

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/339289936>

Coupling kinetic and continuum using data-driven maximum entropy distribution

Preprint · February 2020

DOI: 10.13140/RG.2.2.20019.99367

CITATIONS

0

READS

345

3 authors:



Mohsen Sadr

RWTH Aachen University

7 PUBLICATIONS 17 CITATIONS

[SEE PROFILE](#)



Qian Wang

École Polytechnique Fédérale de Lausanne

14 PUBLICATIONS 109 CITATIONS

[SEE PROFILE](#)



Hossein Gorji

École Polytechnique Fédérale de Lausanne

31 PUBLICATIONS 306 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Robust and efficient parametric model reduction [View project](#)



Fokker-Planck Model for Multiphase Flows [View project](#)



Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp



Coupling kinetic and continuum using data-driven maximum entropy distribution

Mohsen Sadr^{a,*}, Qian Wang^b, M. Hossein Gorji^b

^a*MATHCCES, Department of Mathematics, RWTH Aachen University, Schinkestrasse 2, D-52062 Aachen, Germany*

^b*Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

ARTICLE INFO

Article history:
Received ?
Received in final form ?
Accepted ?
Available online ?

Communicated by ?

ABSTRACT

An important class of multi-scale flow scenarios deals with an interplay between kinetic and continuum phenomena. While hybrid solvers provide a natural way to cope with these settings, two issues restrict their performance. Foremost, the inverse problem implied by estimating distributions has to be addressed, to provide boundary conditions for the kinetic solver. The next issue comes from defining a robust yet accurate switching criterion between the two solvers. This study introduces a data-driven kinetic-continuum coupling, where the Maximum-Entropy-Distribution (MED) is employed to parametrize distributions arising from continuum field variables. Two regression methodologies of Gaussian-Processes (GPs) and Artificial-Neural-Networks (ANNs) are utilized to predict MEDs efficiently. Hence the MED estimates are employed to carry out the coupling, besides providing a switching criterion. To achieve the latter, a continuum breakdown parameter is defined by means of the Fisher information distance computed from the MED estimates. We test the performance of our devised MED estimators by recovering bi-modal densities. Next, MED estimates are integrated into a hybrid kinetic-continuum solution algorithm. Here Direct Simulation Monte-Carlo (DSMC) and Smoothed-Particle Hydrodynamics (SPH) are chosen as kinetic and continuum solvers, respectively. The problem of monatomic gas inside Sod's

shock tube is investigated, where DSMC-SPH coupling is realized by applying the devised MED estimates. Very good agreements with respect to benchmark solutions along with a promising speed-up are observed in our reported test cases.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

It is often the case that in real-world flow phenomena, the underlying thermo-fluid processes cover a wide range of scales. In particular, gas dynamic problems may encounter a huge variation of the Knudsen number, resulting in a failure of common numerical approaches. While the Navier-Stokes-Fourier (NSF) system of equations provides an accurate description of the flow at the hydrodynamic limit, they fail when significant departures from the equilibrium are encountered. On the other hand, the kinetic framework becomes more relevant once the breakdown of the continuum is observed. In particular, the Boltzmann equation offers a high fidelity governing equation for the evolution of the molecular distribution function. Even though the statistical description provided by the Boltzmann equation holds accurate in the whole range of the Knudsen number, numerical deficiencies have to be dealt with near the continuum. Either using Direct Simulation Monte-Carlo (DSMC) [1, 2, 3], discrete velocity methods [4, 5, 6] or spectral discretizations [7, 8, 9], the efficiency of resulting simulations significantly drops as the collision integral becomes dense. While simplified kinetic models such as Fokker-Planck type approximations [10, 11, 12, 13] have been proposed to improve the efficiency in the near continuum and early transitional regimes, still the resulting models are not computationally competitive with respect to the more mature continuum solvers. These all motivate coupling of the continuum and kinetic solvers in order to tackle a larger set of flow problems, where both continuum and kinetic scales are present. Examples include but are not limited to, unconventional gas reservoirs [14], re-entry problem [15], Stratospheric aerosols [16], high -pressure shock tubes [17], Sonoluminescence [18], and

*Corresponding author:

e-mail: sadr@mathcces.rwth-aachen.de (Mohsen Sadr)

liquid-vapor interface [19]. An extensive set of hybrid algorithms to address the kinetic-continuum coupling has been devised in past decades. Hybrid numerical schemes have been proposed which operate on particles [20, 21, 22, 23], Partial Differential Equations (PDEs) [24, 25, 26] or distributions [27]. While each approach aims at providing universal accuracy and efficiency, two fundamental issues are common themes in the majority of the algorithms. The first issue comes from the fact that kinetic and continuum equations deal with different types of variables. The variable of interest is the distribution in the former, whereas the latter relies on its moments. Clearly, extracting moments from a given distribution is a straightforward task. Yet the inverse, known as the *moment problem*, is not well-posed in general. Common recipes to address the moment problem in hybrid approaches, include Chapman-Enskog (CE) [28, 29] or Grad-Hermite distributions [30, 31]. The second issue arises due to the demand for a relevant switching criterion which preserves the efficiency and accuracy of the underlying schemes. Typically, some norm of non-equilibrium moments or their gradients are employed to provide an optimal criterion for the breakdown of the continuum [32, 33].

Adopting CE [20] or Grad-Hermite distributions [34], comes with a risk of running to distributions with negative parts, which leads to the breakdown of the notion of entropy. On the other hand, the Maximum-Entropy-Distribution (MED) can provide a well-defined probability density while honoring a given set of moments. Furthermore, it can be shown that MED leads to the least-biased estimation of the distribution subject to the given moments. It is important to note that having a well-defined distribution in hand, one can also compute relevant distances of the distribution from the equilibrium to obtain a switching criterion.

By virtue of the method of Lagrange multipliers, it is straightforward to show that MED takes a closed form of the exponential function [35]. Therefore, the moment problem reduces to finding the Lagrange multipliers for a given set of moments. Although the resulting optimization problem is well-posed in bounded domains subject to realizable moments [36, 37], the iterative gradient descent-type algorithms have to be employed. This comes with a significant cost since the Hessian matrix of the objective function can become ill-conditioned. Hence, alternatives such as adaptive basis methods have been pursued [38, 39].

A different perspective is realized by resetting the MED problem into the regression framework. More precisely, the idea becomes to compute the Lagrange multipliers as a map from input moments using regression schemes. Recently, a data-driven approach was introduced by the authors [40], where Gaussian-Processes (GPs) have been adopted to provide a regression map for estimating MEDs. Either using GP regressions or Artificial Neural Networks (ANN), first, the model is trained over a data-set that fills relevant subsets of the moment space. Then, efficient estimations of MDEs are obtained for a given input of moments.

In this study, after training GP and ANN for a MED data-set, a hybrid kinetic-continuum solution algorithm based on efficient MED estimators is devised. Using efficient estimates of MEDs and by choosing a threshold, then a criterion for the breakdown of the continuum can be found, arising as a distance between the estimated MED and the equilibrium. Therefore based on the computed distance, either continuum or kinetic solver is adopted for a given computational cell and time step.

To narrow down our context, we address a hybrid particle-particle solution algorithm, where the NSF system is solved using Smoothed-Particle Hydrodynamics (SPH) [41], and DSMC is employed for the Boltzmann equation. However, note that the methodologies presented here, are quite general and can be applied to a wider class of multi-scale solvers. Once computational cells are switched from SPH to DSMC, the DSMC particles need to be sampled from the estimated MED. Although more advanced sampling techniques such as slice sampling [42] and adaptive rejection sampling [43] could be utilized, for simplicity we rely on the Metropolis-Hastings algorithm [44] for generating DSMC particles. On the other hand, when DSMC cells convert to the SPH ones, the SPH particles are generated based on moments obtained from the DSMC particles.

The content of this manuscript is distributed among the following sections. First, the NSF system of equations, along with the Boltzmann equation, is reviewed in § 2, to the extent needed in our study. Then in § 3, SPH and DSMC solution algorithms required for solving the adopted governing equations are considered. Next in § 4, the basic framework

of the hybrid DSMC-SPH algorithm is discussed. Afterward, in § 5, the Maximum-Entropy method and the regression approaches based on GP and ANN are devised. A hybrid solution algorithm is then proposed in § 6, where a data-driven coupling based on our MED estimator equipped with the switching criterion colormsis adopted. In order to evaluate the regression machinery, the trained MED estimators based on GP and ANN are first examined by estimating a bi-modal distribution. Then the devised hybrid solution algorithm is deployed to solve the well-known Sod's shock tube problem in § 7. In the end, a conclusion along with a projection for next studies is presented in § 8.

2. Governing equations

Before proceed to the solution algorithms and data-driven coupling, let us recall the governing equations relevant for our continuum and kinetic settings.

2.1. Navier-Stokes-Fourier system of equations

The NSF equations provide a system of conservation laws for evolution of the density ρ , the momentum $\rho\mathbf{U}$, and the total energy $E := \frac{1}{2}\rho\mathbf{U} \cdot \mathbf{U} + \rho c_v T$. Here $c_v := 3k_b/(2m)$ indicates the heat capacity of a monatomic gas at constant volume, k_b is the Boltzmann constant, T is the temperature, and m is the molecular mass. A general form of the conservation laws can be cast to

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}(\rho U_i) = 0, \quad (1)$$

$$\frac{\partial(\rho U_i)}{\partial t} + \frac{\partial}{\partial x_j}(\rho U_i U_j + p_{ij}) = 0 \quad (2)$$

$$\text{and} \quad \frac{\partial E}{\partial t} + \frac{\partial}{\partial x_i}(EU_i + q_i + p_{ik}U_k) = 0, \quad (3)$$

where \mathbf{p} stands for the pressure tensor and \mathbf{q} denotes the heat flux vector. Note that Einstein's summation notation is used throughout this study, and δ_{ij} denotes the Kronecker delta. By assuming the ideal gas law besides the Stokes closure, we get $p_{ij} = \pi_{ij} + p\delta_{ij}$, where

$$p = nk_b T \quad (4)$$

$$\text{and} \quad \pi_{ij} = -2\mu S_{ij} \quad (5)$$

with n the number density, μ the viscosity and

$$S_{ij} = \frac{1}{2} \left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} \right) - \frac{1}{3} \frac{\partial U_k}{\partial x_k} \delta_{ij}. \quad (6)$$

Moreover, the heat flux follows the Fourier law

$$q_i = -\kappa \frac{\partial T}{\partial x_i}, \quad (7)$$

where κ denotes the heat conductivity. For further details on the NSF equations along with the corresponding derivations, see e.g. [45].

As the fluid experiences strong departures from the equilibrium, it is well established that the NSF system fails to provide an accurate physical description. This is due to the fact that closure assumptions applied to the pressure tensor and heat fluxes may not hold precise once the Knudsen number becomes large. This failure is pronounced in many non-equilibrium settings including cold-to-hot heat fluxes in the lid-driven cavity [46], slip velocity [47], shock profile [48], and inverted temperature gradients [49].

To cope with these shortcomings, a higher level of closure is necessary. The kinetic theory provides a mesoscale closure by considering the dynamics of the molecular velocity distribution, which is discussed in the following.

2.2. The Boltzmann equation

The statistical account of an ideal monatomic gas can be fully described by the probability density $f(\mathbf{V}|\mathbf{x}, t)$, which gives the probability $f(\mathbf{V}|\mathbf{x}, t)d^3\mathbf{V}$ for finding a particle with a velocity in the vicinity of \mathbf{V} , at a given position \mathbf{x} and instant in time t . For convenience, let us consider the mass density function $\mathcal{F}(\mathbf{V}, \mathbf{x}, t) := \rho(\mathbf{x}, t)f(\mathbf{V}|\mathbf{x}, t)$ (MDF). Assuming that the gas is *dilute* and the *molecular chaos* holds [28], the evolution of MDF follows the Boltzmann equation

$$\frac{\partial \mathcal{F}}{\partial t} + \frac{\partial(\mathcal{F}V_j)}{\partial x_j} = S^{\text{Boltz.}}(\mathcal{F}) \quad (8)$$

with the Boltzmann collision operator

$$S^{\text{Boltz.}}(\mathcal{F}) = \int_{\mathbb{R}^3} \int_0^{2\pi} \int_0^{+\infty} \left[\mathcal{F}(\mathbf{V}^*, \mathbf{x}) \mathcal{F}(\mathbf{V}_1^*, \mathbf{x}) - \mathcal{F}(\mathbf{V}, \mathbf{x}) \mathcal{F}(\mathbf{V}_1, \mathbf{x}) \right] g \hat{b} d\hat{b} d\hat{e} d^3\mathbf{V}_1. \quad (9)$$

The collision operator acts as a source term in the evolution of $\mathcal{F}(\mathbf{V}, \mathbf{x}, t)$, when a particle with velocity \mathbf{V} is produced as the outcome of colliding particles with velocities $(\mathbf{V}^*, \mathbf{V}_1^*)$, and as a sink term once one of the colliding pair has the pre-collision velocity of \mathbf{V} . Note that the magnitude of the relative velocity is indicated by $g = |\mathbf{V} - \mathbf{V}_1|$. Furthermore, the collision plane is determined with the impact parameter \hat{b} and the scattering angle \hat{e} .

Once the Boltzmann equation is solved, the macroscopic properties such as ρ , \mathbf{U} , \mathbf{p} , T and \mathbf{q} can be easily obtained by taking the moments of MDF. Since, in contrast to the NSF system, here no assumption on the distance of \mathcal{F} from the equilibrium is adopted, the Boltzmann equation can also be employed for flow scenarios far from the equilibrium.

3. Particle solution algorithms

There exists a broad set of numerical schemes to tackle flow phenomena governed by the NSF system as well as the Boltzmann equation. However, here we focus on particle approaches, mainly because the coupling of two particle systems is technically more convenient. Therefore first, a particle method based on SPH as the solution algorithm for the continuum scale is reviewed. Then, DSMC is explained as our method of choice for the Boltzmann equation. At the end of each subsection, the algorithms used in this study are summarized explicitly.

3.1. Smoothed-Particle Hydrodynamics

As a particle algorithm for solving the NSF system, the main idea behind SPH relies on notions of the weight m_{SPH} and the kernel $W(r, h)$ carried by each particle [50, 41]. Here r is the distance and h the smoothing parameter. Let the superscript $(\cdot)^{(\alpha)}$ denote the quantity evaluated at the location of the particle α , and $(\tilde{\cdot})^{(\alpha)}$ the corresponding approximation introduced by the scheme. Furthermore $W_h^{(\alpha\beta)} = W(r^{(\alpha\beta)}, h)$ is the value of the kernel evaluated at $r^{(\alpha\beta)} := |\mathbf{x}^{(\alpha)} - \mathbf{x}^{(\beta)}|$ which is the distance between particles with index α and β . The density thus can be estimated by

$$\tilde{\rho}^{(\alpha)} = \sum_{\beta} m_{\text{SPH}}^{(\beta)} W^{(\alpha\beta)} . \quad (10)$$

Next, a quantity of interest A at $\mathbf{x}^{(\alpha)}$ is estimated by

$$\tilde{A}^{(\alpha)} = \sum_{\beta} \frac{m_{\text{SPH}}^{(\beta)}}{\tilde{\rho}^{(\beta)}} \tilde{A}^{(\beta)} W^{(\alpha\beta)} . \quad (11)$$

The accuracy and efficiency of SPH method depend on the choice of the kernel function and many such interpolation kernels are introduced in the literature (see e.g. [51, 52]). For simplicity here, we adopt the Gaussian kernel which takes the following form in the one-dimensional space

$$W(r^{(\alpha\beta)}, h) = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{r^{(\alpha\beta)}}{2h^2}\right). \quad (12)$$

Typically a fixed small h and a large cut-off r_{cut} are used. Here they are set based on the grid size Δx via

$$r_{\text{cut}} = 2\Delta x \quad (13)$$

$$\text{and } h = \frac{1}{30}r_{\text{cut}}. \quad (14)$$

One of the main advantages of the SPH framework is the smooth estimation of the derivatives which are required in the governing equations. It is easy to see that the derivative of A at the position of a particle with index α becomes

$$\nabla \tilde{A}^{(\alpha)} = \sum_{\beta} \frac{m_{\text{SPH}}^{(\beta)}}{\tilde{\rho}^{(\beta)}} \tilde{A}^{(\beta)} \nabla W^{(\alpha\beta)}. \quad (15)$$

Therefore, by considering the Lagrangian framework for the NSF system of equations, the forces and heat transfer experienced by each particle can be computed by a time-integration scheme. We adopt the following discretization of momentum and energy equations

$$\frac{D}{Dt} \tilde{U}_i^{(\alpha)} = - \sum_{\beta} \frac{m_{\text{SPH}}^{(\beta)}}{\tilde{\rho}^{(\alpha)} \tilde{\rho}^{(\beta)}} \left(\tilde{p}_{ij}^{(\alpha)} + \tilde{p}_{ij}^{(\beta)} \right) \frac{\partial W^{(\alpha\beta)}}{\partial x_j^{(\beta)}}, \quad (16)$$

$$\begin{aligned} \text{and } \frac{D}{Dt} \tilde{e}^{(\alpha)} = & -\frac{1}{2} \sum_{\beta} \frac{m_{\text{SPH}}^{(\beta)}}{\tilde{\rho}^{(\alpha)} \tilde{\rho}^{(\beta)}} \left\{ (\tilde{p}^{(\alpha)} + \tilde{p}^{(\beta)}) \tilde{U}_i^{(\alpha\beta)} + (\kappa^{(\alpha)} + \kappa^{(\beta)}) \tilde{T}^{(\alpha\beta)} \right\} \frac{\partial W^{(\alpha\beta)}}{\partial x_i^{(\beta)}} \\ & + \frac{2\mu^{(\alpha)}}{\tilde{\rho}^{(\alpha)}} \tilde{S}_{ij}^{(\alpha)} \tilde{S}_{ij}^{(\alpha)}, \end{aligned} \quad (17)$$

where $e = c_v T$ is the internal energy, $U_i^{(\alpha\beta)} = U_i^{(\alpha)} - U_i^{(\beta)}$ and $T^{(\alpha\beta)} = T^{(\alpha)} - T^{(\beta)}$. For details of derivations see [41, 51]. Adopting a first-order explicit time integration, the particles evolve in the solution domain accordingly. After initializing particles at every cell, the particles evolve according to Algorithm 1.

```

while  $t < T_{\text{final}}$  do
  -Compute density of each particle from (10);
  -Estimate  $\mathbf{S}$  and  $\mathbf{p}$  for every particle;
  -Evolve velocity and internal energy of each particle (16)-(17);
  -Stream particles with their velocities;
  -Increment  $t$ ;
end

```

Algorithm 1: SPH solution algorithm for the NSF system

3.2. Direct Simulation Monte-Carlo

Due to the high-dimensionality of the solution space associate with MDF, particle Monte-Carlo schemes become attractive for numerical solutions of the Boltzmann equation. Consider an ensemble of particles with velocities $\mathbf{M}^{(i)}$, positions $\mathbf{X}^{(i)}$ and weights $w^{(i)}$. The MDF $\mathcal{F}(\mathbf{V}, \mathbf{x}, t)$ is related to particles states through

$$\mathcal{F}(\mathbf{V}, \mathbf{x}, t) = \sum_i \delta(\mathbf{X}^{(i)} - \mathbf{x}) \delta(\mathbf{M}^{(i)} - \mathbf{V}) w^{(i)}, \quad (18)$$

where $\delta(\cdot)$ is the Dirac delta [53, 54, 55]. For simplicity, we consider the same weight for all DSMC particles. Using the particle description, the idea behind DSMC is that instead of updating \mathcal{F} according to the Boltzmann equation Eq. (8), the underlying jump process is simulated. Hence, the evolution of positions $\mathbf{X}^{(i)}$ and velocities $\mathbf{M}^{(i)}$ are derived consistent with the Boltzmann equation in a two-step manner: streaming and collision. While the streaming phase is simply the free flight of particles, the collision follows the No-Time-Counter (NTC) method of Bird [2]. The collision probability is found via

$$\mathcal{P}_{\text{coll.}} = \frac{\sigma_T c_r}{(\sigma_T c_r)_{\text{max}}}, \quad (19)$$

where σ_T indicates the collision cross-section, c_r is the magnitude of the relative velocity for a colliding pair, and the subscript $(\cdot)_{\text{max}}$ indicates the maximum value. In case of the Hard-Sphere molecular potential, the collision cross-section is $\sigma_T = \pi\sigma^2$ where σ is the diameter of the molecules. Collisions occur for the colliding pair (i, j) in a manner that guarantees the conservation of mass, momentum, and energy. Let $\mathbf{c}_c = (\mathbf{M}^{(i)} + \mathbf{M}^{(j)})/2$ be the center of mass velocity and $\mathbf{c}_r = (\mathbf{M}^{(i)} - \mathbf{M}^{(j)})$ the relative velocity. According to the Hard-Sphere scattering law, the post-collision relative velocity is isotropic. Hence the angles

$$\theta = \arccos(2\alpha_1 - 1) \quad (20)$$

$$\text{and } \phi = 2\pi\alpha_2 \quad (21)$$

with uniformly distributed random numbers $\alpha_{1,2} \in [0, 1]$, result in the orientation of the post-collision relative velocity

$$\mathbf{c}_r^* = c_r \left(\cos(\theta), \sin(\theta) \cos(\phi), \sin(\theta) \sin(\phi) \right)^T. \quad (22)$$

The post-collision particles velocities are fully determined by relative and center of mass velocities, following

$$\mathbf{M}^{(i)} = \mathbf{c}_c + \frac{\mathbf{c}_r^*}{2} \quad (23)$$

$$\text{and } \mathbf{M}^{(j)} = \mathbf{c}_c - \frac{\mathbf{c}_r^*}{2}. \quad (24)$$

A short description of DSMC algorithm for Hard-Spheres, after initializing the particles is summarized in Algorithm 2.

```

while  $t < T_{\text{final}}$  do
  -Sample the moments;
  for  $i = 1, \dots, N_{\text{cells}}$  do
     $N_{\text{Cand}} = \frac{1}{2} N_{\text{p/cell}}^2 F_N(\sigma_T c_r)_{\text{max}} \Delta t / V_{\text{cell}};$ 
    for  $j = 1, \dots, N_{\text{Cand}}$  do
      -Pick two samples from the cell;
      -Draw a random number  $r$  with a uniform distribution in  $[0, 1]$ ;
      if  $\sigma_T c_r / (\sigma_T c_r)_{\text{max}} < r$  then
        -Perform the collision (22)-(21);
      end
    end
  end
  -Stream position of particles  $\mathbf{X}$  with their velocities  $\mathbf{M}$ ;
  -Increment  $t$ ;
end

```

Algorithm 2: DSMC algorithm for Hard-Spheres. $F_N := w/m$ indicates the statistical weight, V_{cell} is the volume of the computational cell, $N_{\text{p/cell}}$ is the number of particles per cell, and N_{Cand} is number of candidates to be considered for collisions at each time step.

4. Upscaling and refinement

We focus on a domain decomposition coupling approach, where each of the DSMC/SPH solvers will be employed in the designated sub-domains. Therefore two types of communications between continuum and kinetic solvers have to be addressed. First, we discuss moment

recovery which is upscaling of the kinetic information to the continuum scale. Next, we address the moment problem arising from the refinement of the continuum information into the kinetic scale. The former gives us the information passage from kinetic to continuum sub-domains, whereas the latter accounts for the opposite.

4.1. Moment estimations

The evolution of moments can be easily obtained from the Boltzmann equation by taking the velocity moments of Eq. (8). By comparing the flux terms with respect to the conservation laws i.e. Eqs. (1)-(3), the corresponding macroscopic quantities are readily obtained

$$U_i = \frac{1}{\rho} \int_{\mathbb{R}^3} V_i \mathcal{F} d^3 \mathbf{V} \quad (25)$$

$$p_{ij} = \int_{\mathbb{R}^3} \xi_i \xi_j \mathcal{F} d^3 \mathbf{V} \quad (26)$$

$$\text{and } q_i = \frac{1}{2} \int_{\mathbb{R}^3} \xi_i \xi_j \xi_j \mathcal{F} d^3 \mathbf{V}, \quad (27)$$

where $\boldsymbol{\xi} := \mathbf{V} - \mathbf{U}$ denotes the fluctuating velocity. By applying the identity (18), the above-mentioned moments can be computed from DSMC particles. Yet since a finite number of particles and finite size of computational cells are applied in DSMC, the corresponding equalities turn to estimates; assigned here by the superscript $(\tilde{\cdot})$. The field variables at position \mathbf{x} and time t are found according to

$$\tilde{\rho} = \frac{1}{\delta\Gamma} \sum_{\mathbf{k} \in \mathcal{S}_\Gamma} w^k, \quad (28)$$

$$\tilde{U}_i = \frac{1}{\tilde{\rho}} \sum_{\mathbf{k} \in \mathcal{S}_\Gamma} M_i^k w^k, \quad (29)$$

$$\tilde{p}_{ij} = \sum_{\mathbf{k} \in \mathcal{S}_\Gamma} M_i'^k M_j'^k w^k \quad (30)$$

$$\text{and } \tilde{q}_i = \frac{1}{2} \sum_{\mathbf{k} \in \mathcal{S}_\Gamma} M_i'^k M_j'^k M_j'^k w^k, \quad (31)$$

where \mathcal{S}_Γ is the set of particles residing in the computational cell Γ around position \mathbf{x} at time t , and $\delta\Gamma$ denotes the volume of that cell. Furthermore $\mathbf{M}' := \mathbf{M} - \tilde{\mathbf{U}}$ is the fluctuating particle velocity. The above estimates commit two types of errors. First, the statistical noise arising from using a limited number of particles and second, the spatial homogenization due to the spatial averaging over the computational cell. Note that the former error can be significantly reduced once the time averaging can be utilized relevant for the stationary flows.

4.2. Moment problem

The more challenging part of the coupling is when the information provided by the continuum solver should be transferred to the kinetic sub-domain. In other words, a distribution should be assigned to the moments given by the continuum model. Note that in general this moment problem is ill-posed and needs a regularization for tractability. There exist many approaches addressing the moment problem, including the quadrature method [56], Grad-Hermite distributions [57, 58] and Maximum-Entropy methods [37, 36]. Assuming realizable moments, the quadrature method gives a well-defined probability density, yet suffers from non-uniqueness of the distribution. On the other hand, the Grad-Hermite ansatz gives a unique solution which can be computed cheaply, while it can not guarantee positivity of the probability density [58]. Finally MED gives rise to a unique and positive probability density, providing a bounded domain and realizable set of moments [59]. From the information-theory stand-point, MED provides the least-biased distribution subject to the moment constraints. In the following section, we first review some basics of the Maximum-Entropy approach. Furthermore, we explain and elaborate the recently devised data-driven method as an efficient way to compute MED probability densities [40].

5. Maximum-Entropy Distribution

For simplicity but without loss of generality, in the following we focus on the one-dimensional velocity space. Let us consider finding a probability density that fulfils the moments

$$\int_{\Omega} f \phi dV = \mathbf{p}, \quad (32)$$

where $\mathbf{p} = (\rho, \rho U, p, \dots)^T$ is the vector of given moments and $\phi = (1, V, \xi^2, \dots)^T$ the corresponding vector of velocity polynomials. The Maximum-Entropy approach provides a regularization of the moment problem by minimizing the entropy functional

$$S[f] = \int_{\Omega} f \ln(f) dV, \quad (33)$$

subject to the moment constraints (32). As shown in [59], the resulting optimization problem is well-posed on a bounded domain Ω and for realizable moments \mathbf{p} . Yet note that in case of an unbounded domain, such distribution still uniquely exists once further restrictions on

higher order moments are imposed [60, 61]. Following the method of Lagrange multipliers, the objective functional becomes

$$C_N^\lambda[f] := \int_{\Omega} f \ln(f) dV - \lambda_k \left(\int_{\Omega} f \phi_k dV - p_k \right) \quad (34)$$

where $\boldsymbol{\lambda}$ is vector of Lagrange multipliers owing to the moment constraints. By taking the variational derivative of C_N^λ , it is easy to see that the extremum of the functional becomes

$$f_N^\lambda = Z_\lambda^{-1} \exp(-\lambda_k \phi_k), \quad (35)$$

where the denominator $Z_\lambda := \int_{\Omega} \exp(-\lambda_k \phi_k) dV$ normalizes the MED f_N^λ [35]. Note that at this stage the moment problem is reduced to finding the vector $\boldsymbol{\lambda}$. Next, considering the dual problem, we arrive at the following unconstrained minimization problem

$$\boldsymbol{\lambda}(\mathbf{p}) = \arg \min_{\boldsymbol{\lambda}^* \in \mathbb{R}^N} \{ Z_{\boldsymbol{\lambda}^*} - \lambda_j^* p_j \}, \quad (36)$$

which can be directly employed in order to find the Lagrange multipliers $\boldsymbol{\lambda}$.

5.1. Direct method

The common approach for solving the minimization problem given in Eq. (36), follows Newton's method. The resulting iterative solution algorithm updates the estimate $\boldsymbol{\lambda}^n$ by $\boldsymbol{\lambda}^{n+1}$, according to

$$H_{ij}(\boldsymbol{\lambda}^n) \Delta \lambda_j^n = g_i(\boldsymbol{\lambda}^n) \quad (37)$$

$$\text{and} \quad \lambda_i^{n+1} = \lambda_i^n + \beta^n \Delta \lambda_i^n. \quad (38)$$

Here β^n is a damping factor at n th iteration, and $H(\lambda)$ and $g(\lambda)$ are the Hessian and the gradient of the objective function, respectively. Since the Hessian matrix is a function of the previous Lagrange multipliers estimate, there is no guarantee that H is well-conditioned, hence high computational cost can be expected [62]. Alternatively, an adaptive basis provides a mechanism to avoid the ill-conditioning issue. While the change of the basis, e.g., using Hermit polynomials to enforce a close to diagonal Hessian [62], or devising a Hessian for the intermediate probability density at each iteration [38, 39], can tackle the problem with regards to ill-conditioned Hessian, the overhead computational cost still constraints the use of direct methods.

5.2. Gaussian-Process Regression

Following [40], instead of directly solving for the Lagrange multipliers as explained in § 5.1, numerical advantages can be obtained by considering a regression approach. The idea is to approximate the unique map $\Psi_i : \mathbf{p} \rightarrow \lambda_i$, $i = 1, \dots, N$ by GP via

$$\tilde{\Psi}_i \sim \mathcal{GP}(0, \mathcal{K}_i) \quad (39)$$

given a positive semi-definite kernel function $\mathcal{K}(\mathbf{p}, \mathbf{p}')$. Note that here and henceforth, $A \sim B$ reads A is a sample drawn from B . For details of Gaussian-Process regressions see e.g. [63, 64]. The kernel introduces a set of hyper-parameters Θ_i to be fitted to the data, which can be optimized by maximizing the log-likelihood

$$\ln \left[\tilde{f} \left(\tilde{\Psi}_i(\mathbf{p}) \mid (\mathbf{p}, \lambda_i) \in \mathcal{D}_i \right) \right] \quad (40)$$

over the input/output data-set $D_i = \{(\mathbf{p}, \lambda_i)_k\}_{k=1}^{N_{\text{data}}}$, where $\tilde{f} \left(\tilde{\Psi}_i(\mathbf{p}) \mid (\mathbf{p}, \lambda_i) \in \mathcal{D}_i \right)$ denotes the probability density of $\tilde{\Psi}_i$ conditional on the training data.

Once the hyper-parameters of the chosen kernel are found, the Lagrange multipliers can be predicted based on the values of input moments \mathbf{p}^* according to

$$\left(\tilde{\Psi}_i(\mathbf{p}^*) \mid \tilde{\Psi}_i(\mathbf{p}) = \Psi_i(\mathbf{p}) \right) \sim \mathcal{N}(\bar{m}_i, \bar{\Sigma}_i), \quad (41)$$

where

$$\bar{m}_i = \mathcal{K}_{\Theta_i}(\mathbf{p}^*, \mathbf{p}') \mathcal{K}_{\Theta_i}(\mathbf{p}, \mathbf{p}')^{-1} \Psi_i(\mathbf{p}) \quad \text{and} \quad (42)$$

$$\bar{\Sigma}_i = \mathcal{K}_{\Theta_i}(\mathbf{p}^*, \mathbf{p}^*) - \mathcal{K}_{\Theta_i}(\mathbf{p}^*, \mathbf{p}) \mathcal{K}_{\Theta_i}(\mathbf{p}, \mathbf{p}')^{-1} \mathcal{K}_{\Theta_i}(\mathbf{p}, \mathbf{p}^*). \quad (43)$$

Note that $\mathcal{N}(A, B)$ denotes the normal distribution with mean A and variance B . Also note that the Einstein summation over repeated indices is abandoned if the brackets $(.)$ are used over the index. Similar to [40], a kernel based on the Gaussian radial basis function

$$\mathcal{K}_{\Theta_i}(\mathbf{p}, \mathbf{p}') = \sigma_{(i)} \exp \left(-r_{(i)}^2 / 2 \right) + \epsilon \delta(\mathbf{p} - \mathbf{p}') \quad (44)$$

is used in this work, where $\epsilon \delta(\mathbf{p} - \mathbf{p}')$ with $\epsilon = 10^{-6}$ guarantees that the outcome covariance matrix is diagonally dominant and invertible. Moreover r_i indicates an Euclidean scaled

distance from the other input point. The Broyden-Fletcher-Goldfarb-Shanno algorithm (BFGS) is employed to optimize the hyperparameters as the training step of our GP model [65]. Since here only a prototype study is considered, a simple GP regression using GPflow [66] is employed for estimating the Lagrange multipliers. Yet since the cost of Cholesky's factorization needed for the covariance matrix is cubic with respect to the number of training data points, more elaborate GP schemes are needed for larger data-sets. In this study, $N_{\text{data}} = 1000$ data points were used for training GP. Motivated by limitations of GP and in order to get a better insight into pros and cons of data-driven approaches, we also consider regression based on ANN as discussed in the following.

5.3. Artificial-Neural-Network

As an alternative data-driven approach and universal function approximator, ANN is employed here to address the regression problem of Lagrange multipliers. The objective is to make a comparison between ANN and GP regressions in terms of both accuracy and computational efficiency in the context of our MED problem. Therefore here we are concerned with learning the map $\Psi_i : \mathbf{p} \rightarrow \lambda_i$, $i = 1, \dots, N$ using an ANN approximator. For further details on ANNs and their design principles see e.g. [67, 68].

We start with a small network, including only three hidden layers, and increase numbers of neurons in hidden layers (H_1, H_2, H_3) until a reasonable accuracy in predicting the testing data points $(\mathbf{p}, \boldsymbol{\lambda})$ is achieved. A schematic of ANN is presented in Fig. 1. The forward propagation of the network with hyperbolic tangent activation function reads

$$h_i^{(1)} = \tanh(W_{ij}^{(1)} p_j + b_i^{(1)}), \quad (45)$$

$$h_i^{(2)} = \tanh(W_{ij}^{(2)} h_j^{(1)} + b_i^{(2)}), \quad (46)$$

$$h_i^{(3)} = \tanh(W_{ij}^{(3)} h_j^{(2)} + b_i^{(3)}) \quad (47)$$

$$\text{and } \hat{\lambda}_i = W_{ij}^{(o)} h_j^{(3)} + b_i^{(o)}, \quad (48)$$

where $\mathbf{W}^{(1)} \in \mathbb{R}^{H_1 \times \dim(\mathbf{p})}$, $\mathbf{b}^{(1)} \in \mathbb{R}^{H_1}$, $\mathbf{W}^{(2)} \in \mathbb{R}^{H_2 \times H_1}$, $\mathbf{b}^{(2)} \in \mathbb{R}^{H_2}$, $\mathbf{W}^{(3)} \in \mathbb{R}^{H_3 \times H_2}$, $\mathbf{b}^{(3)} \in \mathbb{R}^{H_3}$, $\mathbf{W}^{(o)} \in \mathbb{R}^{\dim(\boldsymbol{\lambda}) \times H_3}$ and $\mathbf{b}^{(o)} \in \mathbb{R}^{\dim(\boldsymbol{\lambda})}$. Note that $\hat{\boldsymbol{\lambda}} \in \mathbb{R}^{\dim(\boldsymbol{\lambda})}$ indicates the ANN prediction of the Lagrange multipliers. Coefficients of the network are found by minimizing

the loss function $\hat{\mathcal{C}}$ defined as a modified mean squared error (MSE) on the N_{tr} training data points, i.e.,

$$\hat{\mathcal{C}} := \frac{1}{N_{\text{tr}}} \sum_{i=1}^{N_{\text{tr}}} \|\hat{\boldsymbol{\lambda}}^{(i)} - \boldsymbol{\lambda}^{(i)}\|_{2, \mathbb{R}^{\dim(\boldsymbol{\lambda})}}^2 + \sum_{i=1}^3 \mathcal{R}(W^{(i)}) + \mathcal{R}(W^{(o)}), \quad (49)$$

where $\|\cdot\|_{2, \mathbb{R}^N}$ indicates the L^2 -norm of N -dimensional vector, and \mathcal{R} regulates the weights of the network to avoid over-fitting during the training step [68]. In this paper, the L^2 regularizer is used, corresponding to

$$\mathcal{R}(W) = \gamma \|W\|_{\text{F}, \mathbb{R}^{\dim(W)}}^2, \quad (50)$$

where subscript F indicates Frobenius norm and the hyper-parameter γ is used to control the regularization effect. The training of the network is implemented in Keras [69], with TensorFlow [70] as the backend. The optimal weights and biases of the network are obtained using the Adam stochastic optimizer [71], which uses mini-batches of size $N_{\text{batch}} < N_{\text{tr}}$ of the training data to take a single optimization step by minimizing the loss function. More precisely, the full training data set with N_{tr} data-points is shuffled, and $N_{\text{tr}}/N_{\text{batch}}$ mini-batches are extracted to take $N_{\text{tr}}/N_{\text{batch}}$ optimization steps. Once the entire training data set is exhausted, the training is said to complete one training epoch. The training is performed for a sufficient number of epochs to obtain a converged network. The convergence speed of the training is controlled by the learning rate of η . A feature scaling technique in which all the components of the input are scaled to the same range is applied to the data-sets to accelerate the training process [72]. An input scalar x is scaled by the mean normalization

$$x^* = \frac{x - \mu}{\sigma}, \quad (51)$$

where μ is the mean and σ is the standard deviation of x . At the beginning of the training, the weights and biases of the network are randomly initialized using normal distributions [73]. Therefore, the training needs to be performed several times, following a multiple restarts approach [74], to prevent the training results from depending on the initialization of the weights. In this paper, ten restarts are performed for the training of the network, and the trained model with the best validation accuracy is selected as the final model. The validation accuracy metric is the mean squared error (MSE). The optimal network in this paper is obtained with the hyper-parameters: $(H_1, H_2, H_3) = (10, 20, 40)$, $\gamma = 10^{-8}$, $N_{\text{batch}} = 1000$, and $\eta = 0.005$.

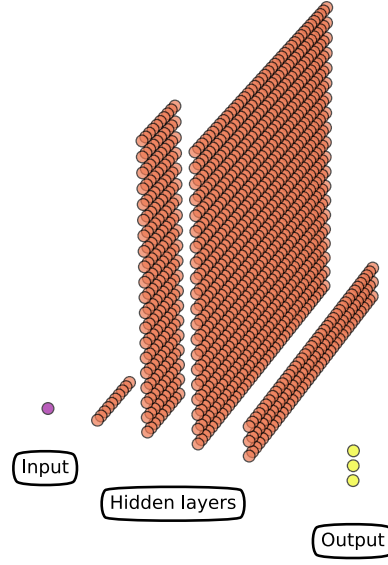


Fig. 1: A schematic representation of the devised ANN for the case $\dim(\lambda) = 3$

5.4. Generating training data-set

Either applying GP regression or ANN approximator, we need to first produce a relevant data-set containing the areas of interest for the pair $(\mathbf{p}, \boldsymbol{\lambda})$. In the context of this study we only consider regressions on MEDs with one-dimensional normalized velocity domain. The corresponding \mathbb{R} -domain hence is truncated to $\Omega = [-10, 10]\sigma$, where σ^2 is the input variance. Since in practice mean and variance of any distribution can be modified by scaling and shifting, the data-set is generated on the distributions with mean zero and unity variance. Although data generation is more efficient by randomly generating Lagrange multipliers and then taking the integrals [40], here we would like to have more control over the domain of the moments. Therefore in this study, the training data-set is generated by sampling from the space of moments and then directly solving for the Lagrange multipliers of MED, as explained in § 5.1. Let \mathcal{L}_i be the subspace from which the moment p_i is sampled. We only focus on MEDs with maximum number of eight moments. These moments are uniformly sampled according to $\mathcal{L}_3 = [-2, 2]$, $\mathcal{L}_4 = [-5, 5]$, $\mathcal{L}_6 = \mathcal{L}_5 = [-10, 10]$ and $\mathcal{L}_8 = \mathcal{L}_7 = [-20, 20]$.

6. Hybrid DSMC-SPH solution algorithm

In this section, the explained data-driven MED machinery is integrated into a DSMC-SPH hybrid algorithm. First, a switching criterion based on MED estimates is proposed to

dynamically assign DSMC/SPH simulator to each computational cell. Having trained an efficient MED estimator, one can approximate the distance from the equilibrium. Furthermore, MED estimates provide a reasonable closure needed at the DSMC-SPH interface; giving the information fed into DSMC from SPH cells.

Before proceed to further details, note that since we are concerned with DSMC/SPH solvers, only the MED with three moments will be relevant. Furthermore, we employ a normalized vector of central moments

$$\hat{\mathbf{p}} = \left(0, 1, \frac{2q}{\rho(k_b T/m)^{3/2}} \right)^T \quad (52)$$

as the input to MED estimators. Notice that since we have only presented the case with a one-dimensional velocity for our MED estimators, only one-dimensional non-equilibrium flow will be considered. Let V_1 be the coordinate along the non-equilibrium dimension, therefore the main assumption here is that the MDF can be decomposed as

$$\mathcal{F}(\mathbf{V}, \mathbf{x}, t) = \rho(\mathbf{x}, t) f(V_1 | \mathbf{x}, t) f^{\text{eq.}}(V_2 | \mathbf{x}, t) f^{\text{eq.}}(V_3 | \mathbf{x}, t), \quad (53)$$

where $f^{\text{eq.}}$ is a Maxwellian with the local mean velocity $\mathbf{U}(\mathbf{x}, t)$ and the local temperature $T(\mathbf{x}, t)$.

6.1. Switching

A relevant metric to obtain the distance between two densities $f_1(V)$ and $f_2(V)$ is given by the Fisher information distance

$$\mathcal{I}(f_1 | f_2) := \int_{\Omega} f_1 [\nabla \ln(f_1/f_2)]^2 dV. \quad (54)$$

For a detailed survey on physical importance of the Fisher information see [75].

In this study, we adopt the Fisher information distance as a measure to see how far the distribution stands from the equilibrium. First $\hat{\mathbf{p}}$ is estimated in every cell leading to an approximate f_3^λ based on GP and ANN regressions. Then, the switching criterion is found from $\mathcal{I}(f^{\text{eq.}} | f_3^\lambda)$ which has an analytical form. By setting a tolerance ϵ , the SPH cells are switched to DSMC if $\mathcal{I}(f^{\text{eq.}} | f_3^\lambda) \geq \epsilon$, and vice-versa if $\mathcal{I}(f^{\text{eq.}} | f_3^\lambda) < \epsilon$.

To switch a continuum cell to the kinetic one, all SPH particles are removed, and new DSMC particles are sampled from f_3^λ and positioned in the cell randomly with a uniform distribution. Conversely, DSMC particles are removed and SPH particles with values based on the moments of DSMC particles are placed equidistantly in the cell.

6.2. Treatment of interface

Let us decompose the domain into two parts; the union of continuum cells with SPH particles in them, and the union of kinetic cells occupied by DSMC particles. At the interface between DSMC and SPH cells, a proper boundary condition for each method needs to be derived. In this section, two candidate numerical approaches for interface treatment, with different efficiency/simplicity trade-off will be presented.

6.2.1. Flux method

One way to treat the interface boundary problem is based on the incoming flux. In particular, we consider the flux of incoming particles, as if the neighbouring cells outside of the considered sub-domain were operating with a solver similar to the one employed inside of the sub-domain. This leads to an accurate boundary condition for each scale, although the implementation in higher dimensions can become cumbersome.

For each kinetic cell lying at the boundary, the incoming flux of the neighbour distribution can be determined from the MED estimation of the distribution at the interface. Hence, first f_3^λ and accordingly the MDF \mathcal{F}_3^λ are estimated from moments of the neighbouring continuum cell. Then new particles are generated based on the MDF

$$\mathcal{F}_{\text{in}} = H(V_l n_l^i) V_j n_j^i \mathcal{F}_3^\lambda \quad (55)$$

of incoming particles [2]. Note that here $H(\cdot)$ denotes the Heaviside function and \mathbf{n}^i the normal of the interface inward to the kinetic cell. The particles are sampled using Metropolis-Hasting sampling method [44]. The new DSMC particles are streamed in the kinetic domain starting from the interface position with a uniformly distributed fraction of time step, i.e., $\delta t = r \Delta t$ where r is a random number in $(0, 1)$.

For the continuum cells at the border with the kinetic ones, the flux is approximated using

the moments of the neighbouring kinetic cell. Consider the mean velocity, the temperature and the density estimated at the neighbouring DSMC cell to be $\mathbf{U}^{(b)}$, $T^{(b)}$ and $\rho^{(b)}$. The incoming flux to the SPH domain is non-zero only if $U_j^{(b)} \tilde{n}_j^i > 0$ where \tilde{n}^i is the normal of the interface inward to the continuum cell. Once this condition is met, we consider the volume $\mathbf{U}^{(b)} \Delta t$ at the neighbouring DSMC cell. Let us call the resulting domain containing that volume, Ω_{in} . The idea is to populate Ω_{in} with SPH particles as described in the following. Based on the weights of SPH particles w and $\rho^{(b)}$, new SPH particles with equidistant spacing, the velocity $\mathbf{U}^{(b)}$ and the temperature $T^{(b)}$ are introduced inside the Ω_{in} sub-domain.

Although describing inflow fluxes leads to accurate and efficient treatment of sub-domain boundaries, it requires special care when it is implemented in higher dimensions. Hence an equivalent but simpler approach is suggested next, where the idea of using ghost cells for each sub-domain is explored.

6.2.2. Ghost cells method

From the implementation point of view, a simpler alternative can be obtained by extending each sub-domain with ghost cells.

In the case of DSMC, at every time step and every interface, a ghost DSMC cell on top of the adjacent SPH cell is considered. Next, new DSMC particles are generated based on the MED estimate \mathcal{F}_3^λ given the set of central moments coming from the SPH solution. Using the Metropolis-Hasting approach new velocities are assigned and particles are located in the ghost cell with a uniform distribution. Once the particles in the ghost cells are streamed, the remaining DSMC particles in the ghost cells are removed.

A similar approach can be deployed for treatment of SPH cells at the interface. On top of each neighbouring DSMC cell, a ghost cell is introduced and equidistant SPH particles are generated according to the moments of the local DSMC particles. Once SPH particles are streamed with the velocities obtained from the moments of the beneath DSMC cell, consistent fluxes are obtained. At the end of each time step, the SPH particles remaining in ghost cells are removed.

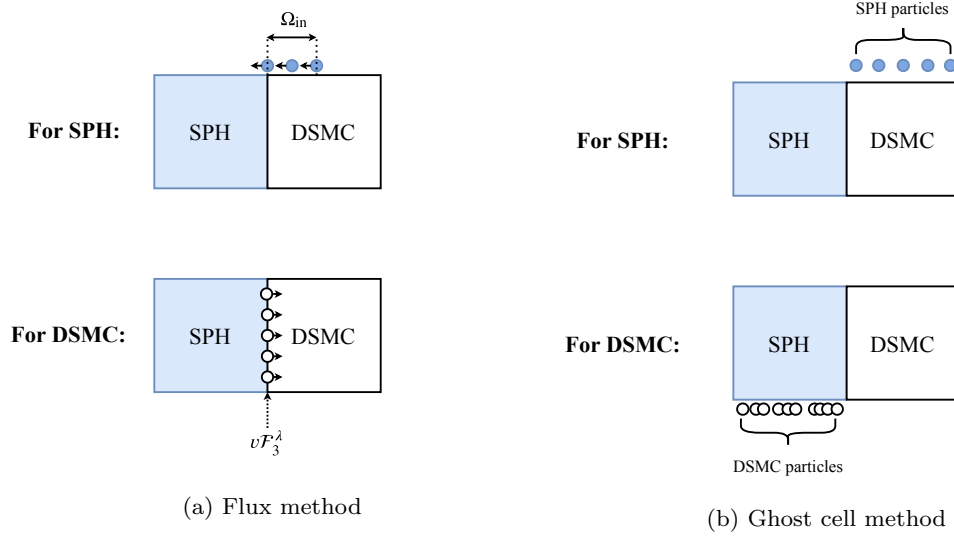


Fig. 2: Schematics of interface treatment

The above-mentioned approaches to cope with the boundary conditions for DSMC/SPH sub-domains are depicted schematically in Fig. 2.

6.3. Algorithm

By assembling all the introduced pieces into a hybrid DSMC-SPH particle method, the solution algorithm is summarized in Algorithm 6.3. Note that while only one-dimensional case is considered here, all methodologies are extendable to higher dimensions in a straightforward manner.

```

Initialization with SPH particles in every cell;
while  $t < T_{\text{final}}$  do
    -Evolve particles states according to SPH or DSMC rules;
    -Apply DSMC-SPH interface treatments;
    -Remove particles in wrong cells;
    for  $cell = 1, \dots, N_{\text{cells}}$  do
        -Estimate  $f_3^\lambda$  from moments using GP or ANN;
        if  $\mathcal{I}(f^{\text{eq}}|f_3^\lambda) > \epsilon$  and  $cell_{\text{flag}} == SPH$  then
            | -Replace SPH particles with samples of  $f_3^\lambda$  ;
        end
        if  $\mathcal{I}(f^{\text{eq}}|f_3^\lambda) < \epsilon$  and  $cell_{\text{flag}} == DSMC$  then
            | -Remove DSMC particles;
            | -Generate SPH particles with local moments;
        end
    end
    -Increment  $t$ ;
end

```

Algorithm 3: Hybrid DSMC-SPH solution algorithm

7. Results

The hybrid solution algorithm based on the data-driven MED estimate devised in previous sections § 4-6 is examined here against the benchmark results. First, the accuracy of GP and ANN regressions to estimate MEDs is studied in § 7.1. Then, in § 7.2, the Sod's shock tube as a test case with evident non-equilibrium effects is simulated using DSMC, SPH and the introduced hybrid algorithm. Before proceed, note that in all MED related computations we truncate the sample space to the domain $\Omega = [-10, 10]\sigma$, where σ^2 is the variance.

7.1. Estimating a bi-modal distribution

Before testing the devised hybrid multi-scale solution algorithm, first, the accuracy of the trained GP and ANN for $N = 3, 4, 6$ and 8 number of moments is evaluated. The reference is the bi-modal probability density which is an archetype of non-equilibrium flows (see e.g. [76]). Consider a bi-modal density obtained by adding two Gaussian ones each of the form $f^N(x|\mu, \sigma)$ with mean μ and standard deviation σ

$$f^{\text{bi}}(x|\mu_1, \sigma_1, \mu_2, \sigma_2) = \frac{1}{2} [f^N(x|\mu_1, \sigma_1) + f^N(x|\mu_2, \sigma_2)] \quad (56)$$

where $\mu_2 = -\mu_1$ and $\sigma_2 = \sqrt{2 - (\sigma_1^2 + 2\mu_1^2)}$ to make sure that mean and variance of $f^{\text{bi}}(x|\mu_1, \sigma_1, \mu_2, \sigma_2)$ is equal to zero and one, respectively. The moments of $f^{\text{bi}}(x|\mu_1, \sigma_1, \mu_2, \sigma_2)$ can be easily obtained by taking the integral in a large enough bounded domain, numerically. Now there are two questions to be addressed here. First, given the set of moments p , how well MED can approximate f^{bi} ? Next, how well GP or ANN can recover the underlying MED estimate?

In order to answer the first question, the direct approach to recover MED is applied, where the converged solution with the error in moments up to the machine accuracy $\epsilon = 10^{-16}$ is considered as the exact solution. Since the direct approach can be expensive if the initial guess is too far from the solution, the output from the trained GP estimate is used as the initial guess here. As shown in Fig. 3, by increasing the number of moments and, consequently, the number of Lagrange multipliers, the exact MED solution converges to the bi-modal distribution. Here, the Kullback-Leibler (KL) divergence

$$D_{KL}(f^{\text{bi}}||f_N^\lambda) = \int_{\mathbb{R}} f^{\text{bi}} \ln(f^{\text{bi}}/f_N^\lambda) dx \quad (57)$$

is employed as an indicator of the distance between the two densities. As it is shown in Fig. 3, the KL divergence between the target and the exact MED decreases as more moments are engaged.

Having investigated the accuracy of the exact MED in recovering the bi-modal density, let us turn to the second question and evaluate the MED estimates using GP and ANN regressions. As depicted in Fig. 3, both GP and ANN provide reasonable accuracy in predicting the exact MED in terms of the KL-divergence point of view. Yet one can observe that the GP regression seems to produce a better estimation overall considering the errors in Lagrange multipliers and the outcome moments. Unlike ANN, GP regressions provide the uncertainty of predictions in terms of the variance. In the considered test cases, the variance of prediction was in the order of 10^{-6} , which also indicates that the prediction point is not far from the training data-set.

In terms of the computational efficiency, the cost of predictions using GP or ANN are fixed. For the former, the cost depends on the number of training data points, whereas

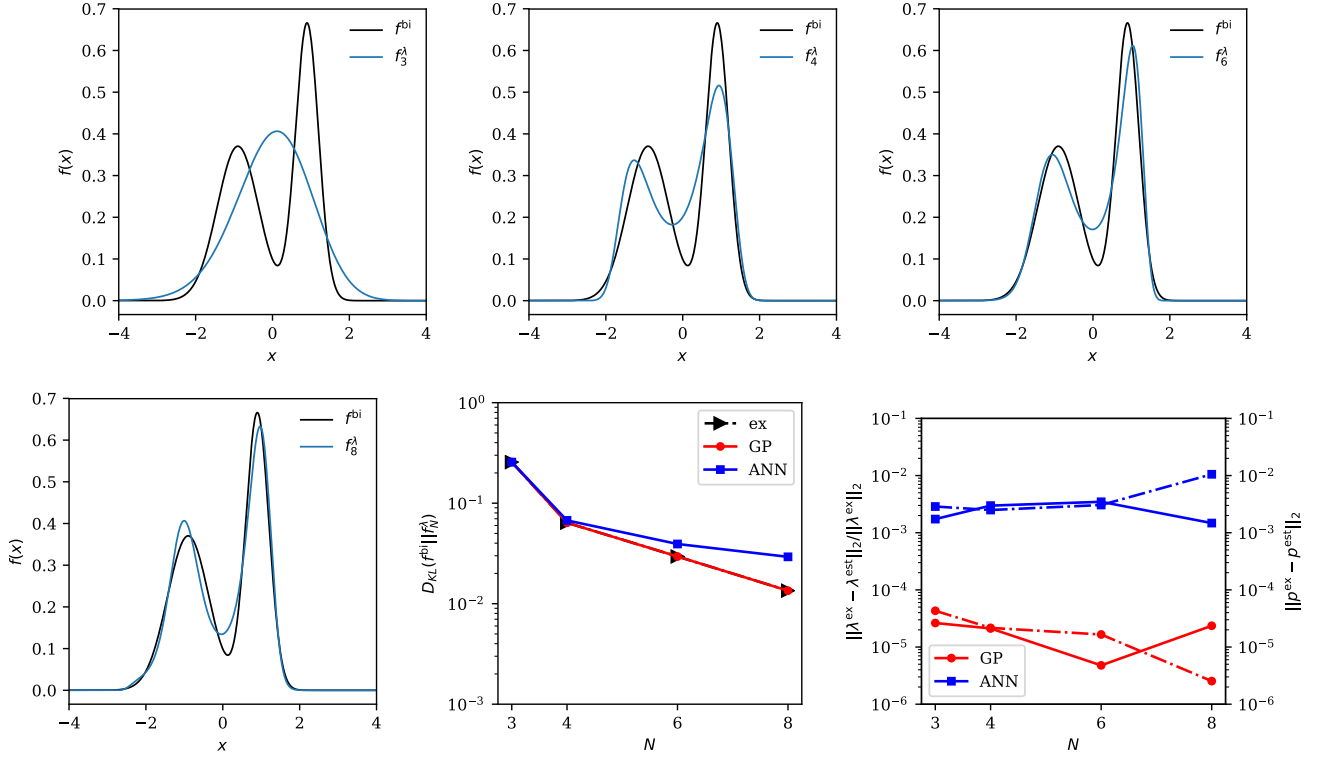


Fig. 3: Estimating $f^{\text{bi}}(x|\mu_1, \sigma_1, \mu_2, \sigma_2) = \frac{1}{2} [f^{\mathcal{N}}(x|\mu_1, \sigma_1) + f^{\mathcal{N}}(x|\mu_2, \sigma_2)]$ with $\mu_2 = -\mu_1$ and $\sigma_2 = \sqrt{2 - (\sigma_1^2 + 2\mu_1^2)}$ using exact, GP and ANN estimates of MED f_N^λ . Here, $\mu_1 = 0.9$ and $\sigma_1 = 0.3$. For the errors, the solid and dashed lines indicate the relative error in estimating Lagrange multipliers and the error in the outcome moments of the estimated MED, respectively.

the latter is a function of the network dimension. In this test case, the direct approach for finding MED is at least one order of magnitude more expensive than data-driven estimates. However note that in this comparison, we did not consider neither the offline costs associated with producing the data-sets, nor tuning/training GP/ANN.

7.2. Sod's shock tube

In this section, the continuum-kinetic hybrid solution algorithm devised in § 6.3 is applied to study of the classical Sod's shock tube [77]. Note that similar settings were used by [20], to evaluate their hybrid algorithm. Consider a domain where the relevant fluid direction is in $x_2 \in (0, L)$ with $L = 1$ m. The other dimensions in the physical space are assumed to be large enough, and consequently the kinetics of the gas along these dimensions can be considered to be at the equilibrium and thus condition (53) is met. For the continuum model, consider the initial values

$$\begin{cases} \rho_L = \rho_0 \\ U_{2,L} = u_0 \\ T_L = T_0 \end{cases} \quad \text{and} \quad \begin{cases} \rho_R = \rho_0/8 \\ U_{2,R} = u_0 \\ T_R = T_0 \end{cases} \quad (58)$$

where $\rho_0 \in \{10^{-4}, 10^{-5}, 10^{-6}\} \text{ kg.m}^{-3}$, the initial temperature $T_0 = 273 \text{ K}$, the initial velocity of $u_0 = 0 \text{ m.s}^{-1}$, and the subscripts L and R indicate the left and right sides of the initial discontinuity, respectively.

In the case of DSMC, while particle number follows the initial densities, the initial velocity in each direction $i \in \{1, 2, 3\}$ follows

$$M_{i,L/R} \sim \mathcal{N}(U_0, kT_{L/R}/m), \quad (59)$$

where the subscript $(.)_{L/R}$ denotes left or right side of the initial discontinuity.

The Neumann boundary conditions are applied at two ends of the physical domain. In the case of DSMC, particles leaving the domain are removed while new particles entering the domain are sampled from the flux of the Maxwell distribution with moments taken from the adjacent DSMC cell. In the SPH case, as particles leave the domain, they are removed while for the inflow duplicates of the adjacent SPH cell are streamed.

After an initial convergence study, a discretization using 200 cells leading to the mesh size of $\Delta x = 0.005 \text{ m}$ is used. The time step is picked based on $\Delta t = 0.01 \times \min(t_{\text{MFT}}, \Delta x/\tilde{U})$, where $\tilde{U} = \max(U, \mu/\rho)$, the mean-free-time is indicated by $t_{\text{MFT}} = \lambda(k_b T/m)^{-1/2}$, and λ denotes the mean-free-path based on the number density of the left side of the initial discontinuity

$$\lambda = \frac{1}{\sqrt{2}n_L\pi\sigma^2}. \quad (60)$$

Properties of Argon with molecular mass $m = 6.633521 \times 10^{-26} \text{ kg}$ and Hard-Sphere diameter of $\sigma = 3.405 \times 10^{-10} \text{ m}$ as an ideal monatomic gas are assumed. The viscosity follows $\mu = 5\sqrt{mk_b T/\pi}/(16\sigma^2)$ and the heat conductivity $\kappa = 15k_b\mu/(4m)$. In all test cases, the time step size is fixed to the value of $\Delta t = 2.01 \times 10^{-9} \text{ s}$. The SPH particles adopt identical weight. Moreover, SPH and hybrid solution algorithms are initialized with 40,000 particles in the left side of the domain and corresponding number to the right side. DSMC simulations are performed using the statistical weights $F_N \in \{10^{13}, 10^{12}, 10^{11}\} \times 9.42184$ which corresponds to $\rho_0 \in \{10^{-4}, 10^{-5}, 10^{-6}\} \text{ kg.m}^{-3}$.

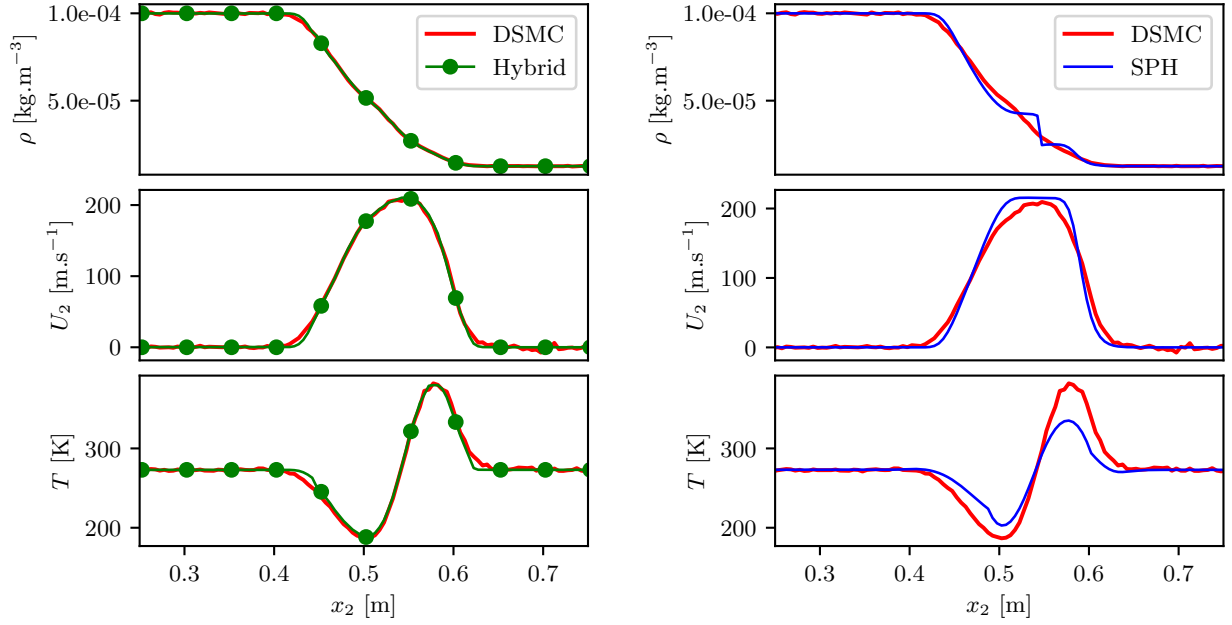


Fig. 4: Sod's shock tube with initial values of $(\rho_L, U_{2,L}, T_L) = (10^{-4} \text{ kg.m}^{-3}, 0 \text{ m.s}^{-1}, 273 \text{ K})$ and $(\rho_R, U_{2,R}, T_R) = (0.125 \times 10^{-4} \text{ kg.m}^{-3}, 0 \text{ m.s}^{-1}, 273 \text{ K})$ at $t = 2.01 \times 10^{-4} \text{ s}$. The profiles of the density ρ , the mean velocity U_2 , and the temperature T are obtained from SPH, DSMC, and the devised hybrid DSMC-SPH equipped with GP estimator of MED.

Three cases of the initial densities with the same termination time, provides different levels of non-equilibrium. First, the solution for the case with $\rho_L = 10^{-4} \text{ kg.m}^{-3}$ is obtained from full DSMC, full SPH, and hybrid DSMC-SPH based on GP/ANN MED estimates are presented in Fig. 4-5. For comparisons, the evolution of the Fisher information distance from equilibrium, along with the cell flags at four time intervals are shown in Fig. 6. Although DSMC solution is subject to noise, overall good agreement between the hybrid solution algorithm and full DSMC results is observed; considering both GP and ANN estimates of MED.

To gain a better insight into the accuracy of data-driven GP/ANN estimators of MED, we extract moments from four probe positions $x_A = 0.45 \text{ m}$, $x_B = 0.5 \text{ m}$, $x_C = 0.55 \text{ m}$, and $x_D = 0.6 \text{ m}$ at the terminal time. Instead of computing the exact MED solution, we have computed the moments of the estimated MEDs and compared them with respect to the input moment. As it can be seen in Fig. 7, the predictions based on the GP regression shows a better accuracy in comparison to the one from ANN. Consequently, for the remaining test cases, only the trained GP method is employed in the hybrid solution algorithm.

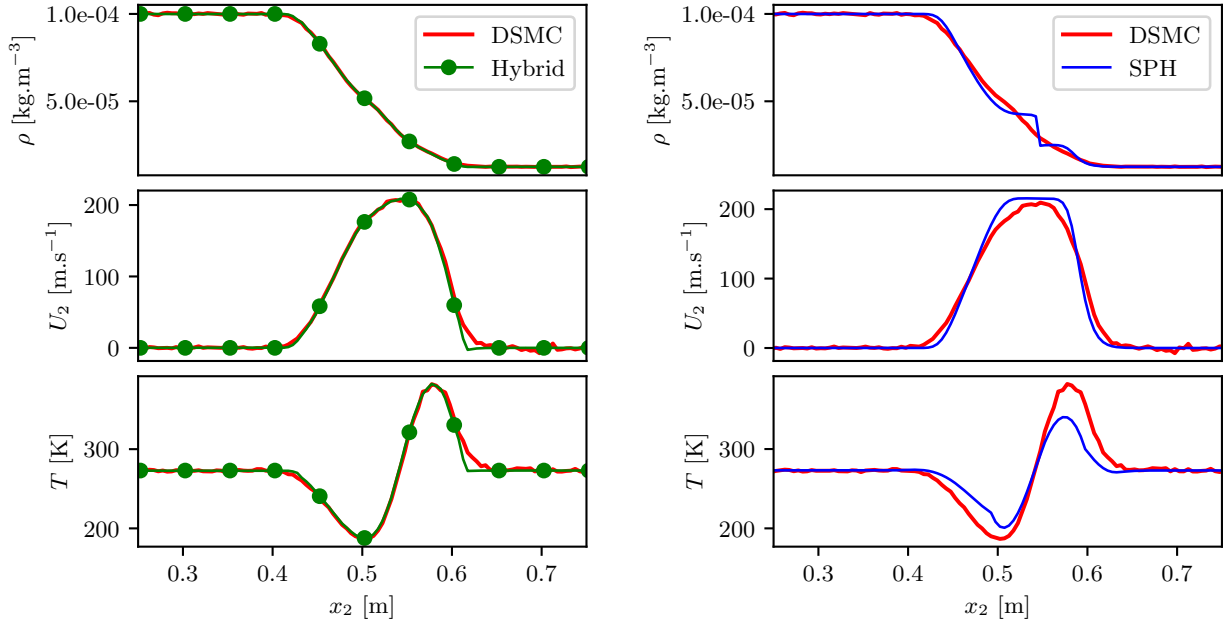


Fig. 5: Sod's shock tube with initial values of $(\rho_L, U_{2,L}, T_L) = (10^{-4} \text{ kg.m}^{-3}, 0 \text{ m.s}^{-1}, 273 \text{ K})$ and $(\rho_R, U_{2,R}, T_R) = (0.125 \times 10^{-4} \text{ kg.m}^{-3}, 0 \text{ m.s}^{-1}, 273 \text{ K})$ at $t = 2.01 \times 10^{-4} \text{ s}$. The profiles of the density ρ , the mean velocity U_2 , and the temperature T are obtained from SPH, DSMC, and the devised hybrid DSMC-SPH equipped with ANN estimator of MED.

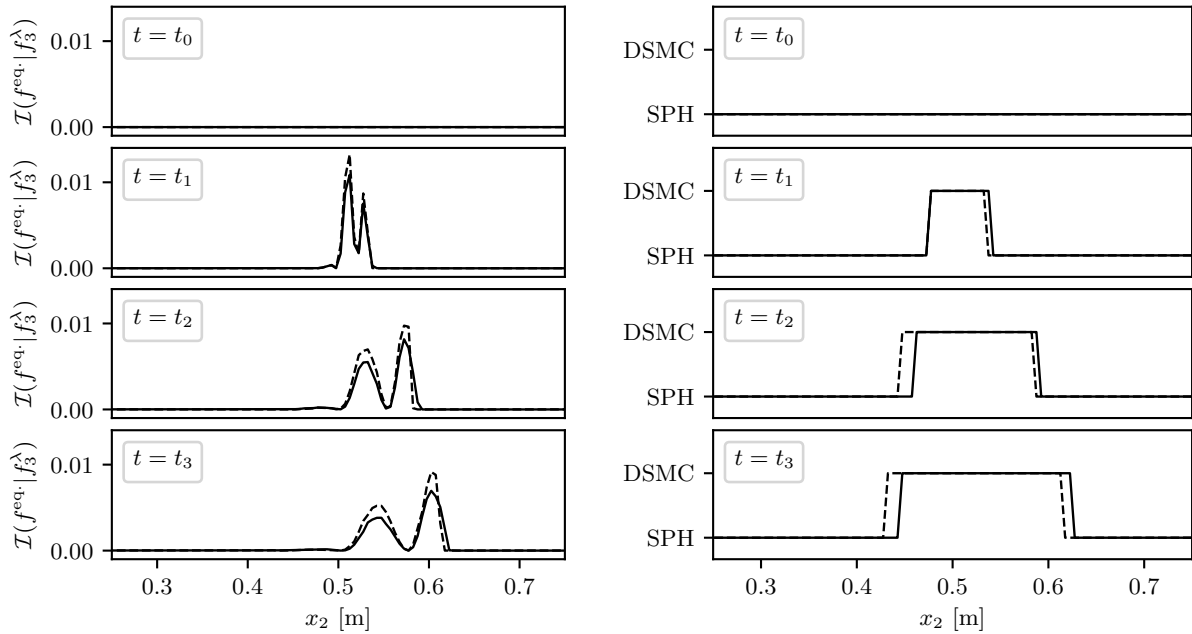


Fig. 6: Evolution of the Fisher information distance $\mathcal{I}(f^{\text{eq.}} | f_3^\lambda)$ at $t \in \{0, 0.5025, 1.5075, 2.01\} \times 10^{-4} \text{ s}$ and the assigned DSMC/SPH solver in the hybrid method for the Sod's shock tube test case with $\rho_0 = 10^{-4} \text{ kg.m}^{-3}$, shown at left and right, respectively. Solid and dashed lines indicate that MED is estimated using GP and ANN, respectively.

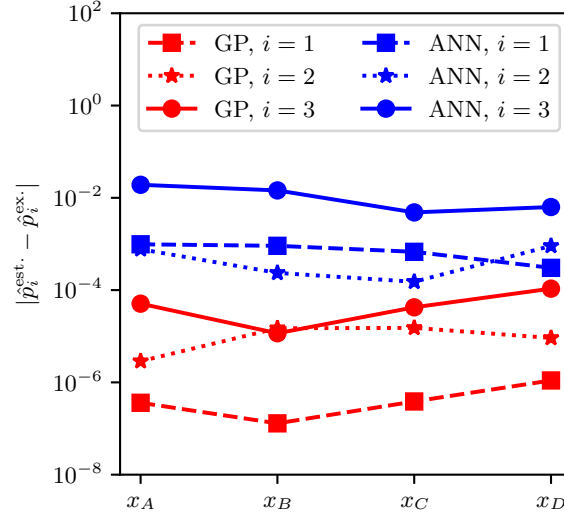


Fig. 7: Absolute value error of moments of the estimated f_3^λ using GP and ANN at four locations $x_A = 0.45$ m, $x_B = 0.5$ m, $x_C = 0.55$ m, and $x_D = 0.6$ m for the Sod's shock tube hybrid simulation at $t = 2.01 \times 10^{-4}$ s.

Next, more rarefied scenarios are considered for the Sod's shock tube with $\rho_0 = 10^{-5}$ kg.m $^{-3}$ and $\rho_0 = 10^{-6}$ kg.m $^{-3}$, as the results are shown in Figs. 8 and 9, respectively. For lower density while keeping ratio ρ_L/ρ_R constant, the solution of SPH departs further from DSMC results. However, the hybrid solution shows a good accuracy in comparison with the benchmark DSMC result.

7.3. Performance of hybrid solution algorithm

As more cells remain in the continuum limit and assigned to the SPH solver in the hybrid approach, much less noise compared to the full DSMC simulation is expected. Moreover, even though the hybrid solution algorithm introduces an overhead mainly due to particle samplings from MED estimates, still good computational efficiency can be obtained since typically only a small fraction of cells switch to DSMC at a given time step. Computational speedups along with noise reductions associated with the hybrid results are reported in Fig. 10.

8. Conclusion and outlook

This study explores some potential of data-driven methodologies to accelerate multi-scale computations, relevant to hybrid kinetic-continuum settings. Two main challenges of interface treatment and switching criterion, associated with coupling of continuum and kinetic

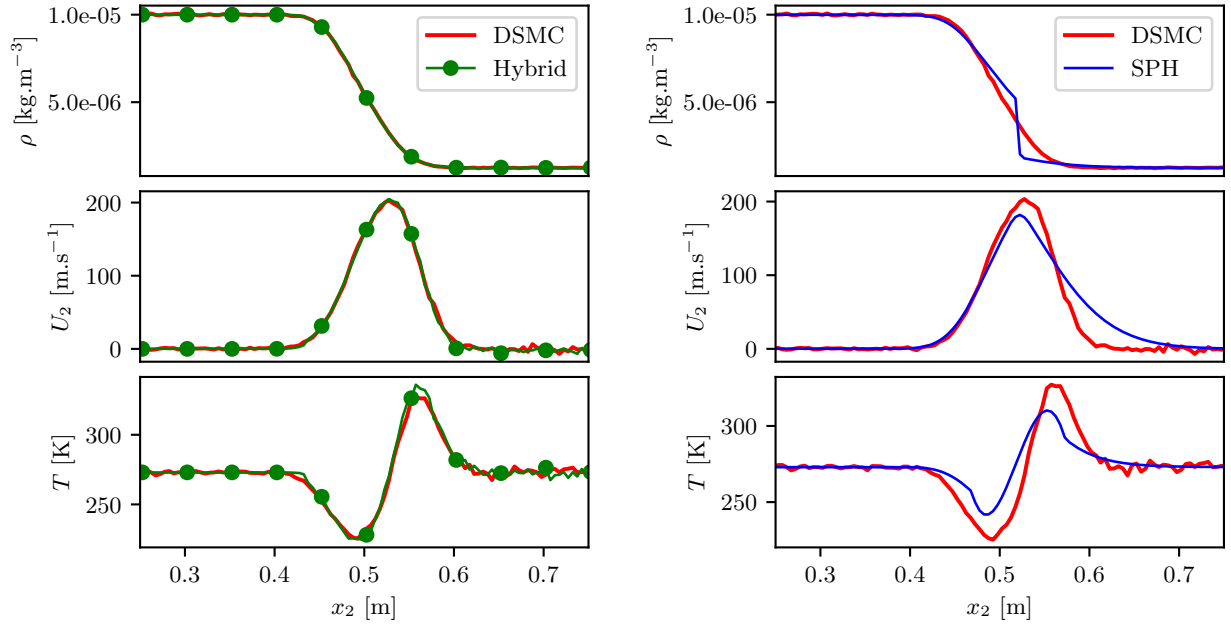


Fig. 8: Sod's shock tube with initial values of $(\rho_L, U_{2,L}, T_L) = (10^{-5} \text{ kg.m}^{-3}, 0 \text{ m.s}^{-1}, 273 \text{ K})$ and $(\rho_R, U_{2,R}, T_R) = (0.125 \times 10^{-5} \text{ kg.m}^{-3}, 0 \text{ m.s}^{-1}, 273 \text{ K})$ at $t = 2.01 \times 10^{-4} \text{ s}$. The profiles of the density ρ , the mean velocity U_2 , and the temperature T are obtained from SPH, DSMC, and the devised hybrid DSMC-SPH equipped with ANN estimator of MED.

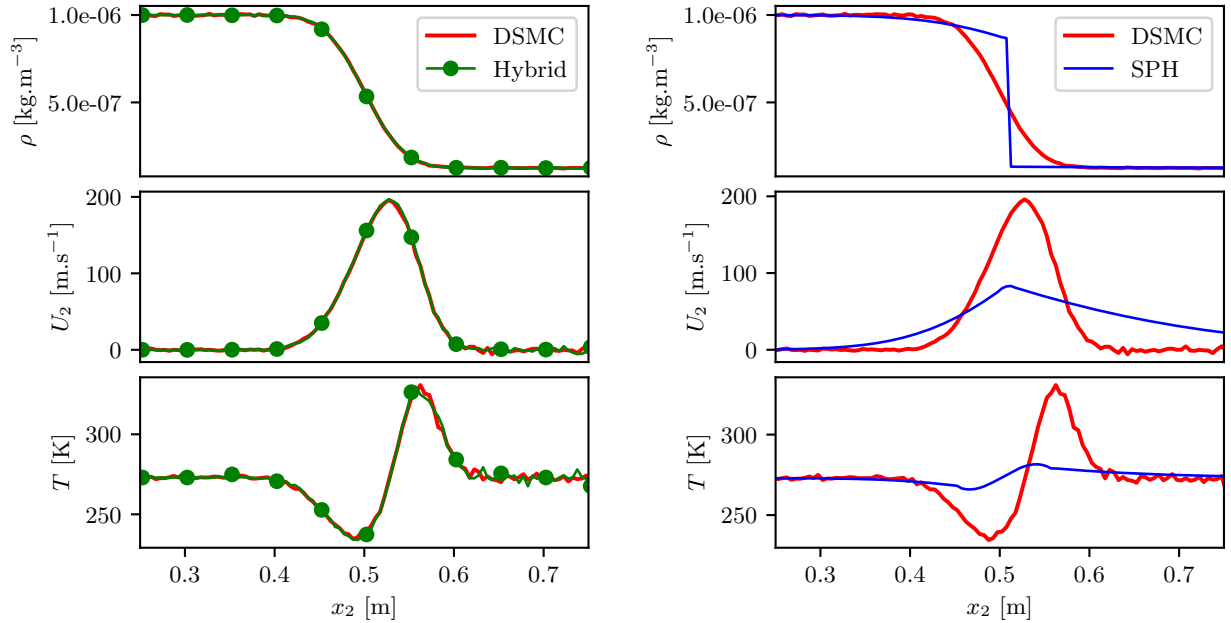
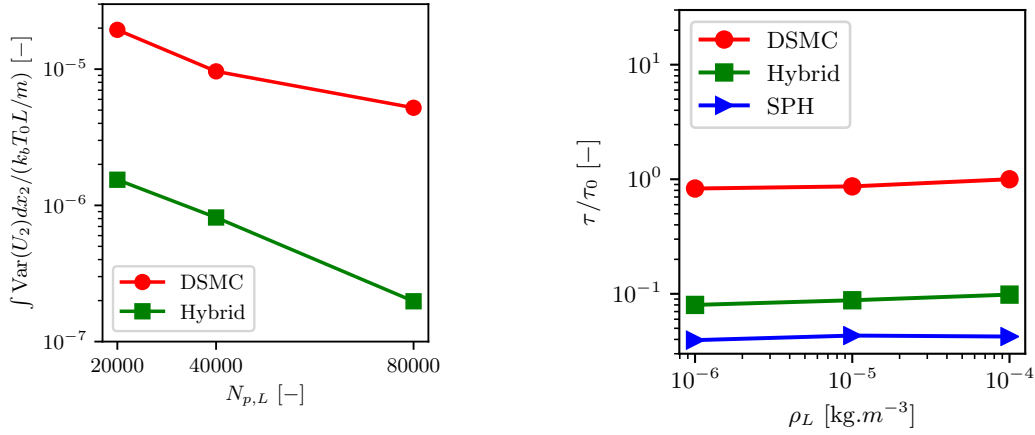


Fig. 9: Sod's shock tube with initial values of $(\rho_L, U_{2,L}, T_L) = (10^{-6} \text{ kg.m}^{-3}, 0 \text{ m.s}^{-1}, 273 \text{ K})$ and $(\rho_R, U_{2,R}, T_R) = (0.125 \times 10^{-6} \text{ kg.m}^{-3}, 0 \text{ m.s}^{-1}, 273 \text{ K})$ at $t = 2.01 \times 10^{-4} \text{ s}$. The profiles of the density ρ , the mean velocity U_2 , and the temperature T are obtained from SPH, DSMC, and the devised hybrid DSMC-SPH equipped with ANN estimator of MED.



(a) Spatial averaged noise of U_2 estimations using DSMC and devised hybrid algorithm for the case of $\rho_0 = 10^{-4} \text{ kg.m}^{-3}$ at $t = 1 \times 10^{-4} \text{ s}$, where 10 independent simulation results were employed to obtain the variance at each point in space. Here $N_{p,L}$ indicates the number of SPH particles initialized at the left side of the discontinuity for the hybrid method, while for the full DSMC simulation 200 times more particles i.e. $N_{p,L}^{\text{DSMC}} = 200N_{p,L}$ were generated at the left portion of the domain.

(b) Normalized CPU time τ for the Sod's shock tube problem using DSMC, SPH and devised hybrid DSMC-SPH with $t_{\text{final}} = 1 \times 10^{-4} \text{ s}$. Here $N_{p,L} = 40,000$ particles were initialized on the left side of the discontinuity for the SPH and hybrid method, while $N_{p,L}^{\text{DSMC}} = 200N_{p,L}$ used initially for full DSMC simulation. The execution time τ is normalized by τ_0 indicating DSMC execution time of the case $\rho_0 = 10^{-4} \text{ kg.m}^{-3}$.

Fig. 10: Noise and cost study of DSMC, SPH and DSMC-SPH solution algorithms for the Sod's shock tube test case

solvers are addressed. It is shown that a practical and accurate probability density estimator can circumvent these challenges and thus make efficient hybrid algorithms more accessible. By leveraging GP and ANN regression schemes, efficient density estimations based on the Maximum-Entropy condition are derived. Accordingly, a hybrid solution algorithm based on fast and accurate MED estimators is devised and tested in one-dimensional shock tube scenarios. Based on detailed comparisons, the GP regression model was picked as the method of choice for the studied scenario. Very good agreement with respect to the DSMC benchmark along with significant a speed-up are observed. While the setting considered here was one-dimensional, more dimensions can be included in a straightforward manner, given that all presented derivations are independent of the number of dimensions. Furthermore, even though only DSMC-SPH coupling was addressed throughout this work, the proposed approach motivates the use of data-driven MED for coupling higher-order moment systems coupled with kinetic solution algorithms.

9. Acknowledgments

Hossein Gorji acknowledges the funding provided by the Swiss National Science Foundation under the grant number 174060. Furthermore, Mohsen Sadr acknowledges the scholarship provided by the German Academic Exchange Service (DAAD) identifiable with number 57438025-91749111.

References

- [1] G. A. Bird, Approach to translational equilibrium in a rigid sphere gas, *Physics of Fluids* 6 (1963) 1518–1519.
- [2] G. A. Bird, *Molecular gas dynamics and the direct simulation of gas flows*, volume 42, Clarendon press Oxford, 1994.
- [3] M. A. Gallis, J. Torczynski, D. Rader, G. A. Bird, Convergence behavior of a new DSMC algorithm, *Journal of Computational Physics* 228 (2009) 4532–4548.
- [4] J. E. Broadwell, Study of rarefied shear flow by the discrete velocity method, *Journal of Fluid Mechanics* 19 (1964) 401–414.
- [5] L. Mieussens, Discrete-velocity models and numerical schemes for the Boltzmann-BGK equation in plane and axisymmetric geometries, *Journal of Computational Physics* 162 (2000) 429–466.
- [6] A. Morris, P. L. Varghese, D. B. Goldstein, Monte Carlo solution of the Boltzmann equation via a discrete velocity model, *Journal of Computational Physics* 230 (2011) 1265–1280.
- [7] L. Pareschi, G. Russo, Numerical solution of the Boltzmann equation i: Spectrally accurate approximation of the collision operator, *SIAM Journal on Numerical Analysis* 37 (2000) 1217–1245.
- [8] L. Wu, C. White, T. J. Scanlon, J. M. Reese, Y. Zhang, Deterministic numerical solutions of the Boltzmann equation using the fast spectral method, *Journal of Computational Physics* 250 (2013) 27–52.
- [9] I. M. Gamba, S. H. Tharkabhushanam, Spectral-Lagrangian methods for collisional models of non-equilibrium statistical states, *Journal of Computational Physics* 228 (2009) 2012–2036.
- [10] P. Jenny, M. Torrilhon, S. Heinz, A solution algorithm for the fluid dynamic equations based on a stochastic model for molecular motion, *Journal of computational physics* 229 (2010) 1077–1098.
- [11] M. H. Gorji, M. Torrilhon, P. Jenny, Fokker–Planck model for computational studies of monatomic rarefied gas flows, *Journal of fluid mechanics* 680 (2011) 574–601.
- [12] J. Mathiaud, L. Mieussens, A Fokker–Planck model of the Boltzmann equation with correct Prandtl number, *Journal of Statistical Physics* 162 (2016) 397–414.
- [13] S. Singh, S. Ansumali, Fokker-Planck model of hydrodynamics, *Physical Review E* 91 (2015) 033303.
- [14] R. Sander, Z. Pan, L. D. Connell, Laboratory measurement of low permeability unconventional gas reservoir rocks: A review of experimental methods, *Journal of Natural Gas Science and Engineering* 37 (2017) 248–279.
- [15] K.-C. Tseng, J.-S. Wu, I. D. Boyd, Simulations of re-entry vehicles by using DSMC with chemical-reaction module, in: *14th AIAA/AHI Space Planes and Hypersonic Systems and Technologies Conference*, 2006, p. 8084.
- [16] S. Kremser, L. W. Thomason, M. Hobe, M. Hermann, T. Deshler, C. Timmreck, M. Toohey, A. Stenke, J. P. Schwarz, R. Weigel, et al., Stratospheric aerosol-observations, processes, and impact on climate, *Reviews of Geophysics* 54 (2016) 278–335.
- [17] E. L. Petersen, R. K. Hanson, Nonideal effects behind reflected shock waves in a high-pressure shock tube, *Shock Waves* 10 (2001) 405–420.
- [18] M. P. Brenner, S. Hilgenfeldt, D. Lohse, Single-bubble sonoluminescence, *Reviews of Modern Physics* 74 (2002) 425.
- [19] R. N. Dahms, J. Manin, L. M. Pickett, J. C. Oefelein, Understanding high-pressure gas-liquid interface phenomena in diesel engines, *Proceedings of the Combustion Institute* 34 (2013) 1667–1675.

- [20] S. Tiwari, A. Klar, S. Hardt, A particle–particle hybrid method for kinetic and continuum equations, *Journal of Computational Physics* 228 (2009) 7109–7124.
- [21] J. M. Burt, I. D. Boyd, A hybrid particle approach for continuum and rarefied flow simulation, *Journal of Computational Physics* 228 (2009) 460–475.
- [22] J. Ye, J. Yang, J. Zheng, X. Ding, I. Wong, W. Li, C. Chen, A multi-scale flow analysis in hydrogen separation membranes using a coupled DSMC-SPH method, *international journal of hydrogen energy* 37 (2012) 894–902.
- [23] M. H. Gorji, P. Jenny, Fokker–Planck–DSMC algorithm for simulations of rarefied gas flows, *Journal of Computational Physics* 287 (2015) 110–129.
- [24] T. E. Schwartzentruber, I. D. Boyd, A hybrid particle-continuum method applied to shock waves, *Journal of Computational Physics* 215 (2006) 402–416.
- [25] W.-L. Wang, I. Boyd, Hybrid DSMC-CFD simulations of hypersonic flow over sharp and blunted bodies, in: 36th AIAA Thermophysics Conference, 2003, p. 3644.
- [26] O. Aktas, N. R. Aluru, A combined continuum/DSMC technique for multiscale analysis of microfluidic filters, *Journal of Computational Physics* 178 (2002) 342–372.
- [27] K. Xu, J.-C. Huang, A unified gas-kinetic scheme for continuum and rarefied flows, *Journal of Computational Physics* 229 (2010) 7747–7764.
- [28] S. Chapman, T. G. Cowling, *The mathematical theory of non-uniform gases: an account of the kinetic theory of viscosity, thermal conduction and diffusion in gases*, Cambridge university press, 1970.
- [29] A. L. Garcia, B. J. Alder, Generation of the chapman–enskog distribution, *Journal of computational physics* 140 (1998) 66–70.
- [30] H. Struchtrup, M. Torrilhon, Regularization of Grad’s 13 moment equations: derivation and linear analysis, *Physics of Fluids* 15 (2003) 2668–2680.
- [31] H. Grad, On the kinetic theory of rarefied gases, *Communications on pure and applied mathematics* 2 (1949) 331–407.
- [32] W.-L. Wang, I. D. Boyd, Predicting continuum breakdown in hypersonic viscous flows, *Physics of fluids* 15 (2003) 91–100.
- [33] G. Bird, Breakdown of translational and rotational equilibrium in gaseous expansions, *Aiaa Journal* 8 (1970) 1998–2003.
- [34] G. Di Staso, H. Clercx, S. Succi, F. Toschi, Lattice Boltzmann accelerated direct simulation Monte Carlo for dilute gas flow simulations, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2016) 20160226.
- [35] J. N. Kapur, *Maximum-entropy models in science and engineering*, John Wiley & Sons, 1989.
- [36] W. Dreyer, Maximisation of the entropy in non-equilibrium, *Journal of Physics A: Mathematical and General* 20 (1987) 6505.
- [37] C. D. Levermore, Moment closure hierarchies for kinetic theories, *Journal of statistical Physics* 83 (1996) 1021–1065.
- [38] R. V. Abramov, An improved algorithm for the multidimensional moment-constrained maximum entropy problem, *Journal of Computational Physics* 226 (2007) 621–644.
- [39] R. V. Abramov, The multidimensional moment-constrained maximum entropy problem: A BFGS algorithm with constraint scaling, *Journal of Computational Physics* 228 (2009) 96–108.
- [40] M. Sadr, M. Torrilhon, M. H. Gorji, Gaussian process regression for maximum entropy distribution, *Journal of Computational Physics* (submitted) (2019).
- [41] G.-R. Liu, M. B. Liu, *Smoothed particle hydrodynamics: a meshfree particle method*, World scientific, 2003.
- [42] R. M. Neal, et al., Slice sampling, *The annals of statistics* 31 (2003) 705–767.
- [43] W. R. Gilks, P. Wild, Adaptive rejection sampling for Gibbs sampling, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 41 (1992) 337–348.
- [44] S. Chib, E. Greenberg, Understanding the metropolis-hastings algorithm, *The american statistician* 49 (1995) 327–335.
- [45] J. D. Anderson, Governing equations of fluid dynamics, in: *Computational fluid dynamics*, Springer, 1992, pp. 15–51.
- [46] A. Rana, M. Torrilhon, H. Struchtrup, A robust numerical method for the R13 equations of rarefied

- gas dynamics: Application to lid driven cavity, *Journal of Computational Physics* 236 (2013) 169–186.
- [47] X.-j. Gu, D. R. Emerson, A high-order moment approach for capturing non-equilibrium phenomena in the transition regime, *Journal of fluid mechanics* 636 (2009) 177–216.
- [48] M. Y. Timokhin, H. Struchtrup, A. Kokhanchik, Y. A. Bondar, Different variants of R13 moment equations applied to the shock-wave structure, *Physics of Fluids* 29 (2017) 037105.
- [49] A. Frezzotti, P. Grosfils, S. Toxvaerd, Evidence of an inverted temperature gradient during evaporation/condensation of a Lennard-Jones fluid, *Physics of fluids* 15 (2003) 2837–2842.
- [50] J. J. Monaghan, Smoothed particle hydrodynamics, *Annual review of astronomy and astrophysics* 30 (1992) 543–574.
- [51] J. J. Monaghan, Smoothed particle hydrodynamics, *Reports on progress in physics* 68 (2005) 1703.
- [52] L. D. G. Sigalotti, H. López, A. Donoso, E. Sira, J. Klapp, A shock-capturing SPH scheme based on adaptive kernel estimation, *Journal of computational Physics* 212 (2006) 124–149.
- [53] S. Rjasanow, W. Wagner, *Stochastic numerics for the Boltzmann equation*, Springer, 2005.
- [54] M. Pfeiffer, A. Mirza, C.-D. Munz, S. Fasoulas, Two statistical particle split and merge methods for particle-in-cell codes, *Computer Physics Communications* 191 (2015) 9–24.
- [55] M. H. Gorji, P. Jenny, An efficient particle Fokker–Planck algorithm for rarefied gas flows, *Journal of Computational Physics* 262 (2014) 325–343.
- [56] R. O. Fox, Higher-order quadrature-based moment methods for kinetic equations, *Journal of Computational Physics* 228 (2009) 7771–7791.
- [57] M. Torrilhon, Modeling nonequilibrium gas flow based on moment equations, *Annual review of fluid mechanics* 48 (2016) 429–458.
- [58] H. Struchtrup, M. Torrilhon, Regularization of Grad’s 13 moment equations: derivation and linear analysis, *Physics of Fluids* 15 (2003) 2668–2680.
- [59] C. D. Hauck, C. D. Levermore, A. L. Tits, Convex duality and entropy-based moment closures: Characterizing degenerate densities, *SIAM Journal on Control and Optimization* 47 (2008) 1977–2015.
- [60] J. Noonan, N. Tzannes, T. Costello, On the inverse problem of entropy maximizations (corresp.), *IEEE Transactions on Information Theory* 22 (1976) 120–123.
- [61] J. Einbu, On the existence of a class of maximum-entropy probability density functions (corresp.), *IEEE Transactions on Information Theory* 23 (1977) 772–775.
- [62] R. P. Schaerer, M. Torrilhon, The 35-moment system with the maximum-entropy closure for rarefied gas flows, *European Journal of Mechanics-B/Fluids* 64 (2017) 30–40.
- [63] C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.
- [64] H. Owhadi, C. Scovel, *Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design*, volume 35, Cambridge University Press, 2019.
- [65] J. Nocedal, S. Wright, *Numerical optimization*, Springer Science & Business Media, 2006.
- [66] A. G. d. G. Matthews, M. van der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. León-Villagrà, Z. Ghahramani, J. Hensman, GPflow: A Gaussian process library using TensorFlow, *Journal of Machine Learning Research* 18 (2017) 1–6.
- [67] S. Haykin, *Neural networks: a comprehensive foundation*, Prentice Hall PTR, 1994.
- [68] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*, MIT press, 2016.
- [69] F. Chollet, et al., Keras, 2015.
- [70] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: <http://tensorflow.org/>, software available from tensorflow.org.
- [71] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [72] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv preprint arXiv:1502.03167* (2015).
- [73] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010, pp. 249–256.

- [74] K.-l. Hsu, H. V. Gupta, S. Sorooshian, Artificial neural network modeling of the rainfall-runoff process, *Water resources research* 31 (1995) 2517–2530.
- [75] B. R. Frieden, *Physics from fisher information: a unification*, 2000.
- [76] M. A. Solovchuk, T. W. Sheu, Prediction of shock structure using the bimodal distribution function, *Physical Review E* 81 (2010) 056314.
- [77] G. A. Sod, A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws, *Journal of computational physics* 27 (1978) 1–31.