EFFICIENT MONTE CARLO FOR NEURAL NETWORKS WITH LANGEVIN SAMPLERS

L.A. BREYER, M. PICCIONI, AND S. SCARLATTI

Abstract. We consider the task of simulating efficiently from the posterior distribution over weight space of a feed-forward neural network using a Langevin-Metropolis sampler, given a finite data set. It is shown that as the number N of hidden neurons increase, the proposal variance must scale as $N^{-1/3}$ in order to get convergence of the underlying discretized diffusions. This generalizes previous results of Roberts and Rosenthal (1998) for the i.i.d. case, shows robustness of their analysis, and has also practical implications.

1. Introduction

When solving non linear regression or classification problems by means of neural networks, It is well known that one must control carefully the complexity of the model (number of nodes of the network) with respect to the amount of training data at disposal, this in order to avoid overfitting, and hence bad or poor generalization performance (see e.g. Haykin, 1999, and updated references therein). To handle this central issue in learning theory researchers have focused their efforts on comparing the network's theoretical and empirical abilities to generalize, and giving quantitative estimates for the discrepancy between these two quantities in the worst case, regardless of the nature of the source generating the data.

At least in case of networks having a treshold activation function these explicit bounds depend on a finite parameter, the Vapnik-Chervonenkis dimension for the family of functions which can in principle be realized by the network. It is shown (Vapnik, 1996, Bishop, 1996) that this discrepancy vanishes in the limit as the available training data increase without bound. When however the amount of

---

Date: March, 1999.

1991 *Mathematics Subject Classification.* Primary 60J, Secondary 60F.

Key words and phrases. Neural Networks, Bayesian Learning, Markov Chain Monte Carlo, Hastings-Metropolis, Langevin Diffusion, Propagation of Chaos.

data stays limited (as in real applications), practical experience suggests that good generalization can be obtained with training sets much smaller than those required by theory. Thus in some cases, regular theory can lead to overfitting.

It should also be noted that these theoretical results usually apply to networks with treshold activation functions. In practice, smooth activation functions are generally used when implementing back propagation (gradient descent) to search for the best set of weights on the nodes of the network. The corresponding theoretical analysis becomes then much more involved (Devroye, Gyorfi and Lugosi, 1996).

A different approach, which avoids the overfitting problem, consists in using a Bayesian methodology (Bishop, 1996, chap. 10, Neal, 1996, and references therein). Here, instead of searching for optimal weights to be fixed for all time, one lets the weights fluctuate. Given a finite observed data set, a prior distribution over weight space is applied and one gets, via Bayes' theorem, an *a posteriori* distribution over weights.

For a single layer network, this may be described more precisely as follows: Let $a_1, \ldots, a_n \in \mathbb{R}^d$ be the input data, and denote the corresponding output data $z_1, \ldots, z_n \in \mathbb{R}$. We set $H : \mathbb{R}^{d+1} \to \mathbb{R}^n$,

$$H^k(y^0, \ldots, y^d) = y^0 g\big(\langle a_k, (y^1, \ldots, y^d)\rangle\big), \quad g \text{ sigmoid function}, \ k = 1, \ldots, n,$$

and thus the posterior with N neurons is in the form [?? more detailed deriv here ??]

$$\pi_N(dx) = C_N^{-1} \exp\left(\sum_{i=1}^N \langle z, H(x_i)\rangle - \frac{1}{2N} \sum_{i,j=1}^N \langle H(x_i), H(x_j)\rangle\right) \bigotimes_{i=1}^N \mu(dx_i), \tag{1}$$

The distribution π_N now expresses the probability that a configuration $x = (x_1, \ldots, x_N)$ of weights generated the training data. We stress that here, the size of the data set stays finite, and we are interested in analysing the behaviour of the network (i.e. the above distribution π_N) as the number of nodes N increases. There are many reasons for studying such a limit (see Neal, 1996): firstly, real life problems of great complexity tend to require networks with a very large number of nodes. Secondly, if the number of nodes is allowed to grow without bound, feed forward neural networks have been shown to be good universal approximators (see Cybenko, 1989, Barron 1993). Finally, the infinite network limit gives insight into the properties of different choices of prior distributions over weights, touching upon a delicate aspect of the Bayesian approach (see Neal, 1996, chap.2).

In our first result (Theorem 1), we show the propagation of chaos for the family of distributions (1), as $N \to \infty$. Since the model is of mean field type, our result is not unexpected, but since the function H is unbounded, standard results (e.g. Kusuoka, 199?) cannot be applied directly. The implication for the neural network is that in the limit, any finite collection of weights behaves as if the individual weights had been draws independently from a single distribution π. This distribution is a member of the exponential family generated by H, which depends on the activation function, and the prior μ. Our second result (Theorem 2) is mainly of a technical nature, and is used in the proof of Theorem 3.

Once the posterior has been specified in the form (1), the problem remains of simulating efficiently from it. This is the essential step for predicting the expected values of the response variables using Markov Chain Monte Carlo integration. Here, one defines a Markov chain $X_t^{(N)}$ which is guaranteed to have π_N as a limiting distribution. Consequently, for any given function g of the responses, we can estimate

$$\frac{1}{T}\sum_{t=1}^{T} g(X_t) \to \int g(x)\pi_N(x)dx \text{ as } T \to \infty. \tag{2}$$

What is more, estimates of the accuracy achieved are obtainable by statistical means, usually the (Markov chain) Central Limit Theorem together with a hypothesis test.

Designing the Markov chain $X_t^{(N)}$ for the target distribution π_N is an easy matter with the Hastings-Metropolis method. However, the enormous flexibility afforded also brings with it a question about the most efficient Markov chain to use. In this paper (Theorem 3), we shall analyse one such chain, the Metropolis adjusted Langevin algorithm, defined below. One way of thinking about it is as a gradient descent method coupled with a random walk, so as to allow escapes from local extrema of the density. Considering that most neural networks implement some form of back propagation (gradient descent), this algorithm offers a simple migration path to individuals wanting to try the Bayesian approach.

We now specify the algorithm $X_t^{(N)}$. We shall describe some asymptotics afterwards, together with a discussion of how to tune the free parameter so as to optimize the speed of convergence of the estimates (2).

For a given distribution π_N in the form (1) and which has a density on $(\mathbb{R}^n)^N$ with respect to Lebesgue measure, to be denoted $\pi_N(x)$, the Metropolis-adjusted Langevin algorithm consists of a Markov chain $X_t^{(N)}$ whose stationary distribution is $\pi_N(x)dx$, and which is of Metropolis-Hastings type: moves are first proposed, and then accepted or rejected by a simple test (see Tierney, 1994, Roberts and Smith, 1994). For reference, we list the details of its implementation:

Choose $X_0^{(N)}$ arbitrarily. Given $X_t^{(N)}$, to compute $X_{t+1}^{(N)}$, first generate

$$Y_\sigma = X_t^{(N)} + \sigma W + \frac{\sigma^2}{2} \nabla \log \pi_N(X_t^{(N)}), \tag{3}$$

where W is a standard Gaussian on $(\mathbb{R}^n)^N$, independent of $X_t^{(N)}$, and therefore the law of Y_σ given $X_t^{(N)} = x$ is proportional to

$$\begin{aligned} q_N(x,y) &= \exp\left(-\frac{1}{2\sigma^2} \left\| y - x - \frac{\sigma^2}{2} \nabla \log \pi_N(x) \right\|^2 \right) \\ &= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N \left\| y_i - x_i - \frac{\sigma^2}{2} \left(\langle z, \nabla U(x_i) \rangle - \frac{1}{N} \sum_{j=1}^N \langle \nabla H(x_i), H(x_j) \rangle \right) \right\|^2 \right). \end{aligned} \tag{4}$$

where $U(x) = \langle z, H(x) \rangle + \log d\mu/dx$. Finally, the proposal Y_σ is accepted or rejected:

$$X_{t+1}^{(N)} = \begin{cases} Y_\sigma & \text{if } \pi_N(Y_\sigma) q_N(Y_\sigma, X_t^{(N)}) > \xi \cdot \pi_N(X_t^{(N)}) q_N(X_t^{(N)}, Y_\sigma), \\ X_t^{(N)} & \text{otherwise.} \end{cases} \tag{5}$$

where $\xi \sim U[0,1]$.

With the algorithm defined as above, we can now generate as many steps as desired to produce accurate Monte Carlo estimates as in (2). Note that in practice, the partial sums do not start from $t = 0$, but typically from some large value $t = t_0 \gg 1$, which ensures that the effect of the (arbitrary) initial values $X_0^{(N)}$ is countered. Indeed, the positive recurrence of the chain ensures that $\mathbb{P}(X_{t_0}^{(N)} = x) \approx \pi_N(x)$ if t_0 is sufficiently large, and the values of the chain prior to t_0 are usually discarded.

For purposes of analysis, we shall therefore assume from now on that $X_0^{(N)} \sim \pi_N$. Our last result, Theorem 3, is a diffusion approximation, as $N \to \infty$. We show that if the variance $\sigma^2 = \ell^2/N^{1/3}$, then for any finite number of nodes $1, \dots, k$ say, the processes

$$(X_{[sN^{1/3}]}^{(N),1}, \dots, X_{[sN^{1/3}]}^{(N),k} : 0 \le s \le T)$$

converge weakly to a diffusion process $Z = (Z_s^1, \dots, Z_s^k : 0 \le s \le T)$ whose form we give explicitly, and which has the stationary distribution $\pi^{\otimes k}$, where π is given by Theorem 1. An implication of this result is that

$$\frac{1}{TN^{1/3}} \sum_{s=1}^{TN^{1/3}} g(X_s^{(N)}) \Rightarrow \frac{1}{T} \int_0^T g(Z_s) ds, \tag{6}$$

if g is bounded and continuous and depends only on a finite number of nodes. This may be loosely interpreted as follows: The Monte Carlo estimate of $\int g(x)\pi_N(x)dx$ requires a number of iterations proportional to $N^{1/3}$. The proportionality factor is the time required to get convergence on the right of (6), and depends on the mixing properties of the diffusion Z. This in turn depends on the value of ℓ, through a function $v(\ell)$ which we compute explicitly. By finding the value $\hat{\ell}$ which maximizes $v(\ell)$, we obtain the diffusion process Z which mixes the fastest among all possible. We can give an analytic expression for $\hat{\ell}$, but this is in practice useless since it cannot be computed easily (except by Monte Carlo methods, which defeats somewhat the purpose). Fortunately, as in Roberts and Rosenthal (1997), the optimal mixing rate is obtained for a choice $\hat{\ell}$ corresponding to an average acceptance probability $\mathbb{P}_{\pi_N}(X_{t+1}^{(N)} \neq X_t^{(N)}) \approx 0.574$.Thus it suffices to monitor the acceptance rate and tune ℓ until it equals 0.574.

2. Statement of the main result

In Propositions 1 and 3 below, we work with a general mean field model of the form

$$\pi_N(dx) = C_N^{-1} \exp\left(\sum_{i=1}^{N} \langle z, H(x_i) \rangle - \frac{1}{2N} \sum_{i,j=1}^{N} \langle H(x_i), H(x_j) \rangle \right) \bigotimes_{i=1}^{N} \mu(dx_i),$$

where u is a (not necessarily bounded) interaction function. In order to prove the first limiting result, we need to introduce the exponential family of probability measures on $\mathbb{R}^n$ generated by μ and u, which is defined by

$$\mu_\theta(dx) = e^{\langle \theta, H(x) \rangle - K(\theta)} \mu(dx), \theta \in \Theta$$

where $K(\theta) = \log \int e^{\langle \theta, H(x) \rangle} \mu(dx)$ is the cumulant generating function of u under μ. We assume that K is finite only in an open set Θ of $\mathbb{R}^n$. Consider now the strictly convex function $J(\theta) = \frac{1}{2} \|\theta\|^2 + K(\theta) - \langle z, \theta \rangle$, where K is extended to the complement of Θ by setting its value equal to $+\infty$. The function has a unique minimum $\theta_* = \theta_*(z)$ in $\mathbb{R}^n$, as it is lower semicontinuous with compact level sets. We can now state

Proposition 1. *Whenever $f : (\mathbb{R}^n)^\infty \to \mathbb{R}$ is a local function (i.e. depends on only a finite number of components) such that $\int e^{tf} d\pi^{\otimes\infty} < \infty$ for $t \in (-\epsilon, \epsilon)$, then*

$$\lim_{N\to\infty} \int f d\pi_N = \int f d\pi^{\otimes\infty} \quad \textit{(propagation of chaos)},$$

where $\pi = \mu_{\theta_*}$.

Proof. Observe that θ_* is the unique solution of the equation

$$\theta + \nabla K(\theta) = z, \tag{7}$$

which implies by the properties of exponential families, that

$$\theta_* = z - \int H d\pi. \tag{8}$$

We can now easily bound the Kullback-Leibler divergence

$$D(\pi_N \| \pi^{\otimes N}) = \int \log(d\pi_N / d\pi^{\otimes N}) d\pi_N.$$

In fact, by using (8), and setting $\widetilde{H}(x) = H(x) - \int H d\pi$,

$$\begin{aligned}
\log(d\pi_N/d\pi^{\otimes N}) &= \log C_N^{-1} + NK(\theta_*) + \sum_{i=1}^N \langle z - \theta_*, H(x_i)\rangle - \frac{1}{2N} \sum_{i,j=1}^N \langle H(x_i), H(x_j)\rangle \\
&= \log C_N^{-1} + NK(\theta_*) + \frac{N}{2} \left\| \int u d\mu \right\|^2 \\
&- \frac{N}{2} \left(\left\| \int u d\mu \right\|^2 + \langle \sum_{i,j=1}^N H(x_i)/N, H(x_j)/N \rangle - 2 \sum_{i=1}^N \langle H(x_i)/N, \int u d\pi \rangle \right) \\
&= \log \widetilde{C}_N + \left(-\frac{1}{2} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \widetilde{H}(x_i) \right\|^2 \right)
\end{aligned}$$

∎

Now clearly

$$\log \widetilde{C}_N = -\log \int \exp\left(-\frac{1}{2} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \widetilde{H}(x_i) \right\|^2 \right) \bigotimes_{i=1}^N \pi(dx_i),$$

By the Central Limit Theorem, the right side above is therefore bounded, uniformly in N, by some constant M say, from which $D(\pi_N \| \pi^{\otimes N}) \le M$. It follows that if we denote by $\pi_{N,k}$ the marginal of π_N for the first k components, then an inequality of (Csiszar, 1984, p.112) gives

$$D(\pi_{N,k}\|\pi^{\otimes k}) \le \frac{1}{[N/k]} D(\pi_N\|\pi^{\otimes N}) \le \frac{M}{[N/k]},$$

and now the stated convergence follows by (Csiszar, 1975, Lemma 3.1). □

In Proposition 3, we shall use the following technical lemma.

Lemma 2. *For any nonnegative definite matrix A, the convex conjugate of $z \mapsto \frac{1}{2}\langle z, Az\rangle$ is given by*

$$M^*(z) = \begin{cases} \frac{1}{2}\langle w, A^- w\rangle & \text{if } w \in RanA \\ +\infty & \text{otherwise,} \end{cases}$$

where A^- is the pseudo-inverse of A.

Proof. Let $A = U^t L U$ with L a diagonal matrix with the diagonal elements equal to the eigenvalues in decreasing order. By definition,

$$M^*(z) = \sup_\theta(\langle z, \theta\rangle - \frac{1}{2}\langle \theta, A\theta\rangle) = \sup_w(\sum_{i=1}^s v_i w_i - \frac{1}{2}\sum_{i=1}^s \lambda_i w_i^2)$$

where $v = Uz$. If there exists i_0 such that $\lambda_{i_0} = 0$ and $v_{i_0} \neq 0$ (which happens if and only if $z \notin \text{Ran}B$) it is immediately seen that $M^*(z) = +\infty$. Otherwise the function between round brackets has a maximum $w_i = \frac{v_i}{\lambda_i}$ for i such that $\lambda_i > 0$, $w_i = 0$ otherwise. Finally it is easily seen that

$$M^*(z) = \frac{1}{2}\sum_{i:\lambda_i>0} \frac{v_i^2}{\lambda_i} = \frac{1}{2}\langle z, A^- z\rangle$$

for $z \in \text{Ran}A$. □

Proposition 3 (Moderate Deviations)**.** *For any function $g : \mathbb{R}^n \to \mathbb{R}^m$ satisfying $\int e^{\langle\theta, g\rangle} d\pi < \infty$ in a neighbourhood of $\theta = 0$, if $\lambda_N \to \infty$ is a sequence such that $\lambda_N^2/N \to 0$, then*

$$\pi_N\Big(\Big|\frac{1}{N}\sum_{i=1}^N g(x_i) - \int g d\pi\Big| > \lambda_N/\sqrt{N}\Big) \le e^{-c\lambda_N^2 + o(\lambda_N^2)},$$

where $c > 0$ is a constant and the limiting distribution π is given by Theorem 1.

Proof of Theorem 2. Define $\widetilde{g}(x_i) = g(x_i) - \int g d\mu$ and

$$(Z_N, Y_N) = (\lambda_N \sqrt{N})^{-1} \sum_{i=1}^{N} (\widetilde{g}(x_i), \widetilde{H}(x_i)).$$

Now it is easy to compute (Dembo and Zeitouni, 1998)

$$\begin{aligned}\Lambda(\theta, \psi) &= \lim_{N\to\infty} \log \int \exp \lambda_N^2 (\langle \theta, Z_N \rangle + \langle \psi, Y_N \rangle) d\pi^{\otimes N} \\ &= \frac{1}{2} \langle (\theta, \psi), \Sigma(\theta, \psi) \rangle\end{aligned}$$

where Σ is the covariance matrix of $(\widetilde{g}(x), \widetilde{H}(x))$ under π. By applying the Gärtner-Ellis theorem and Lemma 2 below, we prove that (Z_N, Y_N) satisfies an LDP with speed λ_N^2 and rate function

$$J(z, y) = \begin{cases} \frac{1}{2} \langle (z, y), \Sigma^{-}(z, y) \rangle & \text{if } (z, y) \in \mathrm{Ran}\Sigma \\ +\infty & \text{otherwise.} \end{cases}$$

We want to prove the same result for the sequence Z_N under π_N. Decompose Σ into blocks as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \cdots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \int \widetilde{g}\widetilde{g}^t d\pi & \vdots & \int \widetilde{g}\widetilde{u}^t d\pi \\ \cdots & \cdots & \cdots \\ \int \widetilde{u}\widetilde{g}^t d\pi & \vdots & \int \widetilde{u}\widetilde{u}^t d\pi \end{pmatrix}$$

and write

$$\widetilde{\Lambda}_N(\theta) = \log \widetilde{C}_N^{-1} \int \exp\Big(\lambda_N^2 (\langle \theta, Z_N \rangle - \frac{1}{2} \| Y_N \|^2) \Big) d\pi^{\otimes N},$$

we apply Varadhan's Lemma (Dembo and Zeitouni, 1998, Theorem 4.3.1, p.137) with the continuous function $\varphi(z, y) = \langle \theta, z \rangle - \frac{1}{2} \| y \|^2$, which satisfies the moment condition

$$\begin{aligned}&\overline{\lim}_{N\to\infty} \frac{1}{\lambda_N^2} \log \int \exp\Big(a\lambda_N^2 \varphi(Z_N, Y_N) \Big) d\pi^{\otimes N} \\ &\leq \overline{\lim}_{N\to\infty} \frac{1}{\lambda_N^2} \log \int \exp\Big(a\lambda_N^2 \langle \theta, Z_N \rangle \Big) d\pi^{\otimes N} \\ &= \overline{\lim}_{N\to\infty} \frac{N}{\lambda_N^2} \log \int \exp\Big(\frac{\lambda_N}{N} \langle a\theta, \widetilde{g}(x_1) \rangle \Big) \pi(dx_1) \\ &= \overline{\lim}_{N\to\infty} \frac{N}{\lambda_N^2} \Big(1 + \frac{\lambda_N^2 a^2}{2N} \langle \theta, \Sigma_{11}\theta \rangle + o(\lambda_N^2/N) \Big) < \infty,\end{aligned}$$

for any constant a. Since $\widetilde{C}_N$ is bounded in N, we obtain

$$\widetilde{\Lambda}(\theta) := \lim_{N\to\infty} \frac{1}{\lambda_N^2} \widetilde{\Lambda}_N(\theta) = \lim_{N\to\infty} \frac{1}{\lambda_N^2} \log \int \exp \lambda_N^2 \varphi(Z_N, Y_N) d\pi^{\otimes N}$$
$$= \sup_{z,y} \{\varphi(z,y) - J(z,y)\}.$$

In order to maximize the right hand side above, write (z, y) as $\Sigma(u, v)$, without loss of generality since J is equal to $+\infty$ out of the range of Σ. Now

$$\sup_{z,y}\{\varphi(z,y) - J(z,y)\} = \sup_{u,v}\{\langle \theta, \Sigma_{11}u + \Sigma_{12}v\rangle - \frac{1}{2}\|\Sigma_{21}u + \Sigma_{22}v\|^2$$
$$- \frac{1}{2}(\langle u, \Sigma_{11}u\rangle + \langle v, \Sigma_{22}v\rangle + 2\langle u, \Sigma_{12}v\rangle)\}.$$

The function to be maximized in concave in (u, v) and it is immediately checked that $(-\theta, (I + \Sigma_{22})^{-1}\Sigma_{21}\theta)$ is a stationary point. Substituting this back into the above expression, we finally arrive at

$$\widetilde{\Lambda}(\theta) = \frac{1}{2}\langle \theta, B\theta\rangle,$$

where $B = \Sigma_{11} - \Sigma_{12}(I + \Sigma_{22})^{-1}\Sigma_{21}$. In order to apply the Gartner Ellis theorem, we need only check that B is nonnegative definite, and to apply Lemma 2: Let $A = \Sigma_{12}\Sigma_{22}^{-}$, where the superscript $-$ is used to denote the *pseudo-inverse* of Σ_{22}. Since $\mathrm{Ker}[\Sigma_{22}] = \mathrm{Ker}[\Sigma_{12}]$, we have $\Sigma_{12} = A\Sigma_{22}$. As a consequence,

$$\mathrm{Var}_\mu[g(x_i) - AH(x_i)] = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21} \geq 0.$$

Now consider the difference

$$D = \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21} - \Sigma_{12}(I + \Sigma_{22})^{-1}\Sigma_{21} = \Sigma_{12}(\Sigma_{22}^{-} - (I + \Sigma_{22})^{-1})\Sigma_{21},$$

and notice that the matrix between rounded brackets is nonnegative definite on $\mathrm{Ran}\Sigma_{22}$. But since $\mathrm{Ran}[\Sigma_{21}] \subset \mathrm{Ran}\Sigma_{22}$ (as a consequence of the inclusion $\mathrm{Ker}\Sigma_{22} \subset \mathrm{Ker}\Sigma_{12}$), D is nonnegative definite, and hence so is B. □

We shall need the following assumptions to prove our main convergence result.

- **(A):** The functions $H(x)$, $U(x) \equiv \log(d\mu/dx) + \langle z, H(x)\rangle$ and their derivatives are smooth, with at most polynomial growth in the variable x.
- **(B):** The prior distribution μ has a finite moment generating function in some neighbourhood of the origin.

In order to ensure that the limiting process is well defined we need to enforce (A) in the following way:

(C): The functions $H'(x)$ and $U'(x)$ have at most linear growth.

Theorem 4. *Let $\{Z_t^i : i = 1, 2, \dots\}$ be independent copies of the process Z_t which is the solution to the SDE*

$$(9)\quad dZ_t \;=\; v(\ell)^{1/2} dB_t + \Big(\nabla U(Z_t) - \langle \nabla H(Z_t), \int H d\pi\rangle\Big) v(\ell) dt, \quad Z_0 \;\sim\; \pi,$$

where $v = v(\ell) = 2\ell^2 \Phi(-\ell\tau/2)$ and τ is a suitable constant depending on π. This process is well defined under condition (C). If $X_t^{(N)} = (X_t^{(N),1}, \dots, X_t^{(N),N})$ is the Metropolis-adjusted Langevin algorithm defined by (5), with $X_0^{(N)} \sim \pi_N$ and jump size $\sigma^2 = \ell^2/N^{1/3}$, we have the weak convergence result

$$(10)\qquad \big(X^{(N),1}_{tN^{1/3}}, \dots, X^{(N),k}_{tN^{1/3}} : t \in [0, T]\big) \Rightarrow \big(Z_t^i : t \in [0, T], i = 1, \dots, k\big),$$

for any integer k.

The rest of the paper is devoted to the proof of Theorem 4. For the sake of simplicity, we take $n = 1$ below. The changes for the general case will affect only the notation.

3. A quantitative CLT for the acceptance ratio.

We begin with the following preliminary remarks. Below, for any $f : R^k \to R$ and any $t \in (R^k)^N$, we shall write $E_N f(t) = \frac{1}{N}\sum_{i=1}^N f(t_i)$. From the definition (3) we have

$$(11)\qquad Y_{\sigma,i}(x, W) = Y_i = x_i + \sigma W_i + \frac{\sigma^2}{2}\Big(U'(x_i) - H'(x_i) E_N H(x)\Big).$$

Now define

$$
\begin{aligned}
G_{\sigma,N}(x,W) &= \log \frac{\pi_N(Y_\sigma)q_N(Y_\sigma,x)}{\pi_N(x)q_N(x,Y_\sigma)} \\
&= \log \pi_N(Y_\sigma(x)) - \log \pi_N(x) \\
&\quad + \frac{1}{2}\Big[\|W\|^2 - \Big\|W + \frac{\sigma}{2}\big(\nabla \log \pi_N(Y_\sigma(x)) + \nabla \log \pi_N(x)\big)\Big\|^2\Big] \\
&= N(E_N U(Y) - \frac{1}{2}E_N H(Y)^2 - E_N U(x) + \frac{1}{2}E_N H(x)^2) \\
&- \frac{1}{2}\sigma N[E_N(U'(Y)W) + E_N(U'(x)W) - E_N(H'(Y)W)E_N H(Y) \\
&- E_N(H'(x)W)E_N H(x)] + \frac{1}{8}\sigma^2 N[E_N U'(Y) + E_N U'(x) \\
&\quad - E_N H'(Y)E_N H(Y) - E_N H'(x)E_N H(x)]
\end{aligned}
\tag{12}
$$

where $Y = Y_\sigma(x,W)$ and we have used the notation

$$
E_N g(x)h(Y)W^l = \frac{1}{N}\sum_{i=1}^{N} g(x_i)h(Y_i)W_i^l.
$$

In order to study the asymptotic behaviour of the function $G_{\sigma,N}$, we wish to expand it into a Taylor series with a suitable number of terms. The following lemma, which we state without proof, is the result of a tedious, but straightforward computation.

Lemma 5. *The first two derivatives of $G_{\sigma,N}(x,W)$ vanish at $\sigma = 0$. Consequently, we have the Taylor expansion*

$$
G_{\sigma,N}(x,W) = \sum_{k=3}^{6} \sigma^k g_{k,N}(x,W) + \frac{1}{6!}\int_0^\sigma (\sigma-u)^6 \frac{d^7}{du^7} G_{u,N}(x,W)du,
\tag{13}
$$

where $g_{k,N}(x,W) = \frac{d^k}{du^k}G_{u,N}(x,W)(0)$ for $k = 3,\dots,6$.

The main result of this section is that, as $N \to \infty$, a Central Limit Theorem holds for $G_{\sigma,N}(x,W)$:

Proposition 6. *Choose $\sigma = \ell/N^{1/6}$. Then there exist sets $F_N \subset \mathbb{R}^N$, with $\pi_N(F_N^c) = o(N^{-k})$ for any $k > 0$ such that*

$$
\lim_{N\to\infty} N^\beta \sup_{x\in F_N} \sup_u \Big|\mathbb{P}\Big(G_{\sigma,N}(x,W) \le u\Big) - \Phi_{-\ell^2\tau^2/2,\ell^2\tau^2}(u)\Big| = 0,
\tag{14}
$$

where $\beta > 0$ is sufficiently small, $\Phi_{m,s^2}(u)$ denotes the cumulative distribution function of a $\mathcal{N}(m, s^2)$ random variable, and

$$\tau^2 = \frac{1}{144}(9E(\psi''^2(X)\psi'^2(X)) + 18E(\psi'(X)\psi''(X)\psi'''(X)) + 15E(\psi'''^2(X))) - 18(E(H''(X) + H'(X)\psi'(X))E(H'(X)(\psi'''(X) + \psi'(X)\psi''(X)) + 9E(H'^2(X))(E(H''(X) + H'(X)\psi'(X))^2). \tag{15}$$

with $\psi = U - H \int H d\pi$.

Proof. To reduce the burden of notation, we set $\ell = 1$ in the following. By (Petrov, 1995, Lemma 1.9, p.20) and (13), we have, using the Taylor expression from Lemma 5, and writing $\Phi = \Phi_{0,1}$ for simplicity,

$$\begin{aligned}
&\sup_u \left|\mathbb{P}\big((G_{\sigma,N} + \tau^2/2)/\tau \le u\big) - \Phi(u)\right| \le \sup_u \left|\mathbb{P}\Big(N^{-1/2} g_{3,N}(x,W)/\tau \le u\Big) - \Phi(u)\right| \\
&\quad + \mathbb{P}\big(N^{-2/3}\left|g_{4,N}(x,W)\right| > \tau\epsilon_N\big) + \mathbb{P}\big(N^{-5/6}\left|g_{5,N}(x,W)\right| > \tau\epsilon_N\big) \\
&\quad + \mathbb{P}\big(\left|N^{-1} g_{6,N}(x,W) + \tau^2/2\right| > \tau\epsilon_N\big) \\
&\quad + \mathbb{P}\Big[\Big|\frac{1}{6!}\int_0^{N^{-1/6}} (N^{-1/6} - u)^6 \frac{d^7}{du^7} G_{u,N}(x,W)du\Big| > \tau\epsilon_N\Big] + \frac{5\tau\epsilon_N}{\sqrt{2\pi}},
\end{aligned} \tag{16}$$

where $\Phi(u)$ is the standard Gaussian distribution function. Lemma 7 gives the bound

$$\sup_u \left|\mathbb{P}\Big(N^{-1/2} g_{3,N}(x,W)/\tau \le u\Big) - \Phi(u)\right| \le \Big(\frac{1}{\sqrt{N}} + \frac{1}{\epsilon_N^2 N}\Big) F_2\Big(E_N \boldsymbol{\phi}_2(x)\Big) + h_\tau(F_3(E_N \boldsymbol{\phi}_3(x))) + \epsilon_N/\sqrt{2\pi}, \tag{17}$$

where $h_\tau(x) = \left|1 \vee \frac{\sqrt{x}}{\tau}\right| \left|1 - \frac{\tau}{\sqrt{x}}\right|$ is a continuous function vanishing at $\tau^2 = F_3(E\boldsymbol{\phi}_3(X))$. By Lemma 10 in the Appendix, we have $g_{k,N}(x,W) = NF_k(E_N \boldsymbol{r}_k(x,W))$, where F_k and $\boldsymbol{r}_k$ are as in Lemma 8, $k = 4, 5, 6$.

Replacing in the formulas of Lemma 10 the empirical averages with expectations with restpect to $\pi \times \mathbb{P}$, the reader can check that

$$F_4(\int \boldsymbol{r}_4 d\pi d\mathbb{P}) = -\frac{1}{24}\{[3E(\psi''(X)\psi'^2(X)) + 3E(\psi''^2(X)) + 6E(\psi'''(X)\psi'(X)) + 3E(\psi''''(X))] - ([H'(X)\psi'(X) - H''(X)])^2\}$$

$$= -\frac{1}{24}[3c\int (e^{\psi}\psi'')''dx - (c\int (e^{\psi}H')'dx)^2] = 0.$$

where X has the density $\pi(x) = ce^{\psi(x)}$ and $c = e^{K(\theta_*)}$. The last equality is obtained since both the summands are zero, by integration by parts. In fact by assumptions (A) and (B), $(e^{\psi}f)^{(k)}$ is integrable whenever f is any linear combination of derivatives of H and U, $k = 1, 2, 3, \ldots$. This implies that

$$\lim_{|x|\to\infty} (e^{\psi(x)}f(x))^{(k-1)} = 0,$$

otherwise $(e^{\psi}f)^{(k)}$ is not integrable. Next, $F_5(\int \boldsymbol{r}_5 d\pi d\mathbb{P}) = 0$ since each monomial in it contains at least one factor which is an odd power of W, hence it has mean zero. Finally,

$$F_6(\int \boldsymbol{r}_5 d\pi d\mathbb{P}) = -\frac{1}{1440}\{45E(\psi''^2(X)\psi'^2(X)) + 60E(\psi'''(X)\psi'^3(X)) + 270E(\psi'''(X)\psi''(X)\psi'(X)) + 135E(\psi'''^2(X)) + 180E(\psi'''(X)\psi'^2(X)) + 180E(\psi'''(X)\psi''(X)) + 180E(\psi'''''(X)\psi'(X)) + 60E(\psi''''''(X))$$

$$- 90[E(H''(X) + H'(X)\psi'(X))E(H'(X)(\psi'''(X) + \psi'(X)\psi''(X)))]$$

$$- 90[2E(H''(X)\psi'^2(X)) + 2E(H''(X)\psi''(X)) + 4E(H'''(X)\psi'(X)) + E(H''''(X))]E(H''(X) + H'(X)\psi'(X)) + + 45E(H'^2(X))(E(H''(X) + H'(X)\psi'(X))^2)\}$$

$$= -\frac{1}{1440}\{(45E(\psi''^2(X)\psi'(X)^2) + 90E(\psi'(X)\psi''(X)\psi'''(X)) + 75E(\psi'''^2(X))) - 90E(H''(X) + H'(X)\psi'(X)) E(H'(X)(\psi'''(X) + \psi'(X)\psi''(X))) + 45E(H'^2(X))E(H''(X) + H'(X)\psi'(X))^2\}$$

$$- \frac{60}{1440}\{E(\psi'''(X)\psi'^3(X)) + 3E(\psi'(X)\psi''(X)\psi'''(X))$$

$$+ E(\psi'''^2(X)) + 3E(\psi''''(X)\psi'^2(X)) +$$

$$+ 3E(\psi'''''(X)\psi'(X)) + E(\psi''''''(X))\}$$

$$+ \frac{90}{1440}\{2E(H''(X)\psi'^2(X)) + 2E(H''(X)\psi''(X))$$

$$+ 4E(H'''(X)\psi'(X)) + E(H''''(X))][E(H''(X) + H'(X)\psi'(X))\},$$

and this simplifies to

$$F_6(\int \boldsymbol{r}_5 d\pi d\mathbb{P}) = -\frac{\tau^2}{2},$$

because the first term in curly braces equals $-\frac{\tau^2}{2}$ by (15), the second term is proportional to

$$E(\psi'''(X)\psi'^3(X) + 3\psi'(X)\psi''(X)\psi'''(X) + \psi'''^2(X)$$
$$+ 3\psi''''(X)\psi'^2(X) + 3\psi'''''(X)\psi'(X) + \psi''''''(X)$$
$$= c\int (e^{\psi}\psi''')'''dx = 0,$$

and the third term in curly braces contains the multiplicative factor

$$E(H''(X) + H'(X)\psi'(X)) = c\int (e^{\psi}H')'dx = 0.$$

The last two displays equal zero by Assumptions (A) and (B), and integration by parts.

Next, using Lemma 8 three times, we find that

$$\mathbb{P}\big(N^{-2/3}\,|g_{4,N}(x,W)| > \tau\epsilon_N\big) = \mathbb{P}\big(N^{1/3}\,|F_4(E_N\boldsymbol{r}_4(x,W))| > \tau\epsilon_N\big)$$
$$\leq \frac{1}{N^{1-4/6}\epsilon_N^2}F_4\Big(E_N\boldsymbol{\phi}_4(x)\Big) \tag{18}$$

(19)

$$\mathbb{P}\big(N^{-5/6}\,|g_{5,N}(x,W)| > \tau\epsilon_N\big) = \mathbb{P}\big(N^{1/6}\,|F_5(E_N\boldsymbol{r}_5(x,W))| > \tau\epsilon_N\big)$$
$$\leq \frac{1}{N^{1-2/6}\epsilon_N^2}F_5\Big(E_N\boldsymbol{\phi}_5(x)\Big) \tag{20}$$

$$\mathbb{P}\big(\big|N^{-1}g_{6,N}(x,W) + \tau^2/2\big| > \tau\epsilon_N\big) = \mathbb{P}\Big(\Big|F_6(E_N\boldsymbol{r}_6(x,W)) - F_6(\int r_6 d\pi d\mathbb{P})\Big| > \tau\epsilon_N\Big)$$
$$\leq \frac{1}{N\epsilon_N^2}F_6\Big(E_N\boldsymbol{\phi}_6(x)\Big), \tag{21}$$

■

for $x \in \widehat{F}_{N,k}(\epsilon_N)$, where

$$\widehat{F}_{N,k}(\epsilon_N) = \Big\{x : \Big|E_N\boldsymbol{\beta}_k(x) - \int \boldsymbol{r}_k d\pi d\mathbb{P}\Big| < \frac{\tau\epsilon_N}{2}N^{k/6-1}\Big\}.$$

The remainder is estimated in Lemma 9 as

(22)

$$\mathbb{P}\Big[\Big|\frac{1}{6!}\int_0^{N^{-1/6}}(N^{-1/6}-u)^6\frac{d^7}{du^7}G_{u,N}(x,W)du\Big| > \tau\epsilon_N\Big] \le \frac{1}{\epsilon_N N^{7/6}}F_7\Big(E_N\boldsymbol{\phi}_7(x)\Big).$$

We now finish the proof of the theorem. Choose $\epsilon_N = N^{-\alpha}$ for $\alpha > 0$ small enough, and set

$$F_N = \Big\{x : \Big|E_N\boldsymbol{\phi}_k(x) - \int \boldsymbol{\phi}_k d\pi\Big| \le N^{-\alpha},$$
$$k = 2,3,4,5,6,7\Big\}\bigcap\Big(\cap_{k=4,5,6}\widehat{F}_{N,k}(N^{-1/12})\Big).$$

From Proposition 3 with $\lambda_N = N^{1/2-2\alpha}$, we deduce that $\pi_N(F_N^c) = o(N^{-k})$ as claimed. Using the bounds (17), (18), (20), (21), (22) and (16), we get that

$$\Big|\mathbb{P}\Big(G_{\sigma,N}(x,W) \le u\Big) - \Phi_{-\ell^2\tau^2/2,\ell^2\tau^2}(u)\Big| = O(N^{-\beta}).$$

The only term which needs to be discussed is the second one on the right hand side of (17), where we use a Lipschitz estimate for h_τ around $x = \tau^2$ and the explicit form of the sets F_N. □

Notice that in the above proof, the form of the sets F_N is suggested by the second term appearing on the right hand side of (17). Proposition 3 was given precisely to control the the probability of these sets.

Before proceeding with the proofs of those Lemmas needed for Proposition 6, we set up some notational conventions. Let us define $\psi_N : R \times R^N \to R$ as

$$\psi_N(t;x) = U(t) - H(t)E_N H(x), \tag{23}$$

and write

$$(E_N\psi_N)(x) = \frac{1}{N}\sum_{i=1}^N \psi_N(x_i;x).$$

We also denote by $\psi_N^{(k)}$ the k-th derivative of $\psi_N(t,x)$ with respect to t, and then define $E_N\psi_N^{(k)}$ similarly.

We write $\mathcal{D}$ for the algebra of polynomials in the absolute value of derivatives of H and U:

$$\mathcal{D} = \text{Algebra}\{f = \left|H^{(\ell)}\right|^q, f = \left|U^{(\ell)}\right|^q : \ell = 0,1,2,\dots, q = 0,1,2,\dots\}.$$

Lemma 7. *There exist polynomials F_2, F_3, and vectors $\boldsymbol{\phi}_2$, $\boldsymbol{\phi}_3$ with components in $\mathcal{D}$, such that for any $\epsilon > 0$,*

$$\sup_u \left|\mathbb{P}\Big(g_{3,N}(x,W)/\tau\sqrt{N} \le u\Big) - \Phi(u)\right| \le (\frac{1}{\sqrt{N}} + \frac{1}{\epsilon^2 N})F_2\Big(E_N\boldsymbol{\phi}_2(x)\Big)$$
$$+ 1 \vee \sqrt{F_3(E_N\boldsymbol{\phi}_3(x))}/\tau \cdot (1 - \tau/\sqrt{F_3(E_N\boldsymbol{\phi}_3(x))}) + \epsilon/\sqrt{2\pi},$$

Proof. Let us define

$$X_N = \sqrt{N}\Big(3E_N(\psi_N''\psi_N'(x)W) + E_N(\psi_N'''(x)W^3)$$
$$- 3E_N(H'\psi_N'(x))E_N H'(x)W - 3E_N H''(x)E_N(H'(x)W)\Big)$$

and $Y_N = 3\sqrt{N}E_N H''(x)(W^2-1)E_N H'(x)W$. From the expression of $g_{3,N}$ in the Appendix we find that

$$\frac{1}{\sqrt{N}}g_{3,N}(x,W) = X_N + Y_N$$

The term Y_N has zero mean, and we bound its variance as follows:

$$\mathbb{E}{Y_N}^2 = \frac{9}{N^3}\sum_{i,j} H''(x_i)^2 H'(x_j)^2 \mathbb{E}(W_i^2-1)^2 W_j^2 \tag{24}$$
$$\le \frac{\text{Const}}{N} E_N(H'')^2 E_N(H')^2.$$

The expression X_N is a sum of independent random variables, whose mean under the measure $\mathbb{P}$ is zero. We compute its variance τ_N^2 directly as follows:

$$
\begin{aligned}
(25)\quad \tau_N^2 = \frac{1}{144}\big\{&E_N(9{\psi_N''}^2(x){\psi_N'}^2(x) + 18\psi_N'(x)\psi_N''(x)\psi_N'''(x) + 15{\psi_N'''}^2(x)) \\
&- 18E_N[H'(x)\psi_N'(x) + H''(x)]E_N[\psi_N'(x)\psi_N''(x)H'(x) + \psi_N'''(x)H'(x)] \\
&+ 9E_N{H'}^2(x)(E_N H''(x))^2 + 18E_N H''(x)E_N(H'(x)\psi_N'(x))E_N{H'}^2(x) \\
&\qquad + 9E_N H'(x)^2(E_N(H'(x)\psi_N'(x)))^2 \\
&+ \frac{1}{N}[-36E_N(\psi_N''(x)\psi_N'(x)H''(x)H'(x)) - 48E_N(\psi_N'''(x)H''(x)H'(x)) \\
&- 12E_N(\psi_N'(x)\psi_N''(x)H'(x)H''(x)) - 48E_N(\psi_N'''(x)H''(x)H'(x)) \\
&+ 36E_N(H'(x)\psi_N'(x))E_N(H''(x){H'}^2(x)) + 18E_N({H'}^2(x)H''(x))E_N H''(x) \\
&\qquad + 36E_N({H''}^2(x){H'}^2(x))] + 72\frac{1}{N^2}E_N({H''}^2(x){H'}^2(x))\big\}.
\end{aligned}
$$

Now, in order to get the representation $\tau_N^2 = F_3(E_N\boldsymbol{\phi}_3(x))$, we insert into the above terms the explicit formula for ψ_N given in (23), expand the products and rearrange terms.

By Proposition 3, we have $\tau_N^2 \to \tau^2$ in distribution, where τ^2 is defined by (15) in Proposition 6. Using the formula (Petrov, 1995, Lemma 1.9, p.20) again and a simple estimate gives

$$
\begin{aligned}
\sup_u |\mathbb{P}(X_N + Y_N \le \tau u) - \Phi(u)| \le \sup |\mathbb{P}(X_N/\tau_N \le u) - \Phi(u)| + 1 \vee (\tau_N/\tau)\cdot(1 - (\tau/\tau_N)) \\
+ \mathbb{P}(Y_N > \epsilon\tau) + \epsilon/\sqrt{2\pi}
\end{aligned}
$$

and Esseen's inequality (Petrov, 1975, Theorem 3, p.111) for X_N/τ_N, Chebyshev's inequality and the estimate (24) for Y_N, we arrive at

$$
\begin{aligned}
\sup_u \left|\mathbb{P}\Big(g_{3,N}(x,W)/\tau\sqrt{N} \le u\Big) - \Phi(u)\right| \le \frac{1}{\sqrt{N}}\frac{\text{Const}}{\tau_N^3}\big(&E_N\left|\psi_N''(X)\psi_N'(X)\right|^3 \\
+ E_N\left|\psi_N'''(X)\right|^3 + E_N\left|H'(X)\right|^3 (E_N\left|H'(X)\psi_N'(X)\right|^3 + \left|H''(X)\right|^3)\big) \\
+ 1 \vee \tau_N/\tau\cdot(1 - \tau/\tau_N) + \frac{\text{Const}}{N\tau^2\epsilon^2}E_N{H''}^2(x)E_N{H'}^2(x) + \epsilon/\sqrt{2\pi},
\end{aligned}
$$

which is clearly in the form claimed. □

Lemma 8. *Let $F : \mathbb{R}^m \to \mathbb{R}$ be a polynomial and $r_h : \mathbb{R}^2 \to \mathbb{R}$, $h = 1, \ldots, m$, be of the form $r_h(x,W) = b_h(x)W^{\alpha_h}$, where b_h belongs to $\mathcal{D}$. Then there exists a* *polynomial $\bar{F}$ and a function $\bar{\boldsymbol{r}} \in \mathcal{D}$ such that, for $0 \leq \gamma < 1/2$,*

$$\mathbb{P}\Big[N^\gamma \left| F\Big(E_N \boldsymbol{r}(x,W)\Big) - F\Big(\int \boldsymbol{r} d\pi d\mathbb{P}\Big)\right| > \epsilon\Big] \leq \frac{1}{N^{(1-2\gamma)}\epsilon^2} \bar{F}\Big(E_N \bar{\boldsymbol{r}}(x)\Big)$$

holds for all $x \in \widehat{F}_N$, where

$$\widehat{F}_N = \Big\{x : \left| E_N \beta_h(x) - \int r_h d\pi d\mathbb{P}\right| < \epsilon N^{-\gamma}/2K, \quad h = 1, \ldots, m\Big\},$$

K is a local Lipschitz constant for F in a neighbourhood of the point $\int \boldsymbol{r} d\pi d\mathbb{P}$ and

$$\beta_h(x_i) = b_h(x_i)\mathbb{E}W^{\alpha_h}, \quad h = 1, \ldots, m.$$

Proof. We first notice that, when $x \in \widehat{F}_N$, we have

$$\mathbb{P}\left(N^\gamma \left| F(E_N \boldsymbol{r}(x,W)) - F(\int \boldsymbol{r} d\pi d\mathbb{P})\right| > \epsilon\right)$$
$$\leq \mathbb{P}\left(N^\gamma \left| F(E_N \boldsymbol{\beta}(x,W)) - F(\int \boldsymbol{r} d\pi d\mathbb{P})\right| > \epsilon/2\right).$$

Now define the random variables $V_{h,i} = r_h(x_i, W_i)$. These are clearly independent and have respective means $\beta_h(x_i)$. Now the difference

$$F\Big(\frac{1}{N}\sum_{i=1}^N V_{h,i}, h = 1, \ldots, m\Big) - F\Big(E_N \beta_h(x), h = 1, \ldots, m\Big)$$

is a polynomial in the centered variables $\frac{1}{N}\sum_{i=1}^N (V_{h,i} - \beta_h(x_i))$, with coefficients which are polynomials in the variables $E_N \beta_h(x)$, as can be immediately seen by taking a Taylor series centered in $E_N \beta_h(x)$. We proceed to bound the second moment of the generic monomials in this representation,

$$M_{h_1,\ldots,h_s} = \mathbb{E}\Big[\Big(\frac{1}{N}\sum_{i=1}^N Y_i^{(h_1)}\Big) \cdots \Big(\frac{1}{N}\sum_{i=1}^N Y_i^{(h_s)}\Big)\Big]^2,$$

where $Y_i^{(h)} = V_{h,i} - b_h(x_i)$. By using Lemma 11, we get the bound

$$M_{h_1,\ldots,h_s} \leq \frac{1}{N} \times \text{Polynomial in empirical averages of elements in } \mathcal{D}_{p,k}.$$

The proof is completed by an application of Chebyshev's inequality. □

Lemma 9. *Choose $\sigma = 1/N^{1/6}$, then there exist a polynomial P and a function $\boldsymbol{\mu}$ with components in $\mathcal{D}$ such that*

$$\mathbb{P}\Big[\left|\frac{1}{6!}\int_0^\sigma (\sigma - u)^6 \frac{d^7}{du^7} G_{u,N}(x,W) du\right| > \epsilon\Big] \leq \frac{1}{\epsilon N^{7/6}} \; P\Big(\frac{1}{N}\sum_{i=1}^N \boldsymbol{\mu}(x_i)\Big)$$

Proof. By Markov's inequality and Lemma 10, we have

$$\begin{aligned}
\mathbb{P}\Big[\Big|\frac{1}{6!}\int_0^\sigma (\sigma-u)^6\frac{d^7}{du^7}G_{u,N}(x,W)du\Big| > \epsilon\Big]& \\
&\le \frac{1}{6!\epsilon}\mathbb{E}\left|\int_0^\sigma (\sigma-u)^6\frac{d^7}{du^7}G_{u,N}(x,W)du\right| \\
&\le \frac{1}{6!\epsilon}\int_0^\sigma (\sigma-u)^6\mathbb{E}\left|\frac{d^7}{du^7}G_{u,N}(x,W)\right|du \\
\le \frac{1}{6!\epsilon}\int_0^\sigma (\sigma-u)^6 N\mathbb{E}\left|\sum_{k=0}^{9} u^k P_k\Big(E_N\rho_\ell(x)\varphi_\ell(Y_u)W^{r_\ell};\ell=1,\dots,m\Big)\right|du& \\
\le \frac{1}{6!\epsilon}\int_0^\sigma (\sigma-u)^6 N\sum_{k=0}^{9} u^k\mathbb{E}\left|P_k\Big(E_N\rho_\ell(x)\varphi_\ell(Y_u)W^{r_\ell};\ell=1,\dots,m\Big)\right|du,& \\
&\le \frac{1}{6!\epsilon}\int_0^\sigma (\sigma-u)^6 N\sum_{k=0}^{9} u^k C_k(1+\mathbb{E}\left|E_N\rho_\ell(x)\varphi_\ell(Y_u)W^{r_\ell}\right|^{s_k})du.
\end{aligned}$$

Now using Jensen's inequality, we have the bound

$$\begin{aligned}
\left|E_N\rho_\ell(x)\varphi_\ell(Y_u)W^{r_\ell}\right|^{s_k} &\le E_N\left|\rho_\ell(x)\varphi_\ell(Y_u)W^{r_\ell}\right|^{s_k} \\
&\le E_N\left|\rho_\ell(x)\right|^{s_k}\left|\varphi_\ell(Y_u)\right|^{s_k}\left|W^{r_\ell}\right|^{s_k} \\
&\le CE_N(1+|x|^{r_k})^{s_k}(1+|Y_u|^{r_k})^{s_k}(1+|W^{r_\ell}|^{r_k})^{s_k} \\
&\le C'(1+E_N|x|^{t_k}+E_N|Y_u|^{t_k}+E_N|W|^{t_k})
\end{aligned}\tag{26}$$

for some sufficiently large t_k and constant C', by the arithmetic-geometric mean inequality. From (11), we have

$$|Y_{u,i}| \le |x_i| + u|W_i| + (u^2/2)|U'(x_i)| + (u^2/2)|H'(x_i)|\left|\frac{1}{N}\sum_{j=1}^{N}H(x_j)\right|,$$

and replacing this into the estimate (26) gives

$$\begin{aligned}
\mathbb{P}\Big[\Big|\frac{1}{6!}\int_0^\sigma (\sigma-u)^6\frac{d^7}{du^7}G_{u,N}(x,W)du\Big| > \epsilon\Big]& \\
&\le \frac{\text{Const}}{\epsilon}\int_0^\sigma (\sigma-u)^6 N\sum_{k=0}^{9}u^k C_k\Big[1+ \\
\sum_\gamma u^\gamma P_\gamma\Big(E_N|x|, E_N|U'(x)|, E_N|H'(x)|, E_N|H(x)|\Big)\Big]du& \\
\le \frac{\text{Const}N\sigma^7}{\epsilon}\sum_\delta \sigma^\delta\widetilde{P}_\delta(E_N|x|, E_N|U'(x)|, E_N|H'(x)|, E_N|H(x)|)&
\end{aligned}$$

$$\leq \frac{\text{Const}N^{7/6}}{\epsilon} P(E_N \eta)$$

where we first summed over a finite number of indices γ, and then integrated over u, using the bound $\int_0^\sigma (\sigma - u)^6 u^\gamma du \leq 2^6 \sigma^{7+\gamma}$, to get the sum over δ. After this, we bounded the derivatives of H and U by Assumption (A), and finally, we simultaneously bounded all the polynomials $\widetilde{P}_\delta$ by a single, much larger polynomial P. Upon substituting $\sigma = 1/N^{1/6}$, we arrived at the stated form. □

4. Proof of the convergence result

For a given configuration $x \in (\mathbb{R}^n)^N$, let us begin by expanding the discrete generator $A_N f(x) = \mathbb{E}[f(X_{t+1}^{(N)}) - f(x) \,|\, X_t^{(N)} = x]$ on a suitably smooth test function f (C^∞ with compact support) which moreover depends on only a finite set $\Delta(f)$ of components. For the sake of simplicity, we suppose that f depends only on one component x_p.

$$\begin{aligned} A_N f(x) &= \mathbb{E}\Big[\big(f(Y_{\sigma,p}) - f(x_p)\big) 1 \wedge e^{G_{\sigma,N}(x,W)}\Big] \\ &= \mathbb{E}\Big[\Big(\sigma f'(x_p) W_p + \frac{\sigma^2}{2} W_p f''(x_p) W_p \\ &\quad + \frac{\sigma^2}{2} f'(x_p)(U'(x_p) - H'(x_p) E_N H(x))\Big) 1 \wedge e^{G_{\sigma,N}(x,W)}\Big] + \sigma^2 r(\sigma, x), \end{aligned} \tag{27}$$

where

$$\begin{aligned} r(\sigma,x) = \frac{\sigma}{3!}\mathbb{E}\Big([f'''(Y_{\widetilde{\sigma}})(W_1 + \widetilde{\sigma}(U'(x_1) - H'(x_1)E_N H(x)))^3 \\ +3f''(Y_{\widetilde{\sigma}})(U'(x_1) - H'(x_1)E_N H(x))(W_1 + \widetilde{\sigma}(U'(x_1) - H'(x_1)E_N H(x)))] 1 \wedge e^{G_{\widetilde{\sigma},N}(x,W)}\Big), \end{aligned}$$

where $0 \leq \widetilde{\sigma} \leq \sigma$. Observe that if we define the set

$$\widetilde{F}_N = \big\{x : |U'(x_p)|\,, |U''(x_p)|\,, |H'(x_p)|\,, |H''(x_p)| \leq N^\kappa\big\},$$

for κ suitably small, then with $\sigma = N^{-1/6}$, the remainder $r(\sigma, x)$ goes to zero uniformly for $x \in \widetilde{F}_N \cap F_N$, as $N \to \infty$, so it is dropped from further consideration. The bounds on the second derivatives in $\widetilde{F}_N$ have been included for later purposes.

If Γ is a smooth function of a real variable, Taylor's formula gives

$$
\begin{aligned}
1\wedge e^{\Gamma(1)} &= 1\wedge e^{\Gamma(0)} + \int_{\{u:\Gamma(u)<0 \text{ and } 0\le u\le 1\}} \Gamma'(u)e^{\Gamma(u)}du \\
&= 1\wedge e^{\Gamma(0)} + \sum_{i=1}^{k}\Gamma'(s_i)e^{\Gamma(s_i)}(t_i-s_i) + \sum_{i=1}^{k}\int_{s_i}^{t_i}\int_{s_i}^{u}\Big(\Gamma''(v)+\Gamma'(v)^2\Big)e^{\Gamma(v)}dvdu \\
&= 1\wedge e^{\Gamma(0)} + \sum_{i=1}^{k}\Gamma'(s_i)e^{\Gamma(s_i)}(t_i-s_i) + \sum_{i=1}^{k}\int_{s_i}^{t_i}(t_i-v)\Big(\Gamma''(v)+\Gamma'(v)^2\Big)e^{\Gamma(v)}dv,
\end{aligned}
$$

where the sums are over a partition

$$\{u:\Gamma(u)<0, 0\le u\le 1\} = \cup_{i=1}^{k}(s_i,t_i),$$

with $0\le s_1<t_1\le s_2<\cdots\le s_k<t_k\le 1$. We apply this formula now to the function $\Gamma(u)=G_{\sigma,N}(x,uW_p,W_{p^c})$, arriving at

$$
\begin{aligned}
(28)\quad & \mathbb{E}(1\wedge e^{G_{\sigma,N}(x,W)}\,|\,W_p) = \mathbb{E}(1\wedge e^{G_{\sigma,N}(x,W)}\,|\,W_p=0) \\
&+ W_p\mathbb{E}\Big(\sum_{i=1}^{k}\frac{d}{d(s_iW_p)}G_{\sigma,N}(x,s_iW_p,W_{p^c})e^{G_{\sigma,N}(x,s_iW_p,W_{p^c})}(t_i-s_i)\,|\,W_p\Big) \\
&+ \mathbb{E}\Big(\sum_{i=1}^{k}\int_{s_i}^{t_i}(t_i-v)\Big(\Gamma''(v)+\Gamma'(v)^2\Big)e^{\Gamma(v)}dv\,|\,W_p\Big)
\end{aligned}
$$

We now substitute (28) into the expression (27), and obtain, placing all unwanted terms into a remainder $R_N(x,f)$,

$$
\begin{aligned}
(29)\quad & \frac{1}{\sigma^2}A_Nf(x) = \frac{1}{2}f''(x_p)\mathbb{E}(1\wedge e^{G_{\sigma,N}(x,W)}\,|\,W_p=0) \\
&+ \frac{1}{2}f'(x_p)(U'(x_p)-H'(x_p)E_NH(x))\mathbb{E}(1\wedge e^{G_{\sigma,N}(x,W)}) + R_N(x,f).
\end{aligned}
$$

Keeping only the leading order term in $R_N(x,f)$ (as the remaining ones can be treated in a similar way), we have

$$
\begin{aligned}
& R_N(x,f) = \\
& \frac{1}{\sigma}f'(x_p)\mathbb{E}\Big[W_p^2\mathbb{E}\Big(\sum_{i=1}^{k}\frac{d}{d(sW_p)}G_{\sigma,N}(x,s_iW_p,W_{p^c})e^{G_{\sigma,N}(x,s_iW_p,W_{p^c})}(t_i-s_i)\,|\,W_p\Big)\Big]
\end{aligned}
$$

Observe that we can bound

$$\left|\mathbb{E}\Big(\sum_{i=1}^{k}\frac{d}{d(sW_p)}G_{\sigma,N}(x,s_iW_p,W_{p^c})e^{G_{\sigma,N}(x,s_iW_p,W_{p^c})}(t_i-s_i)\,|\,W_p\Big)\right| \le \mathbb{E}\Big(\sup_{s\le 1}\left|\frac{d}{d(sW_p)}G_{\sigma,N}(x,sW_p,W_{p^c})\right|\,|\,W_p\Big), \tag{30}$$

since $G_{\sigma,N}(x,W)$ is negative on $\cup_{i=1}^{k}(s_i,t_i)$. Let us write explicitly

$$\frac{d}{d(sW_p)}G_{\sigma,N}(x,W)=\frac{\sigma}{2}\Big(U'(Y_p)-U'(x_p)-H'(Y_p)E_NH(Y)+H'(x_p)E_NH(x)\Big) - \frac{\sigma^2}{2}\Big(W_p(U''(Y_p)-H''(Y_p)E_NH(Y))+H'(Y_p)\frac{1}{N}\sum_{i=1}^{N}W_iH'(Y_i)\Big) + \frac{\sigma^3}{8}\Big(U''(Y_p)-H''(Y_p)E_NH(Y)-H'(Y_p)E_NH'(Y)\Big),$$

then we can rewrite the right hand side of (30) as

$$\mathbb{E}\Big(\sup_{s\le 1}\left|\frac{d}{d(sW_p)}G_{\sigma,N}(x,sW_p,W_{p^c})\right|\,|\,W_p\Big) = \frac{\sigma}{2}\mathbb{E}\Big(\sup_{s\le 1}\left|U'(\widetilde{Y}_{\sigma,p})-U'(x_p)-H'(\widetilde{Y}_{\sigma,p})E_NH(\widetilde{Y_\sigma})+H'(x_p)E_NH(x)\right|\,|\,W_p\Big)+S_N(x,W_p),$$

where

$$\widetilde{Y}_{\sigma,q}=\begin{cases} x_p+\sigma sW_p+\frac{\sigma^2}{2}(U'(x_p)-H'(x_p)E_NH(x)) & \text{if } q=p\\ Y_{\sigma,q} & \text{if } q\neq p.\end{cases}$$

For $x\in\widetilde{F}_N\cap F_N$, we have, using the fundamental theorem of calculus,

$$\begin{aligned}
&\mathbb{E}(W_p^2\mathbb{E}(\sup_{s\le 1}\left|\frac{d}{dsW_p}G_{\sigma,N}(x,sW_p,W_{p^c})\right|\,|\,W_p)) \le \\
&\mathbb{E}\left|W_p^2\mathbb{E}\sup_{s\le 1}\left|U'(\widetilde{Y}_p)-U'(x_p)-H'(\widetilde{Y}_p)E_NH(\widetilde{Y})+H'(x_p)E_NH(x)\right|\,|\,W_p)\right| \\
&\le \mathbb{E}\left|W_p^2\right|\sup_{s\le 1}\left|U'(\widetilde{Y}_p)-U'(x_p)\right|+\mathbb{E}\left|W_p^2\right|\sup_{s\le 1}\left|H'(x_p)-H'(\widetilde{Y}_p)\right|\left|E_NH(x)\right| \\
&\qquad + \mathbb{E}\left|W_p^2\right|\sup_{s\le 1}\left|H'(\widetilde{Y}_p)\right|E_N\left|H(\widetilde{Y})-H(x)\right|.
\end{aligned} \tag{31}$$

Observe that, when T is either U', H' or H, we can write

$$T(Y_p)-T(x_p)=\frac{\sigma^2}{2}(U'(x_p)-H'(x_p)E_NH(x))\int_0^1 T'(x_p+\frac{v\sigma^2}{2}(U'(x_p)-H'(x_p)E_NH(x)))dv + \sigma W_p\int_0^s T'(x_p+\frac{\sigma^2}{2}(U'(x_p)-H'(x_p)E_NH(x))+u\sigma W_p)du.$$

By bounding the derivative of T with a suitable polynomial it is easily obtained that for $x \in \widetilde{F}_N \cap F_N$ we have

$$\mathbb{E} \sup_{s \leq 1} |T(Y_p) - T(x_p)| \leq \sigma \cdot \text{Const} N^{\kappa r},$$

for some integer r, so that for $\sigma = N^{-1/6}$ and $x \in F_N \cap \widetilde{F}_N$ and κr small enough, we find that

$$|T(Y_p) - T(x_p)| \leq \text{Const} \cdot N^{-(1/6 - \kappa r)}.$$

By substituting this bound into (31), it is easy to see that for $x \in F_N \cap \widetilde{F}_N$, the right hand side is bounded by $O(N^{\kappa r - 1/6})$ uniformly over x. With similar arguments, we can show that $\mathbb{E}(S_N(x, W_p) W_p^2)$ is bounded by $o(N^{\kappa r - 1/6})$ uniformly over $x \in F_N \cap \widetilde{F}_N$.

Let us now define an operator A by

$$Af(x) = \frac{\ell^2}{2} \mathbb{E}(1 \wedge e^{\mathcal{N}}) \Big[f''(x_p) + \Big(U'(x_p) - H'(x_p) \int H d\pi \Big) f'(x_p) \Big],$$

where $\mathcal{N}$ is a Gaussian random variable with mean $-\tau^2/2$ and variance τ^2. This is the generator of the diffusion process Z described in (9), where $v(\ell) = 2\ell^2 \mathbb{E}(1 \wedge e^{\mathcal{N}})$, since the latter expected value is equal to $\Phi(-\ell\tau/2)$ by a direct calculation.

With this definition, we have

$$\begin{aligned}
(32) \quad & \left| \sigma_N^{-2} A_N f(x) - \ell^{-2} A f(x) \right| \\
& \leq \frac{1}{2} |f''(x_p)| \left| \mathbb{E}\Big(1 \wedge e^{G_{\sigma,N}(x,W)} \,|\, W_p = 0\Big) - \mathbb{E}\Big(1 \wedge e^{\mathcal{N}}\Big) \right| \\
& \quad + |f'(x_p)| \, |U'(x_p)| \left| \mathbb{E}(1 \wedge e^{G_{\sigma,N}(x,W)}) - \mathbb{E}(1 \wedge e^{\mathcal{N}}) \right| \\
& + |f'(x_p)| \, |H'(x_p)| \Big(|E_N H(x)| \left| \mathbb{E}(1 \wedge e^{G_{\sigma,N}(x,W)}) - \mathbb{E}(1 \wedge e^{\mathcal{N}}) \right| \\
& \quad + \left| E_N H(x) - \int H d\pi \right| \left| \mathbb{E}(1 \wedge e^{\mathcal{N}}) \right| \Big)
\end{aligned}$$

In order to bound the above expression, we note that

$$\left| \mathbb{E}\Big(1 \wedge e^{G_{\sigma,N}(x,W)} \,|\, W_p = 0\Big) - \mathbb{E}\Big(1 \wedge e^{G_{\sigma,N}(x,W)}\Big) \right| \leq \mathbb{E} \left| G_{\sigma,N}(x, 0, W_{p^c}) - G_{\sigma,N}(x, W) \right|.$$

The right hand side goes to zero uniformly on the set $\widetilde{F}_N \cap F_N$, by arguments similar to the above. The first term on the right hand side of (32) goes to zero uniformly on the sets $F_N \cap \widetilde{F}_N$, since

$$\begin{aligned}&\left|\mathbb{E}(1 \wedge e^{G_{\sigma,N}(x,W)}) - \mathbb{E}(1 \wedge e^{\mathcal{N}})\right| \\ &\qquad\qquad \leq \text{Const} \cdot \sup_u \left|\mathbb{P}\Big(G_{\sigma,N}(x,W) \leq u\Big) - \Phi_{-\ell^2\tau^2/2,\ell^2\tau^2}(u)\right|,\end{aligned}$$

an estimate derived by integration by parts. In order to show that the remaining terms tend to zero uniformly on those same sets $F_N \cap \widetilde{F}_N$, we only have to choose κ in the definition of $\widetilde{F}_N$ to be smaller than the β given in Proposition 6.

We conclude the proof by collecting the previous facts together. For any test function f, we have shown that

$$\lim_{N\to\infty} \sup_{x\in F_N\cap\widetilde{F}_N} \left|\sigma_N^{-2} A_N f(x) - \ell^{-2} A f(x)\right| = 0. \tag{33}$$

Now suppose that $X_0^{(N)} \sim \pi_N$, and consider the probabilities

$$\mathbb{P}\left(X^{(N)}_{[N^{1/3}t]} \notin F_N \cap \widetilde{F}_N \text{ for some } t \leq T\right) \leq [N^{1/3}T]\pi_N(x : x \notin F_N \cap \widetilde{F}_N),$$

where we have used the stationarity of π_N. Then immediately by Proposition 6, we see that

$$\lim_{N\to\infty} \mathbb{P}\left(X^{(N)}_{[N^{1/3}t]} \in F_N \cap \widetilde{F}_N \text{ for all } t \leq T\right) = 1.$$

Consequently, by (Ethier and Kurtz, 1986, Corollary 8.9, p.233), the weak convergence (10) holds as stated. This ends the proof of Theorem 4.

5. Appendix

In this appendix, we list the terms in the Taylor expansion given in Lemma 5. We begin with the following lemma:

Lemma 10. *For $h = 0, 1, 2, \ldots$, we have the representation*

$$\frac{d^h}{d\sigma^h} G_{\sigma,N}(x,W) = N \sum_{k=0}^{h+2} \sigma^k P_k\Big(E_N \rho_\ell(x)\varphi_\ell(Y_\sigma) W^{r_\ell}; \ell = 1,\ldots,m\Big), \tag{34}$$

for some integer m, where ρ_ℓ, $\varphi_\ell \in \mathcal{D}$. In particular, for $\sigma = 0$, we have

$$g_{3,N}(x,W) = -\frac{N}{12}\Big(E_N(3\psi''\psi'W + \psi'''W^3) - 3E_N(H'W)E_N(H'\psi'_N) \\ - 3E_N(H''W^2)E_N(H'W)\Big)$$

Moreover,

$$g_{4,N} = -\frac{N}{24}\Big(E_N(3\psi''\psi'^2 + 3\psi''^2W^2 + 6\psi'''\psi'W^2 + \psi''''W^4) - \{3(E_N H'\psi')(E_N H'\psi') \\ + 6(E_N H''W^2)(E_N H'\psi') + 3(E_N H''W^2)(E_N H''W^2)\} + \delta_{4,N}\Big)$$

and

$$g_{6,N} = -\frac{N}{1440}\Big(E_N(45\psi''^2\psi'^2 + 60\psi'''\psi'^3 + 90\psi'''\psi''\psi'W^2 + 180\psi'''\psi''\psi' \\ + 45\psi'''W^4 + 180\psi''''\psi'^2W^2 + 60\psi''''\psi''W^4 + 60\psi'''''\psi'W^4 + 4\psi''''''W^6 \\ - [90(E_N H'\psi'\psi'')(E_N H'\psi') + 180(E_N H''\psi'^2)(E_N H'\psi') \\ + 90(E_N\psi'''H'W^2)(E_N H'\psi') + 90(E_N H'\psi''\psi')(E_N H''W^2) + 180(E_N H''\psi'')(E_N H'\psi') \\ + 90(E_N\psi'''H'W^2)(E_N H''W^2) + 360(E_N H'''\psi'W^2)(E_N H''\psi') + 180(E_N H''W^2)(E_N H''\psi'^2) \\ + 180(E_N H''W^2)(E_N H''\psi''W^2) + 60(E_N H''''W^4)(E_N H'\psi') + 360(E_N H''W^2)(E_N H'''\psi'W^2) \\ + 60(E_N H''''W^4)(E_N H''W^2)] + [45(E_N H'^2)(E_N H'\psi')(E_N H'\psi') \\ + 90(E_N H'^2)(E_N H''W^2)(E_N H'\psi') + 45(E_N H'^2)(E_N H''W^2)(E_N H''W^2)] + \delta_{6,N}\Big),$$

where $\delta_{4,N}$ and $\delta_{6,N}$ are sums of monomials in empirical averages of functions involving at least one odd power of W, that asymptotically have mean zero.

Proof. The representation (34) can be verified by induction, noting that it is true for $h = 0$, by (12). The stated formulas are found by explicit computation. □

Lemma 11. *Define in $\mathbb{R}^r$ the vector valued, centered and independent random variables $Y_i^{(j)} = b^{(j)}(x_i)W_i^{(j)}$, where $W_i^{(j)}$ are i.i.d. standard Gaussians for $i = 1, \dots, r$, and $j = 1, \dots, N$. Then*

$$\mathbb{E}\Big(\prod_{j=1}^{r}\big(\frac{1}{N}\sum_{i=1}^{N}Y_i^{(j)}\big)\Big)^2 \leq \\ \frac{1}{N^r}\sum_{m=1}^{r}\frac{1}{N^{r-m}}\sum_{|\mathcal{P}|=m}\big(\frac{1}{N}\sum_{h_1=1}^{N}b^{A_1}(x_{h_1})\big)\cdots\big(\frac{1}{N}\sum_{h_m=1}^{N}b^{A_m}(x_{h_m})\big), \tag{35}$$

where the sum is taken over partitions $\mathcal{P} = \{A_1, \dots, A_m\}$ *of the set of repeated indices* $I = \{1, \dots, r, 1, \dots, r\}$ *such that* $I = \cup_{k=1}^m A_k$ *and each* A_k *contains at least two elements of* I.

Proof. Begin by writing

$$\mathbb{E}[\prod_{j=1}^r (\frac{1}{N}\sum_{i=1}^N Y_i^{(j)})]^2 = \frac{1}{N^{2r}} \sum_{i_1=1}^N \cdots \sum_{i_r=1}^N \sum_{k_1=1}^N \cdots \sum_{k_r=1}^N \mathbb{E} Y_{i_1}^{(1)} \cdots Y_{i_r}^{(r)} Y_{k_1}^{(1)} \cdots Y_{k_r}^{(r)}.$$

A summand in the last expression is zero as soon as there exists an index ($i_1, \dots, i_r$ or $k_1, \dots, k_r$) whose value is *not* repeated by another. This follows by the independence and zero mean property of the Y_i. Another way of rearranging this sum is therefore as follows: partition the set I into a finite union $I = A_1 \cup \cdots \cup A_m$, where each $|A_k| \geq 2$ for each k. We write $Y_i^{A_k} = \prod_{j \in A_k} Y_i^{(j)}$ to simplify notation. Then the sum on the left is bounded above in absolute value by

$$\sum_{\mathcal{P}} \sum_{|\mathcal{P}|=m} \underbrace{\sum_{h_1=1}^N \cdots \sum_{h_m=1}^N}_{h_i \neq h_k \text{ if } k \neq i} \mathbb{E} \left| Y_{h_1}^{A_1} \right| \cdots \left| Y_{h_m}^{A_m} \right|. \tag{36}$$

Since the sum is over non repeating indices $h_1, \dots, h_m$, we have by independence $\mathbb{E} \left| Y_{h_1}^{A_1} \right| \cdots \left| Y_{h_m}^{A_m} \right| = b^{A_1}(x_{h_1}) \cdots b^{A_m}(x_{h_m})$, where $b^{A_k}(x) = \mathbb{E} \prod_{j \in A_k} \left[b(x) W_1 \right]^{(j)}$. Now the summand in (36) is positive, so we can bound the sum above by a sum over all (possibly repeating) indices $h_1, \dots, h_m$, and after rearrangeing the sums, we obtain (35). □

References

[1] Ethier, S.N. and Kurtz, T.G. *Markov processes. Characterization and convergence.* John Wiley and Sons.

[2] Breyer, L.A. and Roberts, G.O. (1998) *From Metropolis to diffusions: Gibbs states and optimal scaling.*

[3] Deuschel, J.D. and Stroock, D.W. (1989) *Large Deviations.* Academic Press.

[4] Rudin, W. (1991) *Functional Analysis.* McGraw-Hill, New York.

[5] Ellis, R. S. (1985) Entropy, Large Deviations and Statistical Mechanics, Springer Verlag, New York.

[6] Kusuoka, S. and Tamura, Y. (1984) Gibbs measures for mean field potentials, Jour. Fac. Sci. Tokio Sect. IA, 31, 223-245.

[7] Daras,T. (1997) Large and moderate deviations for the empiricla measures of an exchangeable sequence, Stat.Prob.Lett. 36, 91-100

[8] de Acosta, A. (1994) Projective Systems in Large Deviation Theory II:some applications, Bikhaueser Progress in Probability 35, 241-250

[9] Dembo,A. and Zeitouni, O. (1993) Large Deviations Techniques and Applications, Jones abd Bartlett Publishers, Boston-London

Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, 16, Mill Lane, Cambridge CB2 1SB

E-mail address: L.Breyer@statslab.cam.ac.uk

University of L'Aquila, Italy

E-mail address: piccioni@aquila.infn.it

University of Rome 2, Italy

E-mail address: scarlatt@axp.mat.uniroma2.it