Using one Thousand GPUs to Understand the Euler Equations

Kjetil Lye (ETH Zurich)

joint work with: Ulrik S. Fjordholm (University of Oslo) Siddhartha Mishra (ETH Zurich) Geilo Winter School, 2019

Funded by the ERC SPARCCLE project

Compute time on CSCS through s665, s839

We will use a thousand of these



price tag: 15 000 EUR

We will use a thousand of these



... to understand these

$$\rho_t + \nabla_x \cdot (\rho \mathbf{v}) = 0$$
$$(\rho \mathbf{v})_t + \nabla_x \cdot (\rho \mathbf{v} \otimes \mathbf{v} + pI) = 0$$
$$E_t + \nabla_x \cdot ((E + p)\mathbf{v}) = 0$$

The Compressible Euler Equations

A model for gas



$$\rho_t + \nabla_x \cdot (\rho \mathbf{v}) = 0$$
$$(\rho \mathbf{v})_t + \nabla_x \cdot (\rho \mathbf{v} \otimes \mathbf{v} + pI) = 0$$
$$E_t + \nabla_x \cdot ((E + p)\mathbf{v}) = 0$$

with suitable equation of state

Hyperbolic conservation laws

$$u_t +
abla_{ imes} \cdot f(u) = 0$$
 on $D imes \mathbb{R}^+$
here $D \subset \mathbb{R}^d$ and $u : D imes \mathbb{R}^+ o \mathbb{R}^N$.

- *d*: dimension
- N: number of unknowns

Hyperbolic conservation laws

$$u_t +
abla_{\mathbf{x}} \cdot f(u) = 0$$
 on $D imes \mathbb{R}^+$
here $D \subset \mathbb{R}^d$ and $u : D imes \mathbb{R}^+ o \mathbb{R}^N$.

- *d*: dimension
- N: number of unknowns

The compressible **Euler equations** (d = 3, N = 5):

$$\begin{pmatrix} \rho_t \\ (\rho \mathbf{v})_t \\ E_t \end{pmatrix} + \nabla_x \cdot \begin{pmatrix} \rho \mathbf{v} \\ \rho \mathbf{v} \otimes \mathbf{v} + \rho l \\ (E+\rho) \mathbf{v} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Are non-linear hyperbolic conservation laws well-posed?

Burgers' equation with smooth initial data

$$u_t + \left(\frac{1}{2}u^2\right)_x = 0$$
 on $D \times \mathbb{R}^+$

• Discontinuities form



6

Are non-linear hyperbolic conservation laws well-posed?

Go to weak formulation

$$\int_{D\times\mathbb{R}^+} u\phi_t + \nabla_x \phi \cdot f(u) \, dx + \int_D u(x,0)\phi(x,0) \, dx = 0$$

for all $\phi \in C_c^\infty(D\times\mathbb{R}^+)$

• Weak solutions are not unique

$$\int_{D\times\mathbb{R}^+} u\phi_t + \nabla_x \phi \cdot f(u) \, dx + \int_D u(x,0)\phi(x,0) \, dx = 0$$

for all $\phi \in C_c^\infty(D\times\mathbb{R}^+)$

Introduce entropy conditions

$$\eta(u)_t +
abla_x \cdot q(u) \leq 0$$
 in the sense of distributions

• Well-posedness for scalar equations

$$\int_{D\times\mathbb{R}^+} u\phi_t + \nabla_x \phi \cdot f(u) \, dx + \int_D u(x,0)\phi(x,0) \, dx = 0$$

for all $\phi \in C_c^\infty(D\times\mathbb{R}^+)$

Introduce entropy conditions

 $\eta(u)_t +
abla_{x} \cdot q(u) \leq 0$ in the sense of distributions

- Well-posedness for scalar equations
- Well-posedness for systems in 1D with small initial data. [Glimm, 1965; Bressan et al, 2000]

$$\int_{D\times\mathbb{R}^+} u\phi_t + \nabla_x \phi \cdot f(u) \, dx + \int_D u(x,0)\phi(x,0) \, dx = 0$$

for all $\phi \in C_c^\infty(D\times\mathbb{R}^+)$

Introduce entropy conditions

 $\eta(u)_t +
abla_{x} \cdot q(u) \leq 0$ in the sense of distributions

- Well-posedness for scalar equations
- Well-posedness for systems in 1D with small initial data. [Glimm, 1965; Bressan et al, 2000]

However, DeLellis et al constructed *infinitely* many entropy solutions to 2D Euler, **and** the Shallow Water equations

Are non-linear hyperbolic conservation laws well-posed?

Only if N = 1 (scalar) or d = 1 (1 spatial dimension).

Are non-linear hyperbolic conservation laws well-posed?

Only if N = 1 (scalar) or d = 1 (1 spatial dimension).

Does it matter in practice?

Kelvin-Helmholtz: Numerical experiment



Kelvin-Helmholtz: Numerical experiment





Kjetil Olsen Lye, ETHZ

Infinitely small structures appear (16384×16384)



Appears in clouds



\ldots and Norwegian slow TV



Source: NRK.no, Saltstraumen minutt for minutt, Bodø, Norway



















Cauchy convergence $(L^{1}(D))$

$$ar{
ho}(x,t) = rac{1}{2} \left(
ho_1(x,t) +
ho_2(x,t)
ight).$$



17

$$\bar{\rho}(x,t) = \frac{1}{8} \sum_{k=1}^{8} \rho_k(x,t).$$



17

$$\bar{\rho}(x,t) = \frac{1}{64} \sum_{k=1}^{64} \rho_k(x,t).$$



$$\bar{\rho}(x,t) = rac{1}{1024} \sum_{k=1}^{1024} \rho_k(x,t).$$



Statistics of the solution are well-behaved²

²For *one point statistics*, theory and experiments were done in [Fjordholm, Käppeli, Mishra, Tadmor, 2014]

Statistical solutions

The epistemological value of probability theory is based on the fact that chance phenomena, considered collectively and on a grand scale, create non-random regularity.

– Andrey Kolmogorov,

Limit Distributions for Sums of Independent Random Variables (1954)

Statistical solutions

• Classical: Function $u: D \times [0, T) \rightarrow \mathbb{R}^N$



Statistical solutions

- Classical: Function $u: D \times [0, T) \rightarrow \mathbb{R}^N$
- Statistical solution: $\mu_t \in \operatorname{Prob}(L^1(D, \mathbb{R}^N))$

Duality



A formal computation

$$u_t + f(u)_x = 0$$

Second moment evolves as

 $\partial_t(u(x_1,t)u(x_2,t))$

٠

A formal computation

$$u_t + f(u)_x = 0$$

Second moment evolves as

$$\partial_t(u(x_1,t)u(x_2,t)) + \overbrace{\partial_{x_2}(u(x_1,t)f(u(x_2,t)))}^{=u(x_1,t)u(x_2,t)_t} + \underbrace{\partial_{x_1}(u(x_2,t)f(u(x_1,t)))}_{=u(x_1,t)_tu(x_2,t)} = 0.$$

Generalized to

$$\partial_t(u(x_1,t)\cdots u(x_k,t))+\sum_{i=1}^k \partial_{x_i}\Big(u(x_1,t)\cdots f(u(x_i,t))\cdots u(x_k,t)\Big)=0.$$

A
$$u^t = \left(
u^t_{x_1},
u^t_{x_1, x_2}, \ldots
ight)$$
 is a statistical solution if for every $k \in \mathbb{N}$

$$\partial_t \langle \nu_{x_1,\ldots,x_k}^t, \xi_1 \cdots \xi_k \rangle + \sum_{i=1}^k \partial_{x_i} \langle \nu_{x_1,\ldots,x_k}^t, \xi_1 \cdots f(\xi_i) \cdots \xi_k \rangle = 0,$$

in the sense of distributions.

Numerical algorithm

A man provided with paper, pencil, and rubber, and subject to strict discipline, is in effect a universal machine.

- Alan Turing
Let $\mu_0 \in \operatorname{Prob}(L^1(D, \mathbb{R}^N))$ be the initial data. Let \mathcal{S}^{Δ} be a numerical evolution operator

³Based on the FKMT algorithm [Fjordholm, Käppeli, Mishra, Tadmor, 2014]

Numerical algorithm for statistical solutions³

Let $\mu_0 \in \operatorname{Prob}(L^1(D, \mathbb{R}^N))$ be the initial data. Let \mathcal{S}^{Δ} be a numerical evolution operator

1. Draw *M* i.d.d. samples u_0^k from μ_0



Numerical algorithm for statistical solutions³

Let $\mu_0 \in \operatorname{Prob}(L^1(D, \mathbb{R}^N))$ be the initial data. Let \mathcal{S}^{Δ} be a numerical evolution operator

- 1. Draw *M* i.d.d. samples u_0^k from μ_0
- 2. For $k = 1, \ldots, M$ evolve

$$u^{k,\Delta} = \mathcal{S}^{\Delta}(u_0^k)$$



 u_0^k





Let $\mu_0 \in \operatorname{Prob}(L^1(D, \mathbb{R}^N))$ be the initial data. Let S^{Δ} be a numerical evolution operator

- 1. Draw *M* i.d.d. samples u_0^k from μ_0
- 2. For $k = 1, \ldots, M$ evolve

$$u^{k,\Delta} = \mathcal{S}^{\Delta}(u_0^k)$$

3. Estimate statistical solution using Monte-Carlo:

$$\mu^{\Delta,M} = \frac{1}{M} \sum_{k=1}^{M} \delta_{u^{k,\Delta}}.$$

³Based on the FKMT algorithm [Fjordholm, Käppeli, Mishra, Tadmor, 2014]

Numerical algorithm for statistical solutions³

Goal: show that
$$\mu^{\Delta,M}$$
 converges

- 1. Draw *M* i.d.d. samples u_0^k from μ_0
- 2. For $k = 1, \ldots, M$ evolve

$$u^{k,\Delta} = \mathcal{S}^{\Delta}(u_0^k)$$

3. Estimate statistical solution using Monte-Carlo:

$$\mu^{\Delta,M} = \frac{1}{M} \sum_{k=1}^{M} \delta_{u^{k,\Delta}}.$$

³Based on the FKMT algorithm [Fjordholm, Käppeli, Mishra, Tadmor, 2014]

Theory for Systems

I summed up all systems in a phrase, and all existence in an epigram.

- Oscar Wilde

Convergence of correlation measures

Theorem (Fjordholm, Lye, Mishra, in preparation) Assume that ν_n is a sequence of correlation measures satisfying

$$\sup_{n \in \mathbb{N}} \langle \nu_n^1, |\xi|^p \rangle \le c^p$$
$$\lim_{r \to 0} \sup_{n \in \mathbb{N}} \left(\int_D \int_{B_r(x)} \langle \nu_{n,x,y}^2, |\xi_1 - \xi_2|^p \rangle \, dy dx \right)^{1/p} = 0$$

for some c > 0.

Convergence of correlation measures

Theorem (Fjordholm, Lye, Mishra, in preparation) Assume that ν_n is a sequence of correlation measures satisfying

$$\sup_{n \in \mathbb{N}} \langle \nu_n^1, |\xi|^p \rangle \le c^p$$
$$\lim_{r \to 0} \sup_{n \in \mathbb{N}} \left(\int_D \int_{B_r(x)} \langle \nu_{n,x,y}^2, |\xi_1 - \xi_2|^p \rangle \, dy dx \right)^{1/p} = 0$$

for some c > 0.

Then there exists a weak-* convergent subsequence of ν_n

Convergence of correlation measures

Theorem (Fjordholm, Lye, Mishra, in preparation) Assume that ν_n is a sequence of correlation measures satisfying

$$\sup_{n \in \mathbb{N}} \langle \nu_n^1, |\xi|^p \rangle \le c^p$$
$$\lim_{r \to 0} \sup_{n \in \mathbb{N}} \left(\int_D \int_{B_r(x)} \langle \nu_{n,x,y}^2, |\xi_1 - \xi_2|^p \rangle \, dy dx \right)^{1/p} = 0$$

for some c > 0.

Then there exists a weak-* convergent subsequence of ν_n

If we have some boundedness assumptions on ν^n , then observables converge strongly:

$$\lim_{j\to\infty}\int_{D^k}\left|\langle\nu_{n_j,x}^k,g(x)\rangle-\langle\nu_x^k,g(x)\rangle\right|\,dx=0$$

Theorem (Fjordholm, Lye, Mishra, in preparation) Assume $\{\mu_t^{\Delta_\lambda}\}_{\lambda}$ converges weakly-* to μ_t in $Prob(L^1(\mathbb{R}^d, \mathbb{R}^N))$ where $\Delta_{\lambda} \to 0$, and assume

$$\underbrace{\int_{L^{1}(D)} TV(\{\overline{v}_{i}\}) d\mu_{t}^{\Delta_{\lambda}}(u)}_{\text{Average total variation}} \leq C\Delta_{\lambda}^{-s},$$

Average total variation

for some s < 1.

Then μ_t is a statistical solution.

The last two theorems translate to:

$$\lim_{r\to 0}\sup_{n\in\mathbb{N}}\left(\int_D \oint_{B_r(x)} \langle \nu_{n,x,y}^2, |\xi_1-\xi_2|^p\rangle\,dydx\right)^{1/p} = 0$$

then

lf

$$\mu^{\Delta,M}$$
 converges to a statistical solution

The last two theorems translate to:

$$\lim_{r\to 0}\sup_{n\in\mathbb{N}}\left(\int_D \oint_{B_r(x)} \langle \nu_{n,x,y}^2, |\xi_1-\xi_2|^p\rangle\,dydx\right)^{1/p} = 0$$

then

lf

$$\mu^{\Delta,M}$$
 converges to a statistical solution

We will use

$$\left(\int_D \oint_{B_r(x)} \langle \nu_{n,x,y}^2, |\xi_1 - \xi_2|^p \rangle \, dy dx\right)^{1/p} \approx \frac{1}{M} \sum_k S_h^p(u_k^{\Delta})$$

The last two theorems translate to:

$$\lim_{r\to 0}\sup_{n\in\mathbb{N}}\left(\int_D \oint_{B_r(x)} \langle \nu_{n,x,y}^2, |\xi_1-\xi_2|^p\rangle\,dydx\right)^{1/p} = 0$$

then

lf

$$\mu^{\Delta,M}$$
 converges to a statistical solution

We will use

$$\left(\int_D \oint_{B_r(x)} \langle \nu_{n,x,y}^2, |\xi_1 - \xi_2|^p \rangle \, dy dx\right)^{1/p} \approx \frac{1}{M} \sum_k S_h^p(u_k^{\Delta})$$

Recall: S_h^p is a FVM evolution!

Can we do better than Monte-Carlo?

Multilevel Monte Carlo

"Will you walk a little faster?" said a whiting to a snail. "There's a porpoise close behind us, and he's treading on my tail."

- Lewis Carroll,

Alice's Adventures in Wonderland.

Multi-level Monte Carlo [Heinrich, 2001, Giles, 2008]





$$\ell = 1$$



Multi-level Monte Carlo [Heinrich, 2001, Giles, 2008]



$$\ell = 2$$

Levels $\ell = 0, \ldots, L$, approximations u_k^{ℓ} .



$$\mu^{L} = \mu^{0} + \sum_{\ell=1}^{L} (\mu^{\ell} - \mu^{\ell-1})$$

.



 $\ell =$

Multi-level Monte Carlo [Heinrich, 2001, Giles, 2008]



$$\ell = 2$$

Levels $\ell = 0, \ldots, L$, approximations u_k^{ℓ} .



$$\ell = 1 \qquad \mu^{L} = \mu^{0} + \sum_{\ell=1}^{L} (\mu^{\ell} - \mu^{\ell-1}) \\ \approx \frac{1}{M_{0}} \sum_{k=1}^{M_{0}} \delta_{u_{k}^{0}} + \sum_{\ell=1}^{L} \frac{1}{M_{\ell}} \sum_{k=1}^{M_{\ell}} (\delta_{u_{k}^{\ell}} - \delta_{u_{k}^{\ell-1}}) \\ =: \mu_{\Delta \times, L}.$$



 $\ell = 0$

Burgers': Uncertain shock location: MLMC



Burgers': Brownian initial data: MLMC



MLMC test: Shock-Vortex

MLMC test: Shock-Vortex



Mean, MLMC



Variance, MLMC



Mean, MC



Variance, MC

MLMC-EMVS: Error analysis Shock-Vortex

Reference solution: Monte-Carlo 1024x1024 using 1024 samples



Kelvin-Helmholtz + MLMC

Kelvin-Helmholtz + MLMC



Mean, MLMC



Variance, MLMC



Mean, MC



Variance, MC

Convergence

For MLMC to be faster, we need

$$\operatorname{Var}(\rho^{\Delta} - \rho^{2\Delta}) < \operatorname{Var}(\rho)$$

For MLMC to be faster, we need

 $\operatorname{Var}(\rho^{\Delta} - \rho^{2\Delta}) < \operatorname{Var}(\rho)$







Can we use MLMC for systems?

Yes, but we only get speed-up for well-behaved data.

All further experiments are done with single level Monte Carlo

Alsvinn

Árvakr ok Alsviðr, þeir skulu upp héðan svangir sól drag Early-woke, All-swift, hence must these horses wearily draw up the sun – Grímnismál

The Alsvinn Simulator

• alsvinn.github.io [GPLv3]



The Alsvinn Simulator

- alsvinn.github.io [GPLv3]
- Multi-GPU Finite Volume simulator



The Alsvinn Simulator

- alsvinn.github.io [GPLv3]
- Multi-GPU Finite Volume simulator
- Written in C++/CUDA



Reproducibility

The best is the enemy of the good.

- Voltaire

1. Open source [GPLv3]



Reproducibility of Alsvinn experiments

- 1. Open source [GPLv3]
- 2. Unit tests

[RUN] [[0K] [[]	ReconstructionTests/ReconstructionConvergenceTest.ReconstructionTest/11 ReconstructionTests/ReconstructionConvergenceTest.ReconstructionTest/11 (8 12 tests from ReconstructionTests/ReconstructionConvergenceTest (120 ms tota	ns al
[] [RUN] [[OK] [[RUN] [[OK]]	2 tests from MeanVarConvergenceTests/MeanVarTest MeanVarConvergenceTests/MeanVarTest.MeanVarTest/0 MeanVarConvergenceTests/MeanVarTest.MeanVarTest/0 (1768 ms) MeanVarConvergenceTests/MeanVarTest.MeanVarTest/1 MeanVarConvergenceTests/MeanVarTest.MeanVarTest/1 (729 ms) 2 tests from MeanVarConvergenceTests/MeanVarTest (7497 ms) total)	
[] ([======]] [PASSED]]	Global test environment tear-down 139 tests from 40 test cases ran. (3819 ms total) 139 tests.	
- 1. Open source [GPLv3]
- 2. Unit tests
- 3. Essential information stored in output:

NetCDF File



- 1. Open source [GPLv3]
- 2. Unit tests
- 3. Essential information stored in output:



- 1. Open source [GPLv3]
- 2. Unit tests
- 3. Essential information stored in output:
 - commit id
 - CPU/GPU used



- 1. Open source [GPLv3]
- 2. Unit tests
- 3. Essential information stored in output:
 - commit id
 - CPU/GPU used
 - library versions





- 1. Open source [GPLv3]
- 2. Unit tests
- 3. Essential information stored in output:
 - commit id
 - CPU/GPU used
 - library versions
 - solver settings (flux, grid, etc)

- 1. Open source [GPLv3]
- 2. Unit tests
- 3. Essential information stored in output:
 - commit id
 - CPU/GPU used
 - library versions
 - solver settings (flux, grid, etc)
- 4. Docker



Numerical experiments for systems

I think that it is a relatively good approximation to truth – which is much too complicated to allow anything but approximations – that mathematical ideas originate in empirics.

– John von Neumann

What converges?

Kelvin-Helmholtz: 2 point histograms

Histogram at resolution 128, for Kelvin-Helmholtz, between (0.70, 0.70) and (0.70, 0.80)





Histogram at resolution 256, for Kelvin-Helmholtz, between (0.70, 0.70) and (0.70, 0.80) 2.2 2.0 (08.0.02.0) 0 J.6 1.4 1.6 alle 1.2 1.0 0.8 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 Value of p(0.70, 0.70)

$$N = 256$$

Histogram at resolution 512, for Kelvin-Helmholtz, between (0.70, 0.70) and (0.70, 0.80)



N = 512



N = 1024

Kelvin-Helmholtz: Wasserstein convergence



Kelvin-Helmholtz: varying perturbation size (stability)



Brownian initial data



 $u_{\rm x}$ at t=0 $$u_y$ at $t=0$ $\rho_0=4$ and $p_0=2.5$ $$



Brownian initial data: 2 point histograms



N = 128



$$N = 256$$

Histogram at resolution 512, for Brownian motion, between (0.70, 0.70) and (0.70, 0.80)



N = 512

Histogram at resolution 1024, for Brownian motion, between (0.70, 0.70) and (0.70, 0.80)



N = 1024

Brownian initial data: Wasserstein convergence



Richtmeyer-Meshkov: 2 point histograms



N = 128Histogram at resolution 256, for Richtmeyer-Meshkov, between (0.70, 0.70) and (0.70, 0.80) -1.4 2.25 2.00 - 1.2 (0 1.75 1.50 1.25 1.00 1.00 - 1.0 0.8 0.6 0.4 0.75 0.2 0.50 0.0 0.5 1.0 1.5 2.0 Value of p(0.70.0.70)

$$N = 256$$





N = 512

Histogram at resolution 1024, for Richtmeyer-Meshkov, between (0.70, 0.70) and (0.70, 0.80)



N = 1024

Richtmeyer-Meshkov: Wasserstein convergence

Wasserstein convergence for Richtmeyer-Meshkov for second correlation r



Can we measure the assumptions for convergence?

$$\lim_{r \to 0} \sup_{n \in \mathbb{N}} \left(\underbrace{\int_{D} \oint_{B_{r}(x)} \langle \nu_{n,x,y}^{2}, |\xi_{1} - \xi_{2}|^{p} \rangle \, dy dx}_{:=\bar{S}_{n}^{p}(r)} \right)^{1/p} = 0?$$

Scaling for Richtmeyer-Meshkov, $t = 5.0, p = 3, \varepsilon = 0.06$



Time evolution, p = 2



Can we measure the assumptions for Lax-Wendroff?

$\int_{L^1(D)} \mathsf{TV}(\{\overline{v}_{\mathbf{i}}\}_{\mathbf{i}\in\mathbb{Z}^d}) d\mu_t^{\Delta}(u) \leq CN^s, \qquad 0 \leq s < 1?$



Summary:

Statistical solutions are observed to be well-behaved

Statistical solutions are observed to be well-behaved



Kelvin-Helmholtz



Brownian



Richtmeyer-Meshkov



Kelvin-Helmholtz



Brownian



Richtmeyer-Meshkov

Thank you. Kjetil Lye <kjetil.lye@sam.math.ethz.ch>

https://alsvinn.github.io