Computing Partition Functions in Unnormalized Density Models Using Bayesian Thermodynamic Integration

Alexander Lobashev[1]
Alexander.Lobashev@skoltech.ru

Mikhail Tamm[2]
thumm.m@gmail.com

[1]Skolkovo Institute of Science and Technology, Moscow, Russia
[2]ERA Chair for Cultural Data Analytics, School of Digital Technologies, Tallinn University, Tallinn, Estonia

Abstract

We introduce a general method for computing partition functions in unnormalized density models. This approach also allows for the reconstruction of thermodynamic functions and phase boundaries in two-parameter statistical mechanics systems, eliminating the need for knowledge of the system's energy. Our method is based on expressing the Fisher metric in terms of the posterior distributions over a space of external parameters and approximating the metric field by a Hessian of a convex function. We use the proposed approach to reconstruct the partition functions and phase diagrams of the Ising model and the exactly solvable non-equilibrium TASEP without any a priori knowledge about microscopic rules of the models, just assuming it is possible to sample from equilibrium (respectively, steady state) distribution for a given set of external parameters.

1 Introduction

Machine learning methods have been actively used for the study of models of classical and quantum statistical physics (1; 2; 3; 4; 5), mostly focusing on localization boundaries between phases and phase transition locations and extracting the learned order parameters. Every equilibrium statistical mechanics model is a probabilistic mapping from a set of macroscopic external parameters $\mathbf{t}$ to a (very multidimensional) microscopic state s and can be thought about as a conditional probability distribution $\mathbb{P}(s|\mathbf{t})$. If this distribution belongs to the exponential family the problem of determining "order parameters" is equivalent to the search of the so-called sufficient statistics (6; 7).

Existing methods for approximating the partition function or its logarithm, also known as free energy, such as importance sampling (8), annealed importance sampling (9), and variational autoregressive networks (10), require explicit knowledge of the Hamiltonian (energy or unnormalized density) of the system. However, it is easy to imagine situations when one can sample from the distribution of microstate of a system, whose Hamiltonian is either unknown or, as is often the case in the steady-states of nonequilibrium models like Totally Asymmetric Simple Exclusion Process (TASEP), does not exist. For such systems, it is possible to apply neural methods to extract sufficient statistics (6; 7). However, order parameters learned in such a way may not be easy to interpret.

In this paper, we address a broader issue of reconstructing the free energy and thermodynamic functions of a given statistical physics model without explicit knowledge of its Hamiltonian. The

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

solution should comply with conventional thermodynamic relations and subsequently lead to better interpretability.

2 Bayesian Thermodynamic Integration

2.1 Fisher metric and posterior distribution

Given a set of external macroscopic parameters $\mathbf{t}$ the probability distribution over microstates $\mathbb{P}(s|\mathbf{t})$ in equilibrium statistical mechanics is known to be given by the Gibbs distribution $\mathbb{P}(s|\mathbf{t}) = Z^{-1}(\mathbf{t})\exp[H(s|\mathbf{t})]$ where $Z(\mathbf{t}) = \sum_s \exp[H(s|\mathbf{t})]$ is the normalization constant known as the partition function of the model, and the function $H(s|\mathbf{t})$ typically has a clear microscopic meaning and is referred to as the Hamiltonian of the model. The main problem of interest to statistical physics is the study of phase transitions. They correspond to sharp changes in typical microstates with a gradual change in external parameters, i.e., sharp changes of the probability distribution $\mathbb{P}(s|\mathbf{t}))$ with the change of parameter $\mathbf{t}$. It is conventional to quantify the rate of change of a probability distribution with Fisher information metric $G(\mathbf{t})$ defined as

$$G(\mathbf{t}) = \int \mathbb{P}(s|\mathbf{t})\nabla_{\mathbf{t}}\log\mathbb{P}(s|\mathbf{t})(\nabla_{\mathbf{t}}\log\mathbb{P}(s|\mathbf{t}))^T ds \tag{1}$$

Phase transitions can be thought of as points at which, in the limit of a large system size, the Fisher metric becomes singular. For instance, in the case of a 2D Ising model with only a single temperature parameter, the solitary component of the Fisher metric is the magnetic susceptibility, which diverges near the critical temperature.

Here we presume that, despite not knowing the distribution $\mathbb{P}(s|\mathbf{t})$ explicitly, we can sample from it through a given simulation algorithm. Moreover, the labels $\mathbf{t}$ are themselves chosen randomly from some distribution $P(\mathbf{t})$ (in the examples of Section 3 we typically choose $P(\mathbf{t})$ to be a flat distribution over some compact domain in the space of parameters). As a result, we are sampling pairs

$$(s_i, \mathbf{t}_i) \sim \mathbb{P}(s|\mathbf{t})P(\mathbf{t}), \tag{2}$$

where $\sim$ sign here and below designates sampling from the corresponding distribution. Define now, according to the Bayes formula, the posterior distribution $\Pi(\mathbf{t}|s)$ on the space of external parameters given an observed microstate

$$\Pi(\mathbf{t}|s) = \frac{\mathbb{P}(s|\mathbf{t})P(\mathbf{t})}{P(s)} \tag{3}$$

where $P(s) = \sum_{\mathbf{t}'}\mathbb{P}(s|\mathbf{t}')P(\mathbf{t}')$ is the probability of observing microstate s in the sampling procedure (2). This posterior distribution, $\Pi(\mathbf{t}|s)$, functions as a "probabilistic thermometer": it examines a microstate s and provides a probabilistic estimate of the external parameters (such as temperature) under which the microstate was generated.

Taking the derivative of the logarithm of (3) gives $\nabla_{\mathbf{t}}\log\Pi(\mathbf{t}|s) = \nabla_{\mathbf{t}}\log\mathbb{P}(s|\mathbf{t}) + \nabla_{\mathbf{t}}\log P(\mathbf{t})$, i.e., the derivative of $\mathbb{P}(s|\mathbf{t})$ in the definition of the Fisher metric can be expressed in terms of the derivative of the posterior distribution and the known derivative of $P(\mathbf{t})$ (in the simplest case when $P(\mathbf{t})$ is a constant in some compact domain, $\nabla_{\mathbf{t}}\log P(\mathbf{t}) = 0$ except for the domain boundary). Thus, if we manage to approximate the posterior distribution $\Pi(\mathbf{t}|s)$ the Fisher information metric can be estimated as

$$G(\mathbf{t}) \approx \frac{1}{N}\sum_{k=1}^{N}\nabla_{\mathbf{t}}\log\Pi(\mathbf{t}|s_k)(\nabla_{\mathbf{t}}\log\Pi(\mathbf{t}|s_k))^T, \text{ where } s_k \sim \mathbb{P}(s|\mathbf{t}),\ k = 1,\dots,N. \tag{4}$$

2.2 Approximation of the Posterior

Our task now is to estimate the posterior distribution $\Pi(\mathbf{t}|s)$ using the set of samples $\mathbf{t}_k, s_k$. To do that, we are trying to approximate $\Pi(\mathbf{t}|s)$ by representatives from some parametric family of distributions $\Pi_\theta(\mathbf{x}|y)$, choosing θ to maximize some cost function (here as Π_θ we use an image-to-image network with U^2-Net architecture (11), which transforms a 2d microstate into a 2d probability distribution on a discrete uniform grid). It is conventional to choose the cost function to maximize the likelihood of the true external parameters $\mathbf{t}_i$ given s_i over all samples in the set. The minimization of the negative

log-likelihood is equivalent to minimization of the KL divergence between the target distribution $\mathbb{P}_{\text{target}}(\mathbf{t}|\mathbf{t}')$ and the predicted distribution $\mathbb{P}_\theta(\mathbf{t}|\mathbf{t}') = \int \Pi_\theta(\mathbf{t}|s)\mathbb{P}(s|\mathbf{t}')ds$:

$$\mathcal{L}_1(\theta) = \mathbb{E}_{\mathbf{t}'\sim P(\mathbf{t})}\text{KL}\left(\mathbb{P}_{\text{target}}(\mathbf{t}|\mathbf{t}')||\mathbb{P}_\theta(\mathbf{t}|\mathbf{t}')\right), \quad \theta_{\text{optimal}} = \underset{\theta}{\text{argmin}}\, \mathcal{L}_1(\theta) \tag{5}$$

As a target distribution it seems natural to choose $\mathbb{P}_{\text{target}}(\mathbf{t}|\mathbf{t}') = \delta(\mathbf{t}-\mathbf{t}')$ (i.e., the reconstructed labels $\mathbf{t}$ are identical to the input labels $\mathbf{t}'$). However, in practice, to avoid singularities it is convenient to replace this "hard label" target distribution with a smoothened distribution $\mathbb{P}_{\text{target}}(\mathbf{t}|\mathbf{t}') = C \cdot \exp\left(-\sigma^{-2}||\mathbf{t}'-\mathbf{t}||^2/2\right)$, where σ is a smoothening parameter and C is the normalizing constant. In what follows, to avoid overfitting we use tuples of microstates with macroparameters close to a given $\mathbf{t}'$ as a learning sample; in this case it is natural to choose σ of the same order as the uncertainty of $\mathbf{t}'$ within the tuple.

The distribution $\mathbb{P}_\theta(\mathbf{t}|\mathbf{t}')$ obtained as a result of the minimization, can in principle be used as input of formula (4) for the approximation of the Fisher metric. However, its derivative-dependent nature makes it noisy, so to make the approach workable we further approximate $\mathbb{P}_\theta(\mathbf{t}|\mathbf{t}')$ with a distribution from an exponential family $\mathbb{P}_w(s|\mathbf{t}) = \exp[f(s)^T\mathbf{t}]Z_w^{-1}(\mathbf{t})$, parameterized by an Input Convex Neural Network (ICNN) (12) (w here is a set of parameters of the ICNN). If our prior is uniform, the posterior distribution in the case of exponential family takes the form

$$\Pi_w(\mathbf{t}|s) = \frac{\mathbb{P}_w(s|\mathbf{t})\mathbb{P}(\mathbf{t})}{\int \mathbb{P}_w(s|\mathbf{t}')\mathbb{P}(\mathbf{t}')d\mathbf{t}'} = \frac{\exp[f(s)^T\mathbf{t} - \log Z_w(\mathbf{t})]}{\int \exp[f(s)^T\mathbf{t}' - \log Z_w(\mathbf{t}')]d\mathbf{t}'} \tag{6}$$

Notably, when reconstructing $Z_w(\mathbf{t})$ we do not know the functions $f(s)$. We know, however, that their averages at a given **t'** are equal to the derivatives of $\log Z_w$ at this $\mathbf{t}'$:

$$f(\mathbf{t}') \approx \frac{1}{K}\sum_{k=1}^{K} f(s_k) \approx \nabla_{\mathbf{t}'}\log Z(\mathbf{t}'),\ s_k \sim \mathbb{P}(s|\mathbf{t}) \tag{7}$$

This allows us to define the posterior distribution averaged over a set of microstates generated with the same macroscopic parameters $\mathbf{t}'$

$$\mathbb{P}_w(\mathbf{t}|\mathbf{t}') = \frac{\exp\left[\mathbf{t}^T(\nabla_{\mathbf{t}'}\log Z_w(\mathbf{t}')) - \log Z_w(\mathbf{t})\right]}{\int \exp\left[\mathbf{t}''^T(\nabla_{\mathbf{t}'}\log Z_w(\mathbf{t}')) - \log Z_w(\mathbf{t}'')\right]d\mathbf{t}''} \tag{8}$$

To find parameters w we minimize Jensen-Shannon divergence between posterior $\mathbb{P}_\theta(\mathbf{t}|\mathbf{t}')$ predicted using maximum likelihood as described above and posterior $\mathbb{P}_w(\mathbf{t}|\mathbf{t}')$):

$$\mathcal{L}_2(w) = \mathbb{E}_{\mathbf{t}'\sim\mathbb{P}(\mathbf{t})}\text{JSD}\left(\mathbb{P}_\theta(\mathbf{t}|\mathbf{t}'), \mathbb{P}_w(\mathbf{t}|\mathbf{t}')\right), \quad w_{\text{optimal}} = \underset{w}{\text{argmin}}\, \mathcal{L}_2(w) \tag{9}$$

The resulting approximation for the Fisher metric is $G(\mathbf{t}) = \nabla_{\mathbf{tt}}\log Z_{w_{\text{optimal}}}(\mathbf{t})$.

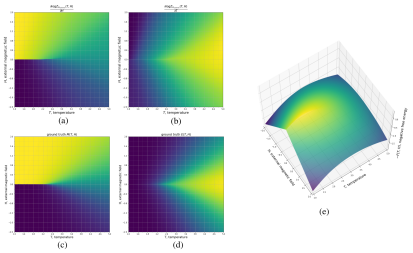
3 Numerical experiments

3.1 Ising model

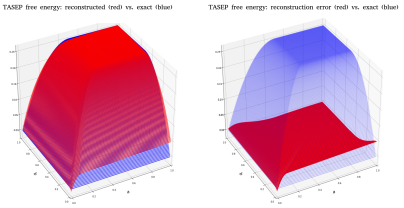
Here we use our approach to reconstruct thermodynamic functions for the 2D Ising model (13), which is an equilibrium statistical mechanics model with exponential distribution over the microstates.

We use image-to-image network with U^2-Net architecture (11) to approximate the posterior $\Pi_\theta(\mathbf{t}|\mathbf{t'})$ in the nodes of a 2d uniform grid $\mathcal{D}$. As an input, the network takes a tuple of K_{bundle} images (concatenated across channel dimension), representing microstates equilibrated at values of external parameters in the viticinty of **t'**. Tuples are used to artificially augment the training data and to reduce over-fitting, separate K_{bundle}-nearest neighbour graphs are computed to construct input tuples for both the train and test datasets. Output consists of the posterior probability densities $\Pi_\theta(\mathbf{t}|\mathbf{t'})$ in discrete grid points. For all experiments the training was performed on a single Nvidia-HGX compute node with 8 A100 GPUs. We trained U^2-Net using Adam optimizer with learning rate 0.00001 and batch size of 2048 for $N_{\text{U2Net steps}} = 20000$ gradient update steps. In all experiments the training set consists of 80% of samples and the other 20% are used for testing.

Moreover, given the posterior distribution, we use ICNN (12) architecture, which has 5 layers with hidden dimension 512, to approximate the free energy. We train it using Adam optimizer with learning rate 0.001. Reconstructed free energy is shown in Figure 1 A. The model was able to correctly determine 1st order phase transition line $T < T_{\text{cr}}$ and $H = 0$.



A:



B:

Figure 1: The results of the free energy reconstruction for (A) the 2D Ising and (B) the TASEP models. (A). (top row) Partial derivatives of the reconstructed free energy with respect to temperature (left) and magnetic field (right), (bottom row) energy $E(H,T) = \sum_{\langle i,j \rangle} s_i(H,T)s_j(H,T)$ (left) and magnetization $M(H,T) = \sum_i s_i$ (right) of the Ising model, (right) reconstructed free energy. (B) (left): reconstructed (red) "free energy" F_{TASEP} compared to the exact solution (10) (blue), (right): reconstruction error (red) vs. exact solution.

3.2 Totally asymmetric simple exclusion process

Totally asymmetric simple exclusion process (TASEP) is a simple model of transport phenomena, where discrete particles jump to the right on a 1d lattice with rate dt provided that exclusion condition (no more than 1 particle per lattice cite) is satisfied. At leftmost site a particle is added with probability αdt per time dt provided that it is empty and at rightmost site particle is removed with probability βdt per time dt provided that it is occupied. This model is exactly solvable, it is known to achieve a steady state with a rich behavior depending on (α, β) (14; 15; 16): three distinct phases - the low-density phase, the high-density phase and the maximal current phase are possible, and the asymptotic "free energy" which is defined as $F_{\text{TASEP}}(\alpha,\beta) = \lim_{M\to\infty} M^{-1} \log Z(\alpha,\beta)$ and coincides with average flow per unit time, equals

$$F_{\text{TASEP}}(\alpha,\beta) = \begin{cases} \frac{1}{4},\ \alpha > \frac{1}{2},\ \beta > \frac{1}{2}; \\ \alpha(1-\alpha),\ \alpha < \beta,\ \alpha < \frac{1}{2}; \\ \beta(1-\beta),\ \beta < \alpha,\ \beta < \frac{1}{2}. \end{cases} \tag{10}$$

where the first, second and third cases correspond to maximal current, high density and low density cases, respectively.

The setups for posterior approximation and free energy approximation are similar to those in the Ising model. The comparison between the reconstructed free energy and the exact analytical solution is shown in Figure 1 B.

4 Discussion

We propose a new approach to reconstructing thermodynamic functions (partition function, free energy and their derivatives) as functions of the macroscopic external parameters, and apply it to exemplary two-parametric statistical mechanics models. Our method is based on expressing the Fisher metric on the manifold of probability distributions over a high-dimensional space of microstates through the posterior distributions over a space of external parameters. Log-partition function is obtained by approximating the metric field by a Hessian of a convex function parametrized by an ICNN.

The proposed approach is used to reconstruct the partition functions and phase diagrams of the equilibrium Ising model and the steady state of the non-equilibrium TASEP model with open boundary conditions. Our approach does not use any a priori knowledge about Hamiltonian of the models, it only needs an algorithm to sample microstates for given values of the external parameters.

Acknowledgements

AL thanks Magicly AI for partially sponsoring this research. MT's work is supported by CUDAN ERA Chair project (Grant No. 810961 of the EU Horizon 2020 program).

References

[1] Van Nieuwenburg, E. P., Y.-H. Liu, S. D. Huber. Learning phase transitions by confusion. *Nature Physics*, 13(5):435–439, 2017.

[2] Carrasquilla, J., R. G. Melko. Machine learning phases of matter. *Nature Physics*, 13(5):431–434, 2017.

[3] Rem, B. S., N. Käming, M. Tarnowski, et al. Identifying quantum phase transitions using artificial neural networks on experimental data. *Nature Physics*, 15(9):917–920, 2019.

[4] Wang, J., W. Zhang, T. Hua, et al. Unsupervised learning of topological phase transitions using the calinski-harabaz index. *Physical Review Research*, 3(1):013074, 2021.

[5] Canabarro, A., F. F. Fanchini, A. L. Malvezzi, et al. Unveiling phase transitions with machine learning. *Physical Review B*, 100(4):045129, 2019.

[6] Jiang, B., T.-y. Wu, C. Zheng, et al. Learning summary statistic for approximate bayesian computation via deep neural network. *Statistica Sinica*, pages 1595–1618, 2017.

[7] Chen, Y., D. Zhang, M. Gutmann, et al. Neural approximate sufficient statistics for implicit models. *arXiv preprint arXiv:2010.10079*, 2020.

[8] Kloek, T., H. K. Van Dijk. Bayesian estimates of equation system parameters: an application of integration by monte carlo. *Econometrica: Journal of the Econometric Society*, pages 1–19, 1978.

[9] Yasuda, M., C. Takahashi. Free energy evaluation using marginalized annealed importance sampling. *Physical Review E*, 106(2):024127, 2022.

[10] Wu, D., L. Wang, P. Zhang. Solving statistical mechanics using variational autoregressive networks. *Physical review letters*, 122(8):080602, 2019.

[11] Qin, X., Z. Zhang, C. Huang, et al. U2-net: Going deeper with nested u-structure for salient object detection. *Pattern Recognition*, 106:107404, 2020.

[12] Amos, B., L. Xu, J. Z. Kolter. Input convex neural networks. In *International Conference on Machine Learning*, pages 146–155. PMLR, 2017.

[13] Ising, E. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

[14] Derrida, B., M. R. Evans, V. Hakim, et al. Exact solution of a 1d asymmetric exclusion model using a matrix formulation. *Journal of Physics A: Mathematical and General*, 26(7):1493, 1993.

[15] Blythe, R. A., M. R. Evans. Nonequilibrium steady states of matrix-product form: a solver's guide. *Journal of Physics A: Mathematical and General*, 40(46):R333, 2007.

[16] Krapivsky, P. L., S. Redner, E. Ben-Naim. *A kinetic view of statistical physics*. Cambridge University Press, 2010.

[17] Onsager, L. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Physical Review*, 65(3-4):117, 1944.

[18] Kac, M., J. C. Ward. A combinatorial solution of the two-dimensional ising model. *Physical Review*, 88(6):1332, 1952.

[19] Baxter, R. J., I. G. Enting. 399th solution of the ising model. *Journal of Physics A: Mathematical and General*, 11(12):2463, 1978.

[20] Walker, N., K.-M. Tam, M. Jarrell. Deep learning on the 2-dimensional ising model to extract the crossover region with a variational autoencoder. *Scientific reports*, 10(1):1–12, 2020.

5 Supplementary Material

5.1 Ising model dataset generation

We consider a 2D Ising model (13), which is an archetypal model of phase transitions in statistical mechanics. A microstate of this model is a set s of spin variables $s_i = \pm 1$ defined on each vertex of a square lattice of size $L \times L$. At equilibrium probability distribution over the space of microstates is

$$\mathbb{P}(s|H,T) = \frac{1}{Z(H,T)} e^{-\beta \sum_{\langle i,j \rangle} s_i s_j - \beta H \sum_i s_i} \tag{11}$$

where H and $T = 1/\beta$ are external parameters called magnetic field and temperature, respectively, the first sum is over all edges of the lattice, and $Z(H,T)$ is a normalization parameter known as a partition function:

$$Z(H,T) = \sum_{s_1=\pm 1} \cdots \sum_{s_N=\pm 1} e^{-\beta \sum_{\langle i,j \rangle} s_i s_j - \beta H \sum_i s_i}. \tag{12}$$

This model is exactly solvable for $H = 0$ (17; 18; 19). In particular, it is known that at $T_{\text{cr}} = \frac{2}{\log(1+\sqrt{2})} \approx 2.269$ a transition occurs between the high-temperature disordered state, where spin variables are on average equal zero, and the low-temperature ordered state in which average value of spin becomes distinct from zero. For general values of $H \neq 0$ partition function is intractable.

Our dataset consists of $N = 540000$ samples of spin configurations on the square lattice of size $L \times L = 128 \times 128$ with periodic boundary conditions. We consider the parameter ranges $T \in [T_{\min}, T_{\max}] = [1,5], H \in [H_{\min}, H_{\max}] = [-2,2]$ similar to the ranges used in (20). Point (T,H) is sampled uniformly from this rectangle, and then a sample spin configuration is created for these values of temperature and external field by starting with a random initial condition and equilibrating is with Glauber (one-spin Metropolis) dynamics for $10^4 \times 128 \times 128 \approx 1.64 \times 10^8$ iterations. We represent spin configuration as a single-channel image with values $+1$ and -1. When constructing target probability distributions we choose $\sigma = \frac{1}{50}$ and set the discretization $\mathcal{D}$ of the square $[T_{\min}, T_{\max}] \times [H_{\min}, H_{\max}] = [1,5] \times [-2,2]$ to be a uniform grid with $L \times L = 128 \times 128$ grid cells.

5.2 TASEP model dataset generation

Totally asymmetric simple exclusion process (TASEP) is a simple model of 1-dimensional transport phenomena. A microscopic configuration is a set of particles on a 1d lattice respecting the condition that there can be no more than one particle in each lattice cell. Each particle can move to the site to the right of it with probability pdt per time dt provided that it is empty (we put $p = 1$ without loss of generality). When complemented with boundary conditions, the TASEP attains a stationary state as time goes to infinity.

One particular case is open boundary conditions, when a particle is added with probability αdt per time dt to the leftmost site provided that it is empty and removed with probability βdt per time dt from the rightmost site provided that it is occupied. For this boundary condition the probability distribution is known exactly (14; 15; 16) and it once again takes the form

$$\mathbb{P}(s|\alpha,\beta) = \frac{f(s|\alpha,\beta)}{Z(\alpha,\beta)}; \quad Z(\alpha,\beta) = \sum_s f(s|\alpha,\beta); \tag{13}$$

where microstate s is a concrete sequence of filled and empty cells, and $f(s|\alpha,\beta)$ is some function of s and external parameters α, β. Importantly, however, the function f, which is known exactly for all system sizes M and all values of s, α, β does not take the form (**??**). TASEP with free boundaries exhibits a rich phase behavior: for large system sizes three distinct phases - the low-density phase, the high-density phase and the maximal current phase are possible depending on the values of α, β, and the asymptotic "free energy" which is defined as

$$F_{\text{TASEP}}(\alpha,\beta) = \lim_{M \to \infty} M^{-1} \log Z(\alpha,\beta) \tag{14}$$

and coincides with average flow per unit time, equals

$$F_{\text{TASEP}}(\alpha,\beta) = \begin{cases} \frac{1}{4}, \ \alpha > \frac{1}{2}, \ \beta > \frac{1}{2}; \\ \alpha(1-\alpha), \ \alpha < \beta, \ \alpha < \frac{1}{2}; \\ \beta(1-\beta), \ \beta < \alpha, \ \beta < \frac{1}{2}. \end{cases} \tag{15}$$

where the first, second and third cases correspond to maximal current, high density and low density cases, respectively.

We generate a dataset of $N = 150000$ stationary TASEP configurations on a 1d lattice with $M = 16384$ sites. The rates $\alpha(\beta)$ of adding (removing) particles at the left(right) boundary are sampled from the uniform prior distribution over a square $[0, 1] \times [0, 1]$. To reach the stationary state we start from a random initial condition and perform $N_{\text{steps}} = 2 \times 10^9 \approx 8M^2$ move attempts, which is known to be enough to achieve the stationary state except for the narrow vicinity of the transition line $\alpha = \beta < 1/2$ between high-density and low-density phases (in this case the stationary state has a slowly diffusing front of a shock wave in it, one needs of order M^2 move attempts to form the shock but of order M^3 move attempts for it to diffusively explore all possible positions of the shock).

We reshape 1d lattice with $M = 16384$ sites into an image of size $L \times L = 128 \times 128$ using raster scan ordering. To construct target probability distributions we set $\sigma = \frac{1}{150}$ and define the discretization $\mathcal{D}$ as a uniform grid on $[\alpha_{\min}, \alpha_{\max}] \times [\beta_{\min}, \beta_{\max}] = [0, 1] \times [0, 1]$ with $L \times L = 128 \times 128$ grid cells.

5.3 Algorithms

We summarize the procedure described in the section 2 by two algorithms outlined below, the first one generates the training dataset (1), the second one finds the approximations for the partition function and the Fisher metric (2).

Algorithm 1: Dataset Generation for Bayesian Thermodynamic Integration

Input:
$\mathbb{P}(s|\mathbf{t})$ - likelihood function represented by an algorithm to sample data,
$\mathbb{P}(\mathbf{t})$ - uniform prior distribution on the space of external parameters,
N_{dataset} - dataset length,
σ^2 - variance of the target gaussian distribution,
$\mathcal{D} = \{\hat{\mathbf{t}}_1, \ldots, \hat{\mathbf{t}}_{N_\mathcal{D}}\}$ - a discretization of the support of the uniform prior distribution $\mathbb{P}(\mathbf{t})$.
L_{sam} = [] - list of input samples,
L_{par} = [] - list of external parameters,
L_{tar} = [] - list of target posterior distributions
for $i = 1$ **to** N_{dataset} **do**
  sample $\mathbf{t}_i \sim \mathbb{P}(\mathbf{t})$ and append $\mathbf{t}_i$ to L_{par}
  sample $s_i \sim \mathbb{P}(s|\mathbf{t}_i)$ and append s_i to L_{sam}
  $p^{(i)}_{\text{tar}}$ = [] - discretized target probability distributions
  for $j = 1$ **to** $N_\mathcal{D}$ **do**
    $p^{(i)}_{\text{tar}}[j] = \exp\left(-\frac{1}{2\sigma^2}||\hat{\mathbf{t}}_j - \mathbf{t}_i||^2\right)$
  end for
  $p^{(i)}_{\text{tar}} = \text{normalize}(p^{(i)}_{\text{tar}})$ and append $p^{(i)}_{\text{tar}}$ to L_{tar}
end for
Output:
L_{sam} - dataset of input samples,
L_{par} - dataset of external parameters,
L_{tar} - dataset of target posterior distributions

Algorithm 2: Bayesian Thermodynamic Integration

Input:

L_{sam} - train dataset of input samples,
L_{par} - train dataset of external parameters,
L_{tar} - train dataset of target posterior distributions,
K_{bundle} - bundle size or the number of nearest neighbours,
$N_{\text{U2Net steps}}$ - steps for posterior approximation.
$N_{\text{ICNN steps}}$ - steps for free energy approximation.
$\Pi_\theta(\mathbf{t}|s_1,\ldots,s_{K_{\text{bundle}}})$ - image-to-image neural network with U^2-Net architecture with output shape equal to the shape of elements of L_{tar},
L_{neib} = [] - list of K_{bundle} nearest neighbors of points in L_{par}.

$\mathrm{L}_{\text{neib}} = \mathrm{KNN}(\mathrm{L}_{\text{par}}, K_{\text{bundle}}, \text{include\_self=True})$
for $i = 1$ **to** $N_{\text{U2Net steps}}$ **do**
  choose $\mathbf{t}'$ randomly from L_{par}
  $s(\mathbf{t}') = (s_1,\ldots,s_{K_{\text{bundle}}})$ shuffle and concatenate K_{bundle} microstates corresponding to the nearest neighbors of $\mathbf{t}'$ taken from L_{neib}
  $\mathbb{P}_{\text{target}} = \frac{1}{K_{\text{bundle}}} \sum_{\hat{\mathbf{t}}\in \text{Neib}(\mathbf{t}')} \mathrm{L}_{\text{tar}}[\text{index}(\hat{\mathbf{t}})]$
  $\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{t}'\sim\mathbb{P}(\mathbf{t})}\mathrm{KL}(\mathbb{P}_{\text{target}}(\mathbf{t}|\mathbf{t}')||\Pi_\theta(\mathbf{t}|s(\mathbf{t}')))$
  $\theta \leftarrow \mathrm{Adam}(\theta, \nabla_\theta \mathcal{L}(\theta))$
end for
  $Z_w(\mathbf{t})$ - input convex neural network,
for $i = 1$ **to** $N_{\text{ICNN steps}}$ **do**
  $\mathbb{P}_w^{(\text{eq})}(\mathbf{t}|\mathbf{t}') = \frac{e^{(\nabla_{\mathbf{t}'}\log Z_w(\mathbf{t}'))^T\mathbf{t} - \log Z_w(\mathbf{t})}}{\int e^{(\nabla_{\mathbf{t}'}\log Z_w(\mathbf{t}'))^T\mathbf{t} - \log Z_w(\mathbf{t})} d\mathbf{t}}$
  $\mathcal{L}(w) = \mathbb{E}_{\mathbf{t}'\sim\mathbb{P}(\mathbf{t})}\mathrm{JSD}\left(\Pi_\theta(\mathbf{t}|s(\mathbf{t}')), \mathbb{P}_w^{(\text{eq})}(\mathbf{t}|\mathbf{t}')\right)$
  $w \leftarrow \mathrm{Adam}(w, \nabla_w \mathcal{L}(w))$
end for
Output:
  $F(\mathbf{t}) = \log Z_w(\mathbf{t})$ - approximated free energy,
  $G(\mathbf{t}) = \nabla_{\mathbf{tt}} \log Z_w(\mathbf{t})$ - approximated Fisher metric.

KNN = k-nearest neighbors
KL = Kullback–Leibler divergence
JSD = Jensen–Shannon divergence

5.4 Examples of microstates

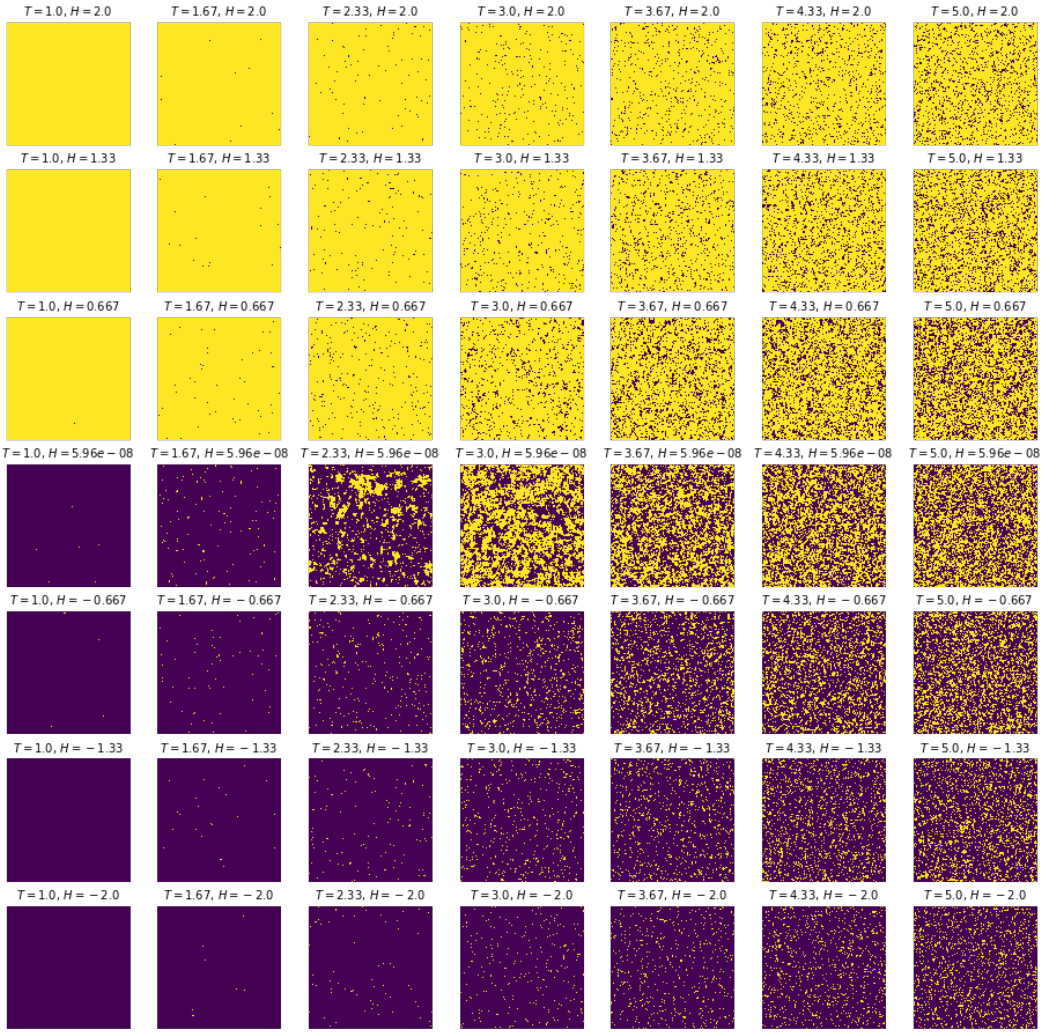


Figure 2: Examples of the microstates of the 2D Ising model, temperature increases from right to left, magnetic field increases from bottom to top.

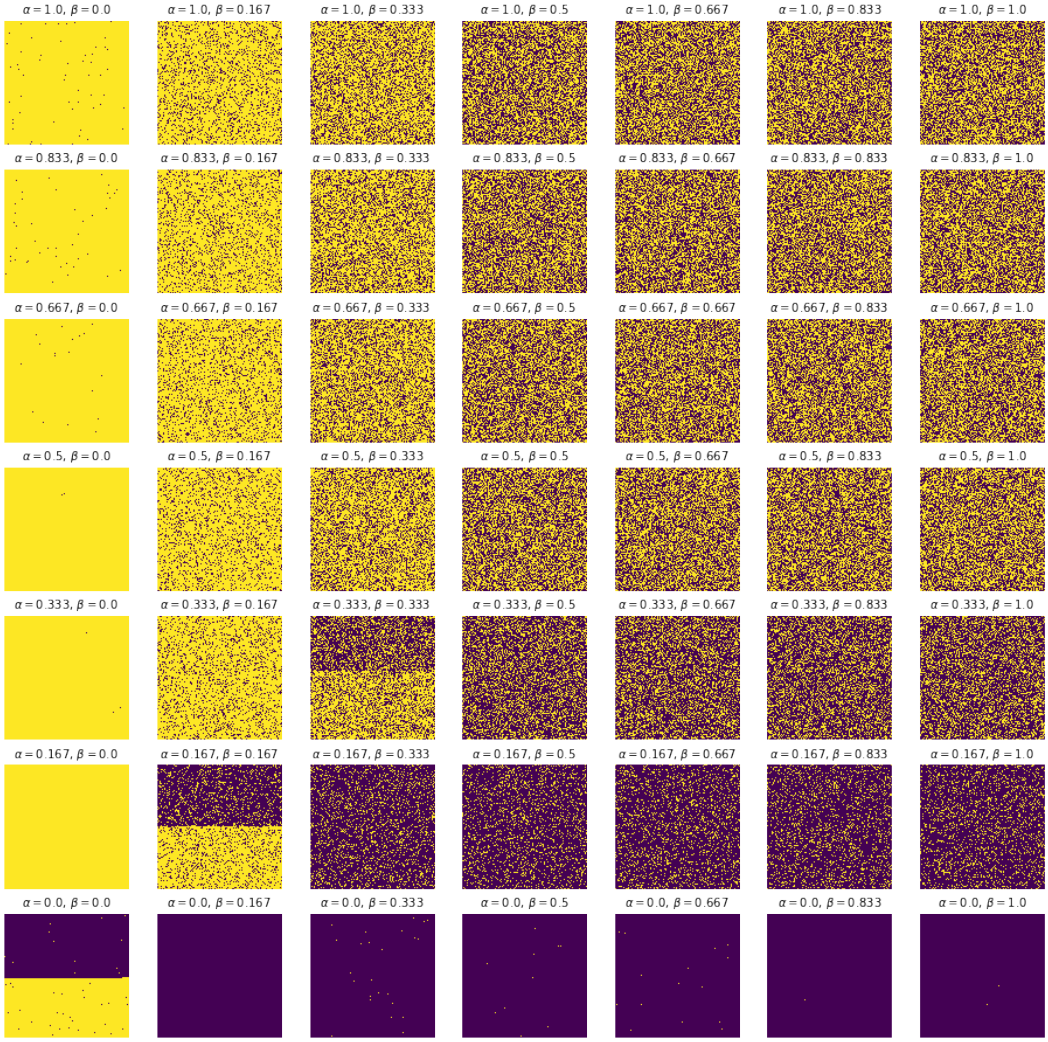


Figure 3: Examples of microstates of the stationary state TASEP. The microstate is presented in raster ordering, the cell numbers increase downwards from row to row, and rightwards within a row. α and β increase from bottom to top and from left to right, respectively. Once can clearly see the high density (left), low-density (bottom) and maximal current (density = 1/2, top right) phases.

5.5 Learned posterior distributions for 2D Ising model

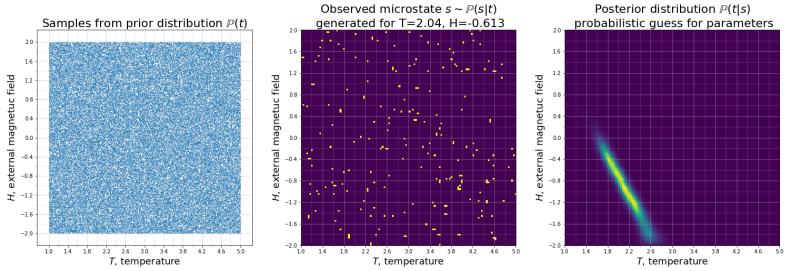


Figure 4: (left) Uniform prior distribution on the square $[T_{\min}, T_{\max}] \times [H_{\min}, H_{\max}] = [1, 5] \times [-2, 2]$. (center) Observed microstate of the Ising model generated for $T = 2.04, H = -0.613$. (right) Posterior distribution on the square $[T_{\min}, T_{\max}] \times [H_{\min}, H_{\max}]$ predicted by U^{2}Net.

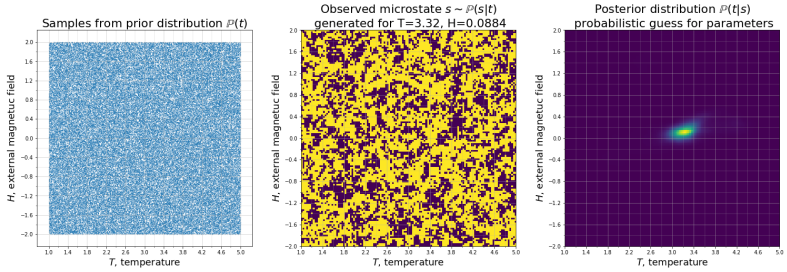


Figure 5: (left) Uniform prior distribution on the square $[T_{\min}, T_{\max}] \times [H_{\min}, H_{\max}] = [1, 5] \times [-2, 2]$. (center) Observed microstate of the Ising model generated for $T = 3.32, H = 0.0884$. (right) Posterior distribution on the square $[T_{\min}, T_{\max}] \times [H_{\min}, H_{\max}]$ predicted by U^{2}Net.

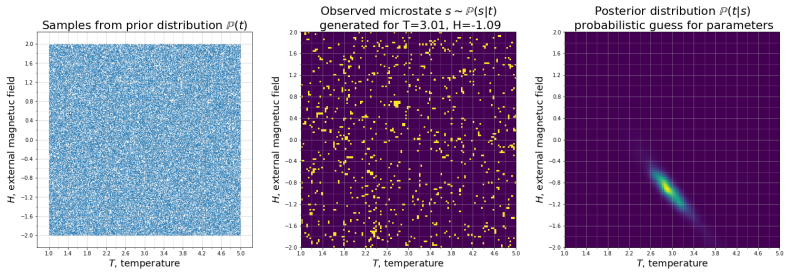


Figure 6: (left) Uniform prior distribution on the square $[T_{\min}, T_{\max}] \times [H_{\min}, H_{\max}] = [1, 5] \times [-2, 2]$. (center) Observed microstate of the Ising model generated for $T = 3.01, H = -1.09$. (right) Posterior distribution on the square $[T_{\min}, T_{\max}] \times [H_{\min}, H_{\max}]$ predicted by U^{2}Net.