UNIVERSALITY FOR LANGEVIN-LIKE SPIN GLASS DYNAMICS

AMIR DEMBO, EYAL LUBETZKY, AND OFER ZEITOUNI

ABSTRACT. We study dynamics for asymmetric spin glass models, proposed by Hertz et al. and Sompolinsky et al. in the 1980's in the context of neural networks: particles evolve via a modified Langevin dynamics for the Sherrington–Kirkpatrick model with soft spins, whereby the disorder is i.i.d. standard Gaussian rather than symmetric. Ben Arous and Guionnet (1995), followed by Guionnet (1997), proved for Gaussian interactions that as the number of particles grows, the short-term empirical law of this dynamics converges a.s. to a non-random law $\mu_\star$ of a "self-consistent single spin dynamics," as predicted by physicists. Here we obtain universality of this fact: For asymmetric disorder given by i.i.d. variables of zero mean, unit variance and exponential or better tail decay, at every temperature, the empirical law of sample paths of the Langevin-like dynamics in a fixed time interval has the same a.s. limit $\mu_\star$.

1. INTRODUCTION

Consider the dynamics for asymmetric spin glass models, studied in the context of neural networks e.g. by Hertz et al. [20] and Cristani and Sompolinsky [12], given by

$$dX_t^{(i)} = dB_t^{(i)} - U_1'(X_t^{(i)})dt + \frac{\beta}{\sqrt{N}} \sum_{j=1}^{N} J_{ij} X_t^{(j)} dt \qquad (i = 1, \ldots, N), \tag{1.1}$$

where B_t is N-dimensional Brownian motion, $X_t \in [-\mathfrak{s}, \mathfrak{s}]^N$ for some finite $\mathfrak{s}$, the potential U_1 is some smooth function satisfying that $U_1(x) \to \infty$ as $|x| \to \mathfrak{s}$ (e.g. a double-well potential at ± 1 with $\mathfrak{s} = 2$), the parameter $\beta > 0$ is the inverse-temperature and the interactions J_{ij} are quenched (frozen) i.i.d. standard Gaussian random variables.

If instead one were to take a symmetric disorder (that is, $J_{ij} = J_{ji}$ i.i.d. standard Gaussian for each pair $\{i, j\}$) then the stochastic differential system (SDS) (1.1) would be precisely Langevin dynamics for the soft-spin Sherrington–Kirkpatrick (SK) model; see, e.g., [23, 27, 28] and [3, 4, 19] for studies of the short-term dynamics in that case.

The asymmetric nature of the disorder J_{ij} aids some aspects of the analysis via the extra independence, yet makes the dynamics non-reversible, whence various useful tools (e.g., the Fluctuation Dissipation Theorem used in [28] to analyze the symmetric case) become unavailable. As argued e.g. in [12] (see also [14, 22] on the related Hopfield model [21]), the asymmetric disorder seems a better model for the interactions between neurons (cf. Remark 1.4 for other flavors of the model in the context of neural networks).

Many of the dynamical quantities of interest, such as spin autocorrelation and response functions, may be read from the thermodynamic limit ($N \to \infty$) of the empirical measure μ_N of sample-paths of the N particles in a given time interval $[0, T]$; that is,

$$\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_\cdot^{(i)}} \in \mathcal{M}_1(C([0, T])). \tag{1.2}$$

Ben Arous and Guionnet [2], followed by [19] (cf. [4]) were able to show that, from an i.i.d. initial state, μ_N converges a.s. to a law $\mu_\star$ of a *self-consistent single-spin dynamics*, a non-Markovian diffusion involving only one spin, as predicted for this system in [12]. The proofs in [2,19] (and in follow-up works on variants of this model, e.g., Glauber-like dynamics [17,26] and the dynamics where the SDS has an extra non-linearity [15]) relied in an essential way on special properties of Gaussian random variables.

In the related SK model, the first rigorous proof [18, 30] that the free energy has an a.s. limit was specific to Gaussian disorder, as was the identification of this limit. Talagrand [29] later proved that the same limit must be obtained under interactions of Bernoulli ± 1 random variables. This universality property was further generalized in [9,10] to any i.i.d. interactions given by a variable J satisfying $\mathbb{E}J = 0$ and $\mathbb{E}J^2 = 1$.

Our goal in this work is to obtain a similar universality result for the system (1.1), where a self-consistent a.s. limit was till now rigorously verified only in the case of Gaussian interactions. To be precise, consider the probability measures $\mathbb{P}_N^\beta$ of the triplet $(\mathbf{J}, B_\cdot, X_\cdot)$ corresponding to the SDS (1.1) with an initial state that is a product $\nu_0^{\otimes N}$ which places no mass on the boundary, i.e.,

$$\nu_0 \in \mathcal{M}_1((-\mathfrak{s}, \mathfrak{s}))\,,$$

and the $C_2((-\mathfrak{s}, \mathfrak{s}))$ potential function $U_1(x) \to \infty$ as $|x| \to \mathfrak{s}$ fast enough to confine the solution of (1.1) within $(-\mathfrak{s}, \mathfrak{s})$. Specifically, suppose (as in [2, p. 458]), that

$$\lim_{|x|\uparrow\mathfrak{s}} \int_0^x e^{2U_1(t)}\Big(\int_0^t e^{-2U_1(v)}dv\Big)dt = \infty\,, \tag{1.3}$$

which is satisfied for instance by $U_1(x) = -\log(\mathfrak{s}^2 - x^2)$. In this context, the heuristic reasoning for the expected universality, as in the case of the free energy in the SK model, is due to the invariance principle, whereby one expects the interaction term $N^{-1/2}\mathbf{J}X_t$ in (1.1) to approximately follow a Gaussian law when $N \to \infty$, irrespective of the marginal laws of the independent disorder variables J_{ij}. However, even for fully independent (i.e. non-symmetric), Gaussian disorder, the limit $\mu_\star$ of μ_N is characterized only as the global minimum of a certain rate function, corresponding to the variational problem of a large deviation principle (LDP). Consequently, one has to establish the sought-after universality at the level of large deviations. For $\{J_{ij}\}$ which are fully i.i.d Gaussian variables and high temperature (i.e. $\beta^2\mathfrak{s}^2T < 1$), such LDP was proved in [2] by relying on exact Gaussian calculus for the Radon–Nykodim derivative (RND) w.r.t. a reference system with independent particles, corresponding to the $\beta = 0$ measure (the corresponding a.s. convergence $\mu_N \to \mu_\star$ was thereafter extended in [19] to all $\beta < \infty$). Unfortunately, such explicit calculus does not exist for any other law of interactions. Moreover, any attempt to control the RND of Gaussian vs. non-Gaussian interactions via an argument such as Lindeberg's method must be done with utmost care, since it typically yields only an N^{-c} additive error term, which is potentially multiplied — and hence outweighed — by an e^{cN} factor from the RND (see Remark 1.3).

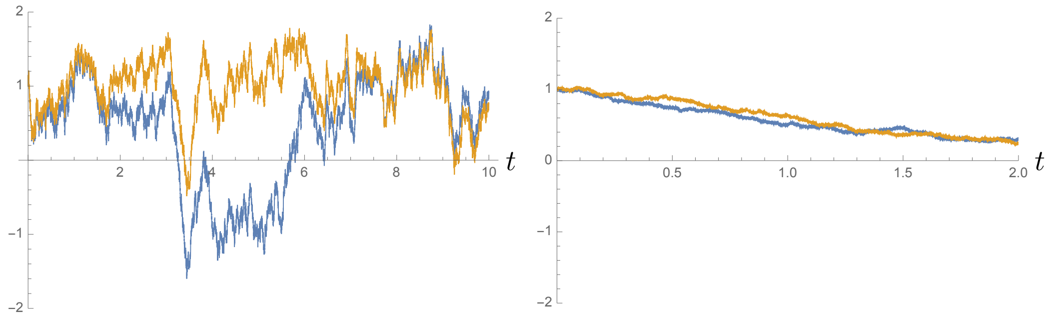

FIGURE 1. Comparison of the diffusions (1.1) under different disorder laws: standard Gaussian J_{ij} (orange) vs. Bernoulli ± 1 set by $\mathrm{sign}(J_{ij})$ (blue), with $N = 100$ particles, $\beta = 1$, $\mathfrak{s} = 2$, the double-well potential at ± 1 given by $U_1(x) = -\log(\mathfrak{s}^2 - x^2) - x^2 + \frac{1}{3}x^4$ and a common N-dimensional Brownian motion B_t driving the systems. Left: sample path of $X_0^{(I)} X_t^{(I)}$ for a uniform particle $I \in \{1, \dots, N\}$. Right: average of 100 samples of $X_0^{(I)} X_t^{(I)}$.

Our results hold for any random interactions consisting of independent $\{J_{ij}\}$, whose laws may depend on i, j, N, subject only to the following moment and tail assumptions:

$$\mathbb{E} J_{ij} = 0\,, \qquad \mathbb{E} J_{ij}^2 = 1\,, \tag{1.4}$$

$$\lim_{\varepsilon \to 0} \sup_{i,j,N} \Big\{ \mathbb{E}\big[e^{\varepsilon |J_{ij}|} \big] \Big\} < \infty\,; \tag{1.5}$$

that is, independent $\{J_{ij}\}$ of zero mean, unit variance and a uniform exponential, or better, tail decay. (In fact, the uniformity over j in (1.5) is not needed and this assumption may be relaxed into the conditions (1.6)–(1.8) stated later; see Remark 1.2.)

Let $W_T^{\mathfrak{s}}$ be the metric space $C([0,T] \to [-\mathfrak{s}, \mathfrak{s}])$ of paths equipped with the distance

$$d_2(x,y) = \left(\frac{1}{T} \int_0^T |x(t) - y(t)|^2 dt \right)^{1/2}.$$

Our main result is the a.s. convergence of μ_N, in the weak topology corresponding to the metric space $W_T^{\mathfrak{s}}$, to the self-consistent limit $\mu_\star$ of [2, 4, 19].

Theorem 1.1. *Let μ_N be the empirical measure defined on (1.2) on sample paths of the Langevin spin glass dynamics (1.1) with independent interactions satisfying (1.4) and (1.5). Then, for every $\beta > 0$, $T < \infty$ and $\mathfrak{s} > 0$, we have that a.s. in the interactions $\mathbf{J}$ and the diffusion, $\mu_N \to \mu_\star$ in $\mathcal{M}_1(W_T^{\mathfrak{s}})$ weakly as $N \to \infty$.*

As a corollary we obtain, for instance, that the dynamics with interactions that are, e.g., centered Bernoulli($\frac{1}{2}$) or centered Exp(1) random variables (see Figure 1) have the same limit as the one derived in [2, 4, 19] for the standard Gaussian case.

Remark 1.2. *Theorem 1.1 remains valid when replacing the assumption (1.5) by the following, less explicit, yet somewhat more relaxed conditions:*

$$\lim_{\varepsilon\to 0}\sup_{\substack{i\le N\\ \theta\in(0,\varepsilon]}} \frac{1}{\theta^2 N}\sum_{j=1}^{N}\log\big(\mathbb{E}[e^{\theta J_{ij}}]\vee\mathbb{E}[e^{-\theta J_{ij}}]\big)<\infty\,, \tag{1.6}$$

$$\frac{1}{N^\gamma}\sum_{i,j=1}^{N}\mathbb{E}(|J_{ij}|^3)\to 0 \quad \textit{for some}\quad \gamma<\frac{5}{2}\,, \tag{1.7}$$

$$\varlimsup_{N\to\infty}\{N^{-1/2}\|\mathbf{J}\|_{2\to 2}\}<\infty \quad \textit{almost surely}. \tag{1.8}$$

The approach of [2] is to establish a weak LDP for the empirical law of the dynamics, in an approximate system of equations where interactions are frozen over a finite number of sub-intervals (see (2.1)), under the topology derived from sup-norm distance between sample paths. One then boosts it via exponential tightness into a full LDP, where the extra assumption $\beta^2\mathfrak{s}^2T<1$ for exponential tightness in [2], is later dispensed of in [19]. The LDP further extends to the original SDS, implying in particular the law of large numbers (LLN). Theorem 1.1 applies beyond Gaussian disorder, albeit for the slightly weaker topology derived from L^2-distance.

Remark 1.3. *As demonstrated in Figure 1, even when using the same Brownian motion, the sample path for a typical (random) coordinate of the solutions of (1.1) under two different disorder laws are not close to one another: one must average over the disorder matrix $\mathbf{J}$ in order to establish the similarity of the limiting μ_N. Going this route, any attempt to control the RND between the average of the measure $\mathbb{P}_N^\beta$ w.r.t. our non-Gaussian interaction $\mathbf{J}$ and the average of such a measure with Gaussian interactions $\widehat{\mathbf{J}}$ requires one to estimate a term of the form $\mathbb{E}_{\mathbf{J}}[e^F]/\mathbb{E}_{\widehat{\mathbf{J}}}[e^G]$ conditioned on the sample paths and Brownian motions. The analysis of this RND becomes particularly delicate since, even upon establishing that $\mathbb{E}_{\mathbf{J},\widehat{\mathbf{J}}}[e^F-e^G]\le(1+N^{-c}\,\Xi)^N$, we must control the effect of the random variable Ξ in order to deduce that the overall ratio is* $\exp(o(N))$.

Remark 1.4. *Various extensions of the model studied here appeared in the context of disordered neural networks. For instance, in [7, 8] (also see [6]) the model allows time delays in the interaction between the particles, a time-dependent self-interaction, and any bounded Lipschitz-continuous pairwise interaction (which in the setup of [2, 4, 19] was a bi-linear map). In studies of networks of Hopefield neurons, e.g. [15] and the references therein, the evolution of $X_t^{(i)}$ has interaction terms J_{ij} as pre-factors of a nonlinear uniformly bounded function of the $X_t^{(j)}$'s, in lieu of a confining potential U. Both of these lines of extensions were studied under the assumption that the interaction variables J_{ij} are Gaussian, while [24, Sec. 4] and [5, Sec. 4.5] follow the same general approach as taken here to establish for such neural networks the universality of the limit of μ_N for sufficiently small T and for* sub-Gaussian *i.i.d.* $\{J_{ij}\}$*. It is plausible that our methods here would be useful in the analysis of these models without those limitations.*

In Section 2 we describe the SDS of piecewise frozen interactions and establish that Theorem 1.1 is a direct consequence of Propositions 2.2 and 2.3. Thereafter, in Section 3 we establish Proposition 2.2, namely the relevant LLN for the approximating SDS, whereas in Section 4 we prove Proposition 2.3, which couples the approximating SDS to the (original) dynamics (1.1) of interest.

2. Proof of Theorem 1.1: piecewise frozen interactions

We start by showing that the conditions in Remark 1.2 indeed relax (1.4)–(1.5).

Lemma 2.1. *Conditions* (1.4) *and* (1.5) *imply the conditions* (1.6)–(1.8)*.*

Proof. Taking the expectation of

$$e^{\theta J_{ij}} - \theta J_{ij} \le 1 + \frac{\theta^2}{\varepsilon^2} e^{\varepsilon |J_{ij}|}\,, \quad \forall |\theta| \le \varepsilon$$

w.r.t. the zero-mean law of J_{ij}, followed by the logarithm of both sides, as $\log(1+y) \le y$ on $\mathbb{R}_+$ it follows that

$$\log \mathbb{E}[e^{\theta J_{ij}}] \le \frac{\theta^2}{\varepsilon^2} \mathbb{E}[e^{\varepsilon |J_{ij}|}]\,, \quad \forall |\theta| \le \varepsilon, i, j, N\,.$$

Thereby, (1.6) follows from (1.5). Similarly, with $|J|^3 \le \frac{6}{\varepsilon^3} e^{\varepsilon |J|}$, upon taking the expectation of both sides w.r.t. the law of J_{ij}, we get (1.7) (for $\frac{5}{2} > \gamma > 2$). As for (1.8), let $\mathbf{A} := \frac{\beta}{\sqrt{N}} \mathbf{J}$ denote the scaled disorder matrix and $\mathbf{Z}_1$, $\mathbf{Z}_2$ be two N-dimensional symmetric matrices, which are independent of $\mathbf{A}$ and of each other, with independent entries above and on their main diagonal satisfying both (1.4) and (1.5). Then, for non-random $\gamma \in \mathbb{R}$ consider the $2N$-dimensional symmetric matrices

$$\mathbf{W}_\gamma := \begin{pmatrix} \frac{\gamma}{\sqrt{N}} \mathbf{Z}_1 & \mathbf{A} \\ \mathbf{A}^{\mathrm{T}} & \frac{\gamma}{\sqrt{N}} \mathbf{Z}_2 \end{pmatrix},$$

noting that

$$\begin{aligned} \|\mathbf{A}\|_{2\to 2} := \sup_{\|z\|_2 = 1} \{\|\mathbf{A} z\|_2\} &= \lambda_{\max}(\mathbf{A}^{\mathrm{T}} \mathbf{A})^{1/2} = \lambda_{\max}(\mathbf{W}_0) \\ &\le \lambda_{\max}(\mathbf{W}_\beta) - \beta \lambda_{\min}(N^{-1/2} \mathbf{Z}_1) - \beta \lambda_{\min}(N^{-1/2} \mathbf{Z}_2)\,. \end{aligned}$$

But $\mathbf{W}_\beta$ is a $\sqrt{2}\beta$ multiple of an $2N$-dimensional Wigner matrix while $N^{-1/2}\mathbf{Z}_i$ for $i = 1, 2$, are a pair of N-dimensional Wigner matrices. The Füredi-Komlós [16] argument applies to each of these three matrices, yielding that $\overline{\lim}_{N\to\infty} \{\lambda_{\max}(\mathbf{W}_\beta)\} \le 2\sqrt{2}\beta$ and $\underline{\lim}_{N\to\infty} \{\lambda_{\min}(N^{-1/2}\mathbf{Z}_i)\} \ge -2$ for $i = 1, 2$. This completes the proof.[1] ■

[1] The result of [16] is for matrices of bounded entries and convergence in probability; the proof remains valid under the condition of uniform boundedness of exponential moments of the entries of $\sqrt{N}A$: see the detailed exposition in [1, Sec. 2.1.6] for the case of i.i.d. entries. The extension to non-identically distributed entries and a.s. convergence, respectively, is immediate (see [1, Ex. 2.1.27 and 2.1.29]).

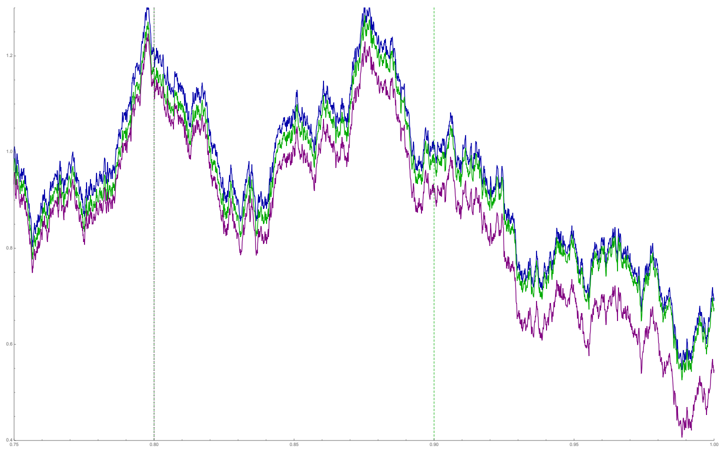

FIGURE 2. Solution of the diffusion (1.1) with Bernoulli ± 1 disorder variables, $N = 100$ and $\beta = 1$ (blue) vs. the approximating system (2.1) with subinterval length 0.1 (green) and subinterval length 0.2 (purple).

A key ingredient in our proof is the analysis of the approximate dynamics of [2, §3], now for a general disorder $\{J_{ij}\}$. Specifically, fixing an integer κ, let

$$\mathsf{t}_k = kT/\kappa \qquad \text{for } k = 0, \ldots, \kappa\,,$$

partitioning the interval $[0, T)$ into κ disjoint sub-intervals $[\mathsf{t}_{k-1}, \mathsf{t}_k)$. We denote by $\widetilde{\mathbb{P}}^{\beta}_{N,\kappa}$ the probability measure of the triplet $(\mathbf{J}, B_\cdot, \widetilde{X}_\cdot)$ corresponding to the diffusion $\widetilde{X}_t$ starting from $\widetilde{X}_0 = X_0$ and given by

$$d\widetilde{X}_t = dB_t - \nabla U(\widetilde{X}_t)dt + \frac{\beta}{\sqrt{N}}\mathbf{J}\widetilde{X}_{\mathsf{t}_{k-1}}dt \qquad (t \in [\mathsf{t}_{k-1}, \mathsf{t}_k],\ 1 \le k \le \kappa)\,, \tag{2.1}$$

i.e., the interaction term between the particles is frozen along each sub-interval $[\mathsf{t}_{k-1}, \mathsf{t}_k)$. (See Figure 2 for a simulation of the approximate dynamics.)

The fact that both (2.1) and the original diffusion (1.1) have unique weak solutions, follows from [2, Proposition 2.1], which established this fact for every (J_{ij}). Furthermore, this solution is in fact strong (see, e.g., [25, exercises (2.10)(1°) and (2.15)(2°) in p. 383 and p. 386]).

Next, for any finite a_2, denote by $\mathbb{P}^{\beta,a_2}_N$ the measure $\mathbb{P}^{\beta}_N$ restricted to the event

$$\mathcal{A}_{a_2} := \{\|\mathbf{A}\|_{2\to 2} \le a_2\}\,. \tag{2.2}$$

We further use Π^{β}_N for the averaged over $\mathbf{A}$ law of the empirical measure μ_N, with Π^{β,a_2}_N similarly standing for the *sub-probability* measure in which the expectation over $\mathbf{A}$ is restricted by an indicator on the event $\mathcal{A}_{a_2}$. In analogy with (1.2), let $\widetilde{\mu}_{N,\kappa}$ be the empirical measure of the solution to (2.1), with $\widetilde{\Pi}^{\beta,a_2}_{N,\kappa}$ denoting its law integrated over the disorder restricted to $\mathcal{A}_{a_2}$.

Recall that $W_T^{\mathfrak{s}}$ is the metric space $C([0,T] \to [-\mathfrak{s}, \mathfrak{s}])$ equipped with the distance

$$d_2(x,y) = \left(\frac{1}{T} \int_0^T |x(t) - y(t)|^2 dt \right)^{1/2}.$$

We further equip the space $\mathcal{M}_1(W_T^{\mathfrak{s}})$ with the corresponding Wasserstein metric

$$d_{W_2}(\phi, \psi) := \inf_{\substack{\xi=(\xi_1,\xi_2) \\ \xi_1=\phi, \xi_2=\psi}} \left\{ \int d_2(x,y)^2 d\xi(x,y) \right\}^{1/2},$$

denoting hereafter by $\mathbb{B}(\mu_\star, \delta)$ the ball of radius δ around $\mu_\star$ in that metric.

Proposition 2.2. *Suppose (1.4), (1.6) and (1.7) hold. Then, for every $T, a_2 < \infty$ and $\delta > 0$ there exists some $\kappa_0 < \infty$ such that for every $\kappa \geq \kappa_0$,*

$$\sum_{N=1}^{\infty} \widetilde{\Pi}_{N,\kappa}^{\beta,a_2}(\mathbb{B}(\mu_\star, \delta)^c) < \infty.$$

Next, let $\mathbb{Q}_N^{\beta,a_2}$ denote the joint law of $\mathbf{J}$, $\widetilde{X}_t$ and X_t, restricted to $\mathcal{A}_{a_2}$, where we use the same N-dimensional Brownian motion B_t for both processes.

Proposition 2.3. *Suppose (1.4), (1.6) and (1.7) hold. Then, for every $T, a_2 < \infty$ and $\delta > 0$, there exists some $\kappa_0 < \infty$ such that for every $\kappa \geq \kappa_0$,*

$$\sum_{N=1}^{\infty} \mathbb{Q}_N^{\beta,a_2}\left(\frac{1}{NT} \int_0^T \|X_t - \widetilde{X}_t\|_2^2 dt > \delta \right) < \infty.$$

Coupling each coordinate of X_t with the corresponding one of $\widetilde{X}_t$, one has that

$$d_{W_2}(\mu_N, \widetilde{\mu}_{N,\kappa})^2 \leq \frac{1}{N} \sum_{i=1}^{N} d_2(X^{(i)}, \widetilde{X}^{(i)})^2 = \frac{1}{NT} \int_0^T \|X_t - \widetilde{X}_t\|_2^2 dt.$$

Thus, combining Proposition 2.2, Proposition 2.3 and the triangle inequality for $d_{W_2}(\cdot,\cdot)$ we have that for any T finite, a_2 finite and $\delta > 0$

$$\sum_{N=1}^{\infty} \mathbb{P}_N^{\beta}\Big(d_{W_2}(\mu_N, \mu_\star) > 2\sqrt{\delta}, \mathcal{A}_{a_2} \Big) < \infty,$$

which by Borel–Cantelli I, implies that for any $\delta > 0$ and a_2 finite,

$$\mathbb{P}^{\beta}\Big[\varlimsup_{N\to\infty} \{ d_{W_2}(\mu_N, \mu_\star) \} > 2\sqrt{\delta}, \ \varlimsup_{N\to\infty} \|\mathbf{A}\|_{2\to 2} < a_2 \ \Big] = 0.$$

In view of (1.8), the proof of Theorem 1.1 is thus complete.

3. Proof of Proposition 2.2

Our proof relies on the following application of the multivariate Lindeberg's method of [11, Theorem 1.1].

Lemma 3.1. *Suppose the random vector $\mathbf{J} \in \mathbb{R}^N$ has independent entries $\{J_j\}$ such that $\mathbb{E}J_j = 0$ and $\mathbb{E}J_j^2 = 1$. Then, for every $N, \kappa \geq 1$, non-random $\mathbf{X} = (x_{kj}) \in \mathbb{R}^{\kappa \times N}$, $\mathbf{b} \in \mathbb{R}^\kappa$, and the quadratic function*

$$h(z) = \frac{1}{2} \|\mathbf{X}z - \mathbf{b}\|_2^2 \qquad (z \in \mathbb{R}^N)\,, \tag{3.1}$$

setting $c_0 = \frac{1}{2} e^{-\sqrt{3}/2}(3^{\frac{1}{4}} + 3^{-\frac{1}{4}})$ one has that

$$|\mathbb{E}e^{-h(\mathbf{J})} - \mathbb{E}e^{-h(\widehat{\mathbf{J}})}| \leq c_0 \sum_{j=1}^N (\mathbf{X}^{\mathrm{T}}\mathbf{X})_{jj}^{3/2} (\mathbb{E}|J_j|^3 + \mathbb{E}|\widehat{J}|^3) \tag{3.2}$$

where $\widehat{\mathbf{J}} = (\widehat{J}_j) \in \mathbb{R}^N$ has i.i.d. standard Gaussian entries. In addition,

$$\mathbb{E}\big[\exp(h(\mathbf{0}) - h(\widehat{\mathbf{J}}))\big] \geq \det(\mathbf{I} + \mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1/2}\,. \tag{3.3}$$

Proof. Having mutually independent entries of $\mathbf{J}$ whose first and second moments match those of $\widehat{\mathbf{J}}$, eliminates the first two terms of the bound on the LHS of (3.2) that we get by applying [11, Theorem 1.1] for the smooth function $f(z) = e^{-h(z)}$. Denoting the first three partial derivatives of a function f w.r.t. z_j by f_j, f_{jj} and f_{jjj}, the proof of [11, Theorem 1.1] provides a sharper bound than stated in its last term, namely

$$|\mathbb{E}f(\mathbf{J}) - \mathbb{E}f(\widehat{\mathbf{J}})| \leq \frac{1}{6} \sum_{j=1}^N \|f_{jjj}\|_\infty (\mathbb{E}|J_j|^3 + \mathbb{E}|\widehat{J}|^3)\,.$$

For $h(z)$ of (3.1), we have $\nabla h = \mathbf{X}^{\mathrm{T}}(\mathbf{X}z - \mathbf{b})$, so $h_{jj} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})_{jj}$ is constant, with $h_{jjj} = 0$ and $|h_j| \leq \sqrt{2hh_{jj}}$ by Cauchy–Schwarz. Substituting $r = \sqrt{2h}$ we thus have that

$$|(e^{-h})_{jjj}| = |h_{jjj} - 3h_j h_{jj} + h_j^3| e^{-h} \leq h_{jj}^{3/2} \sup_{r \geq 0} \Big\{ e^{-\frac{1}{2}r^2}(3r + r^3) \Big\} = 6c_0 h_{jj}^{3/2}\,,$$

from which the RHS of (3.2) follows. To get (3.3) note that the multivariate Gaussian $\mathbf{g} := \mathbf{X}\widehat{\mathbf{J}}$ has zero mean and covariance $\mathbf{X}\mathbf{X}^{\mathrm{T}}$. Consequently,

$$\mathbb{E}\big[e^{h(\mathbf{0}) - h(\widehat{\mathbf{J}})}\big] = \mathbb{E}\big[e^{-\frac{1}{2}\|\mathbf{g}\|_2^2 + \langle \mathbf{g}, \mathbf{b}\rangle}\big] \geq \mathbb{E}\big[e^{-\frac{1}{2}\|\mathbf{g}\|_2^2}\big] = \det\big(\mathbf{I} + \mathbb{E}[\mathbf{g}\mathbf{g}^{\mathrm{T}}]\big)^{-1/2}\,,$$

as claimed in (3.3). ∎

Let $\widehat{\mathbb{P}}^\beta_{N,\kappa}$ and $\widehat{\Pi}^\beta_{N,\kappa}$ be the counterparts of $\widetilde{\mathbb{P}}^\beta_{N,\kappa}$ and $\widetilde{\Pi}^\beta_{N,\kappa}$ when the disorder $\mathbf{J}$ is replaced by $\widehat{\mathbf{J}}$ whose entries are i.i.d. standard Gaussian random variables. Fixing the random variables

$$M_\kappa^{(i)} := \frac{1}{2}\|\mathbf{b}^{(i)}\|_2^2, \qquad b_k^{(i)} := \frac{\widetilde{X}^{(i)}_{\mathsf{t}_k} - \widetilde{X}^{(i)}_{\mathsf{t}_{k-1}} - \int_{\mathsf{t}_{k-1}}^{\mathsf{t}_k} U_1'(\widetilde{X}_s^{(i)})ds}{\sqrt{\mathsf{t}_k - \mathsf{t}_{k-1}}}\,, \quad k = 1, \ldots, \kappa, \tag{3.4}$$

we control $\widetilde{\Pi}^{\beta}_{N,\kappa}$ for some $\delta^{(i)}_N \to 0$ in terms of its counterpart $\widehat{\Pi}^{\beta}_{N,\kappa}$ and

$$\Phi_{N,\kappa} := \frac{1}{N}\sum_{i=1}^{N} \log\big(1+\delta_N^{(i)} e^{M_\kappa^{(i)}}\big), \tag{3.5}$$

where

$$\delta_N^{(i)} = c_1 N^{-3/2}\sum_{j=1}^{N} \mathbb{E}|J_{ij}|^3 . \tag{3.6}$$

Lemma 3.2. *Assume the independent $\{J_{ij}\}$ satisfy (1.4). Then, for any T, β, κ there exist N_0 and c_1 finite, such that for every $N \geq N_0$,*

$$\frac{\mathrm{d}\,\widetilde{\Pi}^{\beta}_{N,\kappa}}{\mathrm{d}\,\widehat{\Pi}^{\beta}_{N,\kappa}} \leq e^{N\Phi_{N,\kappa}} .$$

Proof. Let

$$\Gamma^{\beta}_{N,\kappa}(\widetilde{\mu}_{N,\kappa}) := \frac{1}{N}\sum_{i=1}^{N}\log \mathbb{E}_{\mathbf{J}}\Big[\exp\Big(\langle \mathbf{b}^{(i)}, \mathbf{g}^{(i)}\rangle - \frac{1}{2}\|\mathbf{g}^{(i)}\|_2^2\Big)\Big], \tag{3.7}$$

where J_{ij} are independent and the coordinates of each $\mathbf{g}^{(i)} \in \mathbb{R}^\kappa$ $(i = 1, \dots, N)$ are

$$g_k^{(i)} := \sum_{j=1}^{N} x_{kj} J_{ij}, \quad x_{kj} := \frac{\beta\sqrt{T}}{\sqrt{N\kappa}}\widetilde{X}^{(j)}_{\mathfrak{t}_{k-1}}, \quad k = 1, \dots, \kappa . \tag{3.8}$$

We further define $\widehat{\Gamma}^{\beta}_{N,\kappa}(\widetilde{\mu}_{N,\kappa})$ as in (3.7)–(3.8), except for using $\{\widehat{\mathbf{g}}^{(i)}\}$ and the i.i.d. standard normal variables $\{\widehat{J}_{ij}\}$ instead of $\{\mathbf{g}^{(i)}\}$ and $\{J_{ij}\}$, respectively. Note that by Girsanov's theorem we have the Radon–Nykodim derivative

$$\frac{\mathrm{d}\,\widetilde{\mathbb{P}}^{\beta}_{N,\kappa}}{\mathrm{d}\,\widetilde{\mathbb{P}}^{0}_{N,\kappa}} = \exp\Big(\sum_{i=1}^{N}\big[\langle \mathbf{b}^{(i)}, \mathbf{g}^{(i)}\rangle - \frac{1}{2}\|\mathbf{g}^{(i)}\|_2^2\big]\Big). \tag{3.9}$$

Indeed, under $\widetilde{\mathbb{P}}^{0}_{N,\kappa}$ we get from (3.4) that $\frac{\sqrt{T}}{\sqrt{\kappa}} b_k^{(i)} = B^{(i)}_{\mathfrak{t}_k} - B^{(i)}_{\mathfrak{t}_{k-1}}$, so having in (2.1) the interaction vector $G_t = \frac{\beta}{\sqrt{N}}\mathbf{J}\widetilde{X}_{\mathfrak{t}_{k-1}}$ throughout $[\mathfrak{t}_{k-1}, \mathfrak{t}_k]$, it is easy to verify that then

$$\langle \mathbf{b}^{(i)}, \mathbf{g}^{(i)}\rangle = \int_0^T G_t^{(i)}\, dB_t^{(i)}, \qquad \|\mathbf{g}^{(i)}\|_2^2 = \int_0^T (G_t^{(i)})^2 dt .$$

Further, Novikov's condition holds here since

$$\widetilde{\mathbb{E}}^{0}_{N,\kappa}\Big(\exp\Big(\frac{1}{2}\sum_{i=1}^{N}\|\mathbf{g}^{(i)}\|_2^2\Big)\;\Big|\;\mathbf{J}\Big) < \infty ,$$

due to the uniform bound on $\{x_{kj}\}$ of (3.8) (as $\|\widetilde{X}_t\|_\infty \le \mathfrak{s}$). Under $\widetilde{\mathbb{P}}^0_{N,\kappa} = \mathbb{P}^0_N$ we have that $\mathbf{J}$ is independent of $(\widetilde{X}.)$, thereby yielding that

$$\mathbb{E}_{\mathbf{J}}\Big[\frac{\mathrm{d}\,\widetilde{\mathbb{P}}^\beta_{N,\kappa}}{\mathrm{d}\,\mathbb{P}^0_N}\Big] = \exp(N\Gamma^\beta_{N,\kappa}(\widetilde{\mu}_{N,\kappa}))\,, \tag{3.10}$$

where we also crucially used the independence of the rows of $\mathbf{J}$ to arrive at the specific form on RHS. Being a function of only $\widetilde{\mu}_{N,\kappa}$, the RHS of (3.10) coincides with the Radon–Nykodim derivative restricted to these empirical measures, namely

$$\frac{\mathrm{d}\,\widetilde{\Pi}^\beta_{N,\kappa}}{\mathrm{d}\,\Pi^0_N} = \exp(N\Gamma^\beta_{N,\kappa}(\widetilde{\mu}_{N,\kappa}))\,.$$

The same argument applies for the Radon–Nykodim derivative of $\widehat{\Pi}^\beta_{N,\kappa}$ with respect to Π^0_N. To complete the proof it thus suffices to show that

$$\Gamma^\beta_{N,\kappa}(\widetilde{\mu}_{N,\kappa}) - \widehat{\Gamma}^\beta_{N,\kappa}(\widetilde{\mu}_{N,\kappa}) \le \Phi_{N,\kappa}\,. \tag{3.11}$$

Unraveling (3.4)–(3.8) this follows upon showing that for each $1 \le i \le N$,

$$\mathbb{E}\big[e^{-h^{(i)}(\mathbf{J}^{(i)})}\big] - \mathbb{E}\big[e^{-h^{(i)}(\widehat{\mathbf{J}})}\big] \le \delta^{(i)}_N e^{h^{(i)}(\mathbf{0})}\mathbb{E}\big[e^{-h^{(i)}(\widehat{\mathbf{J}})}\big]\,, \tag{3.12}$$

where $\widehat{\mathbf{J}}$ is a standard multivariate Gaussian, $\mathbf{J}^{(i)} = (J_{i1},\dots,J_{iN}) \in \mathbb{R}^N$ and $h^{(i)}(\cdot)$ of (3.1) with $\mathbf{b}^{(i)}$ of (3.4) and $\{x_{kj}\}$ of (3.8). Since $\widetilde{X}^{(j)}_t \in [-\mathfrak{s},\mathfrak{s}]$, we have that

$$|x_{kj}|^2 \le \frac{(\beta\mathfrak{s})^2 T}{\kappa N} \quad\Longrightarrow\quad (\mathbf{X}^{\mathrm{T}}\mathbf{X})_{jj} \le \frac{(\beta\mathfrak{s})^2 T}{N}, \quad (\mathbf{X}\mathbf{X}^{\mathrm{T}})_{kk'} \le \frac{(\beta\mathfrak{s})^2 T}{\kappa}\,.$$

Thus, from Lemma 3.1 the RHS of (3.3) is bounded below in our case by $1/c_2$ for some $c_2 = c_2(\beta\mathfrak{s}\sqrt{T},\kappa)$ finite, while for some $c_3 = c_3(\beta\mathfrak{s}\sqrt{T})$ finite, the LHS of (3.12) is at most $c_3 N^{-3/2}\sum_{j\le N}(\mathbb{E}|J_{ij}|^3 + \mathbb{E}|\widehat{J}|^3)$. With $\mathbb{E}|J_{ij}|^3 \ge 1$ and $\mathbb{E}|\widehat{J}|^3 = \sqrt{8/\pi}$, taking $c_1 = c_2c_3(1+\sqrt{8/\pi})$ in (3.6) guarantees that (3.12) would hold and thereby completes the proof of the lemma. ∎

The following elementary lemma is needed for proving Lemma 3.4 (namely, to show that $\Phi_{N,\kappa} \to 0$ a.s. when $N \to \infty$).

Lemma 3.3. *Suppose vectors $\mathbf{J} = (J_1,\dots,J_N) \in \mathbb{R}^N$ are composed of independent coordinates $\{J_i\}$ such that for some $\varepsilon > 0$, $v < \infty$, and all $N \ge N_0$,*

$$\sup_{\theta\in(0,\varepsilon]}\Big\{\frac{1}{\theta^2 N}\sum_{j=1}^N \log\big(\mathbb{E}[e^{\theta J_j}] \vee \mathbb{E}[e^{-\theta J_j}]\big)\Big\} \le v\,. \tag{3.13}$$

For any $a < \infty$, if $\alpha < \frac{1}{4v} \wedge \frac{\varepsilon}{4a}$ and $N \ge N_1 := N_0 \vee \frac{2}{\varepsilon a}$, then

$$\sup_{\{\mathbf{u}\in\mathbb{R}^N\,:\,\|\mathbf{u}\|_\infty\le N^{-1/2}\}} \mathbb{E}\Big[\exp\big(\alpha\langle\mathbf{u},\mathbf{J}\rangle^2\big)\,\mathbb{1}_{\{\|\mathbf{J}\|_1\le aN\}}\Big] \le f_\star(\alpha v) < \infty\,.$$

Proof. Fixing $\alpha > 0$, associate with each non-random $\mathbf{u} \in \mathbb{R}^N$ such that $\|\mathbf{u}\|_\infty \le N^{-1/2}$ the variable $Y_{\mathbf{u}} := \sqrt{2\alpha}\langle \mathbf{u}, \mathbf{J}\rangle$, noting that for $\alpha \le \varepsilon/(4a)$ and any such $\mathbf{u}$

$$\{\|\mathbf{J}\| \le aN\} \quad \Longrightarrow \quad |Y_{\mathbf{u}}| \le a\sqrt{2\alpha N} \le \frac{\varepsilon}{2}\sqrt{N/(2\alpha)} := r_N \,.$$

Taking $N \ge N_0$ yields in view of (3.13) (at $\theta = \lambda\sqrt{2\alpha/N}$), that

$$\mathbb{E}\Big[e^{\lambda Y_{\mathbf{u}}}\Big] \le e^{2\alpha v \lambda^2}\,, \qquad \forall |\lambda| \le 2r_N \,. \tag{3.14}$$

Recall the elementary bound, valid for all $r \ge 1$

$$e^{y^2/2}\mathbb{1}_{[-r,r]}(y) \le 2\int_{-2r}^{2r} e^{\lambda y} e^{-\lambda^2/2}\frac{d\lambda}{\sqrt{2\pi}}\,. \tag{3.15}$$

Further, since $r_N \ge \sqrt{\varepsilon a N/2} \ge 1$ for all $N \ge N_1$, upon combining the bounds (3.14), (3.15) with Fubini's theorem, we find that for any such α, $Y_{\mathbf{u}}$ and for all $N \ge N_1$,

$$\mathbb{E}\Big[e^{Y_{\mathbf{u}}^2/2}\mathbb{1}_{\{\|\mathbf{J}\|_1 \le aN\}}\Big] \le 2\int_{-\infty}^{\infty} e^{2\alpha v \lambda^2} e^{-\lambda^2/2}\frac{d\lambda}{\sqrt{2\pi}} := f_\star(\alpha v) < \infty\,,$$

as claimed. ∎

Equipped with Lemma 3.3 we proceed to verify that a.s. $\Phi_{N,\kappa} \to 0$ when $N \to \infty$.

Lemma 3.4. *Suppose the independent variables $\{J_{ij}\}$ satisfy (1.6) and (1.7). Then, for any T, β, a_2, κ and all $\eta > 0$,*

$$\sum_{N=1}^{\infty} \widetilde{\mathbb{P}}^{\beta,a_2}_{N,\kappa}(\Phi_{N,\kappa} > 2\eta) < \infty\,. \tag{3.16}$$

Proof. Set $\widehat{M}^{(i)}_\kappa := \frac{1}{2}\|\widehat{\mathbf{b}}^{(i)}\|_2^2$ for $\widehat{b}^{(i)}_k := b^{(i)}_k - g^{(i)}_k$. Then, for any $q \in (0,1]$ and $r_N \ge 0$,

$$M^{(i)}_\kappa \le (1+q)\widehat{M}^{(i)}_\kappa + q^{-1}\|\mathbf{g}^{(i)}\|_2^2 \le (1+q)\widehat{M}^{(i)}_\kappa + r_N + q^{-1}\|\mathbf{g}^{(i)}\|_2^2\mathbb{1}_{\{\|\mathbf{g}^{(i)}\|_2^2 \ge q r_N\}}\,. \tag{3.17}$$

With $\widehat{\delta}^{(i)}_N := e^{r_N}\delta^{(i)}_N$ we thus get from (3.5) and $\log(1+ye^R) \le R + \log(1+y)$, for $R, y \ge 0$, that

$$\Phi_{N,\kappa} \le \frac{1}{N}\sum_{i=1}^N \log\big(1 + \widehat{\delta}^{(i)}_N e^{(1+q)\widehat{M}^{(i)}_\kappa}\big) + \frac{1}{qN}\sum_{i=1}^N \|\mathbf{g}^{(i)}\|_2^2\mathbb{1}_{\{\|\mathbf{g}^{(i)}\|_2^2 \ge q r_N\}} := \widehat{\Phi}_{N,\kappa} + \Psi_{N,\kappa}\,.$$

Taking $\varphi := 1 - q = \frac{2}{3}(\gamma - 1)$ for $1 < \gamma < 5/2$ of (1.7) and $r_N \to \infty$ slowly enough, we find that as $N \to \infty$,

$$\widehat{\delta}_N := \frac{1}{N}\sum_{i=1}(\widehat{\delta}^{(i)}_N)^\varphi \le (c_1 e^{r_N})^\varphi N^{-\gamma}\sum_{i,j=1}^N \mathbb{E}|J_{ij}|^3 \to 0\,. \tag{3.18}$$

Further, under $\widetilde{\mathbb{P}}^{\beta}_{N,\kappa}$ the variables $\{\widehat{b}^{(i)}_k\}$ are i.i.d. standard Gaussian, independent of $\mathbf{J}$. In particular, $(2\widehat{M}_\kappa)^{1/2}$ is the Euclidean norm of a κ-dimensional, standard Gaussian random vector, which has the density $\hat{c}_\kappa r^{\kappa-1}e^{-r^2/2}$ at $r \in [0,\infty)$ for some $\hat{c}_\kappa < \infty$.

Thus, the elementary inequality $(1+u)^\varphi \le 1+u^\varphi$, valid for $u \ge 0$ and $\varphi < 1$, yields the bound

$$\widetilde{\mathbb{E}}^{\beta}_{N,\kappa}\Big[\big(1+\widehat{\delta} e^{(1+q)\widehat{M}^{(i)}_\kappa}\big)^\varphi\Big] \le 1 + (\widehat{\delta})^\varphi\, \hat{c}_\kappa \int_0^\infty r^{\kappa-1} e^{-(qr)^2/2} dr = 1 + (\widehat{\delta})^\varphi\, q^{-\kappa}\,.$$

Combining this with Markov's inequality, yields for the i.i.d. $\widehat{M}^{(i)}_\kappa$, that

$$\widetilde{\mathbb{P}}^{\beta}_{N,\kappa}(\widehat{\Phi}_{N,\kappa} > \eta) \le e^{-\varphi\eta N} \prod_{i=1}^N \widetilde{\mathbb{E}}^{\beta}_{N,\kappa}\Big[\big(1+\widehat{\delta}^{(i)}_N e^{(1+q)\widehat{M}^{(i)}_\kappa}\big)^\varphi\Big] \le e^{-N(\varphi\eta - q^{-\kappa}\widehat{\delta}_N)}\,.$$

In view of (3.18), we thus deduce that

$$\sum_{N=1}^\infty \widetilde{\mathbb{P}}^{\beta}_{N,\kappa}(\widehat{\Phi}_{N,\kappa} > \eta) < \infty \tag{3.19}$$

and complete the proof of the lemma upon checking that for any $\eta > 0$,

$$\sum_{N=1}^\infty \widetilde{\mathbb{P}}^{\beta,a_2}_{N,\kappa}(\Psi_{N,\kappa} > \eta) < \infty\,. \tag{3.20}$$

To this end, first note that if $\|\mathbf{A}\|_{2\to2} \le a_2$, then necessarily

$$\sum_{i=1}^N \|\mathbf{g}^{(i)}\|_2^2 = \frac{T}{\kappa}\sum_{k=1}^\kappa \|\mathbf{A}\widetilde{X}_{\mathfrak{t}_{k-1}}\|_2^2 \le T a_2^2 \mathfrak{s}^2 N\,. \tag{3.21}$$

Thus, the random set $\mathcal{S}_\star := \{i \le N : \|\mathbf{g}^{(i)}\|_2^2 \ge q r_N\}$ has at most

$$\ell_N := \lceil T a_2^2 \mathfrak{s}^2 N/(q r_N)\rceil = o(N)$$

elements. By the union bound over the at most $\binom{N}{\ell_N} = \exp(o(N))$ ways to choose a non-random set $\mathcal{S} \subseteq [N]$ of size ℓ_N, it suffices for (3.20) to show that

$$\lim_{N\to\infty} \sup_{|\mathcal{S}|=\ell_N} \frac{1}{N} \log \widetilde{\mathbb{P}}^{\beta}_{N,\kappa}(R_{\mathcal{S}} > q\eta N, \mathcal{A}_{a_2}) < 0\,, \quad \text{where} \quad R_{\mathcal{S}} := \sum_{i\in\mathcal{S}} \|\mathbf{g}^{(i)}\|_2^2\,. \tag{3.22}$$

To this end, fixing $\mathcal{S} \subset [N]$ of size ℓ_N, consider the measure $\widetilde{\mathbb{P}}^{\beta;\mathcal{S}}_{N,\kappa}$ where we set $\beta = 0$ at all coordinates $i \in \mathcal{S}$ of (2.1), while not changing the value of β when $i \notin \mathcal{S}$. One then has similarly to (3.9) the following Radon–Nykodim derivative, expressed in terms of $\mathbf{b}^{(i)}$ of (3.4) and $\mathbf{g}^{(i)}$ of (3.8) by

$$\frac{\mathrm{d}\,\widetilde{\mathbb{P}}^{\beta}_{N,\kappa}}{\mathrm{d}\,\widetilde{\mathbb{P}}^{\beta;\mathcal{S}}_{N,\kappa}} = \exp\Big(\sum_{i\in\mathcal{S}} \langle \mathbf{b}^{(i)}, \mathbf{g}^{(i)}\rangle - \frac{1}{2} R_{\mathcal{S}}\Big)\,. \tag{3.23}$$

The RHS of (3.23) is bounded for any $\theta > 0$ (using the trivial bound $xy \le (x^2+y^2)/2$ for $x = (1+\theta)^{-1/2}\mathbf{b}^{(i)}$ and $y = (1+\theta)^{1/2}\mathbf{g}^{(i)}$) by

$$\exp\Big(\frac{M_{\mathcal{S}}}{1+\theta} + \frac{\theta}{2} R_{\mathcal{S}}\Big)\,, \quad \text{where} \quad M_{\mathcal{S}} := \frac{1}{2}\sum_{i\in\mathcal{S}} \|\mathbf{b}^{(i)}\|^2\,.$$

In addition, $\max_i\{N^{-1/2}\sum_{j=1}^N |A_{ij}|\} \le \|\mathbf{A}\|_{2\to 2}$ for any N-dimensional matrix, yielding for $a := a_2/\beta$, that

$$\mathcal{A}_{a_2} \subseteq \mathcal{A}_a^{\mathcal{S}} := \bigcap_{i\in\mathcal{S}} \{\sum_{j=1}^N |J_{ij}| \le aN\}\,.$$

Combining the preceding bounds, we arrive at

$$\begin{aligned}\widetilde{\mathbb{P}}^{\beta}_{N,\kappa}(R_{\mathcal{S}} > q\eta N, \mathcal{A}_{a_2}) &\le \widetilde{\mathbb{E}}^{\beta;\mathcal{S}}_{N,\kappa}\Big[\exp\big(\frac{M_{\mathcal{S}}}{1+\theta} + \frac{\theta}{2} R_{\mathcal{S}}\big)\mathbb{1}_{\{R_{\mathcal{S}}>q\eta N\}}\mathbb{1}_{\mathcal{A}_a^{\mathcal{S}}}\Big] \\ &\le e^{-\theta q\eta N/2}\, \widetilde{\mathbb{E}}^{\beta;\mathcal{S}}_{N,\kappa}\Big[\exp\big(\frac{M_{\mathcal{S}}}{1+\theta} + \theta R_{\mathcal{S}}\big)\mathbb{1}_{\mathcal{A}_a^{\mathcal{S}}}\Big]\,.\end{aligned}$$

Under $\widetilde{\mathbb{P}}^{\beta;\mathcal{S}}_{N,\kappa}$ the variables $\{b_k^{(i)}, i\in\mathcal{S}, k\le\kappa\}$ of (3.4) are i.i.d. standard Gaussian. With $M_{\mathcal{S}}$ being the sum of half the squares of these variables, we clearly have that for some $f_0(\cdot)$ finite, any $\theta>0$ and all $\kappa, \mathcal{S}$,

$$\widetilde{\mathbb{E}}^{\beta;\mathcal{S}}_{N,\kappa}\big[e^{M_{\mathcal{S}}/(1+\theta)}\big] = f_0(\theta)^{\kappa|\mathcal{S}|}\,.$$

Denoting by $\mathcal{F}_N := \sigma(b_k^{(i)}, x_{kj}, k\le\kappa, i,j\le N)$ the σ-algebra generated by $b_k^{(i)}$ of (3.4) and $\{x_{kj}\}$ of (3.8), it thus suffices for (3.22) to show the existence of *non-random* $\theta>0$, N_1 and $f_1 = f_1(\theta) < \infty$, such that for all $N\ge N_1$ and $\mathcal{S}\subseteq[N]$,

$$\widetilde{\mathbb{E}}^{\beta;\mathcal{S}}_{N,\kappa}\Big[e^{\theta R_{\mathcal{S}}}\mathbb{1}_{\mathcal{A}_a^{\mathcal{S}}} \mid \mathcal{F}_N\Big] \le f_1(\theta)^{|\mathcal{S}|}\,. \tag{3.24}$$

To this end, recall that under $\widetilde{\mathbb{P}}^{\beta;\mathcal{S}}_{N,\kappa}$ the vectors $\{\mathbf{J}_i := (J_{ij}) \in \mathbb{R}^N,\ i\in\mathcal{S}\}$ of mutually independent entries are *independent* of $\mathcal{F}_N$. Further, in view of (3.8),

$$R_{\mathcal{S}} = \sum_{i\in\mathcal{S}}\sum_{k=1}^{\kappa}\langle \mathbf{x}_k, \mathbf{J}_i\rangle^2\,,$$

where $\mathbf{x}_k = (x_{kj}) \in \mathbb{R}^N$ is such that $\|\mathbf{x}_k\|_\infty \le \sqrt{T/(N\kappa)}\beta\mathfrak{s}$. We thus have that

$$\begin{aligned}&\widetilde{\mathbb{E}}^{\beta;\mathcal{S}}_{N,\kappa}\Big[e^{\theta R_{\mathcal{S}}}\mathbb{1}_{\mathcal{A}_a^{\mathcal{S}}} \,|\, \mathcal{F}_N\Big] \le f_{N,\kappa}(\theta T\beta^2\mathfrak{s}^2)^{|\mathcal{S}|}\,, \\ &\qquad f_{N,\kappa}(\alpha) := \max_{i\le N}\sup_{\{\mathbf{u}_k\in\mathbb{R}^N: \|\mathbf{u}_k\|_\infty\le N^{-1/2}\}} \mathbb{E}\big[\exp\big(\frac{\alpha}{\kappa}\sum_{k=1}^{\kappa}\langle\mathbf{u}_k, \mathbf{J}_i\rangle^2\big)\mathbb{1}_{\mathcal{A}_a^{\{i\}}}\big]\,.\end{aligned}$$

By Jensen's inequality $f_{N,\kappa}(\alpha) \le f_{N,1}(\alpha)$. Further, thanks to our assumption (1.6), the vectors $\mathbf{J_i}$ of independent coordinates satisfy the condition (3.13) for some $\varepsilon>0$, $v<\infty$ which are independent of i and N. Hence, taking $\theta>0$ so $\alpha = \theta T\beta^2\mathfrak{s}^2$ be as in Lemma 3.3, results for $N\ge N_1$ with $f_{N,\kappa}(\alpha) \le f_\star(\alpha v)$ finite, thus establishing (3.24) and thereby concluding the proof of the lemma. ■

Setting $\widehat{W}_T^{\mathfrak{s}}$ as the metric space $C([0,T]\to[-\mathfrak{s},\mathfrak{s}])$ equipped with the distance

$$d_\infty(x,y) = \sup_{t\in[0,T]} |x(t)-y(t)|\,,$$

we next effectively prove an exponential tightness of $\widetilde{\Pi}^{\beta}_{N,\kappa}$ in the corresponding weak topology (using an entropy bound, this has been proved in [4] for Gaussian disorder).

Lemma 3.5. *Fixing $T, \beta, \kappa, a_2, \alpha$, there exists $\mathcal{K}_\alpha \subset \mathcal{M}_1(\widehat{W}^{\mathfrak{s}}_T)$ compact, with*

$$\limsup_{N\to\infty} \frac{1}{N} \log \widetilde{\Pi}^{\beta,a_2}_{N,\kappa}(\mathcal{K}^c_\alpha) < -\alpha\,.$$

Proof. It follows from (3.9) that for $M^{(i)}_\kappa := \frac{1}{2}\|\mathbf{b}^{(i)}\|_2^2$ and $\mathbf{b}^{(i)}$ of (3.4),

$$\frac{\mathrm{d}\,\widetilde{\mathbb{P}}^{\beta}_{N,\kappa}}{\mathrm{d}\,\mathbb{P}^0_N} \le \exp\Big(\sum_{i=1}^N M^{(i)}_\kappa\Big)\,.$$

By (3.21) and the LHS of (3.17), having $\|\mathbf{A}\|_{2\to 2} \le a_2$ yields that

$$\sum_{i=1}^N M^{(i)}_\kappa \le 2\sum_{i=1}^N \widehat{M}^{(i)}_\kappa + T a_2^2 \mathfrak{s}^2 N\,.$$

Further, under $\widetilde{\mathbb{P}}^{\beta}_{N,\kappa}$ the variables $\{\widehat{b}^{(i)}_k\}$ are i.i.d. standard Gaussian, independent of $\mathbf{A}$, hence for any $\mathcal{A} \subset \mathcal{M}_1(\widehat{W}^{\mathfrak{s}}_T)$ and $r' = (r - T a_2^2\mathfrak{s}^2)/2 \ge 1$,

$$\widetilde{\Pi}^{\beta,a_2}_{N,\kappa}(\mathcal{A}) \le e^{\kappa r N}\,\Pi^0_N(\mathcal{A}) + \widetilde{\Pi}^{\beta}_{N,\kappa}\Big(\sum_{i=1}^N \widehat{M}^{(i)}_\kappa \ge \kappa r' N\Big)\,. \tag{3.25}$$

The CGF $\log \widetilde{\mathbb{E}}^{\beta}_{N,\kappa}[e^{\theta \widehat{M}^{(1)}_\kappa}] = \kappa\Lambda(\theta)$ which is independent of N and $\mathbf{A}$ (hence also on β), is finite at $\theta < 1$ and has $\kappa\Lambda'(0) = \widetilde{\mathbb{E}}^{\beta}_{N,\kappa}\widehat{M}_\kappa = \kappa/2$. Thus, $\theta \ge \Lambda(\theta)$ for small enough $\theta > 0$, so applying Markov's inequality, we get for such $\theta > 0$ and i.i.d. $\widehat{M}^{(i)}_\kappa$, that for some $\widehat{r} = \widehat{r}(\alpha,\kappa)$ finite,

$$\widetilde{\Pi}^{\beta}_{N,\kappa}\Big(\sum_{i=1}^N \widehat{M}^{(i)}_\kappa \ge \kappa \widehat{r} N\Big) \le \exp(-N\kappa[\theta\widehat{r} - \Lambda(\theta)]) < \exp(-\alpha N)\,. \tag{3.26}$$

Thus, thanks to (3.25) and (3.26), it suffices to verify that Π^0_N are exponentially tight in $\mathcal{M}_1(\widehat{W}^{\mathfrak{s}}_T)$. To this end, recall that this is the law of the empirical measure μ_N of independent $(X^{(i)}_\cdot)$ (namely, the solutions of the SDS (1.1) which are uncoupled at $\beta = 0$). These i.i.d. variables take value in a Polish space $\widehat{W}^{\mathfrak{s}}_T$, whence Π^0_N is exponentially tight in the induced weak topology (see [13, Lemma 6.2.6]). ∎

We have the following upon combining [2] and Lemma 3.5.

Lemma 3.6. *For every β, κ, T, a_2 there exists a good rate function I_κ on $\mathcal{M}_1(\widehat{W}^{\mathfrak{s}}_T)$ such that, for any closed $\mathcal{F} \subset \mathcal{M}_1(\widehat{W}^{\mathfrak{s}}_T)$,*

$$\limsup_{N\to\infty} \frac{1}{N} \log \widehat{\Pi}^{\beta,a_2}_{N,\kappa}(\mathcal{F}) \le -I_\kappa(\mathcal{F})\,.$$

In addition, $I_\kappa(\mathcal{F}) \to I(\mathcal{F})$ as $\kappa \to \infty$ and $I(\cdot)$ is a good rate function whose unique minimizer is $\mu_\star$.

Proof. The large deviations upper bound with a good rate function $I_\kappa(\cdot)$ is established in [2, Theorem 3.1(1)–(2)] for $\mathcal{F}$ compact and $\widehat{\Pi}^{\beta,a_2}_{N,\kappa}$. The exponential tightness from Lemma 3.5 applies in particular for the Gaussian disorder $\widehat{\mathbf{J}}$, thereby extending the validity of such upper bound to all closed sets $\mathcal{F}$. Finally, [2, Prop. 4.3(1)] shows the convergence of $I_\kappa(\cdot)$ to some $I(\cdot)$ whose global minimizer $\mu_\star$ is unique. ∎

Proof of Proposition 2.2. We actually prove a slightly stronger statement, where $\mathbb{B}(\mu_\star, \delta)$ denotes instead the ball of radius $\delta > 0$ and center $\mu_\star$ in $\mathcal{M}_1(\widehat{W}^{\mathfrak{s}}_T)$. Fixing T, β, δ, a_2, we have by Lemma 3.2 and the union bound, that for any κ, $N \geq N_0(\kappa)$ and all $\eta > 0$,

$$\widetilde{\Pi}^{\beta,a_2}_{N,\kappa}(\mathbb{B}(\mu_\star,\delta)^c) \leq \widetilde{\mathbb{P}}^{\beta,a_2}_{N,\kappa}(\Phi_{N,\kappa} > 2\eta) + e^{2\eta N}\, \widehat{\Pi}^{\beta,a_2}_{N,\kappa}(\mathbb{B}(\mu_\star,\delta)^c)\,.$$

In view of Lemma 3.4 it thus suffices to show that for any $\delta > 0$ and all $\kappa \geq \kappa_0(\delta)$

$$\limsup_{N\to\infty} \frac{1}{N} \log \widehat{\Pi}^{\beta,a_2}_{N,\kappa}(\mathbb{B}(\mu_\star,\delta)^c) < 0\,. \tag{3.27}$$

Recall from Lemma 3.6, that $I(\mathcal{F}_\delta) > 0$ for the closed set $\mathcal{F}_\delta = \mathbb{B}(\mu_\star,\delta)^c$ and therefore $I_\kappa(\mathcal{F}_\delta) > 0$ for all $\kappa \geq \kappa_0(\delta)$. We thus get (3.27) and thereby complete the proof of the theorem, upon considering the LDP upper bound of Lemma 3.6 for this $\mathcal{F}_\delta$. ∎

4. Proof of Proposition 2.3

For $\beta > 0$, we couple X_t and $\widetilde{X}_t$ using the same Brownian motion: writing

$$\mathcal{E}_t := \widetilde{X}_t - X_t$$

we see that

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_t &= \nabla U(X_t) - \nabla U(\widetilde{X}_t) + \frac{\beta}{\sqrt{N}}\mathbf{J}(\widetilde{X}_{\mathrm{t}_{\lfloor t\kappa/T\rfloor}} - X_t)\,,\\ \mathcal{E}_0 &= 0\,.\end{aligned}$$

Let

$$R_t = \|\mathcal{E}_t\|_2 \quad \text{and} \quad \mathbf{A} = \frac{\beta}{\sqrt{N}}\mathbf{J}\,.$$

Then

$$\begin{aligned}R_t\frac{d}{dt}R_t &= \Big\langle \mathcal{E}_t, \frac{d}{dt}\mathcal{E}_t\Big\rangle\\ &= \langle \mathcal{E}_t, \mathbf{A}\mathcal{E}_t\rangle + \Big\langle \mathcal{E}_t, \nabla U(X_t) - \nabla U(\widetilde{X}_t)\Big\rangle + \Big\langle \mathcal{E}_t, \mathbf{A}(\widetilde{X}_{\mathrm{t}_{\lfloor t\kappa/T\rfloor}} - \widetilde{X}_t)\Big\rangle\,,\end{aligned}$$

which, by the mean value theorem, is at most

$$\|\mathbf{A}\|_{2\to 2}R_t^2 + R_t^2 c'' + \|\mathbf{A}\|_{2\to 2}R_t L_t\,,$$

where, for fixed $\varepsilon > 0$ and $\rho > 0$, we define

$$L_t = \|\widetilde{X}_{\mathrm{t}_{\lfloor t\kappa/T\rfloor}} - \widetilde{X}_t\|_2 \qquad \text{and} \qquad c'' = \sup_{|x|\leq\mathfrak{s}}(-U_1''(x))\,.$$

Restricted to the event $\mathcal{A}_{a_2}$, we have that

$$\frac{d}{dt}R_t \le \big(a_2 + c''\big)\, R_t + 3a_2\rho\sqrt{N}\,, \tag{4.1}$$

up to the stopping time

$$\widetilde{\tau}_\rho := \inf\{t \ge 0 \,:\, L_t \ge 3\rho\sqrt{N}\}\,.$$

Solving the ODE that corresponds to equality in (4.1), starting at $R_0 = 0$, results with

$$R_t \le \frac{3a_2}{a_2 + c''}(e^{(a_2+c'')t} - 1)\rho\sqrt{N} \le \delta\sqrt{N}$$

up to $\widetilde{\tau}_\rho$, provided that

$$\rho \le \delta\left(\frac{a_2 + c''}{3a_2}\right)e^{-(a_2+c'')T}\,.$$

In particular, for such $\rho = \rho(\delta, a_2, T) > 0$ it then follows that

$$\mathbb{Q}_N^{\beta,a_2}\Big(\sup_{t\in[0,T]} R_t \ge \delta\sqrt{N}\Big) \le \mathbb{P}_N^{\beta,a_2}(\widetilde{\tau}_\rho \le T\,, D_\varepsilon^c) + \mathbb{P}_N^{\beta,a_2}(D_\varepsilon)\,,$$

for any $\varepsilon > 0$, where

$$\widetilde{\sigma}_\varepsilon^{(i)} = \inf\{t \ge 0 \,:\, |\widetilde{X}_t^{(i)}| > \mathfrak{s} - \varepsilon\}\,, \qquad D_\varepsilon = \Big\{\sum_{i=1}^N \mathbb{1}_{\{\widetilde{\sigma}_\varepsilon^{(i)} < T\}} > \frac{\rho^2}{\mathfrak{s}^2}N\Big\}\,.$$

Recall from [2, Thm. 4.1(a)] that $\mu_\star$ is absolutely continuous w.r.t. the law $\mathbb{P}_1^0$ of the solution $X_t^{(1)}$, $t \in [0,T]$, of a single SDE (1.1) at $\beta = 0$. Further, $\mathbb{P}_1^0(\{x : \|x\|_\infty = \mathfrak{s}\}) = 0$ thanks to (1.3), hence for any small $\varepsilon > 0$ and the corresponding closed subset

$$\mu_\star \notin \mathcal{F}_\varepsilon := \Big\{\mu : \mu(\|x\|_\infty \ge \mathfrak{s} - \varepsilon) \ge \frac{\rho^2}{\mathfrak{s}^2}\Big\}\,. \tag{4.2}$$

Fixing such $\varepsilon > 0$, we proceed to bound L_t. To this end, recall that for any $t \in [\mathfrak{t}_k, \mathfrak{t}_{k+1}]$,

$$\widetilde{X}_t - \widetilde{X}_{\mathfrak{t}_k} = -\int_{\mathfrak{t}_k}^t \nabla U(\widetilde{X}_\xi)d\xi + B_t - B_{\mathfrak{t}_k} + (t - \mathfrak{t}_k)\mathbf{A}\widetilde{X}_{\mathfrak{t}_k}\,.$$

Hence, with $t - \mathfrak{t}_k \le T/\kappa$, setting $c'_\varepsilon := \sup_{|x|\le\mathfrak{s}-\varepsilon}|U_1'(x)|$ we get that

$$\sum_{i=1}^N \mathbb{1}_{\{\widetilde{\sigma}_\varepsilon^{(i)} \ge T\}}\big|\widetilde{X}_t^{(i)} - \widetilde{X}_{\mathfrak{t}_k}^{(i)}\big|^2 \le \Big[c'_\varepsilon\frac{T}{\kappa}\sqrt{N} + \|B_t - B_{\mathfrak{t}_k}\|_2 + \frac{T}{\kappa}\|\mathbf{A}\|_{2\to2}\,\|\widetilde{X}_{\mathfrak{t}_k}\|_2\Big]^2\,.$$

At the same time, on the event D_ε^c we have that

$$\sum_{i=1}^N \mathbb{1}_{\{\widetilde{\sigma}_\varepsilon^{(i)} < T\}}\big|\widetilde{X}_t^{(i)} - \widetilde{X}_{\mathfrak{t}_k}^{(i)}\big|^2 \le (2\rho)^2 N\,.$$

Consequently, (using that $\|\widetilde{X}_{\mathfrak{t}_k}\|_2 \le \mathfrak{s}\sqrt{N}$ and the restriction to the event $\mathcal{A}_{a_2}$) we deduce that as soon as

$$\kappa \ge \kappa_1 := \big\lceil (c'_\varepsilon + a_2\mathfrak{s})T/\rho \big\rceil$$

we have by the independence of the Brownian increments (and a union bound),

$$\mathbb{P}_N^{\beta,a_2}(\widetilde{\tau}_\rho \le T\,,\, D_\varepsilon^c) \le \kappa\, \mathbb{P}\Big(\sup_{t\in[0,\frac{T}{\kappa}]} \{\|B_t\|_2^2\} \ge \rho^2 N\Big)\,.$$

To bound the latter, note that

$$u(t,x) = \exp\Big[\frac{x^2}{1+2t} - \frac{1}{2}\log(1+2t)\Big]\,,$$

is a positive, smooth solution of the heat equation $u_t + \frac{1}{2}u_{xx} = 0$. Hence, by Ito's formula the integrable $M_t^{(i)} := u(t, B_t^{(i)})$ are i.i.d. positive martingales, starting at $M_0^{(i)} = 1$. Next, increasing κ_1 as needed in order to have

$$\eta := \frac{\rho^2}{1+2T/\kappa_1} - \frac{1}{2}\log(1+2T/\kappa_1) > 0\,,$$

and applying Doob's maximal inequality for the positive martingale $\overline{M}_t = \prod_{i=1}^N M_t^{(i)}$, we deduce that for any $\kappa \ge \kappa_1$,

$$\mathbb{P}\Big(\sup_{t\in[0,\frac{T}{\kappa}]} \{\|B_t\|_2^2\} \ge \rho^2 N\Big) \le \mathbb{P}\Big(\sup_{t\in[0,\frac{T}{\kappa}]} \{\overline{M}_t\} \ge e^{\eta N}\Big) \le e^{-\eta N}\,.$$

Turning now to show that $\mathbb{P}_N^{\beta,a_2}(D_\varepsilon)$ is summable in N for all $\kappa \ge \kappa_1(\varepsilon,\rho)$, note that $D_\varepsilon \subseteq \{\widetilde{\mu}_{N,\kappa} \in \mathcal{F}_\varepsilon\}$ for $\mathcal{F}_\varepsilon$ of (4.2). Thus, proceeding as in the proof of Proposition 2.2, we have by Lemma 3.2 and the union bound, that for any κ, $N \ge N_0(\kappa)$ and all $\eta > 0$,

$$\mathbb{P}_N^{\beta,a_2}(D_\varepsilon) \le \widetilde{\Pi}_{N,\kappa}^{\beta,a_2}(\mathcal{F}_\varepsilon) \le \widetilde{\mathbb{P}}_{N,\kappa}^{\beta,a_2}(\Phi_{N,\kappa} > 2\eta) + e^{2\eta N}\, \widehat{\Pi}_{N,\kappa}^{\beta,a_2}(\mathcal{F}_\varepsilon)\,.$$

Thanks to Lemma 3.4, it thus suffices to establish that for all $\kappa \ge \kappa_1$,

$$\limsup_{N\to\infty} \frac{1}{N}\log \widehat{\Pi}_{N,\kappa}^{\beta,a_2}(\mathcal{F}_\varepsilon) < 0\,, \tag{4.3}$$

which as we have seen before, follows from Lemma 3.6 since $\mu_\star \notin \mathcal{F}_\varepsilon$ (hence $I(\mathcal{F}_\varepsilon) > 0$).

Acknowledgment. We thank G. Ben Arous and A. Guionnet for a valuable feedback on our preliminary draft and for pointing our attention to the references [4, 15, 17, 26]. A.D. was supported in part by NSF grant DMS-1613091 and E.L. was supported in part by NSF grant DMS-1812095. This research was further supported in part by BSF grant 2018088.

References

[1] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[2] G. Ben Arous and A. Guionnet. Large deviations for Langevin spin glass dynamics. *Probab. Theory Related Fields*, 102:455–509, 1995.

[3] G. Ben Arous and A. Guionnet. Symmetric Langevin spin glass dynamics. *Ann. Probab.*, 25:1367–1422, 1997.

[4] G. Ben Arous and A. Guionnet. Langevin dynamics for Sherrington-Kirkpatrick spin glasses. In *Mathematical aspects of spin glasses and neural networks*, volume 41 of *Progr. Probab.*, pages 323–353. Birkhäuser Boston, Boston, MA, 1998.
[5] T. Cabana. *Large deviations for the dynamics of heterogeneous neural networks*. PhD thesis, Université Pierre et Marie Curie - Paris VI, 2016.
[6] T. Cabana and J. Touboul. Large deviations, dynamics and phase transitions in large stochastic and disordered neural networks. *J. Stat. Phys.*, 153(2):211–269, 2013.
[7] T. Cabana and J. D. Touboul. Large deviations for randomly connected neural networks: I. Spatially extended systems. *Adv. in Appl. Probab.*, 50(3):944–982, 2018.
[8] T. Cabana and J. D. Touboul. Large deviations for randomly connected neural networks: II. State-dependent interactions. *Adv. in Appl. Probab.*, 50(3):983–1004, 2018.
[9] P. Carmona and Y. Hu. Universality in Sherrington-Kirkpatrick's spin glass model. *Ann. Inst. H. Poincaré Probab. Statist.*, 42:215–222, 2006.
[10] S. Chatterjee. A simple invariance theorem. 2005. Preprint, available at `arXiv:math/0508213`.
[11] S. Chatterjee. A generalization of the Lindeberg principle. *Ann. Probab.*, 34:2061–2076, 2006.
[12] A. Crisanti and H. Sompolinsky. Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model. *Physical review. A, General physics*, 36:4922–4939, 1987.
[13] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.
[14] B. Derrida, E. Gardner, and A. Zippelius. An exactly solvable asymmetric neural network model. *Europhysics Letters (EPL)*, 4:167–173, 1987.
[15] O. Faugeras, J. Maclaurin, and E. Tanré. The meanfield limit of a network of Hopfield neurons with correlated synaptic weights. 2019. Preprint, available at `arXiv:1901.10248`.
[16] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1:233–241, 1981.
[17] M. Grunwald. Sanov results for Glauber spin-glass dynamics. *Probab. Theory Related Fields*, 106:187–232, 1996.
[18] F. Guerra and F. L. Toninelli. The thermodynamic limit in mean field spin glass models. *Comm. Math. Phys.*, 230:71–79, 2002.
[19] A. Guionnet. Averaged and quenched propagation of chaos for spin glass dynamics. *Probab. Theory Related Fields*, 109:183–215, 1997.
[20] J. A. Hertz, G. Grinstein, and S. A. Solla. Irreversible spin glasses and neural networks. In J. L. van Hemmen and I. Morgenstern, editors, *Heidelberg Colloquium on Glassy Dynamics*, pages 538–546. Springer-Verlag, Berlin, Heidelberg, 1987.
[21] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79:2554–2558, 1982.
[22] R. Kree and A. Zippelius. Asymmetrically diluted neural networks. In E. Domany, J. L. van Hemmen, and K. Schulten, editors, *Models of Neural Networks*, pages 193–212. Springer-Verlag, Berlin, Heidelberg, 1991.
[23] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin glass theory and beyond*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific Publishing Co., Inc., Teaneck, NJ, 1987.
[24] O. Moynot and M. Samuelides. Large deviations and mean-field theory for asymmetric random recurrent neural networks. *Probab. Theory Related Fields*, 123(1):41–75, 2002.
[25] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Fundamental Principles of Mathematical Sciences*. Springer-Verlag, Berlin, third edition, 1999.
[26] H. Rieger, M. Schreckenberg, and J. Zittartz. Glauber dynamics of the asymmetric SK-model. *Zeitschrift für Physik B Condensed Matter*, 74:527–538, 1989.

[27] H. Sompolinsky and A. Zippelius. Dynamic theory of the spin-glass phase. *Phys. Rev. Lett.*, 47:359–362, 1981.

[28] H. Sompolinsky and A. Zippelius. Relaxational dynamics of the Edwards–Anderson model and the mean-field theory of spin-glasses. *Phys. Rev. B*, 25:6860–6875, 1982.

[29] M. Talagrand. Gaussian averages, Bernoulli averages, and Gibbs' measures. *Random Structures Algorithms*, 21:197–204, 2002.

[30] M. Talagrand. *Spin glasses: a challenge for mathematicians*, volume 46 of *A Series of Modern Surveys in Mathematics*. Springer-Verlag, Berlin, 2003.

Amir Dembo
Mathematics Department and Statistics Department, Stanford University, Stanford, CA 94305, USA.
Email address: adembo@stanford.edu

Eyal Lubetzky
Courant Institute, New York University, 251 Mercer Street, New York, NY 10012, USA.
Email address: eyal@courant.nyu.edu

Ofer Zeitouni
Department of Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel, and Courant Institute, New York University, 251 Mercer Street, New York, NY 10012, USA.
Email address: ofer.zeitouni@weizmann.ac.il