Wasserstein-penalized Entropy closure: A use case for stochastic particle methods

Mohsen Sadr[1], Nicolas G. Hadjiconstantinou[2], and M. Hossein Gorji[3]

[1]Paul Scherrer Institute, Forschungsstrasse 111, CH-5232 Villigen, Switzerland
[2]Department of Mechanical Engineering, MIT, Cambridge, MA 02139, USA
[3]Laboratory for Computational Engineering, Empa, Dübendorf, Switzerland

Moment Problem

Given N_m moments corresponding to the vector of basis functions $\boldsymbol{H}$, i.e.

$$\boldsymbol{\mu} = \int f(\boldsymbol{x})\boldsymbol{H}(\boldsymbol{x})d\boldsymbol{x} \tag{1}$$

find the underlying distribution function $f(\boldsymbol{x})$.

Challenge: Solution may not exist and is not unique.

Maximum Entropy Closure

Among all possible solutions to the moment problem, the least bias one can be found by minimizing the Shannon entropy. Consider the cost functional

$$C[\mathcal{F}(\boldsymbol{x})] := \int \mathcal{F}(\boldsymbol{x})\log(\mathcal{F}(\boldsymbol{x}))d\boldsymbol{x} + \sum_{i=1}^{N_m}\lambda_i\left(\int H_i(\boldsymbol{x})\mathcal{F}(\boldsymbol{x})d\boldsymbol{x} - \mu_i(\boldsymbol{x})\right) . \tag{2}$$

The extremum of this functional gives the maximum entropy density function

$$\hat{f}(\boldsymbol{x}) = \frac{1}{Z}\exp\left(\boldsymbol{\lambda}\cdot\boldsymbol{H}(\boldsymbol{x})\right), \qquad \text{where} \;\; Z = \int \exp\left(\boldsymbol{\lambda}\cdot\boldsymbol{H}(\boldsymbol{x})\right)d\boldsymbol{x}. \tag{3}$$

The Lagrange multipliers λ_i, $i = 1...N_m$, may be found using the unconstrained dual formulation $D(\boldsymbol{\lambda})$ with the gradient $\boldsymbol{g} = \nabla D(\boldsymbol{\lambda})$ and Hessian $\boldsymbol{H}(\boldsymbol{\lambda}) = \nabla^2 D(\boldsymbol{\lambda})$ leading to an iterative scheme

$$\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} - \boldsymbol{L}^{-1}(\boldsymbol{\lambda})\boldsymbol{g}(\boldsymbol{\lambda}). \tag{4}$$

Pros	Cons
✓ Least bias	✗ Ill-conditioned Hessian $\boldsymbol{L}$
✓ Convex optimization problem	✗ Requiring an accurate numerical integration method
✓ Matching moments	✗ Cannot guarantee existence in the limit of realizability.

Wasserstein-penalized Entropy closure

Instead of directly inferring f, the idea is to infer a joint probability density $\pi(\boldsymbol{v}, \boldsymbol{w})$ on $\mathbb{R}^{2m}$ such that its marginal

$$f(\boldsymbol{v}) = \int_{\mathbb{R}^m} \pi(\boldsymbol{v}, \boldsymbol{w})\; d\boldsymbol{w} \tag{5}$$

gives a solution to the closure problem (and hence an approximation of $\bar{f}$), whereas the other marginal

$$g(\boldsymbol{w}) = \int_{\mathbb{R}^m} \pi(\boldsymbol{v}, \boldsymbol{w})\; d\boldsymbol{v} \tag{6}$$

is linked to a known density $\bar{g}$, which serves as a mean of introducing some prior knowledge (e.g. a nearby equilibrium state, a prior approximation, etc).

In order to find the optimum solution, consider the cost functional

$$\mathcal{L}_\alpha(\pi) = \alpha\mathcal{W}(\pi) + \mathcal{H}(\pi) \qquad \text{for} \qquad \alpha > 0\;, \tag{7}$$

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

$$\mathcal{H}(\pi) = \int_{\mathbb{R}^m\times\mathbb{R}^m}\Big(\log(\pi(\boldsymbol{v},\boldsymbol{w})) - 1\Big)\;\pi(\boldsymbol{v},\boldsymbol{w})\;d\boldsymbol{v}d\boldsymbol{w}. \tag{8}$$

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

$$\mathcal{W}(\pi) = \int_{\mathbb{R}^m\times\mathbb{R}^m} c(\boldsymbol{v},\boldsymbol{w})\;\pi(\boldsymbol{v},\boldsymbol{w})\;d\boldsymbol{v}d\boldsymbol{w}\;. \tag{9}$$

where $c(\boldsymbol{v},\boldsymbol{w}) = C_0|\boldsymbol{v}-\boldsymbol{w}|^p$.

By setting the variational derivatives to zero, we arrive at the Wasserstein-Entropy joint distribution function

$$\pi(\boldsymbol{v},\boldsymbol{w}) = \exp\left(\sum_i \lambda_i H_i(\boldsymbol{v},\boldsymbol{w}) - \alpha C_0|\boldsymbol{v}-\boldsymbol{w}|^p\right). \tag{10}$$

Challenge: High dimensional integrals need to be take to find the closure.

Finding Lagrange multipliers via Gradient flow

Consider transition of the joint distribution π_t with Gradient flow that reaches π as $t \to \infty$

$$\begin{aligned}\frac{\partial \pi_t}{\partial t} &= \nabla\left[\pi\nabla(\pi_t/\pi)\right]\\ &= -\nabla\cdot\Big[\nabla\left[\log(\pi)\right]\pi_t\Big] + \nabla\Big[\pi_t\Big],\end{aligned} \tag{11}$$

which describes a process for sampling π. The Lagrange multipliers can be found by taking moments of Fokker-Planck equation and enforcing

$$\partial_t\tilde{P} = \frac{1}{\tau}\left(P - \tilde{P}\right), \tag{12}$$

which leads to a linear system of equation to be solved for the Lagrange multipliers

$$\tilde{\boldsymbol{\lambda}} = \boldsymbol{A}^{-1}\boldsymbol{R}. \tag{13}$$

Pros	Cons
✓ Least bias	✗ Cannot guarantee low condition number
✓ Monotone convergence	—
✓ Matching moments	—
✓ Generating samples	—

Limit of realizability

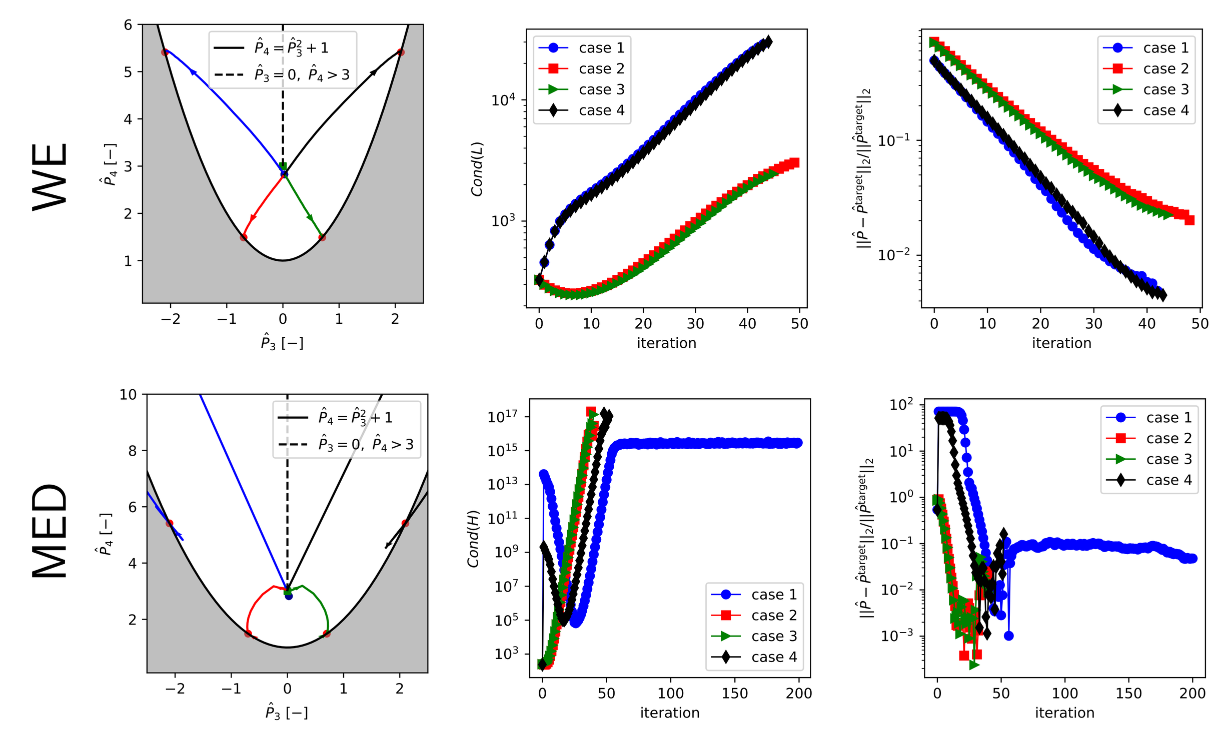

Figure 1: WE convergence path in $(\hat{P}_3, \hat{P}_4)$ plane (left), the evolution of condition number (middle) and relative error in moments (right) for four target points.

Convergence with number of moments

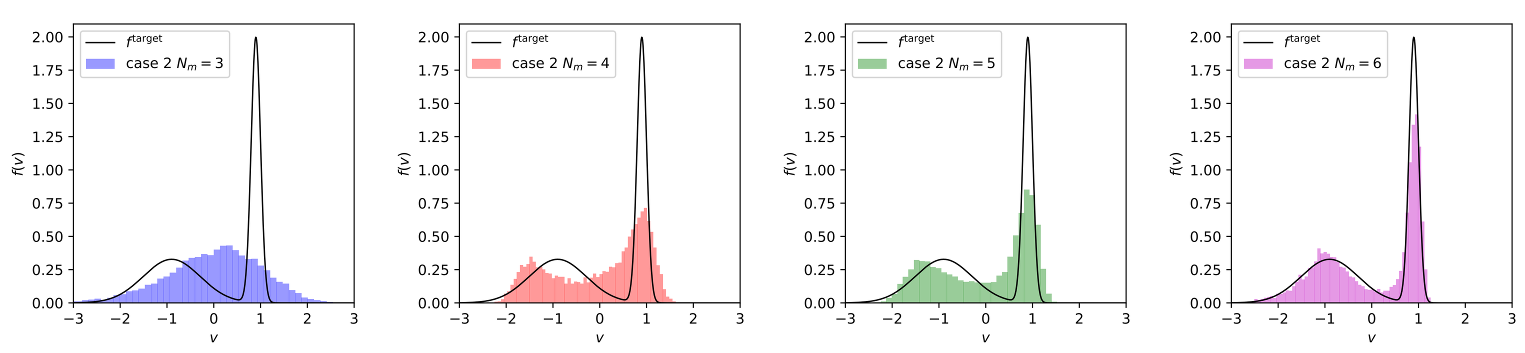

Figure 2: Estimating bi-modal distribution with WE by matching $N_m = 3, 4, 5$ and 6 order moments.

Resampling particles for DSMC simulation

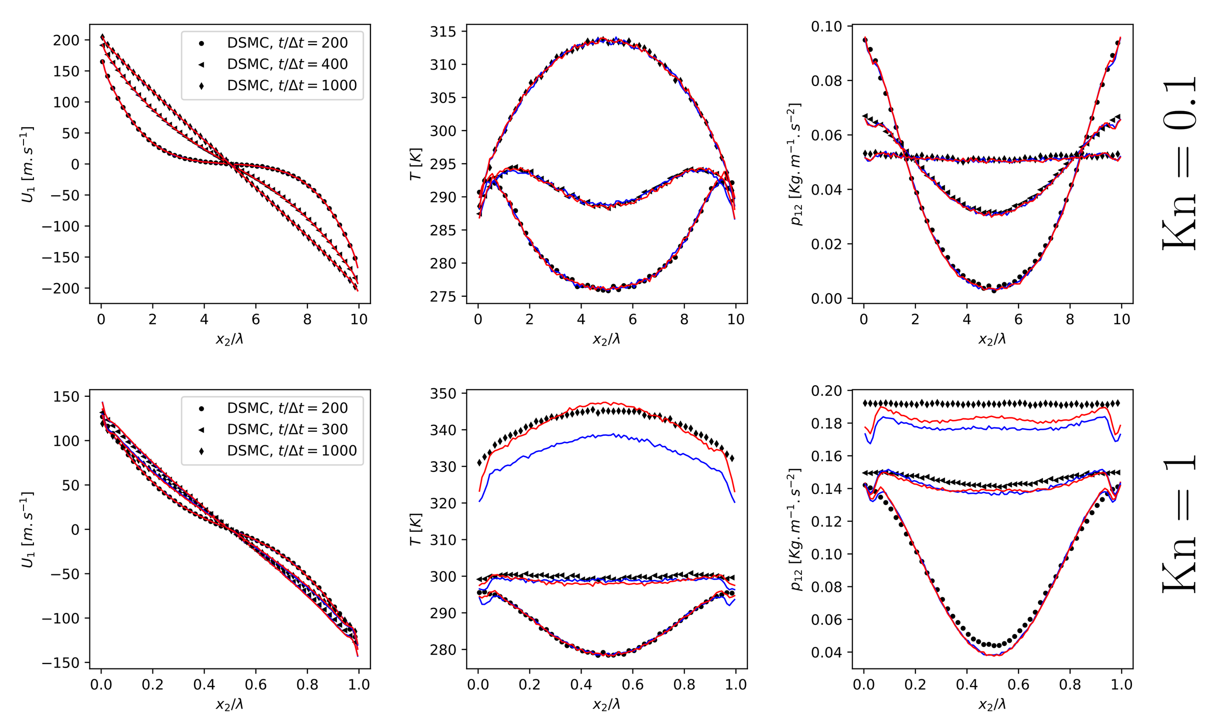

Figure 3: Transient solution for the bulk velocity, temperature, shear stress, and heat flux in a $\mathrm{Ma} = 1$ Couette flow. Comparison between standard DSMC (black) and DSMC with resampling every 100 steps using the WE closure matching up to heat flux (blue) and up to 4th order moment (red).

Preprint at arXiv:2308.02607.