From Variance Reduction to Particle Generation: An Interplay of Entropy and Moments

Mohsen Sadr

ETH, Paul Scherrer Institute,
Switzerland

Collaborators: Prof. Hossein Gorji (ETH-EMPA / USC)
Prof. Nicolas Hadjiconstantinue (MIT)

PSI Massachusetts Institute of Technology

June 24, 2025

DFG Deutsche Forschungsgemeinschaft

Moment Problem

Problem: Given moments corresponding to basis functions H, i.e.

$$\mu = \int f(v)H(v)dv, \tag{1.1}$$

find the underlying distribution function $f(v)$.

Challenges:

- f may not exist.
- f is not unique.

Regularization with Entropy

Find a distribution that maximizes entropy besides matching moments

$$\mathcal{L}[f] := -\int f \log(f) dv + \sum_{i=1}^{N_m} \lambda_i \left(\int H_i f dv - \mu_i \right) .$$

The variational extremum of $\mathcal{L}$ gives

$$f^* = \frac{1}{Z} \exp\left(\lambda \cdot H\right), \qquad \text{where} \quad Z = \int \exp\left(\lambda \cdot H\right) dv. \tag{2.1}$$

The Lagrange multipliers can be found using Newton-Raphson method

$$\lambda \leftarrow \lambda - L^{-1} g \tag{2.2}$$

$$\text{where} \quad g_i = \mu_i - \frac{1}{Z} \int H_i \exp(\lambda \cdot H) dv \tag{2.3}$$

$$\text{and} \quad L_{i,j} = -\frac{1}{Z} \int H_i H_j \exp(\lambda \cdot H) dv. \tag{2.4}$$

$$f^* = \frac{1}{Z} \exp\left(\lambda \cdot H\right), \qquad \text{where} \quad Z = \int \exp\left(\lambda \cdot H\right) dv.$$

Pros	**Cons**
✓ Least bias	✗ Ill-conditioned
✓ Unique distr.	✗ Cannot guarantee existence
✓ Convex/concave optimization problem	✗ High dimensional integrals
✓ Matching moments	
✓ Guaranteed positivity for f	

Importance sampling with cross-entropy

Given a prior f^{prior}, find a distribution that maximizes cross-entropy from it besides matching moments

$$\mathcal{L}_\times[f] := -\int f \log(f/f^{\text{prior}})dv + \sum_{i=1}^{N_m} \lambda_i \left(\int H_i f dv - \mu_i \right) .$$

The variational extremum of $\mathcal{L}_\times[.]$ gives

$$f^* = \frac{f^{\text{prior}}}{Z} \exp\left(\lambda \cdot H\right). \tag{2.5}$$

Using the samples of f^{prior}, the Lagrange multipliers can be found using Newton-Raphson method

$$\lambda \leftarrow \lambda - L^{-1} g \tag{2.6}$$

$$\text{where} \quad g_i = \mu_i - \frac{1}{Z} \int f^{\text{prior}} H_i \exp(\lambda \cdot H) dv \tag{2.7}$$

$$\text{and} \quad L_{i,j} = -\frac{1}{Z} \int f^{\text{prior}} H_i H_j \exp(\lambda \cdot H) dv. \tag{2.8}$$

$$f^* = \frac{f^{\text{prior}}}{Z} \exp\left(\lambda \cdot H\right)$$

Pros	Cons
✓ Least bias	✗ Ill-conditioned
✓ Unique distr.	✗ Cannot guarantee existence
✓ Convex/concave optimization problem	
✓ Matching moments	
✓ Integration with importance sampling	

1. Variance Reduction

When computing moments form particles, consider decomposition

$$\int H(v)f(v|x,t)dv = \underbrace{\int H(v)\left(1 - w(v|x,t)\right)f(v|x,t)dv}_{\text{numerically}} + \underbrace{\int H(v)f^{\text{cont.}}(v|x,t)dv}_{\text{analytically}}, \tag{3.1}$$

where $w(v|x,t) = f^{\text{cont.}}(v|x,t)/f(v|x,t)$.

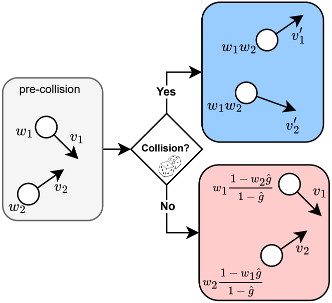
Evolution of weights: we need to find the weight process.
Gain: variance reduction when $\langle H(X)(1 - w(X))\rangle_{f_X} \ll 0$.
Issue: exact weight process becomes unstable for most collision operators.

Given the binary collision operator

$$S(f) = \frac{1}{2}\int (\delta_1^* + \delta_2^* - \delta_1 - \delta_2) f_1 f_2 \hat{g}\, b db d\epsilon dv_1 dv_2$$

the weight process can be found by substituting $f^{\mathrm{cont.}}$ into $S(.)$

$$S(f^{\mathrm{cont.}}) = \frac{C}{2}\int (\delta_1^* + \delta_2^* - \frac{\delta_1}{w_2} - \frac{\delta_2}{w_1}) w_1 w_2 f_1 f_2 \hat{g}\, b db d\epsilon dv_1 dv_2$$
$$+ \frac{C}{2}\int (-\delta_1 - \delta_2 + \frac{\delta_1}{w_2} + \frac{\delta_2}{w_1})(\frac{\hat{g}}{1-\hat{g}}) w_1 w_2 f_1 f_2 \left(1-\hat{g}\right) b db d\epsilon dv_1 dv_2 \ .$$



$$
\begin{aligned}
S(f^{\text{cont.}}) = \frac{C}{2}\int (\delta_1^* + \delta_2^* - \frac{\delta_1}{w_2} - \frac{\delta_2}{w_1}) w_1 w_2 f_1 f_2 \hat{g}\, b db d\epsilon dv_1 dv_2 \\
+ \frac{C}{2}\int (-\delta_1 - \delta_2 + \frac{\delta_1}{w_2} + \frac{\delta_2}{w_1}) \left(\frac{\hat{g}}{1-\hat{g}}\right) w_1 w_2 f_1 f_2 \left(1-\hat{g}\right) b db d\epsilon dv_1 dv_2 \ .
\end{aligned}
$$

$N_{\mathrm{Cand}} = nN_{p/cell}\max(\sigma_T \hat{g})\Delta t$;
for $j = 1, ..., N_{Cand}$ **do**
  Pick two particles from the cell;
  Draw $r \in \mathcal{U}[0,1]$;
  if $\sigma_T \hat{g}/\max(\sigma_T \hat{g}) < r$ **then**
    Collide selected particles;
    $W_1, W_2 = W_1 W_2$;
  end
  else
    $W_1 = W_1(1 - W_2\hat{g})/(1-\hat{g})$;
    $W_2 = W_2(1 - W_1\hat{g})/(1-\hat{g})$;
  end
end

$$S(f^{\mathrm{cont.}}) = \frac{C}{2}\int (\delta_1^* + \delta_2^* - \frac{\delta_1}{w_2} - \frac{\delta_2}{w_1}) w_1 w_2 f_1 f_2 \hat{g}\, b db d\epsilon dv_1 dv_2$$
$$+ \frac{C}{2}\int (-\delta_1 - \delta_2 + \frac{\delta_1}{w_2} + \frac{\delta_2}{w_1}) \left(\frac{\hat{g}}{1-\hat{g}}\right) w_1 w_2 f_1 f_2 (1-\hat{g})\, b db d\epsilon dv_1 dv_2 \ .$$

$N_{\mathrm{Cand}} = n N_{p/cell} \max(\sigma_T \hat{g}) \Delta t$;
for $j = 1, ..., N_{Cand}$ **do**
  Pick two particles from the cell;
  Draw $r \in \mathcal{U}[0,1]$;
  if $\sigma_T \hat{g} / \max(\sigma_T \hat{g}) < r$ **then**
    Collide selected particles;
    $W_1, W_2 = W_1 W_2$;
  end
  else
    $W_1 = W_1(1 - W_2\hat{g})/(1-\hat{g})$;
    $W_2 = W_2(1 - W_1\hat{g})/(1-\hat{g})$;
  end
end

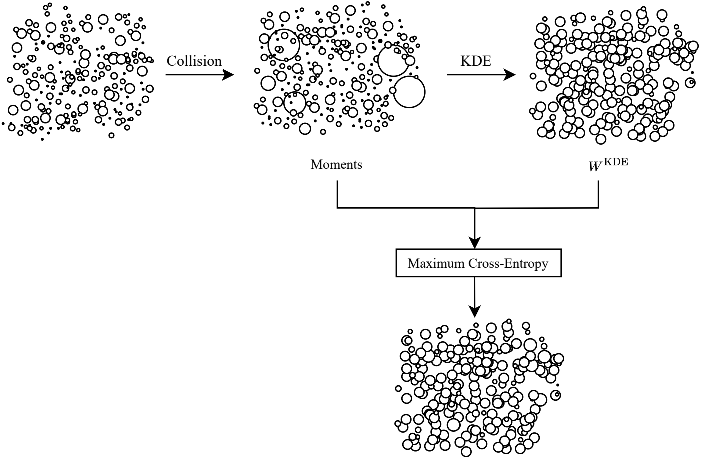
Consistent

Low variance without bias

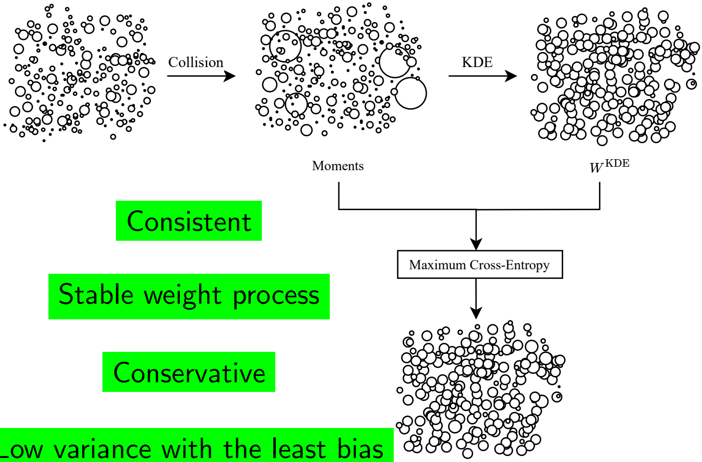
Not-conservative

Unstable weight process

Weight spread, smoothing, and entropic moment conservation.



Weight spread, smoothing, and entropic moment conservation.



Consistent

Stable weight process

Conservative

Low variance with the least bias

Resampling weights with the maximum cross-entropy distribution

$$\mathcal{F}(v|x,t) = \mathcal{F}^{\mathrm{prior}}(v|x,t)\exp\left(\sum_{i=1}^{N_m}\lambda_i(x,t)H_i(v)\right) \tag{3.2}$$

which is the extremum of

$$\mathcal{L}[\mathcal{F}(v|x,t)] := -\int \mathcal{F}(v|x,t)\log\big(\mathcal{F}(v|x,t)/\mathcal{F}^{\mathrm{prior}}(v|x,t)\big)\,dv \tag{3.3}$$

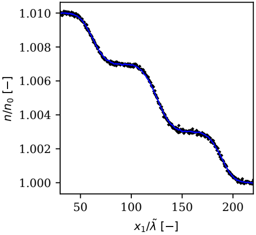
$$+\sum_{i=1}^{N_m}\lambda_i\left(\int H_i(v)\mathcal{F}(v|x,t)dv-\mu_i(x,t)\right)\ . \tag{3.4}$$

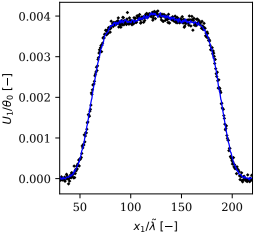
Gains:

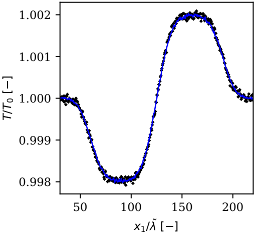
- Convex/concave optimization problem.
- Allows simulation at wide range of rarefaction.

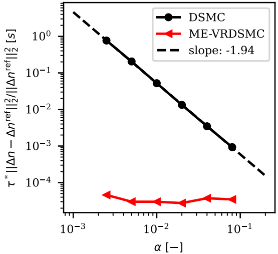
Remaining issue: Ill-conditioned Hessian near the limit of realizability.

Shock tube problem (Boltzmann equation)









2. Sampling particles given moments: relevant for hybrid moment-kinetic models

Moment matching particles

Problem: Given moments corresponding to basis functions H, i.e.

$$\mu = \int f(v)H(v)dv, \tag{4.1}$$

generate samples (particles) of f matching the input moments μ.

Challenges:

- f is not unique.
- Most anstaz such as Grad's $f(v) \approx e^{-v^2/2}(\sum_i \alpha_i H_i(v))$ are not positive preserving.
- How do we match an arbitrary number of moments at the particle level?
- How can we minimize bias?

Revisit Maximum Entropy Distribution

Find a distribution that maximizes entropy besides matching moments

$$\mathcal{L}[f] := -\int f \log(f) dv + \sum_{i=1}^{N_m} \lambda_i \left(\int H_i f dv - \mu_i \right) \ .$$

The variational extremum of $\mathcal{L}$, i.e. $\delta\mathcal{L}/\delta f = 0$, gives

$$f^* = \frac{1}{Z} \exp\left(\lambda \cdot H\right), \qquad \text{where} \ \ Z = \int \exp\left(\lambda \cdot H\right) dv. \tag{4.2}$$

Issue: f^* may **not exist** when H_{N_m} is odd or $\lambda_{N_m} > 0$.
Idea: reformulate optimization problem to enforce existence.

Wasserstein-Penalized Entropy closure

Consider a joint density $\pi(v, w)$ where solution to the closure problem is

$$f(v) \;\; = \;\; \int_{\mathbb{R}^m} \pi(v, w) \; dw \tag{4.3}$$

and other marginal is a known density

$$g(w) \;\; = \;\; \int_{\mathbb{R}^m} \pi(v, w) \; dv \; . \tag{4.4}$$

Then, we find the extremum of the cost functional

$$\mathcal{L}_\alpha(\pi) = \alpha \mathcal{W}(\pi) + \mathcal{S}(\pi) + \sum_i \lambda_i \left(\int H_i(v, w) \pi(v, w) dv dw - \mu_i \right) \tag{4.5}$$

where $\mathcal{S}(\pi)$ enforces least bias

$$\mathcal{S}(\pi) \;\; = \;\; - \int_{\mathbb{R}^m \times \mathbb{R}^m} \Big(\log(\pi(v, w)) - 1 \Big) \; \pi(v, w) \; dv dw \tag{4.6}$$

and $\mathcal{W}$ indicates the transport cost between the two marginals

$$\mathcal{W}(\pi) = \int c(v, w) \; \pi(v, w) \; dv dw \;\; \text{with} \;\; c(v, w) = C_0 |v - w|^p \; . \tag{4.7}$$

WE closure -cont.

By setting $\delta\mathcal{L}_\alpha/\delta f = 0$, we arrive at the Wasserstein-Entropy (WE) joint distribution function

$$\pi(v, w) \quad = \quad \exp\left(\sum_i \lambda_i H_i(v, w) - \alpha C_0 |v - w|^p\right) . \qquad (4.8)$$

Existence is guaranteed by choosing p to be the smallest even number larger that the order of H_{N_m}.
Plug in π back in the loss $\mathcal{L}[.]$, we obtain the dual formulation

$$\lambda = \underset{\lambda^* \in \mathbb{R}^{2n}}{\arg\min} \left\{ \int_{\mathbb{R}^3 \times \mathbb{R}^3} \exp\left(\sum_i \lambda_i^* H_i - \alpha C_0 |v - w|^p\right) \mathrm{dvdw} - \sum_i \lambda_i^* \mu_i \right\}$$

which can be shown to be **convex/concave** optimization problem.

Pros	Cons
✓ Maximizing entropy	✗ Ill-conditioned Hessian L
✓ Unique distr.	✗ High dimensional integrals
✓ Convex/concave optimization problem	
✓ Matching moments	
✓ Guaranteed positivity for f	
✓ Guaranteed existence	

We can write a SDE to sample π, and in return, samples of f and g.

SDE for WE closure

Idea: Instead of Newton-Raphson, let us devise a gradient flow to find λ.

Consider transition of a joint distribution π_t that follows

$$\frac{\partial \pi_t}{\partial t} = \nabla \left[\pi \nabla(\pi_t/\pi)\right] \ . \tag{4.9}$$

From H-Theorem, we know that $\pi_t \to \pi$ monotonically as $t \to \infty$.

The underlying processes for the random variables are

$$\begin{aligned} dV &= \nabla_v[\log(\pi)]\mathrm{d}t + \sqrt{2}\ dB^v \\ dW &= \nabla_w[\log(\pi)]\mathrm{d}t + \sqrt{2}\ dB^w, \end{aligned} \tag{4.10}$$

where B is a six-dimensional Brownian (Wiener) process.

Issue: We still don't know the Lagrange multipliers λ.

SDE for WE closure -cont.

By taking moments of the corresponding Fokker-Planck equation

$$\partial_t \mu^v_{t,l} = \mathbb{E}\left[\left(\tilde{\lambda}^v_k \partial_{v_i} H^v_k(V) - \tilde{\alpha} C_0 p (V_i - W_i)|V - W|^{p-2}\right)\partial_{v_i} H^v_l(V) + \partial^2_{v_i v_i} H^v_l(V)\right]$$

$$\partial_t \mu^w_{t,l} = \mathbb{E}\left[\left(\tilde{\lambda}^w_k \partial_{w_i} H^w_k(W) - \tilde{\alpha} C_0 p (W_i - V_i)|V - W|^{p-2}\right)\partial_{w_i} H^w_l(W) + \partial^2_{w_i w_i} H^w_l(W)\right]$$

and enforcing

$$\partial_t \mu_t \approx \frac{1}{\tau}(\mu - \mu_t) \tag{4.11}$$

we arrive at a linear system to solve for λ, i.e.

$$\tilde{\lambda} = \mathscr{A}^{-1}\mathscr{R}. \tag{4.12}$$

SDE for WE closure -cont.

Given samples of V, W at time t, we compute the Lagrange multipliers by solving

$$\tilde{\lambda} \;=\; \mathscr{A}^{-1}\mathscr{R}. \tag{4.13}$$

where

$$\mathscr{R} = \begin{bmatrix} \mathscr{R}^v \\ \mathscr{R}^w \end{bmatrix} \quad \text{and} \quad \mathscr{A} = \begin{bmatrix} \mathscr{A}^v & 0 \\ 0 & \mathscr{A}^w \end{bmatrix} \tag{4.14}$$

with the right hand side

$$\mathscr{R}_i^v = \tau^{-1}\left(\mu_i^v - \mathbb{E}\left[H_i^v(V)\right]\right) + \mathbb{E}\left[pC_0\tilde{\alpha}\partial_{v_j} H_i^v(V)(V_j - W_j)|V - W|^{p-2} - \partial^2_{v_j v_j} H_i^v(V)\right],$$
$$\mathscr{R}_i^w = \tau^{-1}\left(\mu_i^w - \mathbb{E}\left[H_i^w(W)\right]\right) + \mathbb{E}\left[pC_0\tilde{\alpha}\partial_{w_j} H_i^w(W)(W_j - V_j)|V - W|^{p-2} - \partial^2_{w_j w_j} H_i^w(W)\right]$$

and matrix

$$\mathscr{A}_{ij}^v = \mathbb{E}\left[\partial_{v_k} H_i^v(V)\partial_{v_k} H_j^v(V)\right] \qquad \text{and} \qquad \mathscr{A}_{ij}^w = \mathbb{E}\left[\partial_{w_k} H_i^w(W)\partial_{w_k} H_j^w(W)\right] \; .$$

SDE for WE closure -cont.

- Set Δt and $\tau = a\Delta t$;

while $||\mu_t - \mu||_2/||\mu||_2 > \varepsilon$ **do**

  - Solve the linear system $\tilde{\lambda} = \mathscr{A}^{-1}\mathscr{R}$;
  - Move V and W using Euler–Maruyama solution to $dV = \nabla_v[\log(\pi)]\mathrm{d}t + \sqrt{2}\ dB^v$ and $dW = \nabla_w[\log(\pi)]\mathrm{d}t + \sqrt{2}\ dB^w$;

end

Algorithm 1: Generating samples of $f(v)$ given initial samples of V and W as well as the target moments $\mu = [\mu^v, \mu^w]^T$.

SDE for WE closure -cont.

Pros	Cons
✓ Maximizing entropy	✗ Ill-conditioned Hessian L
✓ Unique distr.	
✓ Convex/concave optimization problem	
✓ Matching moments	
✓ Guaranteed positivity for f	
✓ Guaranteed existence	
✓ Efficient Monte Carlo sampling	

Results: 4 moments system

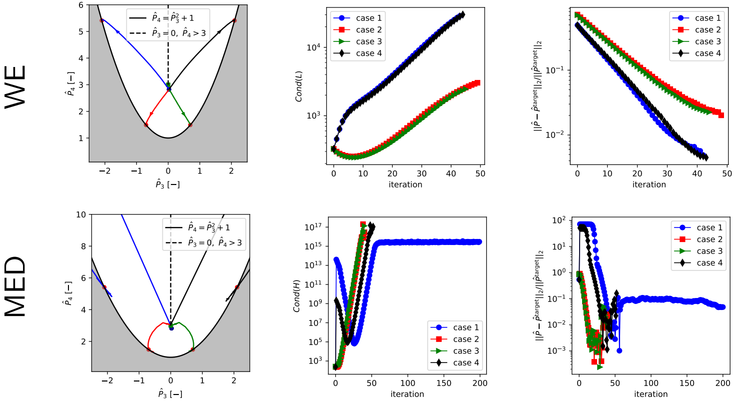


Figure: The convergence path of WE and MED in (μ_3, μ_4) plane (left), the evolution of condition number (middle) and relative error in moments (right) for moment problem near the limit of realizability.

Results: Resampling particles

- Consider simulation of rarefied gas between moving walls.
- Given the moments at time t_0, we re-initialize particles and resume the simulation.
- In each computational cell, we carry out WE algorithm matching moments of the simulation results.
- And compare the evolution of the solution $t > t_0$ with the unperturbed simulation.

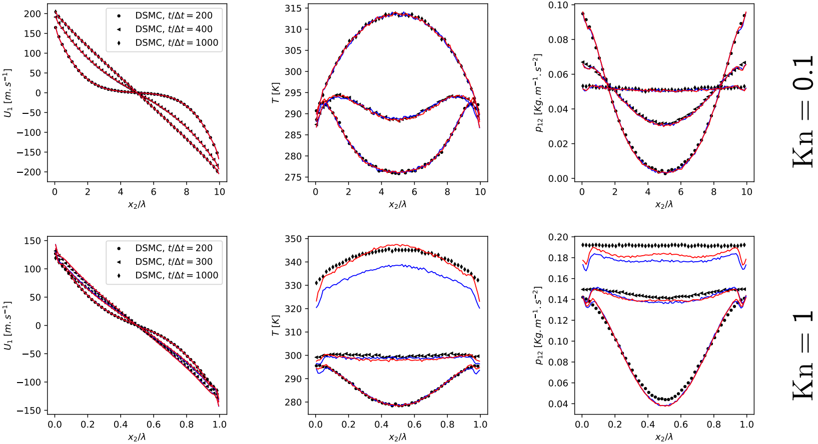


Figure: Transient solution of the Boltzmann eq. for Couette flow at $\mathrm{Ma} = 1$. Here we compare solution from the standard DSMC (black) and DSMC with resampling every 100 steps using the WE closure matching up to heat flux (blue) and up to 4th order moments (red).

Thanks for your attention!

Results: Convergence with number of moments

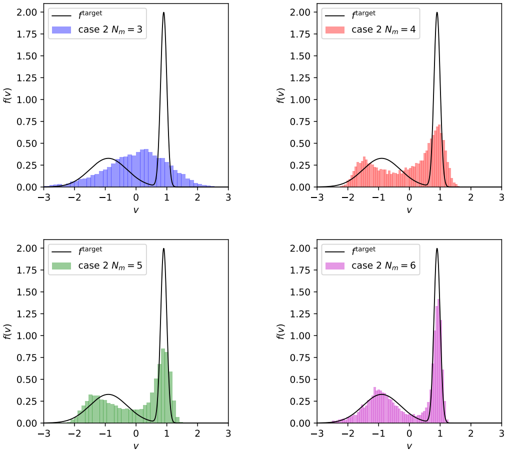


Figure: Estimating a bi-modal distribution with WE by matching $N_m = 3, 4, 5$ and 6 order moments.

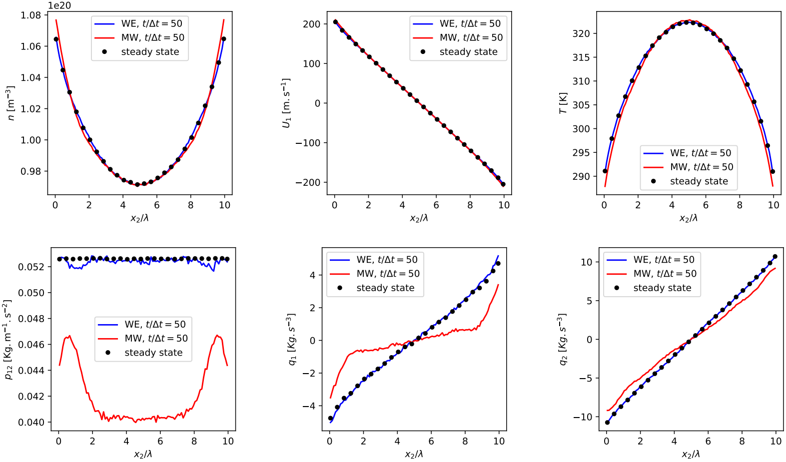
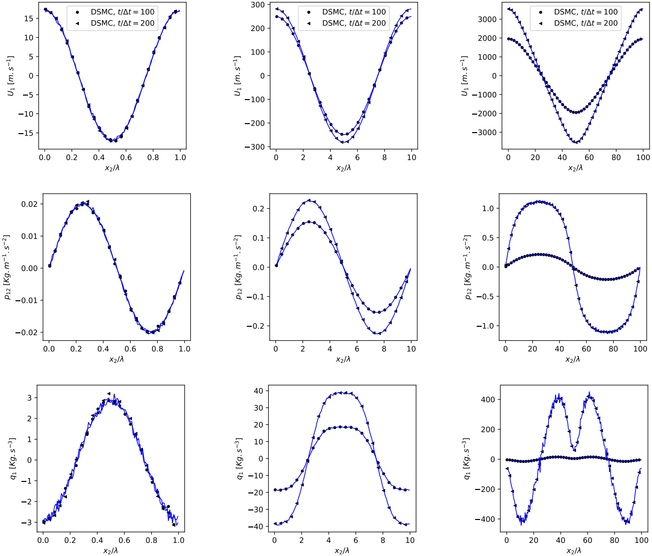


Figure: Resuming steady-state solution of Couette flow (black circles), using the local Maxwellian (red) and the WE matching up to heat flux is shown (blue).



$\mathrm{Kn} = 1$ $\mathrm{Kn} = 0.1$ $\mathrm{Kn} = 0.01$

Figure: Solution of the Boltzmann eq. in the presence of an external force $F_1 = A\cos(\alpha t)\cos(2\beta\pi x_2/L)$.