

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

Mohsen Sadr¹, Nicolas G. Hadjiconstantinou², and M. Hossein Gorji³

¹Paul Scherrer Institute, Forschungsstrasse 111, CH-5232 Villigen, Switzerland

²Department of Mechanical Engineering, MIT, Cambridge, MA 02139, USA

³Laboratory for Computational Engineering, Empa, Dübendorf, Switzerland

Moment Problem

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

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

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

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

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

The extremum of this functional gives the maximum entropy density function

$$\hat{f}(\mathbf{x}) = \frac{1}{Z} \exp(\boldsymbol{\lambda} \cdot \mathbf{H}(\mathbf{x})), \quad \text{where } Z = \int \exp(\boldsymbol{\lambda} \cdot \mathbf{H}(\mathbf{x})) d\mathbf{x}. \quad (3)$$

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

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

Pros

- ✓ Least bias
- ✓ Convex optimization problem
- ✓ Matching moments

Cons

- ✗ Ill-conditioned Hessian \mathbf{L}
- ✗ Requiring an accurate numerical integration method
- ✗ 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 $\pi(v, w)$ on \mathbb{R}^{2m} such that its marginal

$$f(v) = \int_{\mathbb{R}^m} \pi(v, w) dw \quad (5)$$

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

$$g(w) = \int_{\mathbb{R}^m} \pi(v, w) dv \quad (6)$$

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

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

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

$$\mathcal{H}(\pi) = \int_{\mathbb{R}^m \times \mathbb{R}^m} \left(\log(\pi(v, w)) - 1 \right) \pi(v, w) dv dw. \quad (8)$$

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

$$\mathcal{W}(\pi) = \int_{\mathbb{R}^m \times \mathbb{R}^m} c(v, w) \pi(v, w) dv dw. \quad (9)$$

where $c(v, w) = C_0 |v - w|^p$.

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

$$\pi(v, w) = \exp \left(\lambda_i H_i(v, w) - \alpha C_0 |v - w|^p \right). \quad (10)$$

Challenge: High dimensional integrals need to be taken 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 $t \rightarrow \infty$

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

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

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

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

$$\tilde{\lambda}(x, t) = \mathcal{A}^{-1}(x, t) \mathcal{B}(x, t). \quad (13)$$

Pros

- ✓ Least bias
- ✓ Monotone convergence
- ✓ Matching moments
- ✓ Generating samples

Cons

- ✗ Cannot guarantee low condition number

Limit of realizability

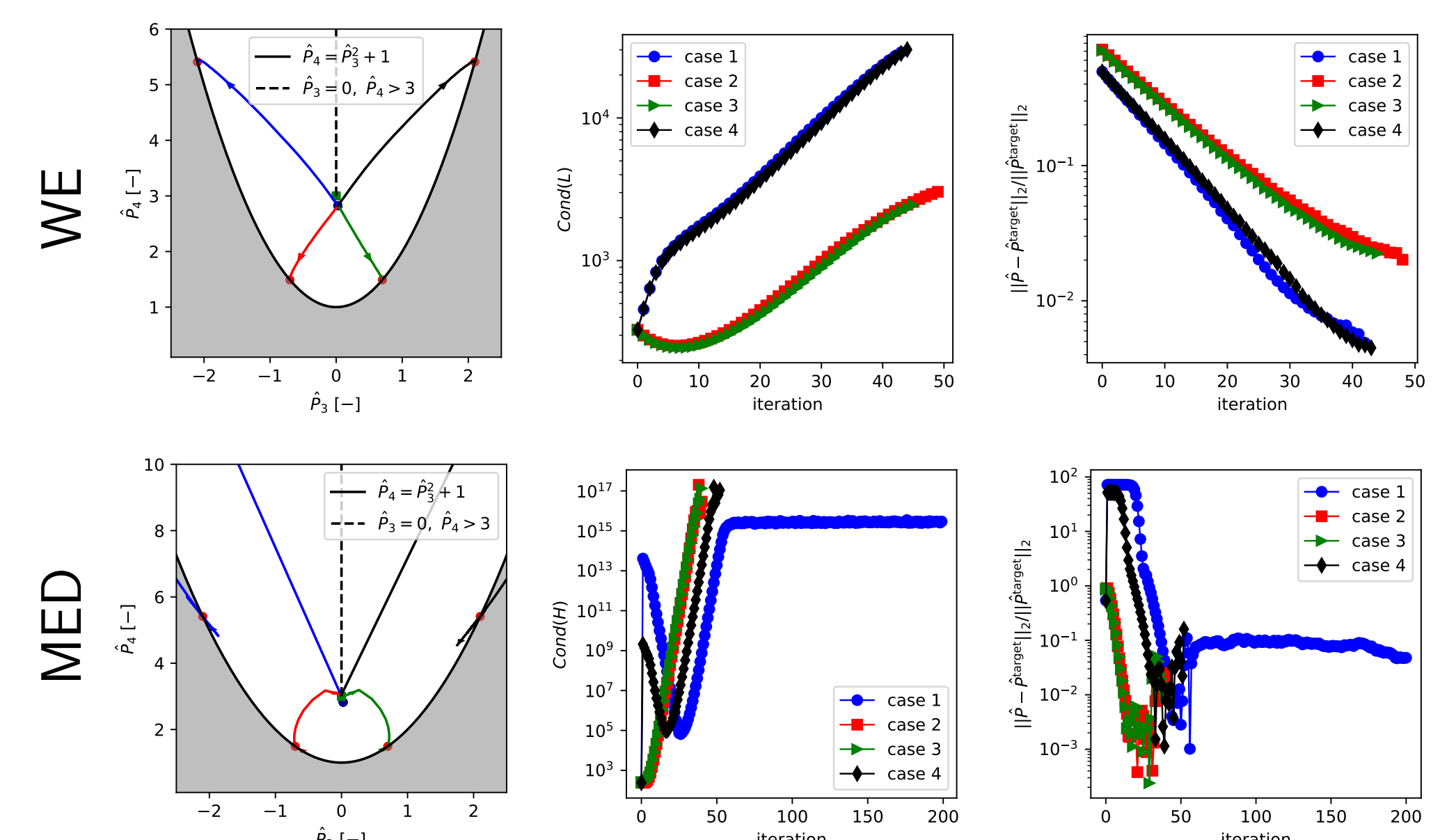


Figure 1: WE convergence path in (\hat{P}_3, \hat{P}_4) plane (left), the evolution of condition number (middle) and relative error in moments (right) for four target points.

Convergence with number of moments

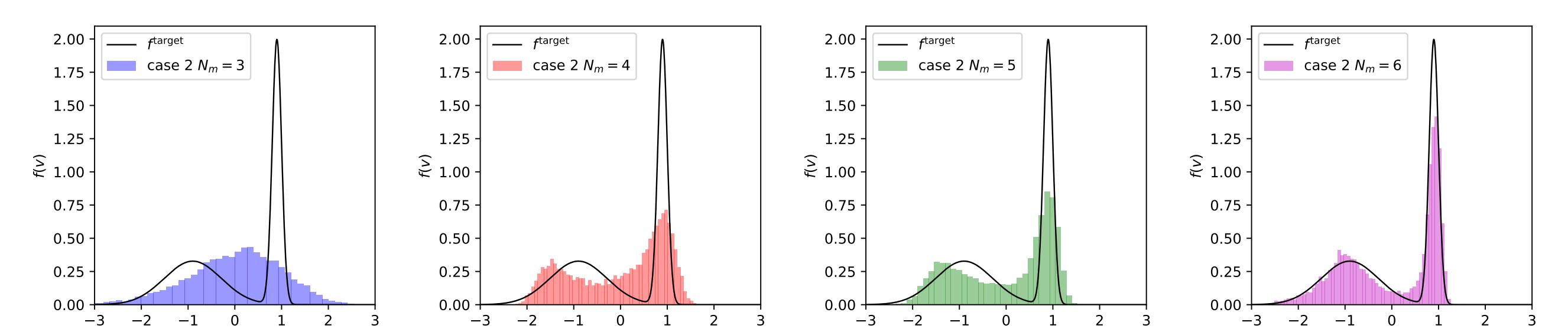


Figure 2: Estimating bi-modal distribution with WE by matching $N_m = 3, 4, 5$ and 6 order moments.

Resampling particles for DSMC simulation

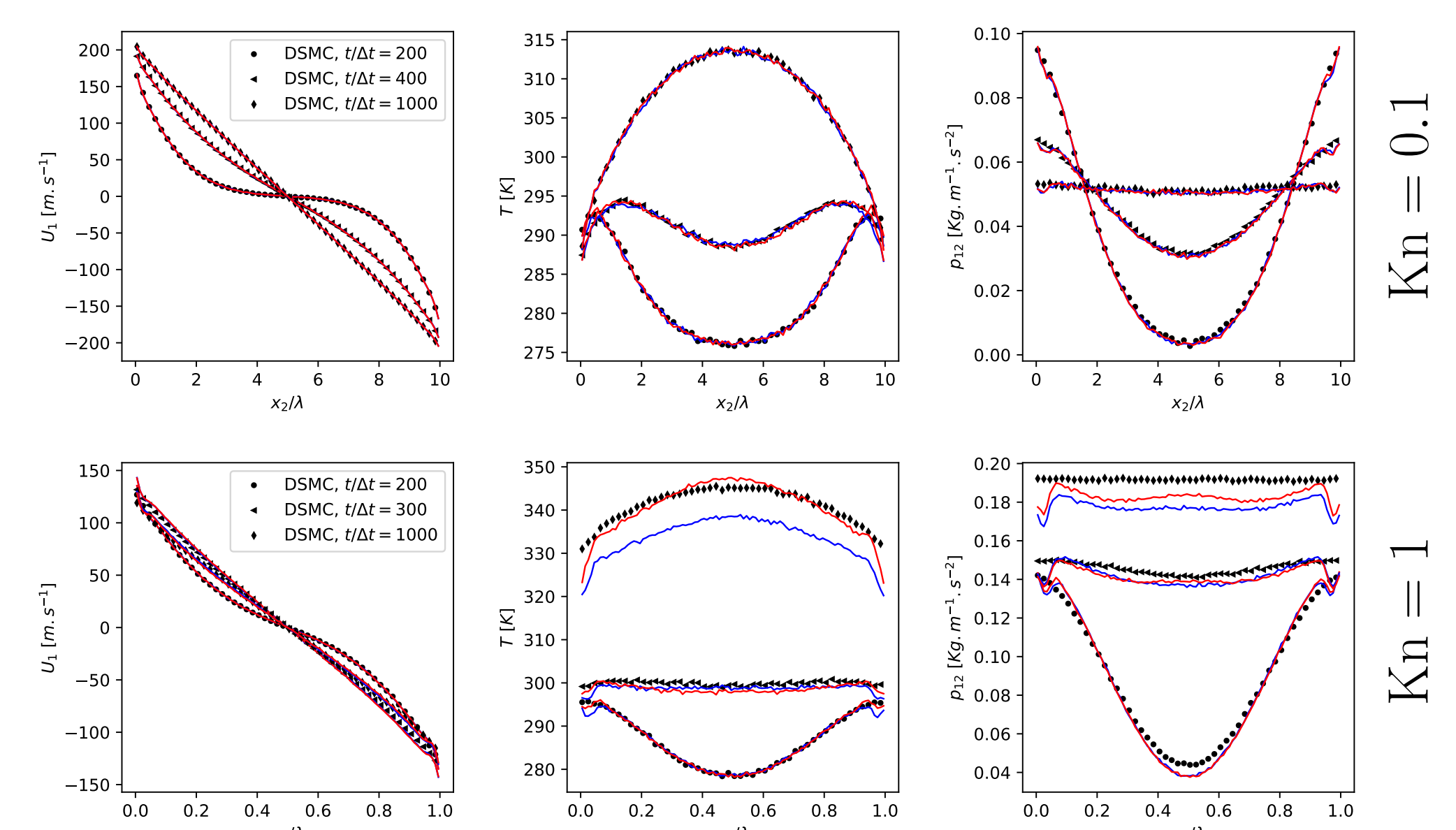


Figure 3: Transient solution for the bulk velocity, temperature, shear stress, and heat flux in a $Ma = 1$ Couette flow. Comparison between standard DSMC (black) and DSMC with resampling every 100 steps using the WE closure matching up to heat flux (blue) and up to 4th order moment (red).