Journal of Computational Physics 336 (2017) 513–545

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

# Scalable information inequalities for uncertainty quantification

Markos A. Katsoulakis, Luc Rey-Bellet, Jie Wang *

*Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01003, USA*

ARTICLE INFO

*Article history:*
Received 20 April 2016
Received in revised form 30 January 2017
Accepted 6 February 2017
Available online 10 February 2017

*Keywords:*
Kullback Leibler divergence
Information metrics
Uncertainty quantification
Statistical mechanics
High dimensional systems
Nonlinear response
Phase diagrams

ABSTRACT

In this paper we demonstrate the only available scalable information bounds for quantities of interest of high dimensional probabilistic models. Scalability of inequalities allows us to (a) obtain uncertainty quantification bounds for quantities of interest in the large degree of freedom limit and/or at long time regimes; (b) assess the impact of large model perturbations as in nonlinear response regimes in statistical mechanics; (c) address model-form uncertainty, i.e. compare different extended models and corresponding quantities of interest. We demonstrate some of these properties by deriving robust uncertainty quantification bounds for phase diagrams in statistical mechanics models.



## 1. Introduction

Information Theory provides both mathematical methods and practical computational tools to construct probabilistic models in a principled manner, as well as the means to assess their validity, [1]. One of the key mathematical objects of information theory is the concept of information metrics between probabilistic models. Such concepts of distance between models are not always metrics in the strict mathematical sense, in which case they are called divergences, and include the relative entropy, also known as the Kullback–Leibler (KL) divergence, the total variation and the Hellinger metrics, the $\chi^2$ divergence, the F-divergence, and the Rényi divergence, [2]. For example, the relative entropy between two probability distributions $P = P(\sigma)$ and $Q = Q(\sigma)$ on $\mathbb{R}^N$ is defined as

$$R(Q \parallel P) = \int_{\mathbb{R}^N} \log\left(\frac{Q(\sigma)}{P(\sigma)}\right) Q(\sigma) d\sigma, \tag{1}$$

when the integral exists. The relative entropy is not a metric but it is a divergence, that is it satisfies the properties: (i) $R(Q \parallel P) \geq 0$, (ii) $R(Q \parallel P) = 0$ if and only if $P = Q$ a.e.

We may for example think of the model $Q$ as an approximation, or a surrogate model for another complicated and possibly inaccessible model $P$; alternatively we may consider the model $Q$ as a misspecification of the true model $P$. When measuring model discrepancy between the two models $P$ and $Q$, tractability depends critically on the type of distance used between models. In that respect, the relative entropy has very convenient analytic and computational properties, in

* Corresponding author.
*E-mail addresses:* markos@math.umass.edu (M.A. Katsoulakis), luc@math.umass.edu (L. Rey-Bellet), wang@math.umass.edu (J. Wang).

http://dx.doi.org/10.1016/j.jcp.2017.02.020
0021-9991/

particular regarding the scaling properties of the system size $N$ which could represent space and/or time. Obtaining bounds which are valid for high dimensional ($N \gg 1$) or spatially extended systems and/or long time regimes is the main topic of the paper and we will discuss these properties in depth in the upcoming sections.

Information metrics provide systematic and practical tools for building approximate statistical models of reduced complexity through variational inference methods [3–5] for machine learning [6,7,4] and coarse-graining of complex systems [8–16]. Variational inference relies on optimization problems such as

$$\min_{Q\in\mathcal{Q}} R(P\,\|\,Q) \quad \text{or} \quad \min_{Q\in\mathcal{Q}} R(Q\,\|\,P)\,, \tag{2}$$

where $\mathcal{Q}$ is a class of simpler, computationally more tractable probability models than $P$. Subsequently, the optimal solution $Q^*$ of (2) replaces $P$ for estimation, simulation and prediction purposes. The choice of order in $P$ and $Q$ in (2) can be significant and depends on implementation methods, availability of data and the specifics of each application, e.g. [3,4,14,5]. In the case of coarse-graining the class of coarse-grained models $\mathcal{Q}$ will also have fewer degrees of freedom than the model $P$, and an additional projection operator is needed in the variational principle (2), see for instance [8,16]. In addition, information metrics provide fidelity measures in model reduction, [17–23], sensitivity metrics for uncertainty quantification, [24–29] and discrimination criteria in model selection [30,31]. For instance, for the sensitivity analysis of parametrized probabilistic models $P^\theta = P^\theta(\sigma)$, $\theta \in \Theta$ the relative entropy $R(P^\theta\,\|\,P^{\theta+\epsilon})$ measures the loss of information due to an error in parameters in the direction of the vector $\epsilon \in \Theta$. Different directions in parameter space provide a ranking of the sensitivities. Furthermore, when $|\epsilon| \ll 1$ we can also consider the quadratic approximation $R(P^\theta\,\|\,P^{\theta+\epsilon}) = \epsilon\mathbf{F}(P^\theta)\epsilon^\top + O(|\epsilon|^3)$ where $\mathbf{F}(P^\theta)$ is the Fisher Information matrix, [27,26,28].

Based on this earlier discussion it is natural and useful to approximate, perform model selection and/or sensitivity analysis in terms of information theoretical metrics between probability distributions. However, one is often interested in assessing model approximation, fidelity or sensitivity on concrete quantities of interest and/or statistical estimators. More specifically, if $P$ and $Q$ are two probability measures and $f = f(\sigma)$ is some quantity of interest or statistical estimator, then we measure the discrepancy between models $P$ and $Q$ with respect to the Quantity of Interest (QoI) $f$ by considering the model bias,

$$E_Q(f) - E_P(f)\,. \tag{3}$$

Indeed, in a statistics context, $f$ could be an unbiased statistical estimator for model $P$ which is either complicated to compute or possibly inaccessible and $Q$ is a computationally tractable nominal or reference model, e.g., a surrogate model. Thus, (3) is the estimator bias when using model $Q$ instead of $P$. Alternatively, $P$ could be the nominal model, for instance a model obtained through a careful statistical inference method, e.g. some type of best-fit approach such as maximum likelihood or variational inference, or just simply our best guess. However, due to the uncertainty whether this is a suitable model for the QoI $f$, we would like to measure the performance on $f$ over a family of alternative models, for example all models within a KL tolerance $\eta$, $\mathcal{Q} = \{Q : R(Q\,\|\,P) \le \eta\}$. In this case, a bound on $\max_{Q\in\mathcal{Q}} |E_Q(f) - E_P(f)|$ will provide a performance guarantee for the model $P$. Our main mathematical goal is to understand how to transfer quantitative results on information metrics into bounds for quantities of interest in (3). In this direction, information inequalities can provide a method to relate quantities of interest (3) and information metrics (1), a classic example being the Csiszar–Kullback–Pinsker (CKP) inequality, [2]:

$$|E_Q(f) - E_P(f)| \le \|f\|_\infty \sqrt{2R(Q\,\|\,P)}\,, \tag{4}$$

where $\|f\|_\infty = \sup_{\sigma\in\mathbb{R}^N} |f(\sigma)|$. In other words relative entropy controls how large the model discrepancy (3) can become for the quantity of interest $f$. More such inequalities involving other probability metrics such as Hellinger distance, $\chi^2$ and Rényi divergences are discussed in the subsequent sections.

In view of (4) and other such inequalities, a natural question is whether these are sufficient to assess the fidelity of complex systems models. In particular complex systems such as molecular or multi-scale models are typically high dimensional in the degrees of freedom and/or often require controlled fidelity (in approximation, uncertainty quantification, etc.) at long time regimes; for instance, in building coarse-grained models for efficient and reliable molecular simulation. Such an example arises when we are comparing two statistical mechanics systems determined by Hamiltonians $H_N$ and $\bar{H}_N$ describing say $N$ particles with positions $X = (x_1, \ldots, x_N)$. The associated canonical Gibbs measures are given by

$$P_N(X)dX = Z_N^{-1} e^{-H_N(X)}dX \quad \text{and} \quad Q_N(X)dX = \bar{Z}_N^{-1} e^{-\bar{H}_N(X)}dX\,, \tag{5}$$

where $Z_N$ and $\bar{Z}_N$ are normalizations (known as partition functions) that ensure the measures (5) are probabilities. Example (5) is a ubiquitous one, given the importance of Gibbs measures in disparate fields ranging from statistical mechanics and molecular simulation, pattern recognition and image analysis, to machine and statistical learning, [32,3,4]. In the case of (5), the relative entropy (1) readily yields,

$$R(Q_N\,\|\,P_N) = E_{Q_N}(H_N - \bar{H}_N) + \log Z_N - \log \bar{Z}_N\,. \tag{6}$$

It is a well known result in classic statistical mechanics [32], that under very general assumptions on $H_N$, both terms in the right hand side of (6) scale like $O(N)$ for $N \gg 1$, therefore we have that

$$R(Q_N \,\|\, P_N) = O(N)\,. \tag{7}$$

Comparing to (4), we immediately realize that the upper bound grows with the system size $N$, at least for nontrivial quantities of interest $f$ and therefore the CKP inequality (4) yields no information on model discrepancy for quantities of interest in (3). In Section 2 we show that other known information inequalities involving other divergences are also inappropriate for large systems in the sense that they do not provide useful information for quantities of interest: they either blow up like (4) or lose their selectivity, in the $N \gg 1$ limit. Furthermore, in Section 2 we also show that similar issues arise for time dependent stochastic Markovian models at long time regimes, $T \gg 1$.

In our main result we address these issues by using the recent information inequalities of [33] which in turn relied on earlier upper bounds in [34]. In these inequalities, the discrepancy in quantities of interest (3) is bounded as follows:

$$\Xi_-(Q \,\|\, P; f) \le E_Q(f) - E_P(f) \le \Xi_+(Q \,\|\, P; f)\,, \tag{8}$$

where

$$\Xi_+(Q \,\|\, P; f) = \inf_{c>0} \left\{ \frac{1}{c} \log E_P\left(e^{c(f - E_P(f))}\right) + \frac{1}{c} R(Q \,\|\, P) \right\}, \tag{9}$$

with a similar formula for $\Xi_-(Q \,\|\, P; f)$. The roles of $P$ and $Q$ in (8) can be reversed as in (2), depending on the context and the challenges of the specific problem, as well as on how easy it is to compute or bound the terms involved in (9); we discuss specific examples in Section 6. The quantities $\Xi_\pm(Q \,\|\, P; f)$ are referred to as a "goal-oriented divergence", [33], because they have the properties of a divergence both in probabilities $P$ and $Q$ and the quantity of interest $f$. More precisely, $\Xi_+(Q \,\|\, P; f) \ge 0$, (resp. $\Xi_-(Q \,\|\, P; f) \le 0$) and $\Xi_\pm(Q \,\|\, P; f) = 0$ if and only if $P = Q$ a.s. or $f$ is constant $P$-a.s.

The bounds (8) turn out to be robust, i.e. the bounds are attained when considering the set of all models $Q$ with a specified uncertainty threshold $\eta$ within the model $P$ given by the distance $R(Q \,\|\, P) \le \eta$; we refer to [34], while related robustness results can be also found in [35]. The parameter $c$ in the variational representation (9) controls the degree of robustness with respect to the model uncertainty captured by $R(Q \,\|\, P)$. In a control or optimization context these bounds are also related to $H^\infty$ control, [36]. Finally, $\Xi_\pm(Q \,\|\, P; f)$ admits an asymptotic expansion in relative entropy, [33]:

$$\Xi_\pm(Q \,\|\, P; f) = \pm\sqrt{Var_P[f]}\sqrt{2R(Q \,\|\, P)} + O(R(Q \,\|\, P))\,, \tag{10}$$

which captures the aforementioned divergence properties, at least to leading order.

In this paper we demonstrate that the bounds (8) scale correctly with the system size $N$ and provide "scalable" uncertainty quantification bounds for large classes of QoIs. We can get a first indication that this is the case by considering the leading term in the expansion (10). On one hand, typically for high dimensional systems we have $R(Q \,\|\, P) = O(N)$, see for instance (6); but on the other hand for common quantities of interest, e.g. in molecular systems non-extensive QoIs such as density, average velocity, magnetization or specific energy, we expect to have

$$Var_P(f) = O(1/N)\,. \tag{11}$$

Such QoIs also include many statistical estimators e.g. sample means or maximum likelihood estimators, [31]. Combining estimates (7) and (11), we see that, at least to leading order, the bounds in (8) scale as

$$\Xi_\pm(Q \,\|\, P; f) \approx O(1)\,,$$

in sharp contrast to the CKP inequality (4). Using tools from statistical mechanics we show that this scaling holds not only for the leading-order term but for the goal oriented divergences $\Xi_\pm(Q \,\|\, P; f)$ themselves, for extended systems such as Ising-type model in the thermodynamic limit. These results are presented in Sections 3 and 4. Furthermore, in [33] it is also shown that such information inequalities can be used to address model error for time dependent problems at long time regimes. In particular our results extend to path-space observables, e.g., ergodic averages, correlations, etc., where the role of relative entropy is played by the relative entropy rate (RER) defined as the relative entropy per unit time. We revisit the latter point here and connect it to nonlinear response calculations for stochastic dynamics in statistical mechanics. Finally, in Section 7 we also show that the bounds (8) are also appropriate for a general class of observables and not just for spatio-temporally averaged quantities that need to satisfy (11).

Overall, the scalability of (8) allows us to address three challenges which are not readily covered by standard numerical (error) analysis, statistics or statistical mechanics calculations: (a) obtain uncertainty quantification (UQ) bounds for quantities of interest in the large degree of freedom limit $N \gg 1$ and/or at long time regimes $T \gg 1$, (b) estimate the impact of large model perturbations, going beyond error expansion methods and providing nonlinear response bounds in the statistical mechanics sense, and (c) address model-form uncertainty, i.e. comparing different extended models and corresponding QoIs.

We demonstrate all three capabilities in deriving robust uncertainty quantification bounds for phase diagrams in statistical mechanics models. Phase diagrams are calculations of QoIs as functions of continuously varying model parameters,

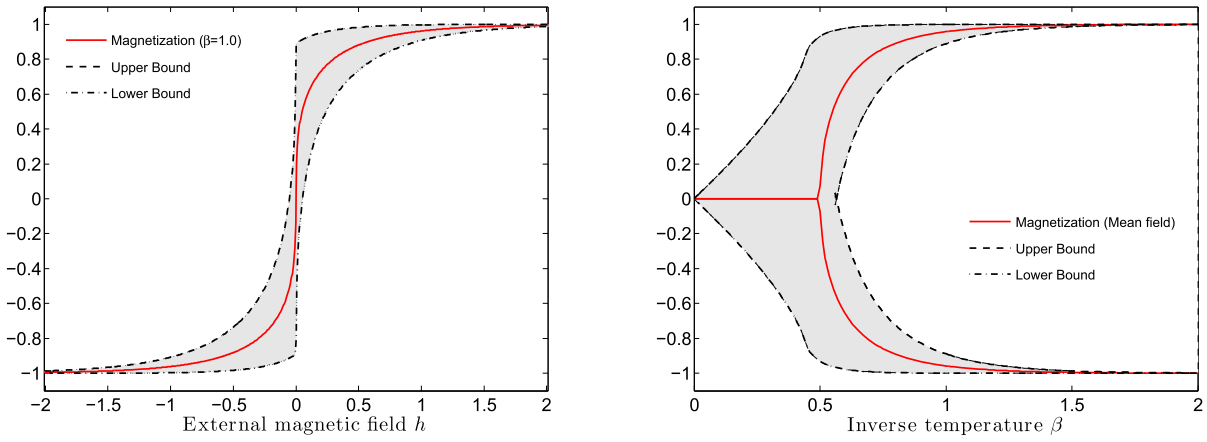



**Fig. 1.** (a) The red solid line is the magnetization of 1-d mean field Ising model with $\beta = 1$, which is the baseline model $P$; the black dashed/dash–dot line is the goal-oriented divergence upper/lower bound of magnetization of the perturbed mean field model $Q$ with $\beta = 1.6$. The gray areas depict the size of the uncertainty region corresponding to the goal-oriented bounds (8). (b) The black dashed/dash–dot line is the goal-oriented divergence upper/lower bound of the magnetization of the nearest neighbor Ising model $P$ with $h = 0$ ($P$ model not shown). The red solid line is the spontaneous magnetization of 2-d mean field Ising model for $h = 0$, which here plays the role of the surrogate, simpler, approximate model $Q$. The gray areas depict the size of the uncertainty region corresponding to the goal-oriented bounds (8). For more details see Fig. 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

e.g. temperature, external forcing, etc. Here we consider a given model $P$ and desire to calculate uncertainty bounds for its phase diagram, when the model $P$ is replaced by a different model $Q$. We note that phase diagrams are typically computed in the thermodynamic limit $N \to \infty$ and in the case of steady states in the long time regime $T \to \infty$; thus, in order to obtain uncertainty bounds for the phase diagram of the model $P$, we necessarily will require scalable bounds such as (8). Similarly, we need such scalable bounds to address any related UQ or sensitivity analysis question for molecular or any other high-dimensional probabilistic model, where $N \to \infty$ and/or $T \to \infty$ regimes are commonplace. To illustrate the potential of our methods, we consider fairly large parameter discrepancies between models $P$ and $Q$ of the order of 50% or more, see for instance Fig. 1(a). We also compare phase diagrams corresponding not just to different parameter choices but to entirely different Gibbs models (5), where $Q$ is a true microscopic model and $P$ is for instance some type of mean field approximation, see Fig. 1(b). These and several other test-bed examples are discussed in detail in Section 5.

This paper is organized as follows. In Section 2 we discuss classic information inequalities for QoIs and demonstrate how they scale with system size or with long time dynamics. We show these results by considering counterexamples such as sequences of independent identically distributed random variables and Markov chains. In Section 3 we revisit the concept of goal oriented divergence introduced earlier in [33] and show that it provides scalable and discriminating information bounds for QoIs. In Section 4 we discuss how these results extend to path-space observables, e.g., ergodic averages, autocorrelations, etc., where the role of relative entropy is now played by the relative entropy rate (RER) and connect to nonlinear response calculations for stochastic dynamics in statistical mechanics. In Section 5 we show how these new information inequalities transfer to Gibbs measures, implying nonlinear response UQ bounds, and how they relate to classic results for thermodynamic limits in statistical mechanics. In Section 6 we apply our methods and the scalability of the UQ bounds to assess model and parametric uncertainty of phase diagrams in molecular systems. We demonstrate the methods for Ising models, although the perspective is generally applicable. Finally, in Section 7 we demonstrate how the bounds (8) can work for a general class of observables and not just for averaged QoIs.

## 2. Poor scaling properties of the classic inequalities for probability metrics

In this section we discuss several classic information inequalities and demonstrate they scale poorly with the size of the system especially when applying the inequalities to ergodic averages. We make these points by considering simple examples such as independent, identically distributed (IID) random variables, as well as Markov sequences and the corresponding statistical estimators.

Suppose $P$ and $Q$ be two probability measures on some measure space $(\mathcal{X}, \mathcal{A})$ and let $f : \mathcal{X} \to \mathbb{R}$ be some QoI. Our goal is to consider the discrepancy between models $P$ and $Q$ with respect to the quantity of interest $f$,

$$E_Q(f) - E_P(f). \tag{12}$$

Our primary mathematical challenge here is to understand what results on information metrics between probability measures $P$ and $Q$ imply for quantities of interest in (12). We first discuss several concepts of information metrics, including divergences and probability distances.

### 2.1. Information distances and divergences

To keep the notation as simple as possible we will assume henceforth that $P$ and $Q$ are mutually absolutely continuous and this will cover all the examples considered here. (Much of what we discuss would extend to general measures by considering a measure dominating $P$ and $Q$, e.g. $\frac{1}{2}(P+Q)$.) For the same reasons of simplicity in presentation, we assume that all integrals below exist and are finite.

**Total Variation** [2]: The total variation distance between $P$ and $Q$ is defined by

$$TV(Q,P) = \sup_{A\in\mathcal{A}} |Q(A) - P(A)| = \frac{1}{2}\int \left|1 - \frac{dP}{dQ}\right| dQ. \tag{13}$$

**Relative entropy** [2]: The Kullback–Leibler divergence, or relative entropy, of $P$ with respect to $Q$ is defined by

$$R(Q \parallel P) = \int \log\left(\frac{dQ}{dP}\right) dQ. \tag{14}$$

**Relative Rényi entropy** [37]: For $\alpha > 0$, $\alpha \neq 1$, the relative Rényi entropy (or divergence) of order $\alpha$ of $P$ with respect to $Q$ is defined by

$$D_\alpha(Q \parallel P) = \frac{1}{\alpha - 1}\int \left(\frac{dP}{dQ}\right)^{1-\alpha} dQ = \frac{1}{\alpha - 1}\int \left(\frac{dQ}{dP}\right)^{\alpha} dP. \tag{15}$$

**$\chi^2$ divergence** [2]: The $\chi^2$ divergence between $P$ and $Q$ is defined by:

$$\chi^2(Q \parallel P) = \int \left(\frac{dQ}{dP} - 1\right)^2 dP. \tag{16}$$

**Hellinger distance** [2]: The Hellinger distance between $P$ and $Q$ is defined by:

$$H(Q,P) = \left(\int \left(1 - \sqrt{\frac{dP}{dQ}}\right)^2 dQ\right)^{1/2}. \tag{17}$$

The total variation and Hellinger distances define proper distances while all the other quantities are merely divergences (i.e., they are non-negative and vanish if and only if $P = Q$). The Rényi divergence of order $1/2$ is symmetric in $P$ and $Q$ and is related to the Hellinger distance by

$$D_{1/2}(Q \parallel P) = -2\log\left(1 - \frac{1}{2}H^2(Q,P)\right).$$

Similarly the Rényi divergence of order 2 is related to the $\chi^2$ divergence by

$$D_2(Q \parallel P) = \log\left(1 + \chi^2(Q \parallel P)\right).$$

In addition the Rényi divergence of order $\alpha$ is an nondecreasing function of $\alpha$ [38] and we have

$$\lim_{\alpha\to 1} D_\alpha(Q \parallel P) = R(Q \parallel P),$$

and thus it thus natural to set $D_1(Q \parallel P) = R(Q \parallel P)$. Using the inequality $\log(t) \le t - 1$ we then obtain the chain of inequalities, [2]

$$H^2(Q,P) \le D_{1/2}(Q \parallel P) \le R(Q \parallel P) \le D_2(Q \parallel P) \le \chi^2(Q \parallel P). \tag{18}$$

### 2.2. Some classic information inequalities for QoIs

We recall a number of classic information-theoretic bounds which use probability distances of divergences to control expected values of QoIs (also referred to as observables). Since

$$\left|E_Q(f) - E_P(f)\right| = \left|\int f(1 - \frac{dP}{dQ})dQ\right| \leq \int \|f\|_\infty |1 - \frac{dP}{dQ}|dQ = 2\|f\|_\infty TV(Q,P),$$

we can readily obtain bounds on QoIs from relationships between $TV(Q,P)$ and other divergences. It is well-known and easy to prove that $TV(Q,P) \leq H(Q,P)$ but we will use here the slightly sharper bound (Le Cam's inequality) [2] given by

$$TV(Q,P) \leq H(Q,P)\sqrt{1 - \frac{1}{4}H^2(Q,P)},$$

which implies

**Le Cam** [2]:

$$\left|E_Q(f) - E_P(f)\right| \leq 2\|f\|_\infty H(Q,P)\sqrt{1 - \frac{1}{4}H^2(Q,P)}. \tag{19}$$

From inequality (18) and $TV(Q,P) \leq H(Q,P)$ we obtain immediately bounds on $TV(Q,P)$ by $\sqrt{D_\alpha(Q\,\|\,P)}$ but the constants are not optimal. The following generalized Pinsker inequality (with optimal constants) was proved in [39] and holds for $0 < \alpha \leq 1$

$$TV(Q,P) \leq \sqrt{\frac{1}{2\alpha}D_\alpha(Q\,\|\,P)},$$

and leads to

**Csiszar–Kullback–Pinsker (CKP)** [2]:

$$\left|E_Q(f) - E_P(f)\right| \leq \|f\|_\infty \sqrt{2R(Q\,\|\,P)}. \tag{20}$$

**Generalized Pinsker** [38]: For $0 < \alpha \leq 1$

$$\left|E_Q(f) - E_P(f)\right| \leq \|f\|_\infty \sqrt{\frac{2}{\alpha}D_\alpha(Q\,\|\,P)}. \tag{21}$$

It is known that the CKP inequality is sharp only if $P$ and $Q$ are close. In particular the total variation norm is always less than 1 while the relative entropy can be very large. There is a complementary bound to the CKP inequality which is based on a result by Scheffé [2]

**Scheffé**:

$$|E_Q(f) - E_P(f)| \leq \|f\|_\infty \left(2 - e^{-R(Q\,\|\,P)}\right). \tag{22}$$

By (18) we have $R(P\,\|\,Q) \leq \chi^2(Q\,\|\,P)$ and thus we can also obtain a bound in terms of the $\chi^2$ divergence and $\|f\|_\infty$. However, an improved bound involving the variance of $f$ and thus not requiring $f$ to be bounded can be readily obtained via the Cauchy–Schwartz inequality:

**Chapman–Robbins** [40]:

$$\left|E_Q(f) - E_P(f)\right| \leq \sqrt{Var_P(f)}\sqrt{\chi^2(Q\,\|\,P)}. \tag{23}$$

**Hellinger-based inequalities:**

Recently a bound using the Hellinger distance and the $L^2$ norm was derived in [41]:

$$\left|E_Q(f) - E_P(f)\right| \leq \sqrt{2}H(Q,P)\sqrt{(E_P(f^2) + E_Q(f^2))}.$$

As we show in Appendix A this bound can be further optimized by using a control variates argument. Note that the left hand side is unchanged by replacing $f$ by $f - \frac{1}{2}(E_P(f) + E_Q(f))$ and this yields the improved bound

$$\left|E_Q(f) - E_P(f)\right| \leq \sqrt{2}H(Q,P)\sqrt{Var_Q(f) + Var_P(f) + \frac{1}{2}(E_Q(f) - E_P(f))^2}. \tag{24}$$

Furthermore, by (24), we can obtain

$$\left|E_Q(f) - E_P(f)\right| \leq \frac{H(Q,P)}{\sqrt{1-H^2(Q,P)}}\sqrt{Var_Q(f) + Var_P(f)}\,, \tag{25}$$

for all $P$, $Q$ such that $0 \leq H(Q,P) < 1$.

### 2.3. Scaling properties for IID sequences

We make here some simple, yet useful observations, on how the inequalities discussed in the previous section scale with system size for IID sequences. We consider the product measure space $\mathcal{X}_N = \mathcal{X} \times \cdots \times \mathcal{X}$ equipped with the product $\sigma$-algebra $\mathcal{A}_N$ and we denote by $P_N = P \times \cdots \times P$ the product measures on $(\mathcal{X}_N, \mathcal{A}_N)$ whose all marginals are equal to $P$ and we define $Q_N$ similarly. From a statistics perspective, this is also the setting where sequences of $N$ independent samples are generated by the models $P$ and $Q$ respectively.

We will concentrate on QoIs which are observables which have the form of ergodic averages or of statistical estimators. The challenge would be to assess based on information inequalities the impact on the QoIs. Next, we consider the simplest such case of the sample mean. For any measurable $g : \mathcal{X} \to \mathbb{R}$ we consider the observable $f_N : \mathcal{X}_N \to \mathbb{R}$ given by

$$f_N(\sigma_1, \cdots, \sigma_N) = \frac{1}{N}\sum_{j=1}^{N} g(\sigma_j)\,. \tag{26}$$

This quantity is also the sample average of the data set $\mathcal{D} = \{\sigma_1, \cdots \sigma_N\}$. We also note that

$$\|f_N\|_\infty = \|g\|_\infty\,, \qquad E_{P_N}(f_N) = E_P(g)\,, \qquad Var_{P_N}(f_N) = \frac{1}{N}Var_P(g)$$

To understand how the various inequalities scale with the system size $N$ we need to understand how the information distances and divergences themselves scale with $N$. For IID random variables the results are collected in the following Lemma.

**Lemma 2.1.** *For two product measures $P_N$ and $Q_N$ with marginals $P$ and $Q$ we have*

$$\begin{aligned}
&\textbf{\textit{Kullback–Leibler:}} && R(Q_N \parallel P_N) = NR(Q \parallel P),\\
&\textbf{\textit{Rényi:}} && D_\alpha(Q_N \parallel P_N) = ND_\alpha(Q \parallel P),\\
&\textbf{\textit{Chi-squared:}} && \chi^2(Q_N \parallel P_N) = \left(1 + \chi^2(Q \parallel P)\right)^N - 1,\\
&\textbf{\textit{Hellinger:}} && H(Q_N \parallel P_N) = \sqrt{2 - 2\left(1 - \frac{H^2(Q,P)}{2}\right)^N}.
\end{aligned} \tag{27}$$

**Proof.** See Appendix B.

Combining the result in Lemma 2.1 with the information bounds in the previous section we obtain a series of bounds for ergodic averages which all suffer from serious defects. Some grow to infinity for $N \gg 1$ while others converge to a trivial bound that is not discriminating, namely provide no new information on the difference of the QoIs $E_{Q_N}(f_N) - E_{P_N}(f_N)$. More precisely we obtain the following bounds:

**Csiszar–Kullback–Pinsker for IID**:

$$\left|E_{Q_N}(f_N) - E_{P_N}(f_N)\right| \leq \|g\|_\infty\sqrt{2NR(Q \parallel P)} = O(\sqrt{N}). \tag{28}$$

**Generalized Pinsker for IID**: For $0 < \alpha < 1$ we have

$$\left|E_{Q_N}(f_N) - E_{P_N}(f_N)\right| \leq \|g\|_\infty\sqrt{\frac{2N}{\alpha}D_\alpha(Q \parallel P)} = O(\sqrt{N}). \tag{29}$$

**Scheffé for IID**:

$$\left|E_{Q_N}(f_N) - E_{P_N}(f_N)\right| \leq \|g\|_\infty\left(2 - e^{-NR(Q \parallel P)}\right) = O(1). \tag{30}$$

**Chapman–Robbins for IID**: We have

$$\left|E_{Q_N}(f_N) - E_{P_N}(f_N)\right| \leq \sqrt{\frac{1}{N}Var_P(g)}\sqrt{\left(1 + \chi^2(Q \parallel P)\right)^N - 1} = O\left(\frac{\sqrt{e^N}}{\sqrt{N}}\right). \tag{31}$$

**Le Cam for IID**:

$$\left|E_{Q_N}(f_N) - E_{P_N}(f_N)\right| \leq 2\|g\|_\infty \sqrt{2 - 2\left(1 - \frac{H^2(Q,P)}{2}\right)^N} \sqrt{\frac{1}{2} + \frac{1}{2}\left(1 - \frac{H^2(Q,P)}{2}\right)^N} = O(1). \tag{32}$$

**Hellinger for IID**:

$$\left|E_{Q_N}(f_N) - E_{P_N}(f_N)\right| \leq \sqrt{2}\sqrt{2 - 2\left(1 - \frac{H^2(Q,P)}{2}\right)^N} \sqrt{\frac{Var_P(g)}{N} + \frac{Var_Q(g)}{N} + \frac{1}{2}(E_P(g) - E_Q(g))^2}. \tag{33}$$

Every single bound fails to capture the behavior of ergodic averages. Note that the left-hand sides are all of order 1 and indeed should be small if $P$ and $Q$ are sufficiently close to each other. The CKP, generalized Pinsker and Chapman–Robbins bounds all diverge as $N \to \infty$ and thus completely fail. The Le Cam bound is of order 1, but as $N \to \infty$ the bound converges to $2\|f\|_\infty$ which is a trivial bound independent of $P$ and $Q$. The Scheffé likewise converges to constant. Finally the Hellinger-based bound converges, as $N \to \infty$, to the trivial statement that $1 \leq \sqrt{2}$.

### *2.4. Scaling properties for Markov sequences*

Next, we consider the same questions as in the previous section, however this time for correlated distributions. Let two Markov chains in a finite state space $S$ with transitions matrix $p(\sigma, \sigma')$ and $q(\sigma, \sigma')$ respectively. We will assume that both Markov chains are irreducible with stationary distributions $\mu_p$ and $\mu_q$ respectively. In addition we assume that for any $\sigma \in S$, the probability measure $p(\sigma, \cdot)$ and $q(\sigma, \cdot)$ are mutually absolutely continuous. We denote by $\nu_p(\sigma)$ and $\nu_q(\sigma)$ the initial distributions of the two Markov chains and then the probability distributions of the path $(\sigma_1, \cdots \sigma_N)$ evolving under $p$ is given by

$$P_N(\sigma_1, \cdots, \sigma_N) = \nu_p(\sigma_0) p(\sigma_0, \sigma_1) \cdots p(\sigma_{N-1}, \sigma_N),$$

and similarly for the distribution $Q_N$ under $q$.

If we are interested in the long-time behavior of the system, for example we may be interested in computing or estimating expectations of the steady state or in our case model discrepancies such as

$$\left|E_{\mu_q}(g) - E_{\mu_p}(g)\right|$$

for some QoI (observable) $g : S \to \mathbb{R}$. In general the steady state of a Markov chain is not known explicitly or it is difficult to compute for large systems. However, if we consider ergodic observables such as

$$f_N(\sigma_1, \cdots, \sigma_N) = \frac{1}{N} \sum_{i=1}^{N} g(\sigma_i), \tag{34}$$

then, by the ergodic theorem, we have, for any initial distribution $\nu_p(\sigma)$ that

$$\lim_{N\to\infty} E_{P_N}(f_N) = E_{\mu_p}(g),$$

and thus can estimate $|E_{\mu_q}(g) - E_{\mu_p}(g)|$ if we can control $|E_{Q_N}(f) - E_{P_N}(f)|$ for large $N$. After our computations with IID sequences in the previous section, it is not surprising that none of the standard information inequalities allow such control. Indeed the following lemma, along with the fact that the variance of ergodic observables such as (34) scales like $Var_{P_N}(f_N) = O(1/N)$ [33], readily imply that the bounds for Markov measures scale exactly as (poorly as) the IID case, derived at the end of Section 2.3.

**Lemma 2.2.** *Consider two irreducible Markov chains with transitions matrix $p$ and $q$. Assume that the initial conditions $\nu_p(\sigma)$ and $\nu_q(\sigma)$ are mutually absolutely continuous and that $p(\sigma, \cdot)$ and $q(\sigma, \cdot)$ are mutually absolutely continuous for each $\sigma$.*

**Kullback–Leibler:** *We have*

$$\lim_{N\to\infty} \frac{1}{N} R(Q_N \,\|\, P_N) = r(q \,\|\, p) := \sum_{\sigma,\sigma'} \mu_q(\sigma) q(\sigma, \sigma') \log\left(\frac{q(\sigma, \sigma')}{p(\sigma, \sigma')}\right),$$

*and the limit is positive if and only if $p \neq q$.*

**Rényi:** *We have*

$$\lim_{N\to\infty} \frac{1}{N} D_\alpha(Q_N \,\|\, P_N) = \frac{1}{\alpha - 1} \log \rho(\alpha),$$

*where $\rho(\alpha)$ is the maximal eigenvalue of the non-negative matrix with entries $q^{\alpha}(\sigma,\sigma')p^{1-\alpha}(\sigma,\sigma')$ and we have $\frac{1}{\alpha-1}\log\rho(\alpha)\geq 0$ with equality if and only if $p\neq q$.*

**Chi-squared:** *We have*

$$\lim_{N\to\infty}\frac{1}{N}\log(1+\chi^2(Q_N\,\|\,P_N))=\log\rho(2)\,,$$

*where $\rho(2)$ is the maximal eigenvalue of the matrix with entries $q^2(x,y)p^{-1}(x,y)$ and we have $\log\rho(2)\geq 0$ with equality if and only if $p=q$.*

**Hellinger:** *We have*

$$\lim_{N\to\infty}H(Q_N\,\|\,P_N)=\sqrt{2}.$$

*if $p\neq q$ and 0 if $p=q$.*

**Proof.** See Appendix B.

## 3. A divergence with good scaling properties

### 3.1. Goal oriented divergence

In this section we will first discuss the goal-oriented divergence which was introduced by [33], following the work in [34]. Subsequently in Sections 3.3 and 4 and Section 5 we will demonstrate that this new divergence provides bounds on the model discrepancy $E_Q(f)-E_P(f)$ between models $P$ and $Q$ which scale correctly with their system size, provided the QoI $f$ has the form of an ergodic average or a statistical estimator.

Given an observable $f:\mathcal{X}\to\mathbb{R}$ we introduce the cumulant generating function of $f$

$$\Lambda_{P,f}(c)=\log E_P(e^{cf}). \tag{35}$$

We will assume $f$ is such that $\Lambda_{P,f}(c)$ is finite in a neighborhood $(-c_0,c_0)$ of the origin. For example if $f$ is bounded then we can take $c_0=\infty$. Under this assumption $f$ has finite moments of any order and we will often use the cumulant generating function of a mean 0 observable

$$\widetilde{\Lambda}_{P,f}(c)=\log E_P(e^{c(f-E_P(f))})=\Lambda_{P,f}(c)-cE_P(f). \tag{36}$$

The following bound is proved in [33] and will play a fundamental role in the rest of the paper.

**Goal-oriented divergence UQ bound** [33]: If $Q$ is absolutely continuous with respect to $P$ and $\Lambda_{P,f}(c)$ is finite in a neighborhood of the origin, then

$$\Xi_-(Q\,\|\,P;f)\leq E_Q(f)-E_P(f)\leq\Xi_+(Q\,\|\,P;f)\,, \tag{37}$$

where

$$\Xi_+(Q\,\|\,P;f)=\inf_{c>0}\left\{\frac{1}{c}\widetilde{\Lambda}_{P,f}(c)+\frac{1}{c}R(Q\,\|\,P)\right\}, \tag{38}$$

$$\Xi_-(Q\,\|\,P;f)=\sup_{c>0}\left\{-\frac{1}{c}\widetilde{\Lambda}_{P,f}(-c)-\frac{1}{c}R(Q\,\|\,P)\right\}. \tag{39}$$

We refer to [33] and [34] for details of the proof but the main idea behind the proof is the variational principle for the relative entropy: for bounded $f$ we have, [42],

$$\log E_P(e^f)=\sup_Q\left\{E_Q(f)-R(Q\,\|\,P)\right\}$$

and thus for any $Q$

$$E_Q(f)\leq\log E_P(e^f)+R(Q\,\|\,P)\,.$$

Replacing $f$ by $c(f-E_P(f))$ with $c>0$ and optimizing over $c$ yields the upper bound. The lower bound is derived in a similar manner.

As the following result from [33] shows, the quantities $\Xi_+$ and $\Xi_-$ are divergences similar to the relative (Rényi) entropy, the $\chi^2$ divergence and the Hellinger distance. Yet they depend on the observable $f$ and thus will be referred to as goal-oriented divergences.

**Properties of the goal-oriented divergence** [33]:

1. $\Xi_+(Q \parallel P; f) \geq 0$ and $\Xi_-(Q \parallel P; f) \leq 0$.
2. $\Xi_\pm(Q \parallel P; f) = 0$ if and only if $Q = P$ or $f$ is constant P-a.s.

It is instructive to understand the bound when $P$ and $Q$ are close to each other. Again we refer to [33] for a proof and provide here just an heuristic argument. First note that if $P = Q$ then it is easy to see that the infimum in the upper bound is attained at $c = 0$ since $R(Q \parallel P) = 0$ and $\widetilde{\Lambda}_{P,f}(c) > 0$ for $c > 0$ (the function is convex in $c$) and we have $\widetilde{\Lambda}_{P,f}(0) = \widetilde{\Lambda}'_{P,f}(0) = 0$ and $\widetilde{\Lambda}''_{P,f}(0) = \frac{1}{2} Var_P(f)$. So if $R(Q \parallel P)$ is small, we can expand the right-hand side in $c$ and we need to find

$$\inf_{c>0} \left\{ c \frac{Var_P(f)}{2} + O(c^2) + \frac{1}{c} R(Q \parallel P) \right\}.$$

Indeed, it has been demonstrated that the minimum has the form $\sqrt{Var_P(f)}\sqrt{2R(Q \parallel P)} + O(R(Q \parallel P))$, [33]. The lower bound is similar and we obtain:

**Linearized UQ bound** [33]: If $R(P \parallel Q)$ is small we have

$$|E_Q(f) - E_P(f)| \leq \sqrt{Var_P(f)}\sqrt{2R(Q \parallel P)} + O(R(Q \parallel P)). \tag{40}$$

**Robustness**: These new information bounds were shown in [34] to be robust in the sense that the upper bound is attained when considering all models $Q$ with a specified uncertainty threshold given by $R(Q \parallel P) \leq \eta$. Furthermore, the parameter $c$ in the variational representations (38) and (39) controls the degree of robustness with respect to the model uncertainty captured by $R(Q \parallel P)$.

### 3.2. Example: exponential family of distributions

Next we compute the goal-oriented divergences for an exponential family which covers many cases of interest including Markov and Gibbs measures (see Sections 4 and 5), as well as numerous probabilistic models in machine learning [4,3].

Given a reference measure $P^0$ (which does not need to be a finite measure) we say that $P^\theta$ is an exponential family if $P^\theta$ is absolutely continuous with respect to $P^0$ with

$$\frac{dP^\theta}{dP^0}(\sigma) = \exp\left(t(\sigma) \cdot \theta - F(\theta)\right),$$

where $\theta = [\theta_1, \cdots, \theta_K]^T \in \Theta \subset \mathbb{R}^K$ is the parameter vector, $t(\sigma) = [t_1(\sigma), ..., t_K(\sigma)]^T$ is the sufficient statistics vector and $F(\theta)$ is the log-normalizer

$$F(\theta) = \log \int e^{t(\sigma)\cdot\theta} dP^0(\sigma).$$

Note that $F(\theta)$ is a cumulant generating function for the sufficient statistics, for example we have $\nabla_\theta F(\theta) = E_{P^\theta}(t)$. The relative entropy between two members of the exponential family is then computed as

$$\begin{aligned} R(P^{\theta'} \parallel P^\theta) = \int \log \frac{dP^{\theta'}}{dP^\theta}(\sigma) dP^{\theta'}(\sigma) &= E_{P^{\theta'}}\left((\theta' - \theta) \cdot t(\sigma)\right) + F(\theta) - F(\theta') \\ &= (\theta' - \theta) \cdot \nabla F(\theta') + F(\theta) - F(\theta'). \end{aligned} \tag{41}$$

If we consider an observable which is a linear function of the sufficient statistics, that is

$$f(\sigma) = t(\sigma) \cdot v \tag{42}$$

for some vector $v \in \mathbb{R}^K$ then the cumulant generating function of $f - E_{P^\theta}(f)$ is

$$\tilde{\Lambda}_{P^\theta, f}(c) = \log E_{P^\theta}[e^{cf}] - cE_{P^\theta}(f) = F(\theta + cv) - F(\theta) - cv \cdot \nabla F(\theta), \tag{43}$$

and thus combining (41) and (43), we obtain the divergence

$$\Xi_\pm(P^{\theta'} \| P^\theta; f) = \pm \inf_{c>0} \frac{1}{c} \{(\theta' - \theta) \cdot \nabla F(\theta') - F(\theta') + F(\theta \pm cv) \mp cv \cdot \nabla F(\theta)\}. \tag{44}$$

### 3.3. Example: IID sequences

To illustrate the scaling properties of the goal-oriented divergence consider first two product measures $P_N$ and $Q_N$ as in Section 2.3 and the same sample mean observable (26). We now apply the bounds (38) and (39) to $Nf_N = \sum_{k=1}^N g(\sigma_k)$ to obtain

$$\frac{1}{N}\Xi_-(Q_N \,\|\, P_N; Nf_N) \le E_{Q_N}(f_N) - E_{P_N}(f_N) \le \frac{1}{N}\Xi_+(Q_N \,\|\, P_N; Nf_N)\,.$$

The following lemma shows that the bounds scale correctly with $N$.

**Lemma 3.1.** *We have*

$$\Xi_\pm(Q_N \,\|\, P_N; Nf_N) = N\Xi_\pm(Q \,\|\, P; g)\,.$$

**Proof.** We have already noted that $R(Q_N \,\|\, P_N) = NR(Q \,\|\, P)$. Furthermore

$$\begin{aligned}\widetilde{\Lambda}_{P_N, Nf_N}(c) &= \log E_{P_N}(e^{cNf_N}) - cE_{P_N}(Nf_N)\\ &= \log \int_{\mathcal{X}_N} e^{c\sum_{i=1}^N g(\sigma_i)} \prod_{i=1}^N dP(\sigma_i) - cE_{P_N}\left(\sum_{i=1}^N g(\sigma_i)\right)\\ &= N\log E_P(e^{cg}) - cNE_P(g) = N\widetilde{\Lambda}_{P,g}(c)\,. \end{aligned} \tag{45}$$

This result shows that the goal oriented divergence bounds capture perfectly the behavior of ergodic average as $N$ goes to infinity. In particular when P and Q are close in the sense of KL divergence, i.e. $R(Q\,\|P)$ is close to 0, by Theorem 2.12 in [33], we can obtain that $\Xi_\pm(Q \,\|\, P; g)$ is also close to 0, which contrasts sharply with all the bounds discussed in Section 2.3.

## 4. UQ and nonlinear response bounds for Markov sequences

In the context of Markov chains, there are a number of UQ challenges which are usually not addressed by standard numerical analysis or UQ techniques: (a) Understand the effects of a model uncertainty on the long-time behavior (e.g. steady state) of the model. (b) Go beyond linear response and be able to understand how large perturbations affect the model, both in finite and long time regimes. (c) Have a flexible framework allowing to compare different models as, for example for Ising model versus mean-field model approximations considered in Section 6.

The inequalities of Section 3.1 can provide new insights to all three questions, at least when the bounds can be estimated or computed numerically or analytically. As a first example in this direction we consider Markov dynamics with the same set-up as in Section 2.4. We have the following bounds which exemplify how the goal-oriented divergences provide UQ bounds for the long-time behavior of Markov chains.

**Theorem 4.1.** *Consider two irreducible Markov chains with transition matrices $p(\sigma, \sigma')$ and $q(\sigma, \sigma')$ and stationary distributions $\mu_p$ and $\mu_q$ respectively. If $p(\sigma, \cdot)$ and $q(\sigma, \cdot)$ are mutually absolutely continuous we have for any observable $g$ the bounds*

$$\xi_-(q \,\|\, p; g) \le E_{\mu_q}(g) - E_{\mu_p}(g) \le \xi_+(q \,\|\, p; g)\,,$$

*where*

$$\begin{aligned}\xi_+(q \,\|\, p; g) &= \inf_{c>0}\left\{\frac{1}{c}\lambda_{p,g}(c) + \frac{1}{c}r(q \,\|\, p)\right\},\\ \xi_-(q \,\|\, p; g) &= \sup_{c>0}\left\{-\frac{1}{c}\lambda_{p,g}(-c) - \frac{1}{c}r(q \,\|\, p)\right\}. \end{aligned} \tag{46}$$

*Here*

$$r(q \,\|\, p) = \lim_{N\to\infty}\frac{1}{N}R(Q_N \,\|\, P_N)$$

*is the relative entropy rate and $\lambda_{p,g}(c)$ is the logarithm of the maximal eigenvalue of the non-negative matrix with entries $p(\sigma, \sigma')\exp(c(g(\sigma') - E_{\mu_p}(g)))$.*

*Moreover, we have the asymptotic expansion in relative entropy rate $r(q \,\|\, p)$,*

$$\left|E_{\mu_q}(g) - E_{\mu_p}(g)\right| \le \sqrt{v_{\mu_p}(g)}\sqrt{2r(q \,\|\, p)} + O(r(q \,\|\, p)) \tag{47}$$

*where*

$$v_{\mu_p(g)} = \sum_{k=-\infty}^{\infty} E_{\mu_p}(g(\sigma_k)g(\sigma_0))$$

*is the integrated auto-correlation function for the observable g.*

**Proof.** We apply the goal-oriented divergence bound to the observable $Nf_N = \sum_{k=1}^{N} g(\sigma_i)$ and have

$$\frac{1}{N}\Xi_-(Q_N \,\|\, P_N; Nf_N) \le E_{Q_N}(f_N) - E_{P_N}(f_N) \le \frac{1}{N}\Xi_+(Q_N \,\|\, P_N; Nf_N).$$

We then take the limit $N \to \infty$. By the ergodicity of $P_N$ we have $\lim_{N\to\infty} E_{P_N}(f_N) = E_{\mu_p}(g)$ and similarly for $Q_N$. We have already established in Lemma 2.2 the existence of the limit $r(q \,\|\, p) = \lim_{N\to\infty} \frac{1}{N} R(Q_N \,\|\, P_N)$. For the cumulant generating function in $\Xi_\pm$ we have

$$\begin{aligned}
\frac{1}{N}\widetilde{\Lambda}_{P_N, Nf_N}(c) &= \frac{1}{N}\log E_{P_N}(e^{cNf_N}) - c\frac{1}{N}E_{P_N}(Nf_N) \\
&= \frac{1}{N}\log \sum_{\sigma_0,\cdots,\sigma_N} \nu_p(\sigma_0)\prod_{k=1}^{N} p(\sigma_{k-1},\sigma_k)e^{cg(\sigma_k)} - cE_{P_N}(f_N) \\
&= \frac{1}{N}\log \nu_p P_{cg}^N - cE_{P_N}(f_N)
\end{aligned}$$

where $P_{cg}$ is the non-negative matrix with entries $p(\sigma,\sigma')e^{cg(\sigma')}$. The Perron–Frobenius theorem gives the existence of the limit [43].

The asymptotic expansion is proved exactly as for the linearized UQ bound (40). It is not difficult to compute the second derivative of $\lambda_{p,g}(c)$ with respect to $c$ by noting all function are analytic function of $c$ and thus we can freely exchange the $N \to \infty$ limit with the derivative with respect to $c$. Therefore we obtain that

$$\frac{d^2}{dc^2}\lambda_{P,g}(0) = \lim_{N\to\infty}\frac{1}{N}Var_{P_N}(Nf_N)$$

and a standard computation shows that the limit is the integrated autocorrelation function $v_{\mu_p}(g)$, we also refer to the derivations in [33]. For details on the numerical computation of the autocorrelation function $v_{\mu_p}$ we refer to [44].

**Remark.** A well studied case of UQ for stochastic models and in particular stochastic dynamics is linear response, also referred to as local sensitivity analysis, which addresses the role of infinitesimal perturbations to model parameters of probabilistic models, e.g. [45,46]. Here (47) provides computable bounds in the linear response regime, as demonstrated earlier in [33] and which can be used for fast screening of uninfluential parameters in reaction networks with a very large number of parameters, [47]. A related linear response bound using relative entropy was carried out earlier in [48] and subsequently in [49].

**Nonlinear response bounds:** Beyond linear response considerations, nonlinear response methods attempt to address the role of larger parameter perturbations. Some of the relevant methods involve asymptotic series expansions in terms of the small parameter perturbation [50,51], which quickly become computationally intractable as more terms need to be computed. However, the inequalities (38) and (39) provide robust and computable nonlinear response bounds.

The main result in Theorem 4.1 was first obtained in [33]. Here we revisit it in the context of scalability in both space and time and connect it to nonlinear response calculations for stochastic dynamics in statistical mechanics. This connection is made more precise in the following Corollaries which follow directly from Theorem 4.1 and provide layers of progressively simpler to compute-and accordingly less sharp-bounds:

**Corollary 4.2.** *Based on the assumptions and definitions in Theorem 4.1, we have the following two bounds that involve two upper bounds of $r(q \,\|\, p)$. Bound (i) is sharper than bound (ii), while the latter is straightforward to calculate analytically.*

*(i) Let $R(q(\sigma,\cdot)\|p(\sigma,\cdot)) = \sum_\sigma q(\sigma,\sigma')\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}$; then,*

$$\begin{aligned}
\xi_+(q \,\|\, p; g) &\le \inf_{c>0}\left\{\frac{1}{c}\lambda_{p,g}(c) + \frac{1}{c}\sup_\sigma R(q(\sigma,\cdot)\|p(\sigma,\cdot))\right\} \\
\xi_-(q \,\|\, p; g) &\ge \sup_{c>0}\left\{-\frac{1}{c}\lambda_{p,g}(-c) - \frac{1}{c}\sup_\sigma R(q(\sigma,\cdot)\|p(\sigma,\cdot))\right\}.
\end{aligned} \tag{48}$$

*(ii) Next, we have the upper bound in terms of the quantity* $\sup_{\sigma,\sigma'} |\log \frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}|$,

$$\xi_+(q\,\|\,p;g) \le \inf_{c>0}\left\{\frac{1}{c}\lambda_{p,g}(c)+\frac{1}{c}\sup_{\sigma,\sigma'}|\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}|\right\}$$
$$\xi_-(q\,\|\,p;g) \ge \sup_{c>0}\left\{-\frac{1}{c}\lambda_{p,g}(-c)-\frac{1}{c}\sup_{\sigma,\sigma'}|\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}|\right\}. \tag{49}$$

**Proof.** We consider the relative entropy rate $r(q\,\|\,p)$; then,

$$\begin{aligned} r(q\,\|\,p) &= \sum_{\sigma,\sigma'}\mu_q(\sigma)q(\sigma,\sigma')\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')} \\ &= E_{\mu_q}\left(\sum_{\sigma'}q(\sigma,\sigma')\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}\right) \\ &= E_{\mu_q}\left(R(q(\sigma,\cdot)\|p(\sigma,\cdot))\right) \\ &\le \sup_{\sigma}R(q(\sigma,\cdot)\|p(\sigma,\cdot)), \end{aligned} \tag{50}$$

where $R(q(\sigma,\cdot)\|p(\sigma,\cdot)) = \sum_{\sigma'}q(\sigma,\sigma')\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}$. Moreover, we have

$$R(q(\sigma,\cdot)\|p(\sigma,\cdot)) = \sum_{\sigma'}q(\sigma,\sigma')\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')} \le \sup_{\sigma'}|\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}|.$$

Therefore we can obtain another bound for $r(q\,\|\,p)$, that is,

$$r(q\,\|\,p) \le \sup_{\sigma,\sigma'}|\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}|. \tag{51}$$

This bound may be not as sharp as the one in (50), but it is more easily computable. Thus, by (46), (50) and (51), it is easy to obtain (i) and (ii). □

If we consider the linearized bound in (47), then combining the bounds (50) and (51) of $r(q\|p)$, we can obtain the following bound, which is a further simplification of Corollary 4.2, again at the expense of the tightness of the bounds.

**Corollary 4.3.** *Under the assumptions and definitions in Theorem 4.1, we have*

$$\xi_\pm(q\,\|\,p;g) \le \pm\sqrt{v_{\mu_p}(g)}\sqrt{2\sup_{\sigma}R(q(\sigma,\cdot)\|p(\sigma,\cdot))}+O(\sup_{\sigma}R(q(\sigma,\cdot)\|p(\sigma,\cdot))) \tag{52}$$
$$\le \pm\sqrt{v_{\mu_p}(g)}\sqrt{2\sup_{\sigma,\sigma'}|\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}|}+O(\sup_{\sigma,\sigma'}|\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}|). \tag{53}$$

**Remark.** By the previous two Corollaries, we get some simplified ways to replace the calculation of $\xi_\pm(q\,\|\,p;g)$ since it is much easier to calculate $\sup_{\sigma}R(q(\sigma,\cdot)\|p(\sigma,\cdot))$ or $\sup_{\sigma,\sigma'}|\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}|$ than $r(q\|p)$ itself, especially the latter one. In practice, we can first attempt to estimate $\xi_\pm(q\,\|\,p;g)$ by calculating the leading term in (52) or (53). If the linearization assumptions in the last Corollary fail, then we can try to use Corollary 4.2 or Theorem 4.1 which can also give computable bounds or estimates of $\xi_\pm(q\,\|\,p;g)$.

Finally, the bound in (51) is the Markov chain analogue of the triple norm $|||\cdot|||$ used in the estimation of UQ bounds for QoIs of Gibbs measures, which we discuss in depth in Section 5.

## 5. UQ and nonlinear response bounds for Gibbs measures

The Gibbs measure is one of the central objects in statistical mechanics and molecular dynamics simulation, [32,52]. On the other hand Gibbs measures in the form of Boltzmann Machines or Markov Random Fields provide one of the key classes of models in machine learning and pattern recognition, [3,4]. Gibbs measures are probabilistic models which are inherently high dimensional, describing spatially distributed systems or a large number of interacting molecules. In this Section we derive scalable UQ bounds for Gibbs measures based on the goal oriented inequalities discussed in Section 3.1.

Gibbs measures can be set on a lattice or in continuum space, here for simplicity in the presentation we focus on lattice systems.

**Lattice spins systems.** We consider Gibbs measures for lattice systems on $\mathbb{Z}^d$. If we let $S$ be the configuration space of a single particle at a single site $x \in \mathbb{Z}^d$, then $S^X$ is the configuration space for the particles in $X \subset \mathbb{Z}^d$; we denote by $\sigma_X = \{\sigma_x\}_{x\in X}$ an element of $S^X$. We will be interested in large systems so we let $\Lambda_N = \{x \in \mathbb{Z}^d, |x_i| \le n\}$ denote the square lattice with $N = (2n+1)^d$ lattice sites. We shall use the shorthand notation $\lim_N$ to denote taking limit along the increasing sequence of lattices $\Lambda_N$ which eventually cover $\mathbb{Z}^d$.

**Hamiltonians, interactions, and Gibbs measures.** To specify a Gibbs measure we specify the energy $H_N(\sigma_{\Lambda_N})$ of a set of particles in the region $\Lambda_N$. It is convenient to introduce the concept of an interaction $\Phi = \{\Phi_X : X \subset \mathbb{Z}^d, X\text{finite}\}$ which associates to any finite subset $X$ a function $\Phi_X(\sigma_X)$ which depends only on the configuration in $X$. We shall always assume that interactions are translation-invariant, that is for any $X \subset \mathbb{Z}^d$ and any $a \in \mathbb{Z}^d$, $\Phi_{X+a}$ is obtained by translating $\Phi_X$. For translation-invariant interactions we have the norm [32]

$$|||\Phi||| = \sum_{X\ni 0} |X|^{-1} \|\Phi_X\|_\infty \tag{54}$$

and denote by $\mathcal{B}$ the corresponding Banach space of interactions. Given an interaction $\Phi$ we then define the Hamiltonian $H_N^\Phi$ (with free boundary conditions) by

$$H_N^\Phi(\sigma_{\Lambda_N}) = \sum_{X\subset\Lambda_N} \Phi_X(\sigma_X), \tag{55}$$

and Gibbs measure $\mu_N^\Phi$ by

$$d\mu_N^\Phi(\sigma_{\Lambda_N}) = \frac{1}{Z_N^\Phi} e^{-H_N(\sigma_{\Lambda_N})} dP_N(\sigma_{\Lambda_N}), \tag{56}$$

where $P_N$ is the counting measure on $S^{\Lambda_N}$ and $Z_N^\Phi = \sum_{\sigma_{\Lambda_N}} e^{-H_N(\sigma_{\Lambda_N})}$ is the normalization constant. The norm $|||\Phi|||$ provides a bound on the energy per lattice site since we have $|||H_N^\Phi||| \le N|||\Phi|||$, see Proposition II.1.1C in [32]. In a similar way one can consider periodic boundary conditions or more general boundary conditions, see [32] for details.

**Example: Ising model.** For the $d$-dimensional nearest neighbor Ising model at inverse temperature $\beta$ we have

$$H_N(\sigma_{\Lambda_N}) = -\beta J \sum_{\langle x,y\rangle\subset\Lambda_N} \sigma(x)\sigma(y) - \beta h \sum_{x\in\Lambda_N} \sigma(x)$$

where $\langle \sigma, \sigma' \rangle$ denotes a pair of neighbors with $\sup_i |x_i - y_i| = 1$. So we have

$$\Phi_X = \begin{cases} -\beta J\sigma(x)\sigma(y), & X = \{x, y\}, \\ -\beta h\sigma(x), & X = \{x\}, \\ 0 & \text{otherwise}, \end{cases}$$

and it is easy to see that (54) becomes

$$|||\Phi||| = \beta(d|J| + |h|).$$

**Observables.** We will consider observables of the form

$$f_N(\sigma_{\Lambda_N}) = \frac{1}{N} \sum_{x\in\Lambda_N} g(\sigma_x) \tag{57}$$

for some observable $g$. It will be useful to note that $Nf_N$ is nothing but Hamiltonian $H_N^{\Gamma^g}$ for the interaction $\Gamma^g$ with

$$\Gamma^g_{\{x\}} = g, \quad \text{and} \quad \Gamma^g_X = 0 \quad \text{if} \quad X \neq \{x\}. \tag{58}$$

**UQ bounds for Gibbs measures in finite volume.** Given two Gibbs measures $\mu_N^\Phi$ and $\mu_N^\Psi$ straightforward computations show that for the relative entropy we have

$$R(\mu_N^\Psi \,\|\, \mu_N^\Phi) = \log Z_N^\Phi - \log Z_N^\Psi + E_{\mu_N^\Psi}(H_N^\Phi - H_N^\Psi), \tag{59}$$

while for the cumulant generating function we have

$$\widetilde{\Lambda}_{\mu_N^\Phi, Nf_N}(c) = \log Z_N^{\Phi - c\Gamma^g} - \log Z_N^\Phi - cE_{\mu_N^\Phi}(Nf_N) \tag{60}$$

and thus we obtain immediately from the results in Section 3.1

**Proposition 5.1.** *(Finite volume UQ bounds for Gibbs measures) For two Gibbs measures $\mu_N^\Phi$ and $\mu_N^\Psi$ we have the bound*

$$\frac{1}{N}\Xi_-(\mu_N^\Psi \,\|\, \mu_N^\Phi; Nf_N) \le E_{\mu_N^\Psi}(f_N) - E_{\mu_N^\Phi}(f_N) \le \frac{1}{N}\Xi_+(\mu_N^\Psi \,\|\, \mu_N^\Phi; Nf_N) \tag{61}$$

*where*

$$\Xi_+(\mu_N^\Psi \,\|\, \mu_N^\Phi; Nf_N) = \inf_{c>0} \frac{1}{c}\left\{\log Z_N^{\Phi - c\Gamma^g} - \log Z_N^\Psi + E_{\mu_N^\Psi}(H_N^\Phi - H_N^\Psi) - cE_{\mu_N^\Phi}(Nf_N)\right\} \tag{62}$$

$$\Xi_-(\mu_N^\Psi \,\|\, \mu_N^\Phi; Nf_N) = \sup_{c>0}(-\frac{1}{c})\left\{\log Z_N^{\Phi + c\Gamma^g} - \log Z_N^\Psi + E_{\mu_N^\Psi}(H_N^\Phi - H_N^\Psi) + cE_{\mu_N^\Phi}(Nf_N)\right\}. \tag{63}$$

**UQ bounds for Gibbs measures in infinite volume.** In order to understand how the bounds scale with $N$ we note first (see Theorem II.2.1 ∈ [32]) that the following limit exists

$$p(\Phi) = \lim_N \frac{1}{N}\log Z_N^\Phi, \tag{64}$$

and $p(\Phi)$ is called the pressure for the interaction $\Phi$ (and is independent of the choice of boundary conditions). The scaling of the other terms in the goal-oriented divergence $\Xi_\pm$ is slightly more delicate. In the absence of first order transition for the Gibbs measure for the interaction $\Psi$ the finite volume Gibbs measures $\mu_N^\Psi$ have a well-defined unique limit $\mu^\Psi$ on $S^{\mathbb{Z}^d}$ which is translation invariant and ergodic with respect to $\mathbb{Z}^d$ translations. In addition we have (see Section III.3 in [32])

$$\lim_N \frac{1}{N}E_{\mu_N^\Psi}(H_N^\Phi) = E_{\mu^\Psi}(A^\Phi) \quad \text{with } A^\Phi = \sum_{X \ni 0}\frac{1}{|X|}\Phi_X$$

and moreover, by [32], $E_{\mu^\Psi}(A^\Phi)$ can also be interpreted in terms of the derivative of the pressure functional

$$E_{\mu^\Psi}(A^\Phi) = -\frac{d}{d\alpha}p(\Psi + \alpha\Phi)|_{\alpha=0}.$$

We obtain therefore the following theorem which is valid in the absence of first order phase transitions.

**Theorem 5.2** *(Infinite-volume UQ bounds for Gibbs measures). Assume that both $\Phi$ and $\Psi$ have corresponding unique, transition invariant and ergodic infinite-volume Gibbs measures $\mu^\Phi$ and $\mu^\Psi$. Then we have the bound*

$$\xi_-(\mu_\Psi \,\|\, \mu_\Phi; g) \le E_{\mu^\Psi}(g) - E_{\mu^\Phi}(g) \le \xi_+(\mu_\Psi \,\|\, \mu_\Phi; g)$$

*where $\Gamma^g$ is given by (58) and,*

$$\xi_+(\mu_\Psi \,\|\, \mu_\Phi; g) = \inf_{c>0}\frac{1}{c}\left\{p(\Phi - c\Gamma^g) - p(\Psi) - \frac{d}{d\alpha}p(\Psi + \alpha(\Phi - \Psi))|_{\alpha=0} + c\frac{d}{d\alpha}p(\Phi + \alpha\Gamma^g)|_{\alpha=0}\right\}$$

$$\xi_-(\mu_\Psi \,\|\, \mu_\Phi; g) = \sup_{c>0}(-\frac{1}{c})\left\{p(\Phi + c\Gamma^g) - p(\Psi) - \frac{d}{d\alpha}p(\Psi + \alpha(\Phi - \Psi))|_{\alpha=0} - c\frac{d}{d\alpha}p(\Phi + \alpha\Gamma^g)|_{\alpha=0}\right\}$$

**Remark.** We now readily obtain that the UQ bounds in Theorem 5.2 are applicable for any observables $g = g(\sigma_x)$, and not just for averaged quantities of the type (57). This is the case under the conditions of Theorem 5.2, i.e. that $\mu_\Phi$ and $\mu_\Psi$ are transition invariant and ergodic and that the corresponding pressure terms $p(\Phi \pm c\Gamma^g)$ are finite for some $c > 0$; of course the bounds hold trivially true when they are infinite.

**Phase transitions.** The bound in Theorem 5.2 is useful even in the presence of first order phase transition which manifests itself by the existence of several infinite volume Gibbs measure consistent with the finite volume Gibbs measure (via the DLR condition [53]) or equivalently by the lack of differentiability of the pressure functional $p(\Phi + \alpha\Upsilon)$ for some interaction $\Upsilon$. For example in the 2-d Ising model discussed in Section 6, below the critical temperature the pressure $p(\Phi)$ is not differentiable in $h$ at $h = 0$: there are two ergodic infinite volume Gibbs measures which correspond to the two values of left and right derivatives of the pressure (aka the magnetization). If necessary, in practice one will select a particular value of the magnetization, see the examples in Section 6. Finally, we also note that in the phase transition regime the

linearized bounds (10) and (67) fail since the variance blows up in the $N \to \infty$ limit, [53]; in contrast, the full UQ bounds of Theorem 5.2 remain finite.

**UQ bounds and the use of the triple norm $|||\Phi|||$.** It is not difficult to show (see Proposition II.1.1C and Lemma II.2.2C in [32] and the definition of the triple norm in (54)), that

$$|\log Z_N^{\Psi} - \log Z_N^{\Phi}| \le \|H_N^{\Psi} - H_N^{\Phi}\|_{\infty} \le N|||\Psi - \Phi|||\,, \tag{65}$$

and thus by (59) we have

$$\frac{1}{N} R(\mu_N^{\Psi} \,\|\, \mu_N^{\Phi}) \le 2|||\Psi - \Phi|||\,. \tag{66}$$

Therefore, we obtain the bounds

$$\Xi_+ \le \inf_{c>0}\left\{\frac{1}{c}\widetilde{\Lambda}_{\mu_N^{\Phi},Nf_N}(c) + \frac{2}{c}|||\Psi - \Phi|||\right\},$$
$$\Xi_- \ge \sup_{c>0}\left\{-\frac{1}{c}\widetilde{\Lambda}_{\mu_N^{\Phi},Nf_N}(-c) - \frac{2}{c}|||\Psi - \Phi|||\right\}.$$

These new upper and lower bounds, although they are less sharp, they still scale correctly in system size, while they are intuitive in capturing the dependence of the model discrepancy on the fundamental level of the interaction discrepancy $|||\Psi - \Phi|||$; finally the bounds do not require the computation of the relative entropy, due to upper bound (66).

**Remark.** On the other hand, it is tempting but nevertheless misguided to try to bound $\widetilde{\Lambda}_{\mu_N^{\Phi},Nf_N}(c)$ in terms of interaction norms. Indeed we have the bound $\frac{1}{c}\widetilde{\Lambda}_{\mu_N^{\Phi},Nf_N}(c) \le \|Nf_N - E_{\mu_N^{\Phi}}(Nf_N)\|_{\infty}$. But this bound becomes trivial: since the infimum over $c$ is then attained at $c = \infty$ with the trivial result that $\Xi_+(\mu_N^{\Phi} \,\|\, \mu_N^{\Psi}; Nf_N) \le \|Nf_N - E_{\mu_N^{\Phi}}(Nf_N)\|_{\infty}$ which is independent of $\Psi$ and thus useless.

**Linearized bounds.** Applying the linearized bound (40) to the Gibbs case gives the bound

$$\frac{1}{N}\Xi_{\pm}(\mu_N^{\Psi}\|\mu_N^{\Phi}; Nf_N) = \pm\sqrt{\frac{1}{N}Var_{\mu_N^{\Phi}}\Big(\sum_{x\in\Lambda_N} g(\sigma_x)\Big)}\sqrt{\frac{2}{N}R(\mu_N^{\Psi}\|\mu_N^{\Phi})} + O\big(\frac{1}{N}R(\mu_N^{\Psi}\|\mu_N^{\Phi})\big). \tag{67}$$

In the large $N$ limit, in the absence of first order transition, and if the spatial correlations in the infinite volume Gibbs state decays sufficiently fast then the variance term converges to the integrated auto-correlation function [32]

$$\begin{aligned}\lim_N \frac{1}{N} Var_{\mu_N^{\Phi}}\Big(\sum_{x\in\Lambda_N} g(\sigma_x)\Big) &= \sum_{x\in\mathbb{Z}^d} E_{\mu^{\Phi}}\left((g(\sigma_x) - E_{\mu^{\Phi}}(g))(g(\sigma_0) - E_{\mu^{\Phi}}(g))\right)\\ &= \frac{d^2}{dc^2} P(\Phi - c\Gamma^g)|_{c=0}\end{aligned} \tag{68}$$

which is also known as susceptibility in the statistical mechanics literature.

Finally, we get a simple and easy to implement linearized UQ bound when we replace (66) in (67), namely

$$\frac{1}{N}\Xi_{\pm}(\mu_N^{\Psi}\|\mu_N^{\Phi}; Nf_N) = \pm 2\sqrt{\frac{1}{N}Var_{\mu_N^{\Phi}}\Big(\sum_{x\in\Lambda_N} g(\sigma_x)\Big)}\sqrt{|||\Psi - \Phi|||} + O(|||\Psi - \Phi|||). \tag{69}$$

Each one of terms on the right hand side of (69) can be either computed using Monte Carlo simulation or can be easily estimated, see for instance the calculation of $|||\Psi - \Phi|||$ in the Ising case earlier.

## 6. UQ for phase diagrams of molecular systems

In this section, we will consider the Gibbs measures for one and two-dimensional Ising and mean field models, which are exactly solvable models, see e.g. [54]. We also note that mean field models can be obtained as a minimizer of relative entropy in the sense of (2), where $P$ is an Ising model and $\mathcal{Q}$ is a parametrized family of product distributions, [3].

Here we will demonstrate the use of the goal-oriented divergence, discussed earlier in Section 3.1 and Section 5, to analyze uncertainty quantification for sample mean observables such as the mean magnetization

$$f_N = \frac{1}{N}\sum_{x\in\Lambda_N}\sigma(x), \tag{70}$$

in different phase diagrams based on these models. We use exactly solvable models as a test bed for the accuracy of our bounds, and demonstrate their tightness even in phase transition regimes. In Appendix C.1, we give some background about one/two-dimensional Ising models and mean field models and recall some well-known formulas.

**Towards the evaluation of the UQ bounds.** The results in Sections 3.3 and 5 demonstrate mathematically that the bounds relying on the goal oriented divergences $\Xi_{\pm}$ are the only available ones that scale properly for long times and high dimensional systems. Therefore we turn our attention to the evaluation of these bounds. First we note that the bounds depending of the triple norms $|||\cdot|||$, as well as the linearized bounds of Section 5 provide implementable upper bounds, see also the strategies in [47] for the linearized regime, which are related to sensitivity screening.

By contrast, here we focus primarily on exact calculations of the goal oriented divergences $\Xi_{\pm}$, at least for cases where either the Ising models are exactly solvable or in the case where the known (surrogate) model is a mean field approximation. We denote by $\mu_N$ the Gibbs measures of the model we assume to be known and $\mu'_N$ the Gibbs measure of the model we try to estimate. Then from (61)–(63), recalling that $\Lambda_{\mu_N,Nf_N}(c) = \tilde{\Lambda}_{\mu_N,Nf_N}(c) + cE_{\mu_N}(Nf_N)$, we can rewrite the bounds as

$$E_{\mu'_N}(f_N) \geq \sup_{c>0}\left\{-\frac{1}{cN}\Lambda_{\mu_N,Nf_N}(-c) - \frac{1}{cN}R(\mu'_N\,\|\,\mu_N)\right\},$$
$$E_{\mu'_N}(f_N) \leq \inf_{c>0}\left\{\frac{1}{cN}\Lambda_{\mu_N,Nf}(c) + \frac{1}{cN}R(\mu'_N\,\|\,\mu_N)\right\},$$

and obtain an explicit formula for each term in the large $N$ limit in terms of the pressure, mean energy and magnetization for the models. In the figures below we will display the upper and lower bounds using simple optimization algorithm in Matlab to find the optimal $c$ in the bounds. Note that in the absence of exact formulas we would need to rely on numerical sampling of those quantities, an issue we will discuss elsewhere.

For completeness and for comparison with the exact bounds we will also use and display the approximate linearized bounds

$$E_{\mu'_N}(f_N) \geq E_{\mu_N}(f_N) - \sqrt{\frac{1}{N}Var_{\mu_N}(Nf_N)}\sqrt{\frac{2}{N}R(\mu'_N\,\|\,\mu_N)} + O(\frac{1}{N}R(\mu'_N\,\|\,\mu_N)),$$
$$E_{\mu'_N}(f_N) \leq E_{\mu_N}(f_N) + \sqrt{\frac{1}{N}Var_{\mu_N}(Nf_N)}\sqrt{\frac{2}{N}R(\mu'_N\,\|\,\mu_N)} + O(\frac{1}{N}R(\mu'_N\,\|\,\mu_N)),$$

where each term is computable in the large $N$ limit in terms of the pressure, susceptibility, magnetization, and so on.

### 6.1. Three examples of UQ bounds for phase diagrams

Next we consider three cases where our methods provide exact UQ bounds for phase diagrams between two high dimensional probabilistic models. Here we compare three classes of Gibbs measures for Ising models. (1) Mean field models with different parameters well beyond the linear response regime, (2) Ising models compared to their mean field approximations, and (3) Ising models with vastly different parameters. All these examples cannot be handled with conventional arguments such as linear response theory because they fall into two categories: either, (a) the models have parameters differing significantly, for instance by at least 50%, or (b) the comparison is between different models, e.g. a complex model and a simplified surrogate model which is a potentially inaccurate approximation such as the mean field of the original Ising model.

**(1) Mean field versus mean field models.** Firstly, we consider two mean field models, assume $\mu_{N;mf}$ and $\mu'_{N;mf}$ are their Gibbs measures (probabilities) defined in Appendix C.1 with $h_{mf} = h + dJm$ and $h'_{mf} = h' + dJ'm'$, respectively. By some straightforward calculation in Appendix C.2, we obtain the ingredients of the UQ bounds discussed earlier in the Section:

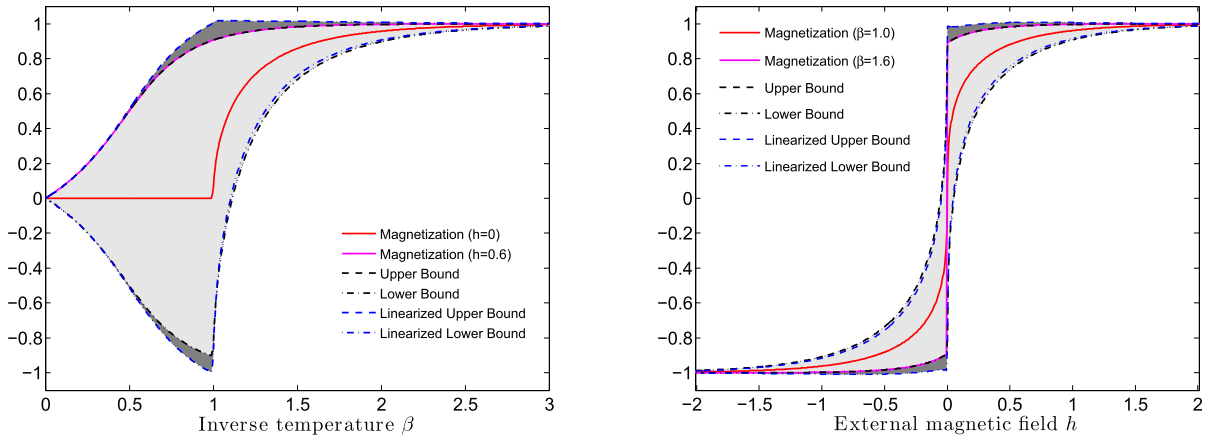
$$\frac{1}{N}R(\mu'_{N;mf}\|\mu_{N;mf}) = \log\frac{e^{\beta h_{mf}} + e^{-\beta h_{mf}}}{e^{-\beta' h'_{mf}} + e^{\beta' h'_{mf}}} + (\beta' h'_{mf} - \beta h_{mf})m', \tag{71}$$

$$\frac{1}{N}\Lambda_{\mu_{N;mf},Nf_N}(c) = \log\frac{e^{(c+\beta h_{mf})} + e^{-(c+\beta h_{mf})}}{e^{-\beta h_{mf}} + e^{\beta h_{mf}}}, \tag{72}$$

and

$$\frac{1}{N}Var_{\mu_{N;mf}}(\sum_{x\in\Lambda_N}\sigma(x)) = 1 - m^2, \tag{73}$$

where $m$ and $m'$ are the magnetizations (70) of these two mean field models and can be obtained by solving the implicit equation (C.21). Here we note that the solution of the equation (C.21) when $h = 0$ has a super-critical pitchfork bifurcation.




(a) (b)

**Fig. 2.** (a): The red solid line is the magnetization for $h = 0$; the magenta solid line is the magnetization for $h = 0.6$. The black dashed/dash–dot line is the upper/lower UQ bound given by the goal-oriented divergences of the magnetization (70) for $h = 0.6$. The blue dashed/dash–dot line is the linearized upper/lower bound. All gray areas (both light and darker gray) depict the size of the uncertainty region corresponding to linearized bounds. The narrower lighter gray area corresponds to the goal-oriented bounds. (b): The red solid line is the magnetization for $\beta = 1$; the magenta solid line is the magnetization for $\beta = 1.6$. The black dashed/dash–dot line is the upper/lower goal-oriented divergence bound of the magnetization for $\beta = 1.6$. The blue dashed/dash–dot line is the linearized upper/lower bound. All gray areas (both light and darker gray) depict the size of the uncertainty region corresponding to linearized bounds. The narrower lighter gray area corresponds to the goal-oriented bounds. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In our discussion regarding mean field vs mean field and 1-d Ising vs mean field models we only consider the upper branch of the stable solution. But, in our discussion about 2d Ising vs mean field, we consider both upper and lower branches.

In Appendix C.1, we can calculate the magnetizations, the goal-oriented divergence bounds and their corresponding linearized bounds which we use in deriving exact formulas for the UQ bounds. Indeed, for Fig. 2(a), we set $J = 2$ and consider the Gibbs measure of the 1-d mean field model with $h = 0$ as the benchmark and plot the magnetization based on this distribution as a function of inverse temperature $\beta$. Then, we perturb the external magnetic field to $h = 0.6$ and consider the Gibbs measure with this external magnetic field. We plot the goal-oriented divergence bounds of the magnetization of the Gibbs measure with $h = 0.6$ as a function of $\beta$ as well as their corresponding linearized approximation in this figure. To test the sharpness of these bounds, we also give the magnetization with $h = 0.6$ in the figure. We can see that the upper bound almost coincides with the magnetization. The linearized approximation works well at low temperature, but, it does not work as well as the goal-oriented bound around the critical point. The reason for this is that relative entropy between those two measures is bigger here due to the bigger perturbation of $h$ and linearization is a poor approximation of the bounds. Also, we can see that the magnetization vanishes at high temperatures for $h = 0$. At low temperatures it approaches its maximum value $m = 1$. For non-zero $h$, we see that there is no phase transition and the magnetization increases gradually from close to $m = 0$ at high temperatures ($\beta \ll 1$) to $m = 1$ at low temperatures ($\beta \gg 1$).

In Fig. 2(b), we set $J = 1$ and consider the Gibbs measure of the 1-d mean field model with $\beta = 1$ as the benchmark and plot the magnetization based on this measure as a function of $h$. Then we perturb $\beta$ by 60% and obtain another Gibbs measure with $\beta = 1.6$ that has a phase transition at $h = 0$. In the figure, we give the upper/lower goal-oriented divergence bounds of the magnetization based on the Gibbs measure with $\beta = 1.6$ as well as their corresponding linearized bounds. To test the sharpness of the bounds, we also plot the magnetization with $\beta = 1.6$ as a function of $h$. We can see the upper bound almost coincide with the magnetization when $h$ is positive and the lower bound almost coincide with the magnetization when $h$ is negative. Similarly with Fig. 2(a), the linearized bounds make a relatively poor estimation around the critical point $h = 0$ because of the bigger relative entropy between these two measures.

**(2a) One-dimensional Ising model versus mean field.** Consider the 1-d Ising model and mean field model and assume $\mu_N$ and $\mu_{N;mf}$ are respectively their Gibbs distributions, which are defined in Appendix C.1. Then, by straightforward calculations, we obtain

$$\lim_N \frac{1}{N} R(\mu_N \| \mu_{N;mf}) = \log \frac{e^{\beta[h+Jm]} + e^{-\beta[h+Jm]}}{e^{\beta J}\cosh(\beta h) + k_1} + \frac{\beta J}{k_1}\left(k_1 - \frac{2e^{-2\beta J}}{e^{\beta J}\cosh(\beta h) + k_1} - m e^{J\beta}\sinh(h\beta)\right) \tag{74}$$

where $k_1 = \sqrt{e^{2J\beta}\sinh^2(h\beta) + e^{-2J\beta}}$; detailed calculations can be found in Appendix C.2. By (72) and (73), we have

$$\frac{1}{N}\Lambda_{N;mf,Nf_N}(c) = \log \frac{e^{[c+\beta(h+Jm)]} + e^{-[c+\beta(h+Jm)]}}{e^{-\beta[h+Jm]} + e^{\beta[h+Jm]}}, \tag{75}$$

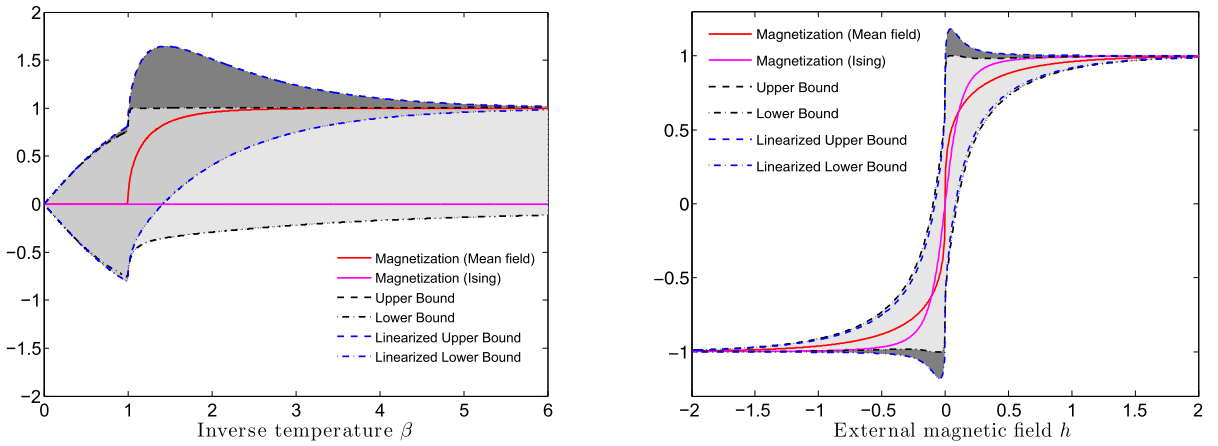



**Fig. 3.** (a): The red solid line is the magnetization of 1-d mean field model for $h = 0$; the magenta solid line is the magnetization of 1-d Ising model for $h = 0$. The black dashed/dash–dot line is the upper/lower goal-oriented divergence bound of the magnetization of the Ising model. The blue dashed/dash–dot line is the linearized upper/lower bound. The sum of the light gray and medium gray areas depict the size of the uncertainty region corresponding to linearized bounds. The sum areas of medium gray and dark dray corresponds to the goal-oriented bounds. (b): The red solid line is the magnetization of 1-d mean field model for $\beta = 1.0$; the magenta solid line is the magnetization of 1-d Ising model $\beta = 1.0$. The black dashed/dash–dot line is the upper/lower goal-oriented divergence bound of the magnetization of the Ising model. The blue dashed/dash–dot line is the linearized upper/lower bound. All gray areas (both light and darker gray) depict the size of the uncertainty region corresponding to linearized bounds. The narrower lighter gray area corresponds to the goal-oriented bounds. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and

$$\frac{1}{N} Var_{\mu_{N;mf}}\Big(\sum_{x\in\Lambda_N} \sigma(x)\Big) = 1 - m^2. \tag{76}$$

Combining with Appendix C.1, for given parameters, we can calculate the magnetizations, the goal-oriented divergence bounds and their corresponding linearized approximation.

In Fig. 3(a), we set $h = 0$ and $J = 1$ and consider the Gibbs measure of the mean field model as the benchmark, that is we use it as a surrogate model for the Ising model. In the figure, we see that its magnetization vanishes at high temperatures. At low temperatures it goes to its maximum value $m = 1$, exhibiting spontaneous magnetization and a phase transition at the inverse temperature $\beta = 1$. We plot the upper/lower goal-oriented divergence bound as well as their corresponding linearized bounds of the magnetization as a function of $\beta$. To test the sharpness of these bounds, we also plot the magnetization of the Ising model in the figure. The magnetization of the Ising model vanishes for all temperatures, exhibiting no phase transitions. In this sense the mean field approximation of the Ising model is a very poor one and the UQ bounds depicted in Fig. 3(a) capture and quantify the nature of this approximation. Furthermore, the linearized lower bound fails in the sense that is not informative for low temperatures because of the considerable difference between the models $\mu_N$ and $\mu_{N;mf}$. In Fig. 3(b), we set $\beta = 1$ and $J = 1$ and consider the bounds and the magnetizations as a function of the external field $h$. Similarly with Fig. 3(a), we take the Gibbs measure of the mean field model as the benchmark. To test the sharpness of the bounds, we also plot the magnetization of the Ising model in the figure. The goal-oriented divergence bounds work well since the uncertainty region given by them is in this case small. And the upper bound almost coincides with it for positive $h$ and the lower bound almost coincide with it for negative $h$. The uncertainty region given by the linearized bounds is wider around $h = 0$, which shows the linearized bounds do not give as good an approximation to the UQ region around $h = 0$.

**(2b) Two-dimensional Ising model versus mean field.** We revisit the example in (2a) above but this time in two dimensions where the Ising model exhibits phase transitions at a finite temperature. WE denote by $\mu_N$ and $\mu_{N;mf}$ the Gibbs distributions for the two-dimensional zero-field Ising model and two-dimensional mean field model with $h_{mf} = 2Jm$, respectively. Then, by straightforward calculations, we obtain

$$\lim_N \frac{1}{N} R(\mu_N \| \mu_{N;mf}) = \log[e^{-2\beta Jm} + e^{2\beta Jm}] - \frac{\log 2}{2} - \frac{1}{2\pi}\int_0^\pi \log[\cosh^2(2\beta J) + k(\theta)]d\theta$$
$$+ \beta J \frac{\sinh(4\beta J)}{\pi}\int_0^\pi \frac{1}{k(\theta)}\Big[1 - \frac{1+\cos(2\theta)}{\cosh^2(2\beta J) + k(\theta)}\Big]d\theta - 2\beta JmM_0, \tag{77}$$

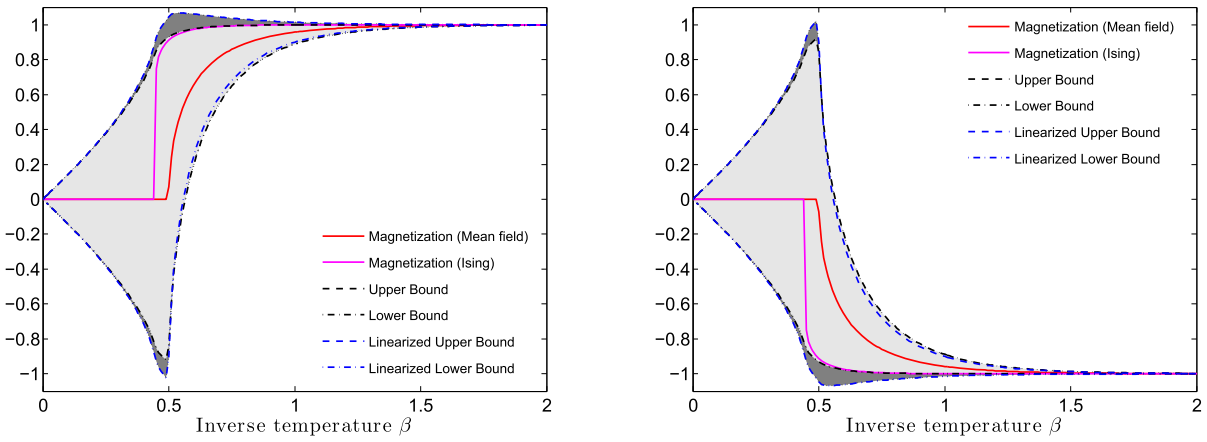



**Fig. 4.** (a) The red solid line is the spontaneous magnetization of the 2-d mean field model with $h = 0^+$; the magenta solid line is the spontaneous magnetization of 2-d Ising model with $h = 0^+$; the black dashed/dash–dot line is the upper/lower goal-oriented divergence bound of the magnetization for Ising model; the blue dashed/dash–dot line is the linearized upper/lower bound. All gray areas (both light and darker gray) depict the size of the uncertainty region corresponding to linearized bounds. The narrower lighter gray area corresponds to the goal-oriented bounds. (b) The red solid line is the spontaneous magnetization of 2-d mean field model with $h = 0^-$; the magenta solid line is the spontaneous magnetization of 2-d Ising model with $h = 0^-$; the black dashed/dash–dot line is the upper/lower goal-oriented divergence bound of the magnetization for Ising model; the blue dashed/dash–dot line is the linearized upper/lower bound. All gray areas (both light and darker gray) depict the size of the uncertainty region corresponding to linearized bounds. The narrower lighter gray area corresponds to the goal-oriented bounds. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\frac{1}{N}\Lambda_{\mu_{N;mf},Nf_N}(c) = \log\frac{e^{(c+2\beta Jm)} + e^{-(c+2\beta Jm)}}{e^{-2\beta Jm} + e^{2\beta Jm}} \tag{78}$$

and

$$\frac{1}{N}Var_{\mu_{N;mf}}\Big(\sum_{x\in\Lambda_N}\sigma(x)\Big) = 1 - m^2, \tag{79}$$

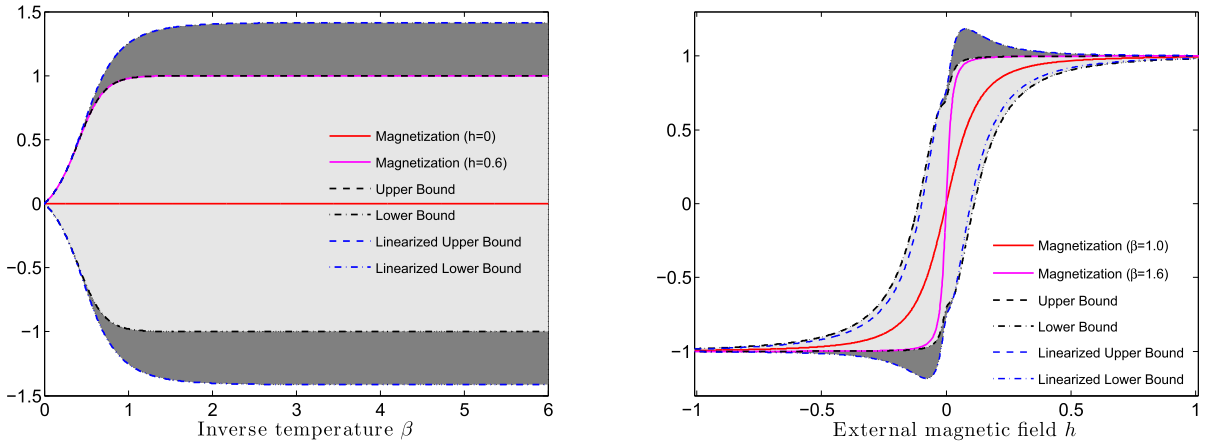
where $m$ and $M_0$ are the spontaneous magnetizations of the two-dimensional mean field model and Ising models, respectively and can be obtain by solving (C.21) and (C.13). Detailed calculations can be found in Appendix C.2. Combining with Appendix C.1, for given parameters, we can calculate the magnetizations, the goal-oriented divergence bounds and their corresponding linearized approximation.

In Fig. 4(a), we set $h = 0$ and $J = 1$ and plot the bounds and the magnetizations as a function of inverse temperature $\beta$. Similarly with Fig. 3, we take the Gibbs measure of the mean field as the benchmark and consider the bounds for the magnetization of the Ising model. We can readily see that the goal-oriented bounds work well in low temperatures. Notice the large uncertainty prior to the onset of the spontaneous magnetization (phase transition) which is due to a pitchfork bifurcation and the two branches (upper and lower) reported in Fig. 1b, as well as in the panels in Fig. 4. The linearized bounds also work well, but they are not as sharp as the goal-oriented divergence bounds around the critical points because of the larger value of the relative entropy $R(\mu_N\|\mu_{N;mf})$. There are phase transitions for both mean field model and Ising model. The critical points are $1/2$ and $\log(1+\sqrt{2})/2$ for mean field model and Ising model, respectively. Both their magnetizations vanish at high temperatures and go to their maximum values 1 at low temperature.

Actually, the spontaneous magnetizations we consider in Fig. 4(a) are both based on the definition $M = \lim_{h\to 0^+}\langle\sigma(x)\rangle$. If we consider the definition $M = \lim_{h\to 0^-}\langle\sigma(x)\rangle$, we can obtain another figure which is Fig. 4(b). We can see the quantities in Fig. 4(b) are just the opposite of the corresponding quantities in Fig. 4(a). Combining both figures gives us the uncertainty region reported in the Introduction.

**(3) 1-d Ising model versus 1-d Ising model.** Consider two one-dimensional Ising models and $\mu_N$ and $\mu'_N$ are their Gibbs distributions defined in Appendix C.1. By straightforward calculation, we have

$$\lim_N \frac{1}{N}R(\mu'_N\|\mu_N) = \log\frac{e^{\beta J}\cosh(\beta h) + \sqrt{e^{2J\beta}\sinh^2(h\beta) + e^{-2J\beta}}}{e^{\beta' J'}\cosh(\beta' h') + \sqrt{e^{2J'\beta'}\sinh^2(h'\beta') + e^{-2J'\beta'}}}$$




(a) (b)

**Fig. 5.** (a) The red solid line is the magnetization of 1-d Ising model for $h = 0$; the magenta solid line is the magnetization of 1-d Ising model for $h = 0.6$; the black dashed/dash–dot line is the upper/lower bound by goal-oriented divergence; the blue dashed/dash–dot line is the linearized upper/lower bound. All gray areas (both light and darker gray) depict the size of the uncertainty region corresponding to linearized bounds. The narrower lighter gray area corresponds to the goal-oriented bounds. (b) The red solid line is the magnetization of 1-d Ising model for $\beta = 1$; the magenta solid line is the magnetization of 1-d Ising model for $\beta = 1.6$; the black dashed/dash–dot line is the upper/lower bound by goal-oriented divergence; the blue dashed/dash–dot line is the linearized upper/lower bound. All gray areas (both light and darker gray) depict the size of the uncertainty region corresponding to linearized bounds. The narrower lighter gray area corresponds to the goal-oriented bounds. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$+ (\beta' J' - \beta J)(1 - \frac{1}{k_1'} \frac{2e^{-2\beta' J'}}{e^{\beta' J'} \cosh(\beta' h') + k_1'}) + (\beta' h' - \beta h) \frac{1}{k_1'} e^{J'\beta'} \sinh(h'\beta') \tag{80}$$

and

$$\lim_N \frac{1}{N} Var_{\mu_N}(\sum_{x \in \Lambda_N} \sigma(x)) = \frac{1}{k_1^3} e^{-J\beta} \cosh(h\beta), \tag{81}$$

where $k_1' = \sqrt{e^{2J'\beta'} \sinh^2(h'\beta') + e^{-2J'\beta'}}$. The cumulant generating function is

$$\lim_N \frac{1}{N} \Lambda_{\mu_N}(c) = \log \frac{e^{\beta J} \cosh(\beta h + c) + \sqrt{e^{2J\beta} \sinh^2(h\beta + c) + e^{-2J\beta}}}{e^{\beta J} \cosh(\beta h) + \sqrt{e^{2J\beta} \sinh^2(h\beta) + e^{-2J\beta}}}, \tag{82}$$

detailed calculations can be found in Appendix C.2. Combining with Appendix C.1, for given parameters, we can calculate the magnetizations, the bounds given by goal-oriented divergence and their corresponding linearized approximation.

In Fig. 5(a), we set $J = 1$ and plot the magnetizations of 1-d Ising model as a function of inverse temperature $\beta$ for $h = 0$ and $h = 0.6$, respectively. For the zero-field Ising model, used here as our benchmark, the magnetization vanishes for all temperatures. For $h = 0.6$, the magnetization increases gradually to its maximum 1. Clearly the models are far apart but the UQ bounds work remarkably well. Indeed, we plot the upper/lower goal-oriented divergence bound of the magnetization for the nonzero-field Ising model. The upper bound almost coincides with the magnetization itself. The lower bound is poor due to the symmetry of the bounds in $h$. If we break the symmetry by comparing models for different positive external fields both bounds become much sharper (not shown). The linearized bounds give a good approximation at high temperatures. However, at low temperatures, they are not as sharp as the goal-oriented divergence bounds. This is due to the larger relative entropy $R(\mu\|\mu')$ between $\mu$ and $\mu'$. In Fig. 5(b), we plot the magnetization of the one-dimensional Ising model as a function of $h$ for two different inverse temperatures $\beta = 1$ and $\beta = 1.6$. The parameter $J$ was set to 1. We also plot the upper/lower goal-oriented divergence bounds for $\beta = 1.6$. Similarly with Fig. 5(a), we also plot the linearized upper/lower bound in the figure. The goal-oriented divergence bounds work well here. We can see the upper bound almost coincides with the magnetization when $h$ is positive and the lower bound almost coincides with the magnetization when $h$ is negative. For $\beta = 1.6$, there is a phase transition at the point $h = 0$ and the linearized bounds make a relatively poor estimation around $h = 0$ since there the models are far apart, see Fig. 5(b).

## 7. Applicability to more general quantities of interest

As we discussed in the Remark following Theorem 5.2, the UQ bounds (8) are also applicable to observables which are not limited to averaged quantities such as (57), or for QoIs satisfying (11). Indeed this is the case under the condition that the probability measures in the UQ bounds of Theorem 5.2 are transition invariant and ergodic, and for general observables for which the pressure terms in the UQ bounds are finite.

Furthermore, an important special case of applicability of our bounds arises when we consider localized perturbations to statistical models such as the Gibbs measures in Section 5. Then, the corresponding interaction $\Psi$ includes only local perturbations to the interaction $\Phi$, e.g. local defects encoded in the interaction potential $J = J(x, y)$, or localized perturbations to the external field $h = h(x)$ in the case of Ising-type models. Defects of finite temperature multiscale models are a continuous source of interest in the computational materials science community, see for instance [55], and the lattice models considered in Section 5 constitute an important class of simplified prototype problems. In the case of localized perturbations to the interaction $\Phi$ in (55) we do not have anymore the scaling (7); in fact the Hamiltonians scale as

$$H_N^{\Psi}(\sigma_{\Lambda_N}) = H_N^{\Phi}(\sigma_{\Lambda_N}) + O(1),$$

and thus the corresponding relative entropy satisfies

$$R(\mu_N^{\Psi} \,\|\, \mu_N^{\Phi}) = O(1),$$

uniformly in the system size $N$. Then, the asymptotic expansion (10) readily yields that the only necessary condition for the QoIs $f_N$ is that

$$Var_P(f_N) = O(1), \tag{83}$$

uniformly in the system size $N$. A corresponding result can be easily obtained for $\Xi_{\pm}(Q \,\|\, P; f)$ and the full UQ bound (8). In this context, we can obtain UQ bounds for wide class of observables, for example purely local observables pertinent to the defects, that do not need to be averages such as (57) satisfying the condition (11).

## 8. Conclusions

In this paper we first showed that the classic information inequalities such as Pinsker-typer inequalities and other inequalities based on the Hellinger distance, the $\chi^2$-divergence, or the Rényi divergence perform poorly for the purpose of controlling QoIs of systems with many degrees of freedom, and/or in long time regimes. On the other hand we demonstrated that the goal oriented divergence introduced in [33] scales properly and allows to control QoIs provided they can be written as ergodic averages or spatial averages, e.g. quantities like autocorrelation, mean magnetization, specific energy, and so on. We illustrated the potential of our approach by computing uncertainty quantification bounds for phase diagrams for Gibbs measures, that is for systems in the thermodynamic limit. We showed that the bounds perform remarkably well even in the presence of phase transitions.

Although we provided computable bounds and exact calculations, there is still a lot to be done towards developing efficient Monte Carlo samplers for the goal oriented divergences $\Xi_{\pm}$, which is a central mathematical object in our approach. An additional strength of our approach is that it also applies to non-equilibrium systems which do not necessarily satisfy detailed balance, providing robust nonlinear response bounds. The key insight here is to study the statistical properties of the paths of the systems and to use the thermodynamic formalism for space–time Gibbs measures. Our results can be applied to a wide range of problems in statistical inference, coarse graining of complex systems, steady states sensitivity analysis for non-equilibrium systems and Markov random fields.

## Acknowledgements

The research of MAK and JW was partially supported by the Defense Advanced Research Projects Agency (DARPA) EQUiPS program under the grant W911NF1520122. The research of LRB was partially supported by the National Science Foundation (NSF) under the grant DMS-1515712.

## Appendix A. Hellinger-based inequalities

**Lemma A.1.** *Suppose $P$ and $Q$ be two probability measures on some measure space $(\mathcal{X}, \mathcal{A})$ and let $f : \mathcal{X} \to \mathbb{R}$ be some quantity of interest (QoI), which is measurable and has second moments with respect to both $P$ and $Q$. Then*

$$\left|E_Q(f) - E_P(f)\right| \le \sqrt{2} H(Q, P) \sqrt{Var_P(f) + Var_Q(g) + \frac{1}{2}(E_Q(f) - E_P(f))^2}. \tag{A.1}$$

**Proof.** By Lemma 7.14 in [41], we have

$$\left|E_Q(f) - E_P(f)\right| \le \sqrt{2} H(Q, P) \sqrt{E_P(f^2) + E_Q(f^2)}.$$

For any $c \in \mathbb{R}$, replace $f$ by $f-c$, we have

$$|E_Q(f) - E_P(f)| = |E_P(f-c) - E_Q(f-c)| \leq \sqrt{2}H(Q,P)\sqrt{E_P((f-c)^2) + E_Q((f-c)^2)}.$$

Thereby,

$$|E_Q(f) - E_P(f)| \leq \inf_c \sqrt{2}H(Q,P)\sqrt{E_P((f-c)^2) + E_Q((f-c)^2)}$$

By some straight calculations, we can find the optimal c is:

$$c^* = \frac{E_P(f) + E_Q(f)}{2}.$$

Thus, we have

$$|E_Q(f) - E_P(f)| \leq \sqrt{2}H(Q,P)\sqrt{E_Q[(f-c^*)^2] + E_P[(f-c^*)^2]} = \sqrt{2}H(Q,P)\sqrt{Var_P(f) + Var_Q(g) + \frac{1}{2}(E_Q(f) - E_P(f))^2}. \quad \square$$

## Appendix B. Proof of Lemmas 2.1 and 2.2

### B.1. I.I.D. sequences

**Proof of Lemma 2.1.** Assume $\sigma^N = (\sigma_1, ..., \sigma_N)$, since $P_N$ and $Q_N$ are product measures we have $\frac{dQ_N}{dP_N}(\sigma^N) = \prod_{j=1}^N \frac{dQ}{dP}(\sigma_j)$.

For the relative entropy we have

$$R(Q_N \| P_N) = \int_{\mathcal{X}_N} \log \frac{dQ_N}{dP_N} dQ_N = \int_{\mathcal{X}_N} \sum_{j=1}^N \log \frac{dQ}{dP}(\sigma_j) dQ_N(\sigma^N) = \sum_{j=1}^N \int_{\mathcal{X}} \log \frac{dQ}{dP}(\sigma_j) dQ(\sigma_j) = NR(Q \| P). \tag{B.1}$$

For the Reny relative entropy we have

$$D_\alpha(Q_N \| P_N) = \frac{1}{\alpha-1} \log \int_{\mathcal{X}_N} \left(\frac{dQ_N}{dP_N}(\sigma^N)\right)^\alpha dP_N(\sigma^N) = \frac{1}{\alpha-1} \log \int_{\mathcal{X}_N} \prod_{j=1}^N \left(\frac{dQ}{dP}(\sigma_j)\right)^\alpha dP_N(\sigma^N) = \sum_{j=1}^N \frac{1}{\alpha-1} \log \int_{\mathcal{X}} \left(\frac{dQ}{dP}(\sigma_j)\right)^\alpha dP(\sigma_j) = ND_\alpha(Q \| P). \tag{B.2}$$

For the $\chi^2$ distance we note first that

$$\chi^2(Q \| P) = \int \left(\frac{dQ}{dP} - 1\right)^2 dP = \int \left(\left(\frac{dQ}{dP}\right)^2 - 2\frac{dQ}{dP} + 1\right) dP = \int \left(\frac{dQ}{dP}\right)^2 dP - 1,$$

and therefore we have

$$\chi^2(Q_N \| P_N) = \int_{\mathcal{X}_N} \left(\prod_{j=1}^N \frac{dQ}{dP}(\sigma_j)\right)^2 dP_N(\sigma^N) - 1 = \prod_{j=1}^N \int_{\mathcal{X}} \left(\frac{dQ}{dP}(\sigma_j)\right)^2 dP(\sigma_j) - 1 = \left(1 + \chi^2(Q \| P)\right)^N - 1. \tag{B.3}$$

For the Hellinger distance we note first that

$$H^2(Q,P) = \int \left(\sqrt{\frac{dQ}{dP}} - 1\right)^2 dP = \int \left(\frac{dQ}{dP} - 2\sqrt{\frac{dQ}{dP}} + 1\right) dP = 2 - 2\int \sqrt{\frac{dQ}{dP}} dP,$$

and thus $\int \sqrt{\frac{dP}{dQ}} dQ = 1 - \frac{1}{2}H^2(P,Q)$. Therefore we have

$$H^2(Q_N, P_N) = 2 - 2\int_{\mathcal{X}_N} \sqrt{\prod_{j=1}^N \frac{dQ}{dP}(\sigma_j) dP(\sigma^N)} = 2 - 2\prod_{j=1}^N \int_{\mathcal{X}} \sqrt{\frac{dQ}{dP}(\sigma_j) dP(\sigma_j)}$$
$$= 2 - 2\left(1 - \frac{H^2(Q,P)}{2}\right)^N. \tag{B.4}$$

This concludes the proof of Lemma 2.1. □

### B.2. Markov sequences

**Proof of Lemma 2.2**: The convergence of the relative entropy rate is well known and we give here a short proof for the convenience of the reader.

Recall that $\nu_p$ and $\nu_q$ are the initial distributions of the Markov chain at time 0 with transition matrices $p$ and $q$ respectively. We write $\nu_p^k$ the distribution at time $k$ as a row vector and we have then $\nu_p^k(\sigma) \equiv \nu_p p^k(\sigma)$ where $p^k$ is the matrix product.

By expanding the logarithm and integrating we find

$$\begin{aligned}
&\frac{1}{N}\int \log \frac{dQ_N}{dP_N} dQ_N \\
&= \frac{1}{N} \sum_{\sigma_0,\cdots\sigma_N} \log\left(\frac{\nu_q(\sigma_0)q(\sigma_0,\sigma_1)\cdots q(\sigma_{n-1},\sigma_n)}{\nu_p(\sigma_0)p(\sigma_0,\sigma_1)\cdots p(\sigma_{n-1},\sigma_n)}\right) \nu_q(\sigma_0)q(\sigma_0,\sigma_1)\cdots q(\sigma_{n-1},\sigma_n),\sigma_n) \\
&= \frac{1}{N}\sum_{\sigma_0} \log\frac{\nu_q(\sigma_0)}{\nu_p(\sigma_0)}\nu_q(\sigma_0) + \frac{1}{N}\sum_{k=1}^N \sum_{\sigma_0,\cdots,\sigma_k} \nu_q(\sigma_0)p(\sigma_0,\sigma_1)\cdots q(\sigma_{k-1},\sigma_k)\log\frac{q(\sigma_{k-1},\sigma_k)}{p(\sigma_{k-1},\sigma_k)} \\
&= \frac{1}{N}\sum_{\sigma_0} \log\frac{\nu_q(\sigma_0)}{\nu_p(\sigma_0)}\nu_q(\sigma_0) + \frac{1}{N}\sum_{k=1}^N \sum_{\sigma,\sigma'} \nu_q^k(\sigma)q(\sigma,\sigma')\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}.
\end{aligned} \tag{B.5}$$

The first term goes to 0 as $N \to \infty$ while for the second term, by the ergodic theorem we have that for any initial condition $\nu_q$, $\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^N \nu_q^k = \mu_q$ where $\mu_q$ is stationary distribution. Therefore we obtain that

$$\lim_{N\to\infty} \frac{1}{N}\int \log\frac{dQ_N}{dP_N} dQ_N = \sum_{\sigma,\sigma'} \mu_q(\sigma)q(\sigma,\sigma')\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')}.$$

Finally we note that the limit can be written as a averaged relative entropy, since

$$\sum_{\sigma,\sigma'} \mu_q(\sigma)q(\sigma,\sigma')\log\frac{q(\sigma,\sigma')}{p(\sigma,\sigma')} = \sum_{\sigma} \mu_q(\sigma) R\left(q(\sigma,\cdot) \,\|\, p(\sigma,\cdot)\right).$$

As a consequence the relative entropy rate vanishes if and only if $R\left(q(\sigma,\cdot) \,\|\, p(\sigma,\cdot)\right) = 0$ for every $\sigma$ that is if and only if $q(\sigma,\sigma') = p(\sigma,\sigma')$ for every $\sigma$ and $\sigma'$.

We turn next to Rényi entropy. As it will turn out understanding the scaling properties of the Rényi entropy will allow us immediately to understand the scaling properties of the chi-squared and Hellinger divergences as well. We have

$$\frac{1}{N} D_\alpha(Q_N \,\|\, P_N) = \frac{1}{N}\frac{1}{\alpha-1}\log \sum_{\sigma_0,\cdots\sigma_N} \nu_p(\sigma_0)^{1-\alpha}\nu_q(\sigma_0)^\alpha \prod_{j=1}^N q(\sigma_{j-1},\sigma_j)^\alpha p(\sigma_{j-1},\sigma_j)^{1-\alpha}.$$

Let $F_\alpha$ be the non-negative matrix with entries

$$F_\alpha(\sigma,\sigma') = q(\sigma,\sigma')^\alpha p(\sigma,\sigma')^{1-\alpha}.$$

Since $p$ and $q$ are irreducible and mutually absolutely continuous the matrix $F_\alpha$ is irreducible as well. Let $v$ be the row vector with entries $v(\sigma) = \nu_p(\sigma)^{1-\alpha}\nu_q(\sigma)^\alpha$ and 1 the column vector with all entries equal to 1. Then we have

$$\frac{1}{N} D_\alpha(Q_N \,\|\, P_N) = \frac{1}{\alpha-1} v F_\alpha^N 1,$$

and thus by the Perron–Frobenius Theorem [43], we have

$$\lim_{N\to\infty}\frac{1}{N}D_\alpha(Q_N\,\|\,P_N)=\frac{1}{\alpha-1}\log\rho(\alpha),$$

where $\rho(\alpha)$ is the maximal eigenvalue of the non-negative matrix $F_\alpha$.

It remains to show that the limit is 0 only if $p=q$. In order to do this we will use some convexity properties of the Rényi entropy [38]. For $0<\alpha\le 1$ the Rényi entropy $D_\alpha(Q\,\|\,P)$ is jointly convex in $Q$ and $P$, i.e. for any $\epsilon\in[0,1]$ we have

$$D_\alpha(\epsilon Q_0+(1-\epsilon)Q_1\,\|\,\epsilon P_0+(1-\epsilon)P_1)\le\epsilon D_\alpha(Q_0\,\|\,P_0)+(1-\epsilon)D_\alpha(Q_1\,\|\,P_1).$$

For $\alpha>1$ the Rényi entropy is merely jointly quasi-convex, that is

$$D_\alpha(\epsilon Q_0+(1-\epsilon)Q_1\,\|\,\epsilon P_0+(1-\epsilon)P_1)\le\max\{D_\alpha(Q_0\,\|\,P_0),D_\alpha(Q_1\,\|\,P_1)\}.$$

In any case let us assume that $p\neq q$ is such that

$$\lim_{N\to\infty}\frac{1}{N}D_\alpha(Q_N\,\|\,P_N)=0.$$

Then by convexity, or quasi-convexity we have for any $\epsilon\in[0,1]$

$$\lim_{N\to\infty}\frac{1}{N}D_\alpha(\epsilon Q_N+(1-\epsilon)P_N\,\|\,P_N)=0.$$

On the other hand, for any smooth parametric family $P_\theta$ we have that, [38],

$$D_\alpha(P^{\theta'}\,\|\,P^\theta)=\frac{\alpha}{2}(\theta-\theta')^2\mathbf{F}(P^\theta)+O((\theta'-\theta)^3)$$

where $\mathbf{F}(P^\theta)$ is the Fisher information. If $P^\theta$ is a discrete probability distribution then the Fisher information is $\mathbf{F}(P^\theta)=\sum_\sigma P^\theta(\sigma)(\frac{d}{d\theta}\log P^\theta(\sigma))^2$.

To compute $\mathbf{F}(P_N^\theta)$ we can use the relative entropy $R(P_N^{\theta'}\,\|\,P_N^\theta)=D_1(P_N^{\theta'}\,\|\,Q_N^\theta)$ and from (B.5) with $q=p^{\theta'}$ and $p=p^\theta$ we obtain

$$\begin{aligned}R(P_N^{\theta'}\,\|\,P_N^\theta)&=(\theta'-\theta)^2\sum_\sigma\nu_{p^\theta}(\sigma)\left(\frac{d}{d\theta}\log\nu_{p^\theta}(\sigma)\right)^2\\&\quad+\frac{1}{2}(\theta-\theta')^2\sum_{k=1}^N\sum_{\sigma,\sigma'}(\nu_{p^\theta})^k(\sigma)p^\theta(\sigma,\sigma')\left(\frac{d}{d\theta}\log p^\theta(\sigma,\sigma')\right)^2+O((\theta-\theta')^3).\end{aligned}$$

So as $N\to\infty$ we obtain

$$\lim_{N\to\infty}\frac{1}{N}R(P_N^{\theta'}\,\|\,P_N^\theta)=\frac{1}{2}(\theta-\theta')^2\sum_{\sigma,\sigma'}\mu_{p^\theta}(\sigma)p^\theta(\sigma,\sigma')\left(\frac{d}{d\theta}\log p^\theta(\sigma,\sigma')\right)^2+O((\theta-\theta')^3)\tag{B.6}$$

If we now apply this to the family $P^\epsilon=P_N+\epsilon(Q_N-P_N)$ we have that

$$\lim_{N\to\infty}\frac{1}{N}R(P_N+\epsilon(Q_N-P_N)\,\|\,P_N)=\frac{1}{2}\epsilon^2\sum_{\sigma,\sigma'}\mu_p(\sigma)\frac{(q(\sigma,\sigma')-p(\sigma,\sigma'))^2}{p(\sigma,\sigma')}+O(\epsilon^3)$$

since the term of order $\epsilon^2$ is strictly positive unless $p=q$ this contradicts our assumption that $\lim_{N\to\infty}\frac{1}{N}D_\alpha(\epsilon Q_N+(1-\epsilon)P_N\,\|\,P_N)=0$.

We can now easily deduce the scaling of the $\chi^2$ divergence from the Rényi relative entropy because of the relation $\chi^2(Q_N\,\|\,P_N)=e^{D_2(Q_N\,\|\,P_N)}-1$. This implies that $\chi^2(Q_N\,\|\,P_N)$ grows exponentially in $N$ unless $\lim_{N\to\infty}\frac{1}{N}D_2(Q_N\,\|\,P_N)=0$ which is possible if and only if $p=q$.

Similarly for the Hellinger distance we use the relation $H^2(Q_N,P_N)=2-2e^{-\frac{1}{2}D_{\frac{1}{2}}(Q_N\,\|\,P_N)}$ and the scaling of the Rényi entropy to see $H(Q_N,P_N)$ converges to $\sqrt{2}$ unless $p=q$. This concludes the proof of Lemma 2.2.

## Appendix C. Background for Section 5

### C.1. Ising models and mean field models

**One-dimensional Ising model.** Consider an Ising model on the lattice $\Lambda_N$ which a line of N sites, labeled successively $x=1,2,\ldots,N$. At each site there is a spin $\sigma(x)$, with two possible values: $+1$ or $-1$. The Hamiltonian is given by

$$H_N(\sigma_{\Lambda_N}) = -\beta J \sum_{x=1}^{N-1} \sigma(x)\sigma(x+1) - \beta h \sum_{x\in\Lambda_N} \sigma(x). \tag{C.1}$$

The configuration probability is given by the Boltzmann distribution with inverse temperature $\beta \geq 0$:

$$d\mu_N(\sigma_{\Lambda_N}) = \frac{1}{Z_N} e^{-H_N(\sigma_{\Lambda_N})} dP_N(\sigma_{\Lambda_N}), \tag{C.2}$$

where

$$Z_N = \sum_{\sigma_{\Lambda_N}} e^{-H_N(\sigma_{\Lambda_N})} \tag{C.3}$$

is the partition function and $P_N(\sigma_{\Lambda_N})$ is the counting measure on $\Lambda_N$.

By [54], the magnetization is

$$M = \frac{e^{J\beta}\sinh(h\beta)}{\sqrt{e^{2J\beta}\sinh^2(h\beta) + e^{-2J\beta}}}, \tag{C.4}$$

and the pressure is

$$P = \lim_N \frac{1}{N} \log Z_N = \log[e^{\beta J}\cosh(\beta h) + \sqrt{e^{2J\beta}\sinh^2(h\beta) + e^{-2J\beta}}]. \tag{C.5}$$

Differentiating (C.3) with respect to $J$ and using (C.5), one obtain

$$\lim_N \frac{1}{N} E_{\mu_N}[\sum_{x\in\Lambda_N} \sigma(x)\sigma(x+1)] = \lim_N \frac{1}{\beta}\frac{\partial}{\partial J}(\frac{1}{N}\log Z_N) = 1 - \frac{1}{k_1}\frac{2e^{-2\beta J}}{e^{\beta J}\cosh(\beta h) + k_1}, \tag{C.6}$$

where

$$k_1 = \sqrt{e^{2J\beta}\sinh^2(h\beta) + e^{-2J\beta}}. \tag{C.7}$$

Consider the susceptibility $\mathcal{X}$, by Section 1.7 in [54], we have

$$\mathcal{X} = \frac{\partial M}{\partial h} = \beta \lim_N \frac{1}{N} Var_{\mu_N}(\sum_{x\in\Lambda_N} \sigma(x)). \tag{C.8}$$

Thus, by differentiating (C.4) with respect to h, we obtain

$$\lim_N \frac{1}{N} Var_{\mu_N}(\sum_{x\in\Lambda_N} \sigma(x)) = \frac{e^{-J\beta}\cosh(h\beta)}{k_1^3}. \tag{C.9}$$

**Square lattice zero-field Ising model.** Consider an Ising model on the square lattice $\Lambda_N$ with $|\Lambda| = N$. Similarly with the 1-d Ising model, the spins $\{\sigma(x)\}_{x=1}^N \in \{-1, 1\}^N$. Assume there is no external magnetic field, then Hamiltonian for the 2-d zero-field Ising model is given by

$$H_N(\sigma_{\Lambda_N}) = -\beta J \sum_{\langle x,y\rangle\subset\Lambda_N} \sigma(x)\sigma(y), \tag{C.10}$$

where the first sum is over pairs of adjacent spins (every pair is counted once). The notation $\langle x, y\rangle$ indicates that sites $x$ and $y$ are nearest neighbors. Then the configuration probability is given by:

$$d\mu_N(\sigma_{\Lambda_N}) = \frac{1}{Z_N} e^{\beta J \sum_{\langle x,y\rangle\subset\Lambda_N} \sigma(x)\sigma(y)} dP_N(\sigma_{\Lambda_N}), \tag{C.11}$$

where

$$Z_N = \sum_{\sigma_{\Lambda_N}} e^{\beta J \sum_{\langle x,y\rangle\subset\Lambda_N} \sigma(x)\sigma(y)} \tag{C.12}$$

is the partition function and $P_N(\sigma_{\Lambda_N}) = \prod_{x=1}^N P(\sigma_{\Lambda_N})$ is the prior distribution with $P(\sigma(x) = 1) = P(\sigma(x) = -1) = 0.5$. By Section 7.10 in [54], the spontaneous magnetization is

$$M_0 = \begin{cases} [1-\sinh^{-4}(2\beta J)]^{1/8} & \beta > \beta_c\,, \\ 0 & \beta < \beta_c, \end{cases}$$

where $\beta_c = \frac{\log(1+\sqrt{2})}{2J}$. Actually, this formula for the spontaneous magnetization is given by the definition $M_0 = \lim_{h\to 0^+}\langle\sigma(x)\rangle$. Sometimes, we can also consider the spontaneous magnetization by using the other definition $M = \lim_{h\to 0^-}\langle\sigma(x)\rangle$, which actually is the opposite of (C.13).

And the pressure is also given by [54]

$$P = \lim_{N\to\infty}\frac{1}{N}\log Z_N = \frac{\log 2}{2} + \frac{1}{2\pi}\int_0^\pi \log[\cosh^2(2\beta J) + k(\theta)]d\theta, \tag{C.13}$$

where

$$k(\theta) = \sqrt{\sinh^4(2\beta J) + 1 - 2\sinh^2(2\beta J)\cos(2\theta)}. \tag{C.14}$$

And, by (C.12) and (C.13), we obtain

$$\lim_{N\to\infty}\frac{1}{N}E_{\mu_N}\Big(\sum_{\langle x,y\rangle\subset\Lambda_N}\sigma(x)\sigma(y)\Big) = \frac{1}{\beta}\frac{\partial}{\partial J}\Big(\frac{1}{N}\log Z_N\Big) = \frac{\sinh(4\beta J)}{\pi}\int_0^\pi \frac{1}{k(\theta)}\Big[1 - \frac{1+\cos(2\theta)}{\cosh^2(2\beta J)+k(\theta)}\Big]d\theta. \tag{C.15}$$

**Mean field model.** Given the Lattice $\Lambda_N$ in d-dimension and set $|\Lambda| = N$, consider the Hamiltonian for d-dimensional Ising model

$$H_N(\sigma_{\Lambda_N}) = -\beta J\sum_{\langle x,y\rangle\subset\Lambda_N}\sigma(x)\sigma(y) - \beta h\sum_{x\in\Lambda_N}^{N}\sigma(x) = -\sum_{x\in\Lambda_N}\sigma(x)\{\frac{1}{2}\beta J\sum_{y}^{n.n}\sigma(y) + \beta h\},$$

where the first sum is over pairs of adjacent spins (every pair is counted once). The notation $\langle x, y\rangle$ indicates that sites $x$ and $y$ are nearest neighbors. And, $\{\sigma(x)\}_{x=1}^N \in \{-1,1\}^N$ are Ising spins. Replace $\sum_y^{n.n}\sigma(y)$ by $\sum_y^{n.n}\langle\sigma(y)\rangle$ in (C.16), we obtain the mean field Hamiltonian

$$\begin{aligned} H_{N;mf}(\sigma_{\Lambda_N}) &= -\sum_{x\in\Lambda_N}\sigma(x)\{\frac{1}{2}\beta J\sum_{y}^{n.n}\langle\sigma(y)\rangle + \beta h\} \\ &= -\sum_{x\in\Lambda_N}\sigma(x)\{\frac{1}{2}\beta J 2dm + \beta h\} \\ &= -\sum_{x\in\Lambda_N}\sigma(x)\{\beta J dm + \beta h\} \\ &= -\beta h_{mf}\sum_{x\in\Lambda_N}\sigma(x) \end{aligned} \tag{C.16}$$

where $h_{mf} = h + Jdm$. Then, we have the probability

$$d\mu_{N;mf}(\sigma_{\Lambda_N}) = \frac{1}{Z_{N;mf}}e^{-H_{N;mf}(\sigma_{\Lambda_N})}dP_N(\sigma_{\Lambda_N}) = \frac{1}{Z_{N;mf}}e^{\beta\sum_{x\in\Lambda_N}h_{mf}\sigma(x)}dP_N(\sigma_{\Lambda_N}). \tag{C.17}$$

So the partition function is

$$\begin{aligned} Z_{N;mf} &= \sum_{\sigma(x)} e^{\beta\sum_{x\in\Lambda_N}h_{mf}\sigma(x)} \\ &= \sum_{\sigma(x)}\prod_{x\in\Lambda_N}e^{\beta h_{mf}\sigma(x)} \\ &= \prod_{x\in\Lambda_N}\sum_{\sigma(x)}e^{\beta h_{mf}\sigma(x)} \\ &= \prod_{x\in\Lambda_N}(e^{\beta h_{mf}} + e^{-\beta h_{mf}}) \\ &= (e^{\beta h_{mf}} + e^{-\beta h_{mf}})^N \\ &= {Z_{1;mf}}^N, \end{aligned} \tag{C.18}$$

where $Z_{1;mf} = e^{\beta h_{mf}} + e^{-\beta h_{mf}}$. So the pressure is

$$P_{mf} = \lim_N \frac{1}{N} \log Z_{N;mf} = \log(e^{\beta h_{mf}} + e^{-\beta h_{mf}}) \tag{C.19}$$

And, we can also consider the $\mu_{N;mf}$ as a product measure

$$d\mu_{N;mf}(\sigma_{\Lambda_N}) = \frac{1}{Z_{N;mf}} e^{\beta \sum\limits_{x\in\Lambda_N} h_{mf}\sigma(x)} dP_N(\sigma_{\Lambda_N}) = \prod_{x\in\Lambda_N} \frac{1}{Z_{1;mf}} e^{\beta h_{mf}\sigma(x)} dP(\sigma(x)). \tag{C.20}$$

It is easy to find the magnetization

$$\begin{aligned}
m &= \frac{1}{N} E_{\mu_{N;mf}}[\sum_{x\in\Lambda_N} \sigma(x)] \\
&= \frac{1}{N} \sum_{x\in\Lambda_N} E_{\mu_{N;mf}}[\sigma(x)] \\
&= \frac{1}{N} \sum_{x\in\Lambda_N} \sum_{\sigma(x)} \sigma(x) \frac{1}{Z_{1;mf}} e^{\beta h_{mf}\sigma(x)} \\
&= \frac{1}{N} \sum_{x\in\Lambda_N} \sum_{\sigma(x)} \sigma(x) \frac{1}{Z_{1;mf}} e^{\beta h_{mf}\sigma(x)} \\
&= \frac{1}{N} \sum_{x\in\Lambda_N} \frac{1}{Z_{1;mf}} (e^{\beta h_{mf}} - e^{-\beta h_{mf}}) \\
&= \frac{1}{e^{\beta h_{mf}} + e^{-\beta h_{mf}}} (e^{\beta h_{mf}} - e^{-\beta h_{mf}}) \\
&= \tanh(\beta h_{mf}) \\
&= \tanh(\beta h + \beta J d m)
\end{aligned} \tag{C.21}$$

and

$$\begin{aligned}
\frac{1}{N} Var_{\mu_{N;mf}}(\sum_{x\in\Lambda_N} \sigma(x)) &= \frac{1}{N} Var_{\mu_{N;mf}}(\sum_{x\in\Lambda_N} \sigma(x)) \\
&= \frac{1}{N} (E_{\mu_{N;mf}}[\sum_{x\in\Lambda_N} \sigma(x)]^2 - N^2 m^2) \\
&= \frac{1}{N} (E_{\mu_{N;mf}}[\sum_{x\in\Lambda_N} \sigma^2(x) + \sum_{x\neq y} \sigma(x)\sigma(y)] - N m^2) \\
&= \{\sum_{\sigma(x)} \sigma^2(x) \frac{1}{Z_{1;mf}} e^{\beta h_{mf}\sigma(x)} + (N-1) E_{\mu_{N;mf}}[\sigma(x)\sigma(y)]\} - N m^2 \\
&= \{1 + (N-1)m^2\} - N m^2 \\
&= 1 - m^2.
\end{aligned} \tag{C.22}$$

So we can obtain the magnetization $m$ by solving the implicit equation (C.21).

### C.2. Computation of goal-oriented divergences

**Mean field versus mean field.** Given two mean field models, assume $\mu_{N;mf}$ and $\mu'_{N;mf}$ are their two configuration probabilities with

$$d\mu_{N;mf}(\sigma) = \frac{1}{Z_{N;mf}} e^{-H_{N;mf}(\sigma_{\Lambda_N})} dP_N(\sigma_{\Lambda_N}) = \frac{1}{Z_{N;mf}} e^{\beta \sum\limits_{x\in\Lambda_N} h_{mf}\sigma(x)} dP_N(\sigma_{\Lambda_N}) \tag{C.23}$$

and

$$d\mu'_{N;mf}(\sigma) = \frac{1}{Z'_{N;mf}} e^{-H'_{N;mf}(\sigma_{\Lambda_N})} dP_N(\sigma_{\Lambda_N}) = \frac{1}{Z'_{N;mf}} e^{\beta' \sum\limits_{x\in\Lambda_N} h'_{mf}\sigma(x)} dP_N(\sigma_{\Lambda_N}), \tag{C.24}$$

where $h_{mf} = h + dJm$ and $h'_{mf} = h' + dJ'm'$. Then, by (59), the relative entropy between $\mu'_{N;mf}$ and $\mu_{N;mf}$ is given by

$$\begin{aligned} R(\mu'_{N;mf}\|\mu_{N;mf}) &= \log Z_{N;mf} - \log Z'_{N;mf} + E_{\mu'_{N;mf}}[H_{N;mf}(\sigma_{\Lambda_N}) - H'_{N;mf}(\sigma_{\Lambda_N})] \\ &= \log \frac{Z_{N;mf}}{Z'_{N;mf}} + (\beta' h'_{mf} - \beta h_{mf}) E_{\mu'_{N;mf}}(\sum_{x\in\Lambda_N}\sigma(x)) \\ &= N \log \frac{Z_{1;mf}}{Z'_{1;mf}} + (\beta' h'_{mf} - \beta h_{mf}) N m' \\ &= N \log \frac{e^{\beta h_{mf}} + e^{-\beta h_{mf}}}{e^{-\beta' h'_{mf}} + e^{\beta' h'_{mf}}} + N(\beta' h'_{mf} - \beta h_{mf}) m'. \end{aligned} \tag{C.25}$$

Therefore, we have

$$\frac{1}{N} R(\mu'_{N;mf}\|\mu_{N;mf}) = \log \frac{e^{\beta h_{mf}} + e^{-\beta h_{mf}}}{e^{-\beta' h'_{mf}} + e^{\beta' h'_{mf}}} + (\beta' h'_{mf} - \beta h_{mf}) m'. \tag{C.26}$$

And, the cumulant generating function of $Nf_N = N\frac{1}{N}\sum_{x\in\Lambda_N}\sigma(x) = \sum_{x\in\Lambda_N}\sigma(x)$ is

$$\begin{aligned} \Lambda_{\mu_{N;mf},Nf_N}(c) &= \log E_{\mu_{N;mf}}(e^{cN\frac{1}{N}\sum_{x\in\Lambda_N}\sigma(x)}) \\ &= \log \sum_{\sigma(x)} e^{cN\frac{1}{N}\sum_{x\in\Lambda_N}\sigma(x)} \frac{1}{Z_{N;mf}} e^{\beta h_{mf}\sum_{x\in\Lambda_N}\sigma(x)} \\ &= \log \sum_{\sigma(x)} \frac{1}{Z_{N;mf}} e^{(c+\beta h_{mf})\sum_{x\in\Lambda_N}\sigma(x)} \\ &= \log \sum_{\sigma(x)} \prod_{x\in\Lambda_N} \frac{1}{Z_{1;mf}} e^{(c+\beta h_{mf})\sigma(x)} \\ &= \log \prod_{x\in\Lambda_N} \sum_{\sigma(x)} \frac{1}{Z_{1;mf}} e^{(c+\beta h_{mf})\sigma(x)} \\ &= \log \prod_{x\in\Lambda_N} \frac{1}{Z_{1;mf}} \{e^{(c+\beta h_{mf})} + e^{-(c+\beta h_{mf})}\} \\ &= N \log \frac{e^{(c+\beta h_{mf})} + e^{-(c+\beta h_{mf})}}{e^{-\beta h_{mf}} + e^{\beta h_{mf}}}. \end{aligned} \tag{C.27}$$

Thus,

$$\frac{1}{N}\Lambda_{\mu_{N;mf},Nf_N}(c) = \log \frac{e^{(c+\beta h_{mf})} + e^{-(c+\beta h_{mf})}}{e^{-\beta h_{mf}} + e^{\beta h_{mf}}}. \tag{C.28}$$

Also, by (C.22), we have

$$\frac{1}{N} Var_{\mu_{N;mf}}(Nf_N) = 1 - m^2. \tag{C.29}$$

**One-dimensional Ising model versus mean field.** Consider the Ising model and mean field model in 1-d and assume $\mu_N$ and $\mu_{N;mf}$ are the configuration probabilities for 1-d Ising model and mean field model respectively, which are defined in section C.1. Then, by (59), the relative entropy between $\mu_N$ and $\mu_{N;mf}$ is

$$\begin{aligned} R(\mu_N\|\mu_{N;mf}) &= \log Z_{N;mf} - \log Z_N + E_{\mu_N}(H_{N;mf}(\sigma_{\Lambda_N}) - H_N(\sigma_{\Lambda_N})) \\ &= \log Z_{N;mf} - \log Z_N + \beta J E_{\mu_N}(\sum_{\langle x,y\rangle\subset\Lambda_N}\sigma(x)\sigma(y)) - \beta J m E_{\mu_N}(\sum_{x\in\Lambda_N}\sigma(x)). \end{aligned} \tag{C.30}$$

Thus, by (C.19), (C.5), (C.6) and (C.21), we have

$$\begin{aligned}
&\lim_N \frac{1}{N} R(\mu_N \| \mu_{N;mf}) \\
&= \lim_N \frac{1}{N} \log Z_{N;mf} - \lim_N \frac{1}{N} \log Z_N + \beta J \lim_N \frac{1}{N} E_{\mu_N}\Big(\sum_{\langle x,y\rangle \subset \Lambda_N} \sigma(x)\sigma(y)\Big) - \lim_N \beta J m \frac{1}{N} E_{\mu_N}\Big(\sum_{x\in\Lambda_N} \sigma(x)\Big) \\
&= \log \frac{e^{\beta[h+Jm]} + e^{-\beta[h+Jm]}}{e^{\beta J}\cosh(\beta h) + k_1} + \frac{\beta J}{k_1}\Big(k_1 - \frac{2e^{-2\beta J}}{e^{\beta J}\cosh(\beta h) + k_1} - m e^{J\beta}\sinh(h\beta)\Big)
\end{aligned} \tag{C.31}$$

And, by (72) and by (73),we obtain

$$\frac{1}{N}\Lambda_{\mu_{N;mf},Nf_N}(c) = \log \frac{e^{[c+\beta(h+Jm)]} + e^{-[c+\beta(h+Jm)]}}{e^{-\beta[h+Jm]} + e^{\beta[h+Jm]}} \tag{C.32}$$

and

$$\frac{1}{N} Var_{\mu_{N;mf}}(Nf_N). = 1 - m^2. \tag{C.33}$$

**Two-dimensional Ising model with $h = 0$ versus mean field.** Assuming $\mu_N$ and $\mu_{N;mf}$ are two configuration probabilities for two-dimensions zeros Ising model and two-dimensions zeros mean field model respectively. By Section C.1,

$$\mu_N(\sigma_{\Lambda_N}) = \mu_N(\sigma_{\Lambda_N}) = \frac{1}{Z_N} e^{-H_N(\sigma_{\Lambda_N})} P_N(\sigma_{\Lambda_N}) = \frac{1}{Z_N} e^{\beta J \sum_{\langle x,y\rangle \subset \Lambda_N} \sigma(x)\sigma(y)} P_N(\sigma_{\Lambda_N}) \tag{C.34}$$

and

$$\mu_{N;mf}(\sigma_{\Lambda_N}) = \frac{1}{Z_{N;mf}} e^{-H_{N;mf}(\sigma_{\Lambda_N})} d\sigma = \frac{1}{Z_{N;mf}} e^{\beta \sum_{x\in\Lambda_N} h_{mf}\sigma(x)} P_N(\sigma_{\Lambda_N}), \tag{C.35}$$

where $Z_{N;mf} = (e^{\beta h_{mf}} + e^{-\beta h_{mf}})^N$ and $h_{mf} = 2Jm$.

Then, by (59), the relative entropy between $\mu_N$ and $\mu_{N;mf}$ is

$$\begin{aligned}
R(\mu_N \| \mu_{N;mf}) &= \log Z_{N;mf} - \log Z_N + E_{\mu_N}(H_{N;mf}(\sigma_{\Lambda_N}) - H_N(\sigma_{\Lambda_N})) \\
&= \log Z_{N;mf} - \log Z_N + \beta J E_{\mu_N}\Big(\sum_{\langle x,y\rangle \subset \Lambda_N} \sigma(x)\sigma(y)\Big) - 2\beta J m E_{\mu_N}\Big(\sum_{x\in\Lambda_N} \sigma(x)\Big).
\end{aligned} \tag{C.36}$$

Thus, by (C.19), (C.13), (C.15) and (C.21), we have

$$\begin{aligned}
&\lim_N \frac{1}{N} R(\mu_N \| \mu_{N;mf}) \\
&= \lim_N \frac{1}{N} \log Z_{N;mf} - \lim_N \frac{1}{N} \log Z_N + \beta J \lim_N \frac{1}{N} E_{\mu_N}\Big(\sum_{\langle x,y\rangle \subset \Lambda_N} \sigma(x)\sigma(y)\Big) - \lim_N 2\beta J m \frac{1}{N} E_{\mu_N}\Big(\sum_{x\in\Lambda_N} \sigma(x)\Big) \\
&= \log \frac{e^{\beta[h+Jm]} + e^{-\beta[h+Jm]}}{e^{\beta J}\cosh(\beta h) + k_1} + \frac{\beta J}{k_1}\Big(k_1 - \frac{2e^{-2\beta J}}{e^{\beta J}\cosh(\beta h) + k_1} - m e^{J\beta}\sinh(h\beta)\Big) \\
&= \log[e^{-2\beta Jm} + e^{2\beta Jm}] - \frac{\log 2}{2} - \frac{1}{2\pi}\int_0^\pi \log[\cosh^2(2\beta J) + k(\theta)]d\theta \\
&\quad + \beta J \frac{\sinh(4\beta J)}{\pi}\int_0^\pi \frac{1}{k(\theta)}\Big[1 - \frac{1 + \cos(2\theta)}{\cosh^2(2\beta J) + k(\theta)}\Big]d\theta - 2\beta JmM
\end{aligned} \tag{C.37}$$

And, by (72) and by (C.22), we obtain

$$\frac{1}{N}\Lambda_{\mu_{N;mf},Nf_N}(c) = \log \frac{e^{(c+2\beta Jm)} + e^{-(c+2\beta Jm)}}{e^{-2\beta Jm} + e^{2\beta Jm}} \tag{C.38}$$

and

$$\frac{1}{N} Var_{\mu_{N;mf}}(Nf_N). = 1 - m^2. \tag{C.39}$$

**One-dimensional Ising model versus Ising model.** Consider two Ising models in 1-d and $\mu_N$ and $\mu'_N$ are their configuration probabilities defined in Section C.1. Then, by (C.5), (C.6) and (C.4), we have

$$
\begin{aligned}
&\lim_N \frac{1}{N} R(\mu'_N \| \mu_N) \\
&= \lim_N \frac{1}{N} E_{\mu'_N}(\log \frac{\mu'_N}{\mu_N}) \\
&= \lim_N \frac{1}{N} \log \frac{Z_N}{Z'_{\Lambda,N}} + \lim_N \frac{1}{N} E_{\mu'_N}(H(\sigma_{\Lambda_N}) - H'_N(\sigma_{\Lambda_N})) \\
&= \lim_N \frac{1}{N} \log Z_N - \lim_N \frac{1}{N} \log Z'_{\Lambda,N} + (\beta' J' - \beta J) \lim_N \frac{1}{N} E_{\mu'_N}(\sum_{\langle x,y\rangle \subset \Lambda_N} \sigma(x)\sigma(y)) \\
&\quad + (\beta' h' - \beta h) \lim_N \frac{1}{N} E_{\mu'_N}(\sum_{x \in \Lambda_N} \sigma(x)) \\
&= \log \frac{e^{\beta J}\cosh(\beta h) + \sqrt{e^{2J\beta}\sinh^2(h\beta) + e^{-2J\beta}}}{e^{\beta' J'}\cosh(\beta' h') + \sqrt{e^{2J'\beta'}\sinh^2(h'\beta') + e^{-2J'\beta'}}} + (\beta' J' - \beta J)(1 - \frac{1}{k'_1}\frac{2e^{-2\beta' J'}}{e^{\beta' J'}\cosh(\beta' h') + k'_1}) \\
&\quad + (\beta' h' - \beta h)\frac{1}{k'_1} e^{J'\beta'} \sinh(h'\beta')
\end{aligned} \tag{C.40}
$$

And,

$$
\begin{aligned}
\lim_N \frac{1}{N} \Lambda_{\mu_N}(c) &= \lim_N \frac{1}{N} \log E_{\mu_N}(e^{cN\frac{1}{N}\sum_{x\in\Lambda_N}\sigma(x)}) \\
&= \lim_N \frac{1}{N} \log \sum_{\sigma_{\Lambda_N}} e^{cN\frac{1}{N}\sum_{x\in\Lambda_N}\sigma(x)} \frac{1}{Z_N} e^{\beta J \sum_{\langle x,y\rangle\subset\Lambda_N}\sigma(x)\sigma(y) + \beta h \sum_{x\in\Lambda_N}\sigma(x)} \\
&= \lim_N \frac{1}{N} \log \frac{1}{Z_N} \sum_{\sigma_{\Lambda_N}} e^{\beta J \sum_{\langle x,y\rangle\subset\Lambda_N}\sigma(x)\sigma(y) + \beta(h+\frac{c}{\beta})\sum_{x\in\Lambda_N}\sigma(x)} \\
&= \lim_N \frac{1}{N} \log \frac{1}{Z_N} \tilde{Z}_{\Lambda,N} \\
&= \lim_N \frac{1}{N} \log \tilde{Z}_{\Lambda,N} - \lim_N \frac{1}{N} \log Z_N,
\end{aligned} \tag{C.41}
$$

where $\tilde{Z}_{\Lambda,N} = \sum_{\sigma_{\Lambda_N}} e^{\beta J \sum_{\langle x,y\rangle\subset\Lambda_N}\sigma(x)\sigma(y) + \beta(h+\frac{c}{\beta})\sum_{x\in\Lambda_N}\sigma(x)}$. By [54], we have

$$
\lim_N \frac{1}{N} \log \tilde{Z}_{\Lambda,N} = \log[e^{\beta J}\cosh(\beta h + c) + \sqrt{e^{2J\beta}\sinh^2(h\beta + c) + e^{-2J\beta}}] \tag{C.42}
$$

and

$$
\lim_N \frac{1}{N} \log Z_N = \log[e^{\beta J}\cosh(\beta h) + \sqrt{e^{2J\beta}\sinh^2(h\beta) + e^{-2J\beta}}]. \tag{C.43}
$$

Thus,

$$
\lim_N \frac{1}{N} \Lambda_{\mu_N}(c) = \log \frac{e^{\beta J}\cosh(\beta h + c) + \sqrt{e^{2J\beta}\sinh^2(h\beta + c) + e^{-2J\beta}}}{e^{\beta J}\cosh(\beta h) + \sqrt{e^{2J\beta}\sinh^2(h\beta) + e^{-2J\beta}}}. \tag{C.44}
$$

And, by (C.9)

$$
\lim_N \frac{1}{N} Var_{\mu_N}(\sum_{x\in\Lambda_N}\sigma(x)) = \frac{e^{-J\beta}\cosh(h\beta)}{k_1^3}. \tag{C.45}
$$

## References


[1] T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd edition, Wiley–Interscience, 2006.
[2] A.B. Tsybakov, Introduction to Nonparametric Estimation, Springer Science & Business, Media, 2008.
[3] D.J.C. MacKay, Information Theory, Inference & Learning Algorithms, Cambridge University Press, 2003.
[4] C.M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag, New York, Inc., Secaucus, NJ, USA, 2006.
[5] F.J. Pinski, G. Simpson, A.M. Stuart, H. Weber, Kullback–Leibler approximation for probability measures on infinite dimensional spaces, SIAM J. Math. Anal. 47 (6) (2015) 4091–4122.
[6] M.J. Wainwright, M.I. Jordan, Graphical models, exponential families, and variational inference, Found. Trends Mach. Learn. 1 (1–2) (2008) 1–305.
[7] M.D. Hoffman, D.M. Blei, C. Wang, J. Paisley, Stochastic variational inference, J. Mach. Learn. Res. 14 (1) (2013) 1303–1347.
[8] M.S. Shell, The relative entropy is fundamental to multiscale and inverse thermodynamic problems, J. Chem. Phys. 129 (14) (2008) 4108.
[9] A. Chaimovich, M.S. Shell, Relative entropy as a universal metric for multiscale errors, Phys. Rev. E 81 (6) (2010) 060104.
[10] J.F. Rudzinski, W.G. Noid, Coarse-graining entropy, forces, and structures, J. Chem. Phys. 135 (21) (2011) 214101.
[11] P. Español, I. Zuniga, Obtaining fully dynamic coarse-grained models from MD, Phys. Chem. Chem. Phys. 13 (2011) 10538–10545.
[12] I. Bilionis, P.-S. Koutsourelakis, Free energy computations by minimization of Kullback–Leibler divergence: an efficient adaptive biasing potential method for sparse representations, J. Comput. Phys. 231 (9) (2012) 3849–3870.
[13] I. Bilionis, N. Zabaras, A stochastic optimization approach to coarse-graining using a relative-entropy framework, J. Chem. Phys. 138 (4) (2013) 044313.
[14] Y. Pantazis, M.A. Katsoulakis, A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics, J. Chem. Phys. 138 (5) (2013) 054115.
[15] T.T. Foley, M.S. Shell, W.G. Noid, The impact of resolution upon entropy and information in coarse-grained models, J. Chem. Phys. 143 (24) (2016).
[16] V. Harmandaris, E. Kalligiannaki, M. Katsoulakis, P. Plecháč, Path-space variational inference for non-equilibrium coarse-grained systems, J. Comput. Phys. 314 (2016) 355–383.
[17] M.A. Katsoulakis, D.G. Vlachos, Coarse-grained stochastic processes and kinetic Monte Carlo simulators for the diffusion of interacting particles, J. Chem. Phys. 119 (18) (2003) 9412–9427.
[18] A.J. Majda, R.V. Abramov, M.J. Grote, Information Theory and Stochastics for Multiscale Nonlinear Systems, CRM Monograph Series, American Mathematical Society, Providence, R.I., 2005.
[19] M.A. Katsoulakis, J. Trashorras, Information loss in coarse-graining of stochastic particle dynamics, J. Stat. Phys. 122 (1) (2006) 115–135.
[20] M.A. Katsoulakis, P. Plecháč, L. Rey-Bellet, D.K. Tsagkarogiannis, Coarse-graining schemes and a posteriori error estimates for stochastic lattice systems, ESAIM, Math.l Model. Numer. Anal. 41 (3) (2007) 627–660.
[21] A.J. Majda, B. Gershgorin, Improving model fidelity and sensitivity for complex systems through empirical information theory, Proc. Natl. Acad. Sci. 108 (25) (2011) 10044–10049.
[22] E. Kalligiannaki, M.A. Katsoulakis, P. Plecháč, Spatial two-level interacting particle simulations and information theory-based error quantification, SIAM J. Sci. Comput. 36 (2) (2014) A634–A667.
[23] N. Chen, A.J. Majda, X.T. Tong, Information barriers for noisy Lagrangian tracers in filtering random incompressible flows, Nonlinearity 27 (9) (2014) 2133, http://stacks.iop.org/0951-7715/27/i=9/a=2133.
[24] H. Liu, W. Chen, A. Sudjianto, Relative entropy based method for probabilistic sensitivity analysis in engineering design, J. Mech. Des. 128 (2) (2006) 326–336.
[25] N. Lüdtke, S. Panzeri, M. Brown, D.S. Broomhead, J. Knowles, M.A. Montemurro, D.B. Kell, Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks, J. R. Soc. Interface 5 (19) (2008) 223–235.
[26] M. Komorowski, M.J. Costa, D.A. Rand, M.P. Stumpf, Sensitivity, robustness, and identifiability in stochastic chemical kinetics models, Proc. Natl. Acad. Sci. 108 (21) (2011) 8645–8650.
[27] A.J. Majda, B. Gershgorin, Quantifying uncertainty in climate change science through empirical information theory, Proc. Natl. Acad. Sci. 107 (34) (2010) 14958–14963.
[28] Y. Pantazis, M.A. Katsoulakis, A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics, J. Chem. Phys. 138 (5) (2013) 054115.
[29] H. Lam, Robust sensitivity analysis for stochastic systems, Math. Oper. Res. 41 (4) (2016) 1248–1275.
[30] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Springer Science & Business, Media, 2003.
[31] S. Konishi, G. Kitagawa, Information Criteria and Statistical Modeling, Springer Series in Statistics, Springer, New York, 2008.
[32] B. Simon, The Statistical Mechanics of Lattice Gases, Vol. 1, Princeton University Press, 2014.
[33] P. Dupuis, M.A. Katsoulakis, Y. Pantazis, P. Plecháč, Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics, SIAM/ASA J. Uncertain. Quantif. 4 (1) (2016) 80–111.
[34] K. Chowdhary, P. Dupuis, Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification, ESAIM, Math.l Model. Numer. Anal. 47 (03) (2013) 635–662.
[35] P. Glasserman, X. Xu, Robust risk measurement and model risk, Quant. Finance 14 (1) (2014) 29–58.
[36] I.R. Petersen, M.R. James, P. Dupuis, Minimax optimal control of stochastic uncertain systems with relative entropy constraints, IEEE Trans. Autom. Control 45 (3) (2000) 398–412.
[37] F. Liese, I. Vajda, Convex Statistical Distances, 1987.
[38] T. Van Erven, P. Harremoës, Rényi divergence and Kullback–Leibler divergence, IEEE Trans. Inf. Theory 60 (7) (2014) 3797–3820.
[39] G.L. Gilardoni, On Pinsker's and Vajda's type inequalities for Csiszár's $f$-divergences, IEEE Trans. Inf. Theory 56 (11) (2010) 5377–5386.
[40] E.L. Lehmann, G. Casella, Theory of Point Estimation, Springer Science & Business, Media, 2006.
[41] M. Dashti, A.M. Stuart, The Bayesian approach to inverse problems, arXiv:1302.6989.
[42] P.G. Dupuis, R.S. Ellis, A Weak Convergence Approach to the Theory of Large Deviations, Wiley Series in Probability and Statistis, Wiley, New York, 1997, a Wiley–Interscience publication, http://opac.inria.fr/record=b1092351.
[43] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, Vol. 38, Springer Science & Business, Media, 2009.
[44] R.Y. Rubinstein, D.P. Kroese, Simulation and the Monte Carlo Method, John Wiley & Sons, 2016.
[45] M. Hairer, A.J. Majda, A simple framework to justify linear response theory, Nonlinearity 23 (4) (2010) 909–922.
[46] S. Asmussen, P.W. Glynn, Stochastic Simulation: Algorithms and Analysis, Vol. 57, Springer Science & Business, Media, 2007.
[47] G. Arampatzis, M.A. Katsoulakis, Y. Pantazis, Accelerated sensitivity analysis in high-dimensional stochastic reaction networks, PLoS ONE 10 (7) (2015) 1–24.
[48] Michail Loulakis, Einstein relation for a tagged particle in simple exclusion processes, Commun. Math. Phys. 229 (2) (2002) 347–367.
[49] T. Komorowski, S. Olla, Einstein relation for random walks in random environments, Stoch. Process. Their Appl. 115 (8) (2005) 1279–1301.
[50] G. Diezemann, Nonlinear response theory for Markov processes: simple models for glassy relaxation, Phys. Rev. E 85 (2012) 051502, http://dx.doi.org/10.1103/PhysRevE.85.051502.

[51] U. Basu, M. Krüger, A. Lazarescu, C. Maes, Frenetic aspects of second order response, Phys. Chem. Chem. Phys. 17 (9) (2015) 6653–6666.
[52] M.E. Tuckerman, Statistical Mechanics: Theory and Molecular Simulation, Oxford Graduate Texts, Oxford University Press, Oxford, 2010.
[53] E. Presutti, Scaling Limits in Statistical Mechanics and Microstructures in Continuum Mechanics, Springer Science & Business, Media, 2008.
[54] R.J. Baxter, Exactly Solved Models in Statistical Mechanics, Courier Corporation, 2007.
[55] J. Marian, G. Venturini, B.L. Hansen, J. Knap, M. Ortiz, G.H. Campbell, Finite-temperature extension of the quasicontinuum method using Langevin dynamics: entropy losses and analysis of errors, Model. Simul. Mater. Sci. Eng. 18 (1) (2010) 015003, http://stacks.iop.org/0965-0393/18/i=1/a=015003.