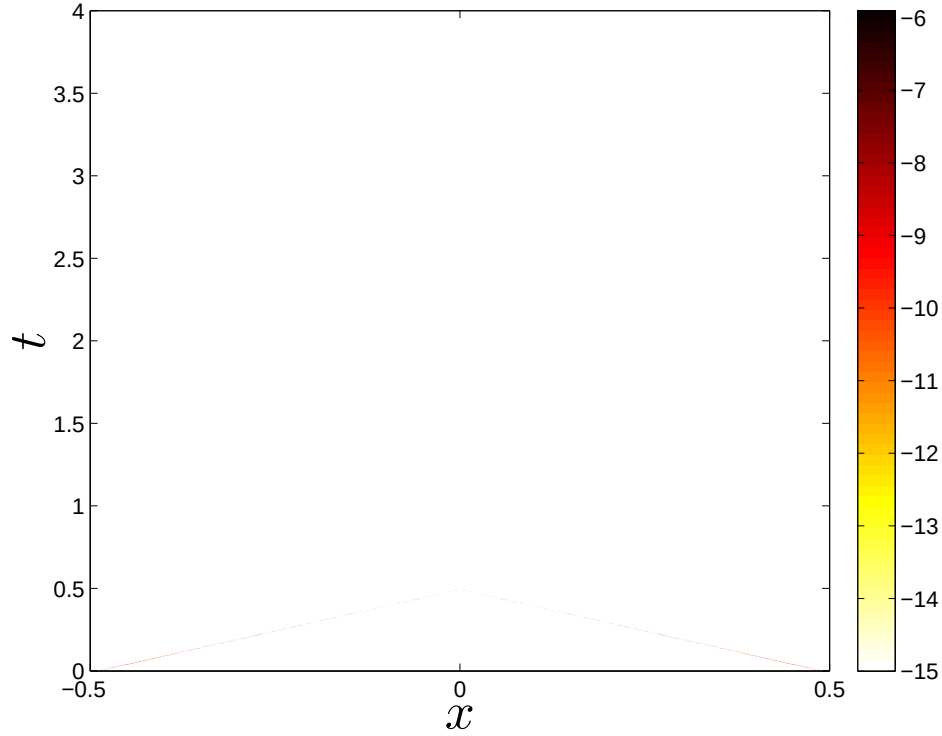
$$\psi(x, \mu, t = 0) = \psi_{\text{floor}}$$

- Purely absorbing medium:

$$\sigma_{\text{a}} = 2\,,\ \sigma_{\text{s}} = 0$$

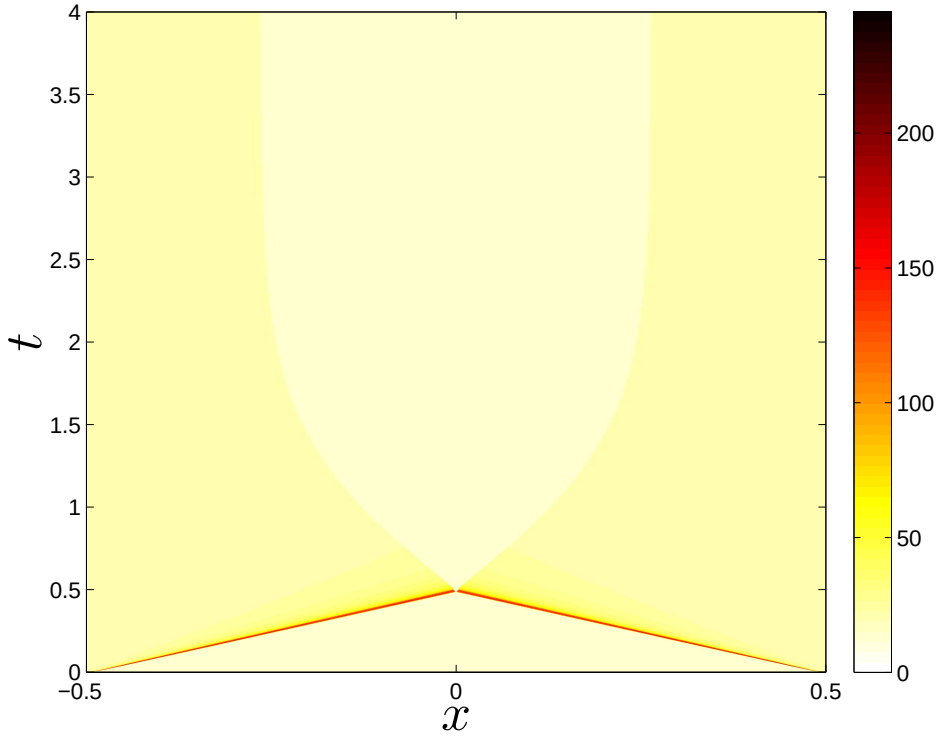


(a) Snapshots of the solution, $u_0(x, t)$.

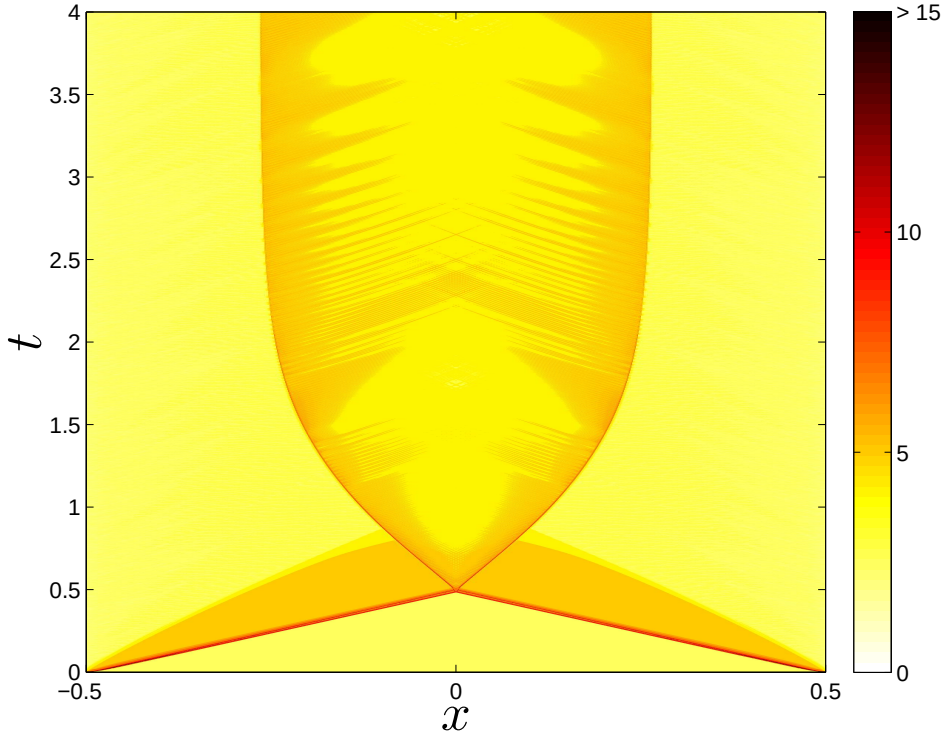


(b) $\log_{10} \|\mathbf{u} - \mathbf{v}(r)\|$.

Figure: A two-beam instability with $N = 1$ and 1000 spatial cells.

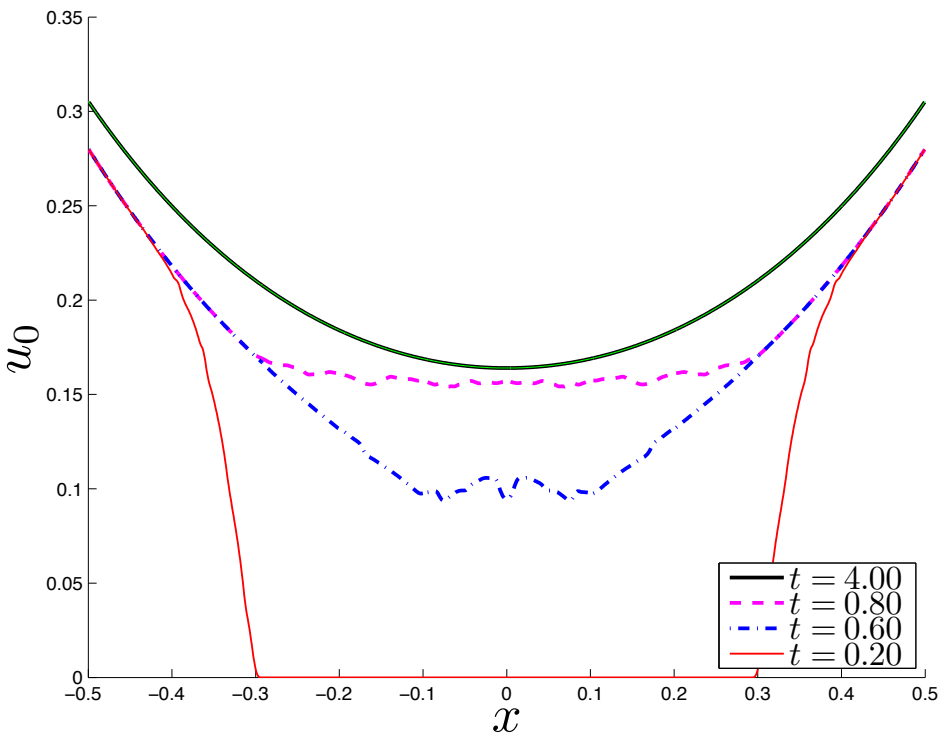


(a) The number of quadrature points.

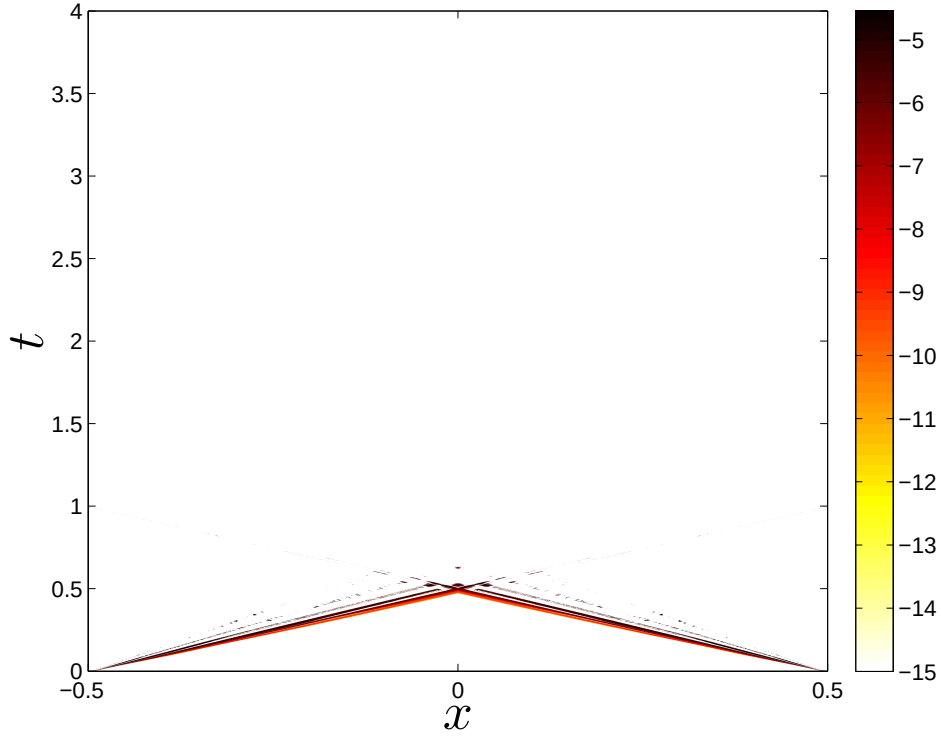


(b) The number of iterations.

Figure: A two-beam instability with $N = 1$ and 1000 spatial cells.

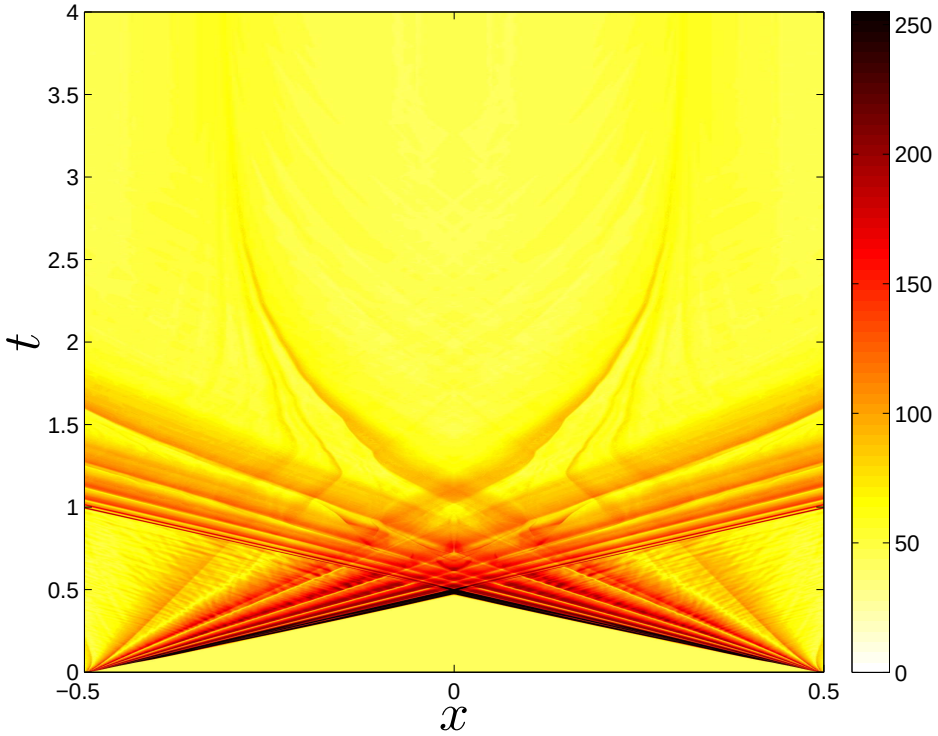


(a) Snapshots of the solution, $u_0(x, t)$.

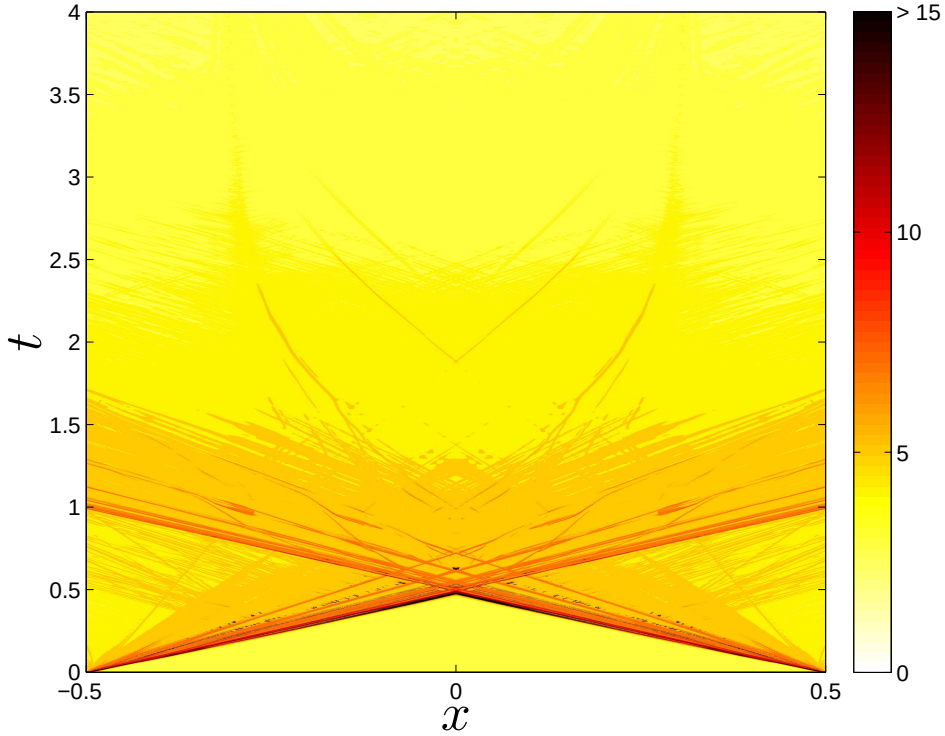


(b) $\log_{10} \|\mathbf{u} - \mathbf{v}(r)\|$.

Figure: A two-beam instability with $N = 15$ and 1000 spatial cells.



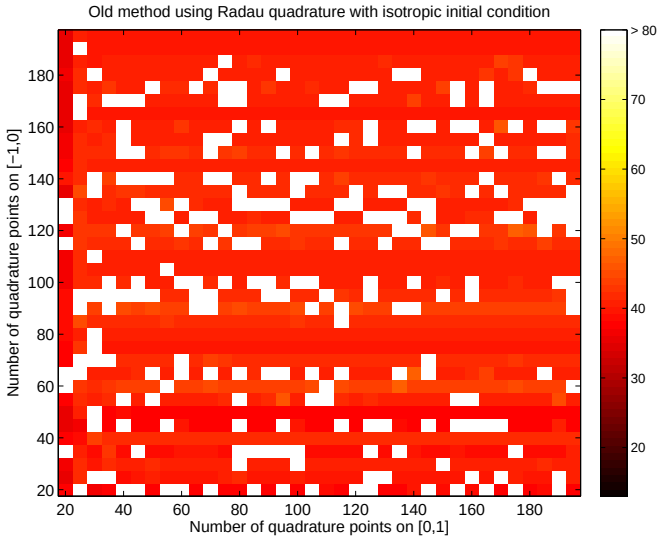
(a) The number of quadrature points.



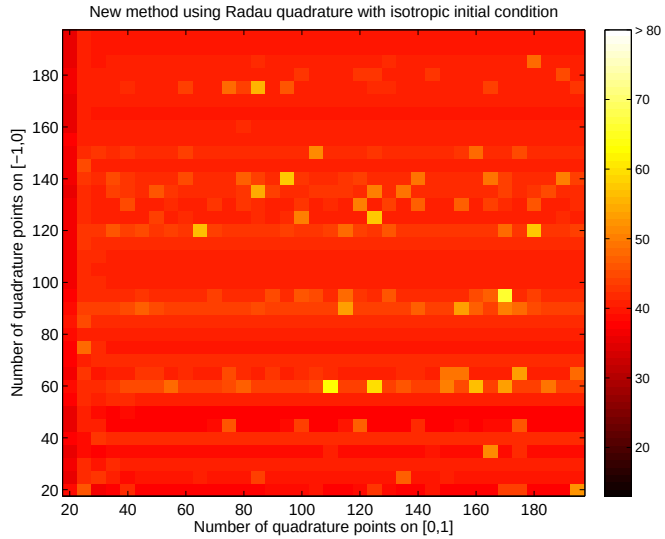
(b) The number of iterations.

Figure: A two-beam instability with $N = 15$ and 1000 spatial cells.

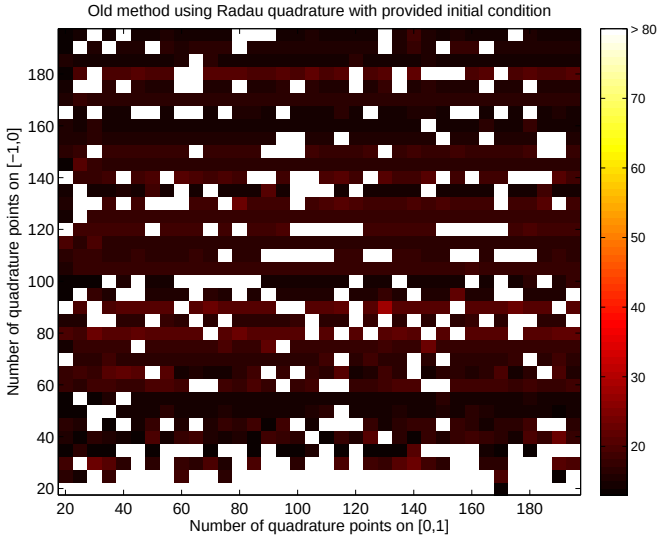
Improving the Hessian: Change of Bases



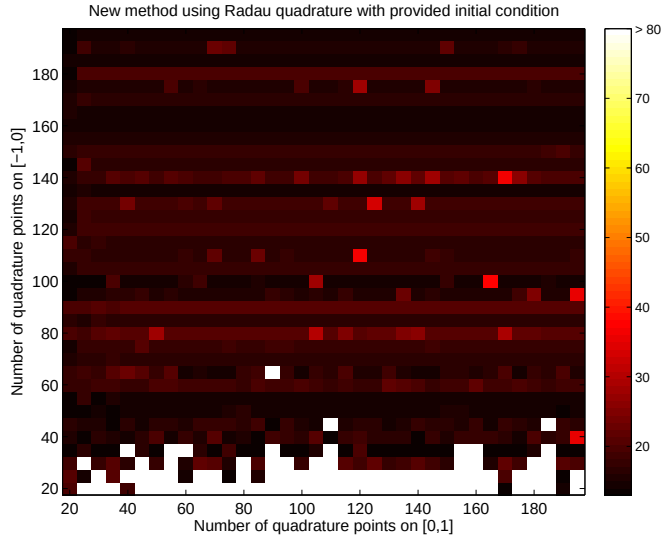
(a) Old, cold start.



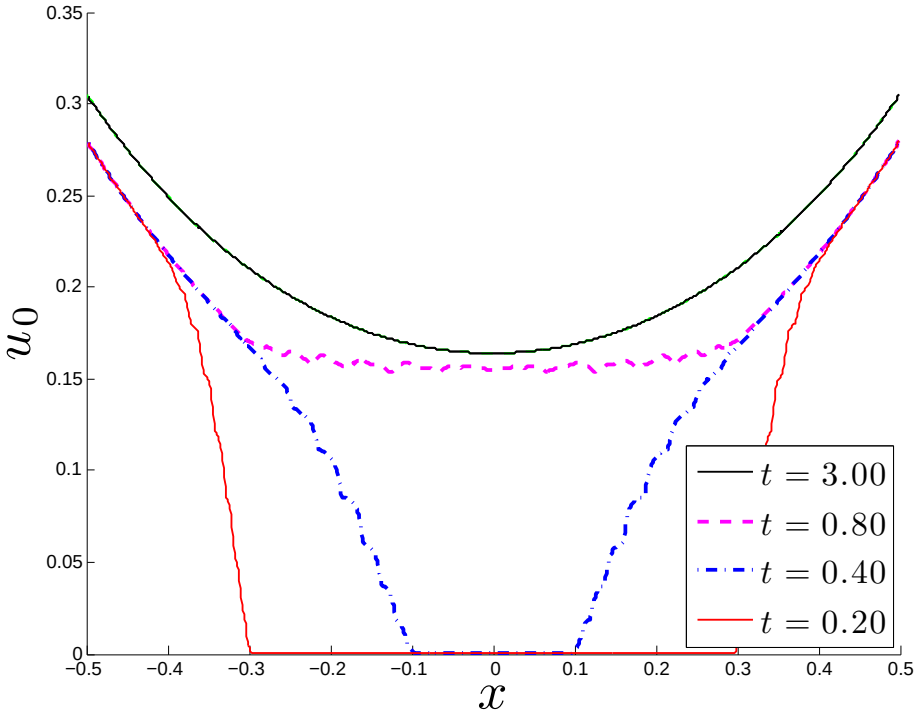
(b) New, cold start.



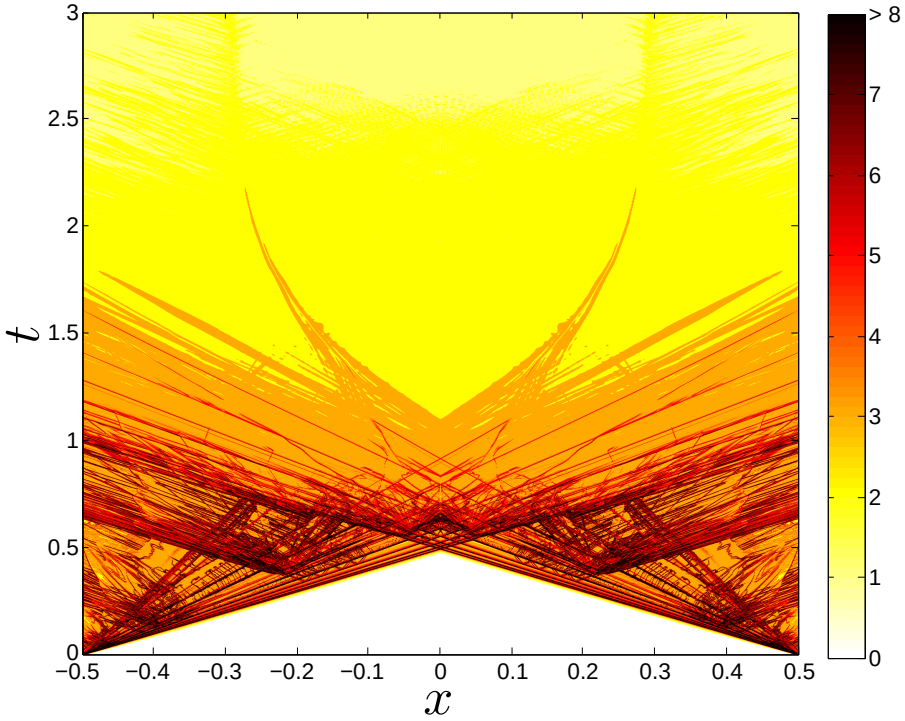
(c) Old, warm start.



(d) New, warm start.



(a) The number of quadrature points.

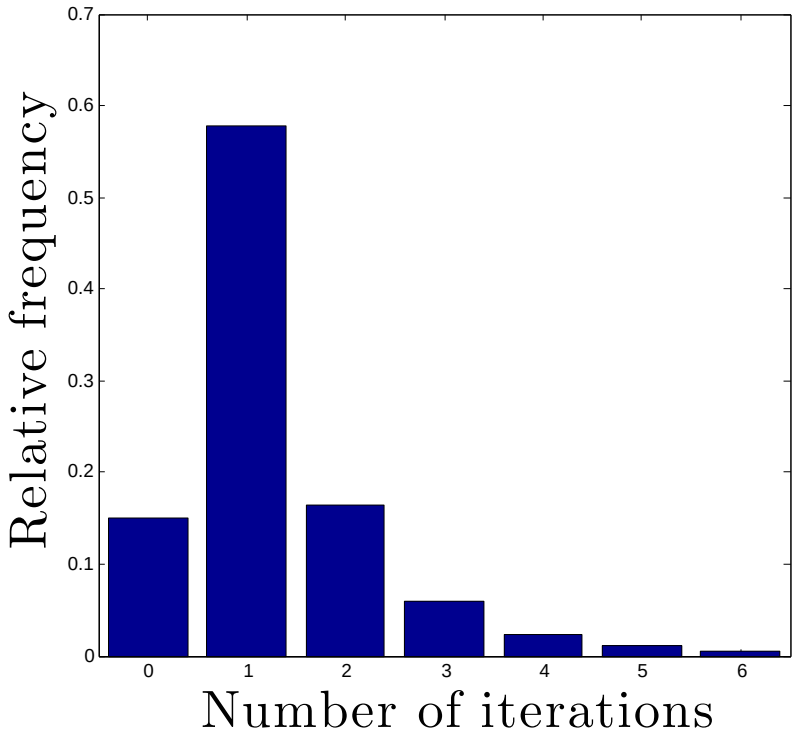


(b) The number of iterations.

Figure: A two-beam instability with $N = 15$, 1000 spatial cells, 40-pt quadrature.

Some Statistics

- The optimization problem is solved 10,851,552 times.
- 94.9031% of problems needed 3 or fewer iterations.
- 1.3266% of problems needed more than 6 iterations.
- average number of iterations was 1.6116
- 0.012772% of problems were regularized.



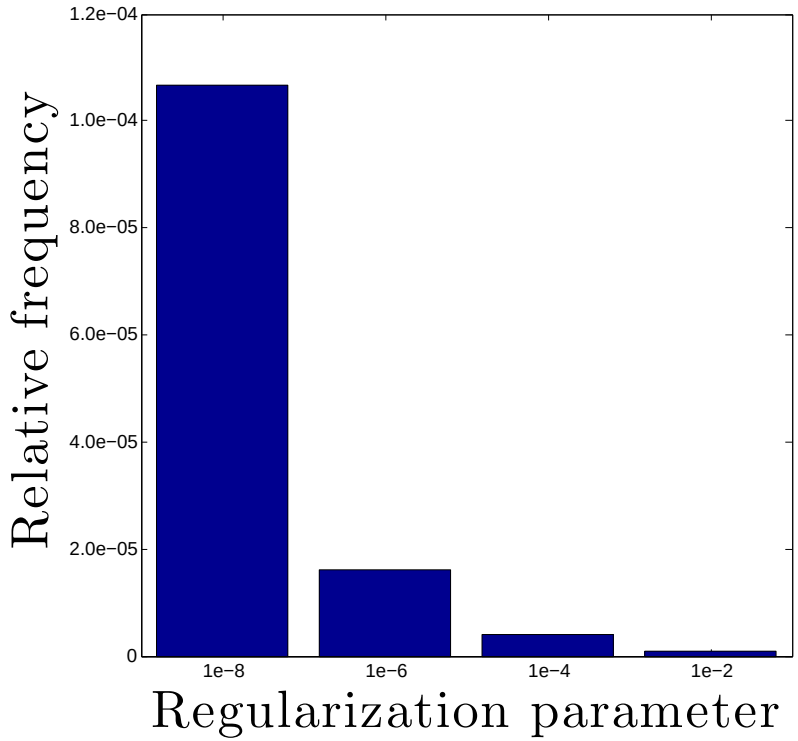


Figure: A two-beam instability with $N = 15$ and 1000 spatial cells.

Using CUDA

- Profiling the old optimization code shows that that integrals evaluation in objective, gradient, and Hessian take 95% of the time.
- Maximum Possible Speedup by Amdahl's Law:

$$\frac{1}{(1 - .95)} = 20.$$

- A More Realistic Estimate
  - 336 Cores at 1.7 GHz vs 1 CPU Core at 3GHz;
  - Double Precision 8 Times Slower Than Single Precision

$$\frac{1}{(1 - .95) + .95/(336 * 1.7/3/8)} \approx 11.$$

- Computational Results
  - Serial Performance: 18.096 sec
  - GPU Performance: 4.376 sec
  - Actual Speedup: 4.135

Outlook

What's been done

- We have a working algorithm in 1-D MATLAB and C which includes
  - A second-order scheme that preserves realizability (in exact arithmetic).
  - An optimization algorithm with change of basis and option for adaptive quadrature.
  - A reasonable regularization procedure when things are really bad.
- A CUDA (GPU) code is also running, but not optimzed and not for the newest algorithm.

What's to be done

- An asymptotic preserving (AP) scheme.
- Implicit time integration.
- Multi-D results.
- Scaling studies with parallel implementation.

Thank you!