Diffusions interacting through a random matrix: universality via stochastic Taylor expansion

Amir Dembo[1] · Reza Gheissari[2]

© The Author(s) 2021

Abstract
Consider $(X_i(t))$ solving a system of N stochastic differential equations interacting through a random matrix $\mathbf{J} = (J_{ij})$ with independent (not necessarily identically distributed) random coefficients. We show that the trajectories of averaged observables of $(X_i(t))$, initialized from some μ independent of $\mathbf{J}$, are universal, i.e., only depend on the choice of the distribution $\mathbf{J}$ through its first and second moments (assuming e.g., sub-exponential tails). We take a general combinatorial approach to proving universality for dynamical systems with random coefficients, combining a stochastic Taylor expansion with a moment matching-type argument. Concrete settings for which our results imply universality include aging in the spherical SK spin glass, and Langevin dynamics and gradient flows for symmetric and asymmetric Hopfield networks.

Keywords Stochastic differential equations · Universality · Markov semi-group · Random matrices, Disordered systems · Langevin dynamics · Gradient flows

Mathematics Subject Classification 60J60 · 60B20 · 60J35 · 60K35 · 82C44

1 Introduction

Markov processes with random coefficients arise in numerous contexts: e.g., dynamics of spin glasses, optimization on random landscapes, and learning with neural networks. In many cases, when the underlying randomness is Gaussian, they have been found to give rise to a rich class of behaviors, including metastability, trapping, and aging. In

✉ Reza Gheissari
gheissari@berkeley.edu

Amir Dembo
adembo@stanford.edu

1 Department of Statistics and Department of Mathematics, Stanford University, Stanford, CA, USA

2 Departments of Statistics and EECS, University of California, Berkeley, Berkeley, CA, USA

this paper, we analyze a class of stochastic differential systems (SDS's) in their high dimensional limit, where the couplings are linear and encoded by a random matrix. We show that trajectories of polynomial statistics of the SDS are universal: they have the same high-dimensional behavior if one replaces the Gaussian interaction matrix by a non-Gaussian one with the same mean and variance profiles.

Universality, can broadly be described as the phenomenon that for high dimensional ensembles $(X_i)_{i \leq N}$ governed by a large number of independent random variables $(Z_i)_{i \leq N}$, macrocopic statistics of the ensemble only depend on the laws of (Z_i) through their low moments. Of course, the most classical example of universality is the central limit theorem (CLT), where $(X_i) = (Z_i)$, and the statistic is the normalized sum. Slightly more involved examples are invariance principles, where the limiting Brownian motion only depends on the distribution of the random walk increments through its first and second moments.

Lindeberg's classical proof of the CLT iteratively replaces Z_i with $\tilde{Z}_i$ (Gaussian with the same mean and variance) and shows that the cumulative effect of these replacements is microscopic. This approach has proven to be very robust, and has been generalized e.g., to polynomials $f(Z_1, \ldots, Z_N)$ in [29,34] and more generally, smooth functions with bounded derivatives in [8,9]. A more combinatorial approach is a *moment matching* argument to compare moments of statistics $f(X_1, \ldots, X_N)$ to moments of $f(\tilde{X}_1, \ldots, \tilde{X}_N)$ and showing that the difference is dominated by the differences in the first few moments of Z_i and $\tilde{Z}_i$.

With these approaches, universality has been proven in a wide range of ensembles where the relationship between (X_i) and (Z_i) is more complicated. A fundamental example is when (X_i) are the eigenvalues of a random matrix with entries (Z_i). There, the empirical distribution of (X_i) is well-known to have the same limit (e.g., the semi-circle law for Wigner matrices [40]). In the last decade, remarkably, universality has been found to extend to local statistics of the ensemble (X_i) e.g., typical size of gaps between eigenvalues, and k-point correlations. Universality in random matrix theory has been a tremendous success and we cannot hope to do justice to the literature therein; we instead refer to the seminal works [19,37] and the surveys [20,38].

Another class of ensembles for which universality has been shown is disordered interacting particle systems from statistical physics, and in particular the family of mean-field spin glass models. A canonical example of these are spin glasses where N particles in states (X_i), interact through a random symmetric coupling matrix (or in the case of higher order interactions, tensor) composed of independent entries Z_i. More precisely, with these interactions, they are endowed with an energy landscape, or Hamiltonian, that is topologically complex, and (X_i) are drawn from the corresponding Gibbs distribution. The statistics of (X_i) in such families of spin glasses have been found to exhibit an extremely rich and varied phase diagram featuring phenomena like breaking of ergodicity and replica symmetry [33]. Most of their analysis, including the calculation of the free energy, and the proof of the celebrated Parisi formula for the overlap distribution, were first carried out in the Gaussian setting [22,32,36]. Talagrand later showed that these also held in the case of Bernoulli (Z_i) in [35]; this universality was extended to general (Z_i) as an application of [9].

The dynamics (Markov processes exploring the Hamiltonian) for such spin glass models are a prototype and motivating force for this paper. The general setting we

consider here is that of a system of N linearly coupled SDE's, where the couplings are encoded in a *random* matrix $\mathbf{J}$, and driven by N independent Brownian motions. That is, $\mathbf{X}_t = (X_1(t), \ldots, X_N(t))$ is the solution to the SDS

$$\begin{cases} d\mathbf{X}_t = \mathbf{J}^T \mathbf{X}_t dt + \mathbf{h} dt + \mathbf{\Sigma}(\mathbf{X}_t) d\mathbf{B}_t \\ \mathbf{X}_0 \sim \mu \in \mathcal{M}_1(\mathbb{R}^N) \end{cases}, \tag{1.1}$$

where $\mathbf{J}$ is a random matrix with independent entries (up to, possibly, a symmetry constraint) and variance profile $\mathbf{m} = (m_{ij})_{i,j}$ scaled such that $\mathbb{E}[\|\mathbf{J}\|_2] = O(1)$, $\mathbf{h}$ is a bounded drift vector, and $\mathbf{\Sigma}$ is an affine transform of $\mathbf{X}_t$. Note that for $\mathbf{\Sigma}(\mathbf{X}_t)$ non-constant, we do not expect to have an explicit closed-form solution to (1.1).

In the $N \to \infty$ limit, the diffusions of (1.1) encompass many interesting and well-studied models of Markov processes with random coefficients, and give rise to rich and varied behavior. This includes metastability, aging, and non-Markovian limiting evolution equations, in e.g., randomly coupled (geometric) Brownian motions, and Langevin dynamics and gradient flows for the spherical Sherrington–Kirkpatrick (SK) spin glass and symmetric and asymmetric Hopfield nets [6,13,25–27]: concrete applications are described in Sect. 1.4. In many such examples, the analysis is more tractable when $\mathbf{J}$ is Gaussian and one can use tools like Gaussian integration by parts, Girsanov, and the rotational invariance of the Gaussian ensemble.

In this paper, we develop a simple combinatorial framework for proving universality for the solution trajectories of SDS's of the form (1.1). Before describing our approach, we explain a few difficulties one encounters when trying to prove universality for solutions of randomly coupled dynamical systems, using some of the approaches described above for other universality results. We begin by considering a Lindeberg approach where we examine the effect that re-sampling one J_{ij} has on an averaged statistic $F(t) = F(X_1(t), \ldots, X_N(t))$. The obstacle in employing such an approach is that changing J_{ij} to $\tilde{J}_{ij}$ on $X_j(t)$, say, beyond affecting the drift

$$\sum_{1 \le i \le N} J_{ij} X_i(t) + h_j,$$

of the j-th coordinate of the SDS, also induces a *highly non-linear* effect both on $X_j(t)$ and on $X_i(t)$ for all $i \ne j$. The problem instead lends itself to comparing the effect of $\mathbf{J} \to \tilde{\mathbf{J}}$ in a more averaged way.

An alternative approach would be to use the linear structure of the problem in a strong way, relying on sharp universality results on the spectra of random matrices to study the problem. This approach, while feasible if $\mathbf{\Sigma}(\mathbf{X}_t)$ is constant, requires one to diagonalize the problem without loss of generality—i.e., it requires an assumption of joint rotational invariance for the laws of $(\mathbf{X}_0, \mathbf{J}, \mathbf{B})$. In [2], such an approach is followed for analyzing the dynamics of the spherical SK model, and their results hold assuming the law of $\mathbf{J}$ is invariant under the orthogonal group, and its spectrum satisfies certain large deviation estimates satisfied by the GOE. However, this restriction would not include the cases of e.g., the uniform measures on $[-1, 1]^N$ and $\{\pm 1\}^N$ absent the rotational symmetry, and could not include the case of non-constant $\mathbf{\Sigma}(\mathbf{X}_t)$.

Very recently, [17] proved a universality result for the dynamics of the *asymmetric* Langevin dynamics for the *soft-spin* SK *model*. There they used large deviations theory to obtain exponential control on the empirical measure on sample paths—as obtained in the Gaussian setting in [6,7]—together with sharp control via Girsanov's theorem on the Radon–Nikodym derivative between the Gaussian paths and those driven by non-Gaussian $\mathbf{J}$ on short time scales, to show universality for the empirical measure $\mathcal{L}_N = \frac{1}{N} \sum_i \delta_{X_i(t)}$. While such an approach allows for a deterministic non-linearity in the drift through a (double-well) confining potential, it cannot handle degenerate diffusions, e.g. the gradient flow. Further, the need for control on the trajectories at the exponential scale forces [17] to consider only asymmetric i.i.d. $\mathbf{J}$ (whereby the Radon–Nikodym derivative is a product of functions of independent rows of $\mathbf{J}^T$).

We introduce a simple combinatorial approach to proving universality for SDS's of the form of (1.1), similar in flavor to the moment method. Namely, we avoid the inherent difficulty of the problem, that the transformation $\mathbf{J} \to \tilde{\mathbf{J}}$ affects $X_j(t)$ through both $(J_{ij})_i \to (\tilde{J}_{ij})_i$ and $(X_i(t))_i \to (\tilde{X}_i(t))_i$. We do so by Taylor expanding the semigroup $P_t f = \mathbb{E}_{\mathbf{X}_0}[f(\mathbf{X}_t)]$ in powers of the infinitesimal generator: each term appearing in this expansion is a polynomial in $(x_i), (J_{ij})$ evaluated at $\mathbf{X}_0$ where, crucially, the initial data is independent of J_{ij}. One then finds that on order one timescales, the predominant contribution to $\mathbb{E}[P_t f]$ is from polynomials whose degree in $(J_{ij})_{i,j}$ is at most two. We refer to Sect. 1.3 for more details.

This approach works quite generally, and is robust to symmetric and asymmetric choices of $\mathbf{J}$ with non-homogenous means and variances, and general choices of diffusion coefficients in (1.1), including $\mathbf{\Sigma}(\mathbf{X}_t)$ non-constant making the diffusion non-linear, and $\mathbf{\Sigma} \equiv 0$ corresponding to a deterministic dynamical system. Lastly, the analysis works for arbitrary initialization independent of $\mathbf{J}$. The assumption of linear drift is, of course, important, and one would like to be able to drop it. We emphasize, though, that this is primarily used in order to justify the absolute convergence of the Taylor expansion of the semigroup, which one could hope to justify by other means for higher order diffusions given that a strong solution exists; the remaining combinatorial framework for moments of the generator may then generalize. We discuss this in Remark 1.5.

We end this section by mentioning two recent results [1,10] showing universality for a Lipschitz family of approximate message passing (AMP) algorithms—a discrete-time state evolution that has found many applications to inference and optimization in high dimensions. Some of the ideas there appear similar in spirit to our approach, using a combinatorial approach to control moments of the final state of the AMP. All the same, the general setting of (1.1) introduces many key differences e.g., the diffusions of (1.1) are in general non-linear, not globally Lipschitz, and have a built-in stochasticity.

1.1 Setup: diffusions with random linear interactions

Consider an N-dimensional stochastic differential system with a mixture of random and deterministic linear interactions, along with possibly, some constant drifts. More precisely, consider the SDS $\mathbf{X}_t := (X_i(t))_{i=1}^N$ driven by the following parameters.

Suppose that for some matrix $\mathbf{m} = (m_{ij})_{i,j}$ we have random interactions given by the random matrix

$$\mathbf{A} = (A_{ij})_{1 \leq i,j \leq N}, \qquad \text{where} \qquad \mathbb{E}[A_{ij}] = 0, \quad \mathbb{E}[A_{ij}^2] = m_{ij}.$$

We assume that the entries A_{ij} are either fully independent, or are independent up to a symmetry constraint $A_{ij} = A_{ji}$. Let $\mathbb{P}_{\mathbf{A}}$ be the law of $\mathbf{A}$. In order to scale the interactions to have an order one cumulative effect, it will be convenient to work with the rescaled interactions matrix $\mathbf{J}$ given by

$$\mathbf{J} := N^{-1/2} \mathbf{A}.$$

We then denote the distribution induced by $\mathbb{P}_{\mathbf{A}}$ on $\mathbf{J}$ by $\mathbb{P}_{\mathbf{J}}$.

We further consider additional deterministic interactions satisfying, for some constant $C_{\mathbf{\Lambda}} < \infty$,

$$\mathbf{\Lambda} = (\Lambda_{ij})_{1 \leq i,j \leq N}, \quad \text{where} \quad \max_i \|(\Lambda_{ij})_j\|_1 \leq C_{\mathbf{\Lambda}} \quad \text{and}$$

$$\sup_{i,j} |\Lambda_{ij}| \leq \frac{C_{\mathbf{\Lambda}}}{\mathcal{N}_{\mathbf{\Lambda}}} \quad \text{for} \quad \mathcal{N}_{\mathbf{\Lambda}} := \max_j \|(\Lambda_{ij})_i\|_0$$

(the $\|\cdot\|_0$-norm of a vector is its number of non-zero entries). We also consider external drift parameters

$$\mathbf{h} = (h_i)_{1 \leq i \leq N}, \qquad \text{where} \qquad \sup_{i \leq N} |h_i| \leq C_{\mathbf{h}} \quad \text{for a constant} \quad C_{\mathbf{h}} < \infty,$$

and diffusion coefficients $\mathbf{\Sigma}(\mathbf{X}_t)$ governed by the matrix

$$\sigma = (\sigma_{ij})_{0 \leq i \leq N, 1 \leq j \leq N} \quad \text{where} \quad \sup_{1 \leq j \leq N} |\sigma_{0j}| \leq C_{\sigma} \quad \text{and}$$

$$\sup_{1 \leq i,j \leq N} |\sigma_{ij}| \leq \frac{C_{\sigma}}{\mathcal{N}_{\sigma}} \quad \text{for} \quad \mathcal{N}_{\sigma} := \max_j \|(\sigma_{ij})_i\|_0.$$

The SDS $(\mathbf{X}_t)_{t \geq 0} = (X_1(t), X_2(t), \ldots, X_N(t))_{t \geq 0}$ initialized from some random $\mathbf{X}_0$ distributed according to a product measure μ is driven by a standard Brownian motion $\mathbf{B}_t = (B_1(t), \ldots, B_N(t))$ as follows

$$dX_j(t) = \sum_{i=1}^N J_{ij} X_i(t) dt + \sum_{i=1}^N \Lambda_{ij} X_i(t) dt + h_j dt + \sqrt{2} \Big(\sum_{i=0}^N \sigma_{ij} X_i(t) \Big) dB_j(t),$$

$$X_i(0) \sim \mu_i, \tag{1.2}$$

where for ease of notation, we hereon set $X_0(t) \equiv 1$ so that $(\sigma_{0j})_{j \geq 1}$ capture the constant diffusion coefficients. We denote the martingale part of $\mathbf{X}_t$ by

$$\mathbf{M}_t = (M_j(t))_{j \leq N}, \quad \text{where} \quad dM_j(t) = \sqrt{2}\Big(\sum_{i=0}^{N} \sigma_{ij} X_i(t)\Big) dB_j(t). \quad (1.3)$$

The process $\mathbf{X}_t$ is well-defined for a.e. $\mathbf{J}$ and all $t \geq 0$ (as we have finite, possibly N-dependent operator norms $\|\mathbf{J}\|_2$, $\|\mathbf{\Lambda}\|_2$ and $\|(\sigma_{ij})_{i\geq 1}\|_2$, see e.g., [31, Theorem 5.2.1]).

Notational comment There are three distinct sources of randomness above dictating the law of the solution $\mathbf{X}_t$ to (1.2): the law of the interaction matrix $\mathbb{P}_\mathbf{J}$, the law of the Brownian motions, denoted $\mathbb{P}_\mathbf{B}$, and the law of the initial data μ—each of these are product measures and we do not distinguish notationally between the law of the individual entries of $\mathbf{J}$, $\mathbf{B}$ or $\mathbf{X}_0$ and the ensembles.

   In proving universality, we consider the difference between $\mathbb{P}_\mathbf{J}$, $\mathbb{P}_{\tilde{\mathbf{J}}}$ induced by two different distributions $\mathbb{P}_\mathbf{A}$ and $\mathbb{P}_{\tilde{\mathbf{A}}}$ over mean-zero random matrices $\mathbf{A}$, $\tilde{\mathbf{A}}$ with independent entries (possibly up to symmetry), having matching variance profiles $\mathbf{m} = \tilde{\mathbf{m}}$. For ease of notation, we will henceforth use

$$\mathbb{P} = \mu \otimes \mathbb{P}_\mathbf{J} \otimes \mathbb{P}_\mathbf{B}, \quad \text{and} \quad \tilde{\mathbb{P}} = \mu \otimes \mathbb{P}_{\tilde{\mathbf{J}}} \otimes \mathbb{P}_\mathbf{B},$$

and denote the corresponding expectations $\mathbb{E}$ and $\tilde{\mathbb{E}}$ respectively.

1.2 Main results

We begin by describing the observables to which our universality results apply. The building blocks of these observables are chosen among the family of vector valued functions,

$$\mathfrak{F} = \begin{cases} \mathbf{1}_t = (1, \ldots, 1) \\ \mathbf{X}_t = (X_1(t), \ldots, X_N(t)) \\ \mathbf{G}_t = (G_1(\mathbf{X}_t), \ldots, G_N(\mathbf{X}_t)) \\ \mathbf{M}_t = (M_1(t), \ldots, M_N(t)) \end{cases}, \quad \text{where} \quad G_j(\mathbf{x}) = \sum_{i=1}^{N} J_{ij} x_i. \quad (1.4)$$

We establish universality in the mean for weighted empirical averages of monomials in functions from $\mathfrak{F}$ evaluated at a finite collection of times. Specifically, fixing an m-tensor $\mathbf{a} = (a_{i_1,\ldots,i_m})$ with entries bounded by $C_\mathbf{a}$ and a p-tuple of times $\mathbf{t} = (t_1, \ldots, t_p)$, for every $\ell \leq m$, fix p observables $\mathcal{Y}^{(\ell,1)}, \ldots, \mathcal{Y}^{(\ell,p)} \in \mathfrak{F}$ which are to be evaluated at these p times. That is,

$$F(\mathbf{t}) = \frac{1}{N^m} \sum_{i_1,\ldots,i_m \leq N} a_{i_1,\ldots,i_m} F_{i_1}^{(1)}(\mathbf{t}) \cdots F_{i_m}^{(m)}(\mathbf{t}), \quad \text{where}$$

$$F_i^{(\ell)}(\mathbf{t}) = \mathcal{Y}_i^{(\ell,1)}(t_1) \cdots \mathcal{Y}_i^{(\ell,p)}(t_p). \quad (1.5)$$

We also need to add a sub-exponential tail constraint on μ and $\mathbb{P}_\mathbf{A}$ beyond the minimal assumptions of zero-mean and matching variances of $\mathbb{P}_\mathbf{A}$ and $\mathbb{P}_{\tilde{\mathbf{A}}}$; this is henceforth referred to as Hypothesis 1.

Hypothesis 1 *Assume that the law μ is a product of μ_i of $X_i(0)$ having finite moments of all order, which are bounded uniformly over i and N. That is, there exist $C_\mu(r) \geq 1$ such that for any r finite,*

$$\sup_N \sup_{i \leq N} \mathbb{E}[|X_i(0)|^r] \leq C_\mu(r) . \tag{1.6}$$

Further assume $\mathbb{P}_\mathbf{A}$ has uniformly bounded exponential tails, i.e., the following equivalent properties hold:

$$\sup_N \sup_{i,j \leq N} \mathbb{E}[e^{\varepsilon|A_{ij}|}] < \infty , \qquad \text{for some} \quad \varepsilon > 0 , \tag{1.7}$$

$$\sup_N \sup_{i,j \leq N} \mathbb{E}[|A_{ij}|^\ell] \leq (\ell - 1)! \, C_\mathbf{A}^{\ell/2} , \qquad \forall \ell \geq 1 \text{ and some } C_\mathbf{A} < \infty . \tag{1.8}$$

For ease of notation for dependencies on constants, we denote by $\mathbf{C}_\star := \max\{C_\mathbf{A}^{1/2}, C_{\tilde{\mathbf{A}}}^{1/2}, C_\Lambda, C_\mathbf{h}, C_\sigma^2\}$ (where $C_{\tilde{\mathbf{A}}}$ is the constant $C_\mathbf{A}$ with respect to distribution $\mathbb{P}_{\tilde{\mathbf{A}}}$), and state our first result, on universality at the level of the mean (hence also of moments), for observables (1.5).

Theorem 1 *Let $\mu, \mathbb{P}_\mathbf{A}, \mathbb{P}_{\tilde{\mathbf{A}}}$ satisfy Hypothesis 1 and suppose that $\mathbf{A}, \tilde{\mathbf{A}}$, symmetric or independent, are mean-zero of matching variance profile $\mathbf{m} = (m_{ij})_{i,j}$. For any $T, m, p < \infty$ and $\mathbf{a} \in \mathbb{R}^{N^m}$ with $\|\mathbf{a}\|_\infty \leq C_\mathbf{a}$, there exists $C(T, m, p, C_\mathbf{a}, \mathbf{C}_\star, C_\mu) < \infty$, such that for every N and F as in (1.5) with $(\mathcal{Y}^{(\ell,1)}, \ldots, \mathcal{Y}^{(\ell,p)}) \in \mathfrak{F}$,*

$$\sup_{\mathbf{t} \in [0,T]^p} \left| \mathbb{E}[F(\mathbf{t})] - \tilde{\mathbb{E}}[F(\mathbf{t})] \right| \leq C N^{-1/2} .$$

In particular, $\left| \mathbb{E}[F(\mathbf{t})] - \tilde{\mathbb{E}}[F(\mathbf{t})] \right| \to 0$ as $N \to \infty$, uniformly in $\mathbf{t} \in [0, T]^p$.

Theorem 1 follows from a more general result bounding the difference in expectations for each individual monomial $F_i^{(\ell)}$ from (1.5) with $(\mathcal{Y}^{(\ell,1)}, \ldots, \mathcal{Y}^{(\ell,p)}) \in \mathfrak{F}$. As a special case, see Proposition 2.1, we find that the moments of each spin $X_i(t)$ are universal. Specifically, for every fixed k,

$$\sup_{t \in [0,T]} \max_{1 \leq i \leq N} \left| \mathbb{E}[X_i(t)^k] - \tilde{\mathbb{E}}[X_i(t)^k] \right| = O(N^{-1/2}) . \tag{1.9}$$

For a more restricted class of observables, with additional restrictions on the distributions μ and $\mathbb{P}_\mathbf{A}$ and $\mathbb{P}_{\tilde{\mathbf{A}}}$, we extend the above to almost sure and L^q convergence for the observable trajectories. Precisely, we restrict the observables of (1.5) to $m = 1$

and $p = 2$, leaving, the following quadratic observables

$$F(\mathbf{t}) = F_{\mathcal{Y}, \mathcal{Y}', \mathbf{a}}(\mathbf{X}_t, \mathbf{X}_{t'}) := \frac{1}{N} \sum_{i=1}^{N} a_i \mathcal{Y}_i(t) \mathcal{Y}_i'(t'). \tag{1.10}$$

In order to extend Theorem 1 to a convergence for the trajectories of these observables, we further need to assume that $\mathbf{\Sigma}$ is constant, so that $\mathbf{M}_t$ is just a scaled Brownian motion, and assume the following concentration property on $\mu, \mathbb{P}_\mathbf{A}, \mathbb{P}_{\tilde{\mathbf{A}}}$, which we refer to as Hypothesis 2.

Hypothesis 2 *A sequence of probability measures* $(\mathbb{P}^{(n)})_{n \geq 1}$ *over* $\mathbf{Z}_n$ *in metric spaces* $(\mathcal{X}_n, d)$ *satisfies* exponential concentration for Lipschitz functions *if there exists some* $C > 0$ *such that for any sequence of 1-Lipschitz functions* $f_n : (\mathcal{X}_n, d) \to (\mathbb{R}, |\cdot|)$ *and all* $\lambda > 0$,

$$\mathbb{P}^{(n)}\big(|f_n(\mathbf{Z}_n) - \mathbb{E}[f_n(\mathbf{Z}_n)]| > \lambda\big) \leq C \exp(-\lambda/C). \tag{1.11}$$

Assume that $\mu, \mathbb{P}_\mathbf{A}$ *respectively satisfy exponential concentration for Lipschitz functions on* $\mathbb{R}^N$ *and* $\mathbb{R}^{N^2}$ *(or* $\mathbb{R}^{N(N+1)/2}$ *if* $\mathbf{A}$ *is symmetric), equipped with their Euclidian norms, for some* $C_\mu, C_\mathbf{A} > 0$.

Remark 1.1 Recall, from the theory of measure concentration, that Hypothesis 2 holds for any distribution on $\mathbb{R}^n$ which satisfy a Poincaré inequality with constant $c > 0$ (independent of n), namely for all nice f one has that $\mathrm{Var}[f(\mathbf{Z}_n)] \leq c\mathbb{E}[|\nabla f(\mathbf{Z}_n)|^2]$ (see [21]). By the tensorization of the Poincaré inequality, if $\mathbf{Z}_n = (Z_1, \ldots, Z_n)$, and each of the laws of Z_i satisfy this inequality, then the product also satisfies it with the worst constant c. Having here product measures $\mu, \mathbb{P}_\mathbf{A}$, the marginal laws can come from any distribution satisfying a Poincaré inequality in $n = 1$. These include (see e.g., [39])

– Exponential, Gaussian, and log-concave measures of the form $\exp(-V(x))$ for $V(x)$ strictly convex,
– Linear functionals of r.v.'s having a Poincaré inequality: e.g., the uniform measure on $[-1, 1]$.

The next theorem shows that under Hypothesis 2, any F of the form (1.10) concentrates around its mean.

Theorem 2 *Suppose* $\mu, \mathbb{P}_\mathbf{A}$ *satisfy Hypotheses 1–2 and the diffusion coefficients have* $\sigma_{ij} = 0$ *if* $i \neq 0$. *Then, for some* $C(T, C_\mathbf{a}, \mathbf{C}_\star, C_\mu) > 0$, *any* $\|\mathbf{a}\|_\infty \leq C_\mathbf{a}$, *every* F *as in* (1.10) *with* $\mathcal{Y}, \mathcal{Y}' \in \mathfrak{F}$, *all* $\lambda > 0$ *and* $N \geq N_0(T, C_\mathbf{a}, \mathbf{C}_\star, C_\mu)$,

$$\mathbb{P}\left(\sup_{\mathbf{t} \in [0,T]^2} |F(\mathbf{t}) - \mathbb{E}[F(\mathbf{t})]| \geq \lambda\right) \leq p_N(\lambda) := \begin{cases} N^C e^{-\lambda \sqrt{N}/C}, & \lambda \leq C \\ e^{-(\log \lambda)\sqrt{N}/C}, & \lambda > C \end{cases}. \tag{1.12}$$

(One might observe that the $\exp(-\Omega(\sqrt{N}))$ concentration in (1.12) differs from the more traditional $\exp(-\Omega(N))$ concentration in e.g. [2,3]; such differences, which recur throughout the paper, are because our Hypothesis 2 allows for merely sub-exponential, as opposed to Gaussian, tails.) Combining Theorems 1 and 2 we get the following strong universality for such quadratic observables.

Corollary 3 *Suppose $\mu, \mathbb{P}_{\mathbf{A}}, \mathbb{P}_{\tilde{\mathbf{A}}}$ satisfy Hypotheses 1–2, where $\mathbf{A}, \tilde{\mathbf{A}}$, symmetric or independent, are mean-zero and have matching variance profile $\mathbf{m} = (m_{ij})_{i,j}$. Let $F(\cdot)$ and $\tilde{F}(\cdot)$ be as in (1.10), for $\mathbf{a} \in \mathbb{R}^N$ such that $\|\mathbf{a}\|_{\infty} \leq C_{\mathbf{a}}$, with respect to the corresponding solutions $\mathbf{X}_t, \tilde{\mathbf{X}}_t$ for (1.2) with constant Σ, i.e., $\sigma_{ij} = 0$ if $i \neq 0$. Then, for every $T < \infty$ we have that as $N \to \infty$,*

$$Z_N := \sup_{\mathbf{t} \in [0,T]^2} \left| F(\mathbf{t}) - \tilde{F}(\mathbf{t}) \right| \to 0 \quad \text{almost surely, and in } L^q \text{ for } q \geq 1.$$

Proof The observables of (1.10) correspond to the $m = 1$ and $p = 2$ case of (1.5), so Theorem 1 applies here with some constant $C_1 = C(T, m, p, C_{\mathbf{a}}, \mathbf{C}_{\star}, C_{\mu})$. For $N \geq (\lambda/C_1)^2$ we then get upon combining the triangle inequality with Theorems 1–2, that

$$\mathbb{P}(Z_N > 3\lambda) \leq 2 p_N(\lambda).$$

Since $\sum_N p_N(\lambda) < \infty$ for any fixed $\lambda > 0$, by Borel-Cantelli $Z_N \overset{a.s.}{\to} 0$ as $N \to \infty$. Similarly, upon using the triangle inequality for $\|\cdot\|_q$ we get from Theorems 1 and 2 that

$$\left(\mathbb{E}[|Z_N|^q]\right)^{1/q} \leq C_1 N^{-1/2} + 2\left(C^q + \int_C^{\infty} q\lambda^{q-1} p_N(\lambda) d\lambda\right)^{1/q}.$$

Further, $N \mapsto p_N(\cdot)$ decrease pointwise on $[C, \infty)$, while for any $q \geq 1$, the preceding integral is finite for all N large enough. With $\{Z_N^q\}_N$ uniformly integrable, it follows that $Z_N \to 0$ also in L^q. □

1.3 Proof strategy

As mentioned in the introduction, traditional approaches to proving universality run into substantial difficulty when we apply them to diffusions with random coefficients. The dependence on specific entries of the random matrix are quite bad, as the dependence applies in the drift both through the J_{ij}, and through its effect on $\mathbf{X}_t$, whose history evidently also depends on J_{ij}: this effect can exponentially amplify small differences; in fact, the exponential amplification is inherent to the problem at hand.

At a high level, our strategy for proving Theorem 1, and the main novelty of the paper, is to leverage the independence of μ from $\mathbb{P}_{\mathbf{J}}, \mathbb{P}_{\tilde{\mathbf{J}}}$ by *pulling back* $f(\mathbf{X}_t)$ and $f(\tilde{\mathbf{X}}_t)$ to properties of (time) derivatives of $f(\mathbf{X}_t)$ evaluated at $t = 0$. At the level of expectations, these derivatives can be seen as iterates of the infinitesimal generator applied to the function F, which can then be controlled by combinatorial moment

methods. The dominant contribution to the drift of F comes from drift terms that are polynomials of degree at most two in $(J_{ij})_{ij}$. Since the first two moments of $\mathbb{P}_{\mathbf{A}}$ and $\mathbb{P}_{\tilde{\mathbf{A}}}$ match, these terms do not contribute to the difference in expectations above. We emphasize that the approach does not need rely on an explicit solution to the SDE of (1.2), nor does it use exponential control, or large deviations theory as in [17], or refined estimates on the spectrum of $\mathbf{A}$ as in the setting of [2] where, crucially, the process has a rotational symmetry.

Recall that the SDE defined in Eq. (1.2) has infinitesimal generator L that we split as follows (see e.g., [31, Theorem 7.3.3]):

$$L := \underbrace{\sum_{1 \le i,j \le N} J_{ij} x_i \partial_j}_{L_{\mathbf{J}}} + \underbrace{\sum_{1 \le i,j \le N} \Lambda_{ij} x_i \partial_j}_{L_{\mathbf{\Lambda}}} + \underbrace{\sum_{1 \le j \le N} h_j \partial_j}_{L_{\mathbf{h}}} + \underbrace{\sum_{1 \le j \le N} \left(\sum_{0 \le i \le N} \sigma_{ij} x_i \right)^2 \partial_j \partial_j}_{L_{\Delta}} .$$

$$(1.13)$$

By Ito's formula, we have for every f, say, in $C^{\infty}(\mathbb{R}^N)$,

$$\left| \mathbb{E}[f(\mathbf{X}_t)] - \tilde{\mathbb{E}}[f(\tilde{\mathbf{X}}_t)] \right| = \left| \mathbb{E}[P_t f(\mathbf{X}_0)] - \tilde{\mathbb{E}}[P_t f(\mathbf{X}_0)] \right| ,$$

where $P_t = P_t(\mathbf{J})$ denotes the *semi-group operator*

$$P_t f(\mathbf{x}) := \mathbb{E}_{\mathbf{B}}[f(\mathbf{X}_t) \mid \mathbf{X}_0 = \mathbf{x}] \quad \text{with formal expansion} \quad P_t = e^{tL} \quad (1.14)$$

in terms of the generator L. In order to reduce the problem to a combinatorial question, we wish to Taylor expand the semi-group operator $P_t f = e^{tL} f$. As long as f is smooth and the Taylor expansion converges absolutely—shown in Sect. 2.2—this formal expansion is valid and we can switch expectations over μ, $\mathbb{P}_{\mathbf{J}}$, $\mathbb{P}_{\tilde{\mathbf{J}}}$ with the sum, and compute expectations of powers of the generator L acting on f. Namely, the difference in expectations is bounded by controlling (1) the size in N, and (2) the growth in k of

$$\left| \mathbb{E}[(L^k f)(\mathbf{X}_0)] - \tilde{\mathbb{E}}[(L^k f)(\mathbf{X}_0)] \right| . \quad (1.15)$$

Expanding these terms as words in $L_{\mathbf{J}}, L_{\mathbf{\Lambda}}, L_{\mathbf{h}}, L_{\Delta}$, we observe that a non-zero difference between the two expectations in (1.15), can only come from the summands (monomials in $\mathbf{J}, \mathbf{X}, \mathbf{\Lambda}, \mathbf{h}, \sigma$) satisfying

– Every J_{ij} that is present, must appear at least twice.
– At least one J_{ij} must appear at least three times.

This is because the means of $\mathbb{P}_{\mathbf{A}}$, $\mathbb{P}_{\tilde{\mathbf{A}}}$ are zero, and the variances of $\mathbb{P}_{\mathbf{A}}$ and $\mathbb{P}_{\tilde{\mathbf{A}}}$ match. A careful analysis of this combinatorial problem for the monomials eventually yields that the contributions from these monomials are, together, $O(N^{-1/2})$ in N, and $o(k!)$ in k: this computation is carried out in Sect. 2.3.

Remark 1.2 One may notice that in the case where $\mathbf{\Sigma}(\mathbf{X}_t)$ is constant so that $\mathbf{M}_t$ is just a Brownian motion, we are left with a linear SDS and one could use this linearity

in a more central way, to explicitly solve expectations of monomials in $(X_i(t))_i$ as Gaussian integrals and time integrals over words in $e^{s\mathbf{J}}$ and $(X_i(t))_i$. If the system $\mathbf{X}_t$ is invariant under rotations, then we can work in the coordinates of $\mathbf{J}$ so that it is diagonal and apply universality results for the spectrum of $\mathbf{J}$. Absent rotational symmetry, however, the natural step would be to Taylor expand $e^{s\mathbf{J}}$, at which point the expansion and the resulting combinatorics will be similar, and perhaps less transparent, than our generator based approach. Of course, for non-constant $\mathbf{\Sigma}(\mathbf{X}_t)$ as in Theorem 1, the SDS is non-linear, and such an approach would not generalize.

In Sect. 3, we extend this bound on the difference in expectations of statistics f to multi-time observables, then to statistics that contain the driving martingale terms and finally establish the universality at the level of expectation for observables of the form of (1.5), as stated in Theorem 1. In Sect. 4, we adapt the approach of [3] to establish Theorem 2, namely, to show that the restricted class of observables of (1.10) concentrate around their expectations, by localizing to a set of large probability where F is $O(N^{-1/2})$-Lipschitz in the triplet $(\mathbf{X}_0, \mathbf{J}, (\mathbf{M}_t)_{t\in[0,T]})$ and using Hypothesis 2.

1.4 Applications

In this section, we discuss systems for which Theorem 1–Corollary 3 imply concrete universality results. All the examples that follow will be in the context of $\mathbf{\Sigma}$ that is constant, i.e., $\sigma_{ij} = 0$ if $i \neq 0$, where both Theorems 1–2 apply. Among the examples with non-constant $\mathbf{\Sigma}$, one which may be of interest is a system of geometric Brownian motions interacting linearly through $\mathbf{J}$.

We next describe two well-studied families of Markov processes/dynamical systems to which our results apply: Langevin dynamics and gradient flows on various energy landscapes (Hamiltonians) or loss functions.

Langevin dynamics

In the case where $\mathbf{J}$ and $\mathbf{\Lambda}$ are symmetric matrices, and σ_{0j} are identically one, (1.2) corresponds exactly to the *Langevin dynamics* for the *Hamiltonian*

$$H(\mathbf{x}) = -\beta \sum_{1 \leq i,j \leq N} (J_{ij} + \Lambda_{ij})x_i x_j - \beta \sum_{1 \leq i \leq N} h_i x_i. \tag{1.16}$$

The linearity of the diffusion here corresponds to having a quadratic Hamiltonian. The Langevin dynamics is a reversible Markov process designed such that, when non-degenerate, its invariant measure on $\mathbb{R}^N$ is given by $d\pi(\mathbf{x}) \propto e^{-H(\mathbf{x})}d\mathbf{x}$. For Hamiltonians coming from spin glass theory, the Langevin dynamics has been analyzed at length in the case of Gaussian disorder, and found to have a varied and rich behavior; in §1.4.1, we explore this further in the context of a simple spin glass model, called the spherical SK model.

Gradient flows

The case where σ_{0j} are identically zero—i.e., besides the randomness of $\mathbf{J}$ and, possibly, the initial data, the dynamics is deterministic— also fits into the framework of the paper. Here, given $\mathbf{J}$ and $\mathbf{X}_0$, the law of the dynamics is taken to be the delta function on the trajectory of the solution to the resulting system of ODE's. This corresponds to the *gradient flow* on $H(\mathbf{x})$: in optimization and learning settings, e.g., the examples of Sects. 1.4.2–1.4.3, gradient descent and its many variants, are favored methods.

We now turn to a few well-studied concrete problems to which our results are applicable.

1.4.1 The (soft) spherical SK model

The dynamics of spin glasses are a canonical setting in which Markov processes with random coefficients are studied in their thermodynamic ($N \to \infty$) limit. The short-time ($N \to \infty$, then $T \to \infty$) behavior of Langevin dynamics, especially, in the context of spin glasses have been extensively studied in both the physics and math literature [2–7,11,12,15,18]. Perhaps the most well-known mean field spin glass is the Sherrington–Kirkpatrick (SK) spin glass, where N spins taking values in $\{+1, -1\}$ interact pairwise with one another, and their interaction strengths are moderated by "coupling" parameters $J_{ij} = J_{ji}$ which are drawn i.i.d., say, Gaussian. We discuss a simplification of this known as the spherical SK model, which has been found to nevertheless exhibit some of the same phenomena.

Take an i.i.d. symmetric matrix $\mathbf{J} = (J_{ij})_{ij}$ with law $\mathbb{P}_\mathbf{J}$. The spherical SK model has *Hamiltonian*

$$H(\mathbf{x}) = \sum_{1 \le i, j \le N} J_{ij} x_i x_j \quad \text{for} \quad \mathbf{x} \in \mathbb{S}^{N-1}(\sqrt{N}). \tag{1.17}$$

To avoid differential geometry on the sphere, it is sometimes preferable to extend the Hamiltonian to all $\mathbf{x} \in \mathbb{R}^N$ (note that the Hamiltonian is homogeneous so that dividing $\mathbf{x}$ by the Euclidean norm $\|\mathbf{x}\|/\sqrt{N}$ gives the same process on $\mathbb{S}^{N-1}(\sqrt{N})$). Instead of adding a non-linear confining force as is done in, e.g., [2], we either add a linear confining force $F_K(x) = Kx$, or have no confinement ($K = 0$) (the linearity of the system ensures no finite time blowup). Consider now the Langevin dynamics at *inverse temperature* $\beta > 0$ for the Hamiltonian of (1.17), corresponding to $\mathbf{X}_t = \mathbf{X}_t^{(\beta)}$ solving the SDS

$$\begin{cases} d\mathbf{X}_t & = -\nabla H(\mathbf{X}_t)dt - F_K'(\|\mathbf{X}_t\|^2/N)\mathbf{X}_t dt + \beta^{-1/2}d\mathbf{B}_t \\ \mathbf{X}_0 & \sim \mu \end{cases}. \tag{1.18}$$

We also consider the *gradient flow* where we take $\beta = \infty$, so that the Brownian motion term drops out: $\mathbf{X}_t$ is then the (deterministic) dynamical system following the (random) gradient vector field of $H(\mathbf{x}) + F_K(\|\mathbf{x}\|^2/N)$. The following universality for the above system is an immediate corollary of Theorem 3.

Corollary 1.3 *Fix $\beta \in (0, \infty]$ and consider the SDS's $\mathbf{X}_t$ and $\tilde{\mathbf{X}}_t$ given by (1.18) for $\mathbf{A}$ and $\tilde{\mathbf{A}}$ having mean zero, matching variance profiles $m_{ij} = \mathbf{1}\{i \neq j\}$. Suppose μ is independent of $\mathbb{P}_{\mathbf{A}}, \mathbb{P}_{\tilde{\mathbf{A}}}$ and these satisfy Hypotheses 1–2. Then for F as in (1.10) with $\mathcal{Y}, \mathcal{Y}' \in \mathfrak{F}$ and $\|\mathbf{a}\|_\infty \leq C_{\mathbf{a}}$, for every $T < \infty$,*

$$\sup_{s,t \in [0,T]} \left| F(\mathbf{X}_s, \mathbf{X}_t) - F(\tilde{\mathbf{X}}_s, \tilde{\mathbf{X}}_t) \right| \to 0 \quad \text{almost surely, and in } L^q \text{ for } q \geq 1 \,.$$

As shown in [14] and rigorously proved in [2], when $\mathbf{J}$ is Gaussian, the spherical SK model, or the soft spherical SK Model with confining potential F satisfying $F(x)/x \to \infty$ as $x \to \infty$, exhibits a sharp *aging* transition. Informally, aging is defined as the notion that the older a system gets, the more it remembers its past; formally, it corresponds to a transition in the behavior of the auto-correlation,

$$C_N(s, t) := \frac{1}{N} \sum_{i \leq N} X_i(s) X_i(t) \,,$$

between a (FDT) regime where $C_N(s, t) \sim \Phi(t - s)$ and an aging regime where $C_N(s, t) \sim \Phi(\frac{t}{s})$ for large s, t. In [2], it was established that for $\mathbf{J}$ having rotationally invariant law, e.g., a GOE matrix, $C_N(s, t)$ solves a non-linear equation [2, Eq. (2.16)], which exhibits exactly this type of transition at some β_{ag}. Our results allow us to read off universality for this limiting behavior, as formalized in the following corollary.

Corollary 1.4 *Consider the Langevin dynamics for the soft spherical* SK *model, as defined in (1.18) where $\mathbb{P}_{\mathbf{A}}$ is a Wigner matrix satisfying* Hypothesis 2*, the confinement is $F_K(x) = Kx$ for some $K > \mathbb{E}[\|\mathbf{J}\|_{2 \to 2}]$, and the initialization μ is e.g., standard Gaussian, independent of $\mathbb{P}_{\mathbf{A}}$. Then, for every $\beta \in (0, \infty]$ and every $T < \infty$, the limit $(\lim_{N \to \infty} C_N(s, t))_{s,t \in [0,T]}$ exists, and satisfies [2, Eq. (2.16)].*

In the specific case of $\beta = \infty$, the conclusions of [2, §3.2.2] apply, and the solution exhibits aging: i.e., there is a $\gamma > 0$ (specified therein) such that for every $\lambda > 1$,

$$\lim_{s \to \infty} \lim_{N \to \infty} \frac{C_N(s, \lambda s)}{\sqrt{C_N(s, s) C_N(\lambda s, \lambda s)}} \approx (\lambda - 1)^{-\gamma} \,.$$

Proof For the first statement, while [2, Theorem 2.6] is stated for confinement F growing super-linearly, following the proof one sees that it is only used to localize the process, for which it suffices for K to exceed $\|\mathbf{J}\|_{2 \to 2}$ (which for Wigner matrices is a.s. less than $2 + \epsilon$ for any $\epsilon > 0$). The first part of the corollary therefore follows from Corollary 1.3 together with the result of [2, Theorem 2.6] showing that for $\mathbf{A}$ standard normal, $C_N(s, t)$ satisfies [2, (2.16)].

For concreteness, the analysis of the limiting equation [2, (2.16)] and the derivation of the aging transition is carried out in [2] only for a specific choice of quadratic F. One could in principle perform the same analyses with other choices of F including $F = F_K$ that is linear, corresponding to the case we consider, and understand the limiting behavior of $C_N(s, t)$ as $N \to \infty$ then $s, t \to \infty$ as β varies. We do not pursue this, and instead notice that in the specific case of $\beta = \infty$, the homogeneity allows us

to disregard the choice of the confining potential and obtain universality for the zero-temperature aging behavior. To see this, since $H(x)$ is a homogeneous polynomial, if $\beta = \infty$, we see that $d\mathbf{X}_t$ is a constant multiple (for a constant depending only on $\|\mathbf{X}_t\|$) of $d(\mathbf{X}_t/\|\mathbf{X}_t\|)$. Therefore, at $\beta = \infty$, the projection of the dynamics (1.18) onto the sphere $\mathbb{S}^{N-1}(\sqrt{N})$ matches the projection of the Langevin SDS of [2], regardless of the choice of confining potential used therein. We apply Corollary 1.3 first to deduce that $\lim_{s\to\infty}\lim_{N\to\infty} C_N(s,s) =: C_\infty$ is the same for Gaussian and non-Gaussian $\mathbb{P}_\mathbf{A}$. Then applying it to $C_N(s,\lambda s)$, we find that the $N \to \infty$ limit of the normalized auto-correlation is the same for Gaussian and non-Gaussian $\mathbb{P}_\mathbf{A}$, and it is further independent of the choice of confining potential: as such for any $\mathbb{P}_\mathbf{A}$, it has the same $N \to \infty$ limit as in [2]. □

Remark 1.5 It would be of interest to consider similar Langevin dynamics for the spherical or soft spherical p-spin glass models for $p > 2$. Permitting higher order interactions gives rise to a wealth of more complicated models and different behavior. At the level of the off-equilibrium Langevin dynamics, these lead to the famous Cugliandolo–Kurchan/Crisanti–Horner–Sommers limit of coupled integro-differential equations for $C_N(s,t)$ and an integrated response $\chi_N(s,t) = \frac{1}{N}\sum_i X_i(s)B_i(t)$ [3,11,12,15,16,18,23], as well as the evolution of other observables e.g., the Hamiltonian and its square gradient [5]. Our combinatorial framework suggests that the differences in expectations (over p-tensors $\mathbf{J}$ and $\tilde{\mathbf{J}}$) of averaged observables are microscopic, as long as there is a non-linear confining potential to prevent finite-time blowup. The complication is in the fact that the two non-linearities (from the interactions, and the confining potential) cancel out, but these cancellations are not easily seen in the Taylor series obtained by expanding in powers of the generator; thus we are not able to show that this series is absolutely summable and exchange the infinite sum with its expectation.

1.4.2 Symmetric and asymmetric Hopfield networks

Let us also mention a different context in which diffusions of the form of (1.1) appear. Hopfield networks were introduced by [26] and have become one of the simplest and most fundamental examples of neural networks. In this model, a set of N neurons $(X_i)_i$ are either active $\{+1\}$ or inactive $\{-1\}$ depending on whether the neuron X_j's input $\sum J_{ij} X_i$, for some weights $\mathbf{J} = (J_{ij})_{i,j}$, exceeds a deterministic threshold h_i. This model was introduced in the symmetric setting, but has since been analyzed extensively both in symmetric and asymmetric setups [13,25,41].

One typically initializes the neurons at some pre-determined state independent of $\mathbf{J}$, e.g., all inactive/active, or uniformly at random, and tracks their time-evolution, whereby each neuron activates/de-activates at some rate, depending on the relationship between its input and threshold. Though there are many ways this is implemented, one is to soften the problem to continuous state space, either to the sphere, or to full-space and add in stochasticity by running some Langevin dynamics. This is the approach pursued in [13] as well as e.g., [41]. Then, with a linear confining force, our results imply universality for both for the symmetric and asymmetric Langevin dynamics (and

gradient flow) of general Hopfield networks: this includes universality for observables capturing the energy/loss in the network, its square gradient, and its "memory".

1.4.3 Rayleigh quotient minimization for random matrices

We conclude with a related optimization problem in high dimensions: that of optimizing the Rayleigh quotient of a random matrix $\mathbf{J}$ with a certain mean and variance profile. Maximizing the Rayleigh quotient is an efficient way to find the top eigenvector and eigenvalue of the random matrix via local iteration, e.g., either gradient descent or Langevin dynamics at low temperatures (large β). To place this in the framework of (1.2), take $H(\mathbf{x}) = \langle \mathbf{x}, \mathbf{Jx} \rangle$ and either no confining force or $F'_K = K$ for some $K > \|\mathbf{J}\|_{2\to 2}$ in (1.18). In the situation where the matrix ensemble is rotationally invariant, e.g., the GOE, the limiting trajectories of, say, $H(\mathbf{X}_t)$ for the gradient flow/Langevin dynamics can be explicitly solved (by diagonalization). Corollary 3 implies these limiting trajectories will be universal, and thus, match the limiting trajectories obtained when $\mathbf{J}$ is not Gaussian. In [1,10], similar universality results were described for an AMP approach to finding the top eigenvalue/eigenvector of $\mathbf{J}$.

2 Universality of expectations of monomial observables

In this section, we prove that two solutions $\mathbf{X}$ and $\tilde{\mathbf{X}}$ of (1.2) driven by $\mathbf{J}$ and $\tilde{\mathbf{J}}$ are such that expectations of observables of the form (1.10) are universal, as long as $\mathbf{A}$ and $\tilde{\mathbf{A}}$ have the same variance profiles. As discussed in Sect. 1.3, we reduce differences in expectations to combinatorial calculations by expanding the Markov transition semigroup of the process $\mathbf{X}_t$ in terms of its generator, an approach for proving universality in randomly driven dynamical systems which is the key contribution of this paper.

For the entirety of this paper, we will take two distributions $\mathbb{P}_{\mathbf{A}}$ and $\mathbb{P}_{\tilde{\mathbf{A}}}$ on $\mathbf{A}$ and $\tilde{\mathbf{A}}$ that are mean zero and have the same, uniformly bounded, variance profiles $\mathbf{m} = \tilde{\mathbf{m}}$. Recall that $\mathbb{P}_{\mathbf{A}}$ and $\mathbb{P}_{\tilde{\mathbf{A}}}$ are either fully independent or symmetric ensembles. For conciseness, we present our results in the case of fully independent (in particular, not symmetric). The case where they are symmetric is handled mutatis mutandis and only induces a few constant factors in certain estimates (see Remark 2.8 for more on these minimal modifications).

2.1 Main result on difference in expectations

The observables in Theorem 1 are composed of polynomials in $\mathbf{J}$ and $\mathbf{X}$, as well as $\mathbf{M}$. We first establish the universality of expectations for general monomials in $\mathbf{J}$ and $\mathbf{X}$ via a combinatorial moment matching type of argument. In Sect. 3 such universality is reduced for monomials that additionally involve the martingale, to that of monomials only in $\mathbf{J}$ and $\mathbf{X}$.

More precisely, the statistics we consider throughout this section are of the following form. Fix any s (not necessarily distinct) pairs $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_s)$ where each $\alpha_k = i_k j_k$, and r-tuple (not necessarily distinct) $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)$ where each $\gamma_i \in \{1, \ldots, N\}$.

Then consider observables $f_{\alpha,\gamma}(\mathbf{x})$ of the form

$$f_{\alpha,\gamma}(\mathbf{x}) = \prod_{k=1}^{s} J_{\alpha_k} \prod_{l=1}^{r} x_{\gamma_l} . \tag{2.1}$$

For an s-tuple of pairs α, let

- I_α count the number of distinct pairs in α, i.e., $I_\alpha = |\{\alpha_1, \ldots, \alpha_s\}|$,
- $I_{\alpha,1}$ count the number of $(\alpha_k)_k$ which appear exactly once in α, and
- $I_{\alpha,1}^+$ equal $I_{\alpha,1}$ plus the indicator that no pair appears more than twice in α.

Our bound on the distance between the expectations of $f_{\alpha,\gamma}(\mathbf{X}_t)$ and $f_{\alpha,\gamma}(\tilde{\mathbf{X}}_t)$ depends on α, γ and the laws μ, $\mathbb{P}_\mathbf{A}$, $\mathbb{P}_{\tilde{\mathbf{A}}}$ only through $\mathbf{C}_\star$, C_μ, s, r and $I_{\alpha,1}^+$. More precisely, we derive here the following.

Proposition 2.1 *There exists $C = C(r, s, T, \mathbf{C}_\star, C_\mu(r))$ such that for every $T, r, s \geq 0$, every s-tuple of pairs α and every r-tuple γ, if $\mathbb{P}_\mathbf{A}$, $\mathbb{P}_{\tilde{\mathbf{A}}}$ and μ satisfy* Hypothesis 1, *then*

$$\sup_{t\in[0,T]} \left| \mathbb{E}[f_{\alpha,\gamma}(\mathbf{X}_t)] - \tilde{\mathbb{E}}[f_{\alpha,\gamma}(\tilde{\mathbf{X}}_t)] \right| \leq C N^{-(s+I_{\alpha,1}^+)/2} .$$

Observe that in the case $s = 0$, the right-hand side is $CN^{-1/2}$.

Remark 2.2 The above theorem shows that having more distinct J's in the observable, decreases the difference in expectations by more than $N^{-s/2}$ as would be expected from the typical size of J_{ij}. This should be expected due to CLT-type cancellations: one way to motivate this scaling is by recalling averaged statistics which have J in them, in the context of the spherical SK model, e.g., the most relevant being

$$\frac{H(\mathbf{x})}{N} = \frac{1}{N^{3/2}} \sum_{1\leq i,j\leq N} A_{ij} x_i x_j \quad \text{and}$$

$$\frac{|\nabla H(\mathbf{x})|^2}{N} = \frac{1}{N} \sum_{1\leq i\leq N} G_i^2(\mathbf{x}) = \frac{1}{N} \sum_{1\leq i\leq N} \left(\sum_{1\leq j\leq N} \frac{1}{\sqrt{N}} A_{ij} x_j \right)^2 .$$

(Notice that these statistics are *not* rescaled by the number of order-one sized monomials; but they remain on the $O(1)$ scale due to additional cancellations from (J_{ij})). This gain in the scaling has to be visible at the level of the difference in expectations under $\mathbb{P}$ and $\tilde{\mathbb{P}}$ in order to hope for universality for such statistics.

Recall from Sect. 1.3 that our high level strategy is to reduce the expectations of statistics of the solution $\mathbf{X}_t$ of the SDS to combinatorial calculations in terms of mixed moments of $\mathbf{J}$ and $\mathbf{X}_0$. This is possible by writing $\mathbb{E}_\mathbf{B}[f(\mathbf{X}_t)]$ as $P_t f(\mathbf{X}_0)$ and then Taylor expanding $P_t = e^{tL}$ where L is the generator for the process $\mathbf{X}_t$ as defined in (1.13). In order for this expansion to be valid, and therefore our approach to be permissible, we need the Taylor expansion for e^{tL} to converge absolutely, for

each fixed N. In the next sub-section, we show that indeed with μ, $\mathbb{P}_\mathbf{A}$, $\mathbb{P}_{\tilde{\mathbf{A}}}$ satisfying Hypothesis 1, for each fixed N, the infinite series corresponding to $P_t f$ converges absolutely, so we can follow this plan.

Before proceeding further, we make the following notational remark.

Notational comment on set and sequence differences For sets $\{b_1, \ldots, b_m\} \subset \{a_1, \ldots, a_n\}$, we let $\{a_1, \ldots, a_n\} \setminus \{b_1, \ldots, b_m\}$ denote the set difference as usual. Frequently we deal with tuples, or sequences in which the order does not matter. For two such tuples $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_m)$ (where of course there may be repetitions in each sequence), we denote by $(a_1, \ldots, a_n) \setminus (b_1, \ldots, b_m)$ the difference wherein for each b_i appearing in $\{a_1, \ldots, a_n\}$ we only remove one of its appearances— say the first one—from $(a_1, \ldots, a_n)$. We also define $(a_1, \ldots, a_n) \sqcup (b_1, \ldots, b_m)$ to be the concatenation given by $(a_1, \ldots, a_n, b_1, \ldots, b_m)$.

2.2 Switching the expectation and the infinite series

The goal of this sub-section is to prove the following absolute convergence result.

Proposition 2.3 *Suppose $\mathbb{P}_\mathbf{A}$ and μ satisfy* Hypothesis 1. *Then, there exists finite $N_o = N_o(r, T, \mathbf{C}_\star)$ such that for every $N \geq N_o$, every $T < \infty$, every s-tuple of pairs $\boldsymbol{\alpha}$, and every r-tuple of indices $\boldsymbol{\gamma}$, we have*

$$\sum_{k \geq 0} \frac{T^k}{k!} \mathbb{E}\big[\big|L^k f_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}(\mathbf{X}_0)\big|\big] < \infty .$$

As a consequence of Proposition 2.3 and Fubini–Tonelli, we may use the following expansion.

Corollary 2.4 *Suppose $\mathbb{P}_\mathbf{A}$, $\mathbb{P}_{\tilde{\mathbf{A}}}$, μ satisfy* Hypothesis 1. *Setting L and $\tilde{L}$ for their generators, we have that*

$$\mathbb{E}[f_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}(\mathbf{X}_t)] - \tilde{\mathbb{E}}[f_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}(\tilde{\mathbf{X}}_t)] = \sum_{k \geq 0} \frac{t^k}{k!} \Big(\mathbb{E}[L^k f_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}(\mathbf{X}_0)] - \tilde{\mathbb{E}}[\tilde{L}^k f_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}(\mathbf{X}_0)] \Big) ,$$

for every $N \geq N_o(r, T, \mathbf{C}_\star)$, every $t < \infty$, and every s-tuple of pairs $\boldsymbol{\alpha}$ and r-tuple of indices $\boldsymbol{\gamma}$.

Proceeding hereafter to prove Proposition 2.3, we fix $r, s, \boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, and set $f = f_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}$. Aiming for upper bounds on $\mathbb{E}[|L^k f(\mathbf{X}_0)|]$ which are summable against $T^k/k!$, we first utilize (1.13) to expand L^k as a sum over the 4^k words W in the letters $\{L_\mathbf{J}, L_\mathbf{\Lambda}, L_\mathbf{h}, L_\Delta\}$ and thereby get the bound

$$\mathbb{E}\big[|L^k f(\mathbf{X}_0)|\big] \leq 4^k \sup_{W \in \{L_\mathbf{J}, L_\mathbf{\Lambda}, L_\mathbf{h}, L_\Delta\}^k} \mathbb{E}\big[|W f(\mathbf{X}_0)|\big], \qquad (2.2)$$

where for every $\mathbf{x} \in \mathbb{R}^N$, $W f(\mathbf{x})$ should be understood as $(W_k \cdots W_2 W_1 f)(\mathbf{x})$. For every word $W \in \{L_\mathbf{J}, L_\mathbf{\Lambda}, L_\mathbf{h}, L_\Delta\}^k$, let $k_\mathbf{J} = k_\mathbf{J}(W)$ denote the number of $L_\mathbf{J}$'s that

appear in W, and similarly define k_Λ, $k_\mathbf{h}$, and k_Δ, so that $k_\mathbf{J} + k_\Lambda + k_\mathbf{h} + k_\Delta = k$ and the following structural decomposition of Wf holds.

Claim For any word $W \in \{L_\mathbf{J}, L_\Lambda, L_\mathbf{h}, L_\Delta\}^k$ with $k_\mathbf{J}, k_\Lambda, k_\mathbf{h}, k_\Delta$ occurrences of the corresponding symbols, Wf can be expressed as a sum of (not necessarily distinct) monomials of the form

$$\phi_{\boldsymbol{\beta}, \boldsymbol{\beta}', \boldsymbol{\zeta}', \boldsymbol{\zeta}, \boldsymbol{\xi}}(\mathbf{x}) = \prod_{i=1}^{s} J_{\alpha_i} \prod_{\ell=1}^{k_\mathbf{J}} J_{\beta_\ell} \prod_{\ell=1}^{k_\Lambda} \Lambda_{\beta'_\ell} \prod_{\ell=1}^{k_\mathbf{h}} h_{\zeta'_\ell} \prod_{\ell=1}^{2k_\Delta} \sigma_{\zeta_\ell} \prod_{\ell=1}^{r} x_{\xi_\ell}, \tag{2.3}$$

$\boldsymbol{\beta}, \boldsymbol{\beta}', \boldsymbol{\zeta}$ denote the collection of pairs $(\beta_\ell)_{\ell \leq k_\mathbf{J}}, (\beta'_\ell)_{\ell \leq k_\Lambda}, (\zeta_\ell)_{\ell \leq 2k_\Delta}$, while $\boldsymbol{\zeta}', \boldsymbol{\xi}$ denote the sequences $(\zeta'_\ell)_{\ell \leq k_\mathbf{h}}, (\xi_\ell)_{\ell \leq r}$ and hereupon we adopt the convention $x_0 \equiv 1$, allowing for $\xi_\ell = 0$ as well as $\zeta_\ell \in (0j)_j$.

In view of Hypothesis 1 on $\mathbb{P}_\mathbf{A}$ we have that for every N, $\ell \geq 0$, and index pair α,

$$\mathbb{E}[|J_\alpha|^{\ell+1}] \leq \ell! \left(\frac{C_\mathbf{A}}{N}\right)^{(\ell+1)/2}.$$

Thus, if $I_{\boldsymbol{\alpha} \sqcup \boldsymbol{\beta}}$ distinct index pairs appear at multiplicities $(n_\ell + 1)_{\ell \leq I_{\boldsymbol{\alpha} \sqcup \boldsymbol{\beta}}}$ in the sequence $\boldsymbol{\alpha} \sqcup \boldsymbol{\beta}$ of length $k_\mathbf{J} + s$, then by the independence of $(J_\alpha)_\alpha$,

$$\mathbb{E}\left[\left|\prod_{i=1}^{s} J_{\alpha_i} \prod_{\ell=1}^{k_\mathbf{J}} J_{\beta_\ell}\right|\right] \leq \left(\frac{C_\mathbf{A}}{N}\right)^{(k_\mathbf{J}+s)/2} \prod_{\ell=1}^{I_{\boldsymbol{\alpha} \sqcup \boldsymbol{\beta}}} n_\ell!.$$

Consequently, with $\mathbf{X}_0$ independent of $\mathbf{J}$ we have in view of the assumed bounds on $(\Lambda_{ij})_{i,j} (\sigma_{ij})_{i,j}$ and $(h_i)_i$, that for any term of the form (2.3) with $I_{\boldsymbol{\zeta}}$ entries such that $\zeta_\ell \notin (0j)_j$,

$$\mathbb{E}\left[|\phi_{\boldsymbol{\beta}, \boldsymbol{\beta}', \boldsymbol{\zeta}', \boldsymbol{\zeta}, \boldsymbol{\xi}}(\mathbf{X}_0)|\right] \leq \left(\frac{C_\mathbf{A}}{N}\right)^{(k_\mathbf{J}+s)/2} \left(\frac{C_\Lambda}{\mathcal{N}_\Lambda}\right)^{k_\Lambda} \left(\frac{C_\sigma^{2k_\Delta}}{\mathcal{N}_\sigma^{I_\zeta}}\right) C_\mathbf{h}^{k_\mathbf{h}} \sup_i\{\mathbb{E}[|X_i(0)|^r]\} \prod_{\ell=1}^{I_{\boldsymbol{\alpha} \sqcup \boldsymbol{\beta}}} n_\ell!$$

$$\leq C_\mu(r) \mathbf{C}_\star^s \frac{\mathbf{C}_\star^k}{N^{(k_\mathbf{J}+s)/2} \mathcal{N}_\Lambda^{k_\Lambda} \mathcal{N}_\sigma^{I_\zeta}} \prod_{\ell=1}^{I_{\boldsymbol{\alpha} \sqcup \boldsymbol{\beta}}} n_\ell!, \tag{2.4}$$

using in the last inequality also (1.6) from Hypothesis 1 on μ, and the definition of $\mathbf{C}_\star$.

Our next result is a first step in controlling the number of monomial terms that can appear in the expansion of each word $W \in \{L_\mathbf{J}, L_\Lambda, L_\mathbf{h}, L_\Delta\}^k$.

Lemma 2.5 *For every $k_\mathbf{J}, k_\Lambda, k_\mathbf{h}, k_\Delta$ and every $\boldsymbol{\beta}, \boldsymbol{\beta}', \boldsymbol{\zeta}', \boldsymbol{\zeta}, \boldsymbol{\xi}$, if we let $\phi = \phi_{\boldsymbol{\beta}, \boldsymbol{\beta}', \boldsymbol{\zeta}', \boldsymbol{\zeta}, \boldsymbol{\xi}}$ be as in (2.3), then $L_\mathbf{h}\phi, L_\mathbf{J}\phi, L_\Lambda\phi$ and $L_\Delta\phi$ can each be expressed as a sum of at most $r, rN, r\mathcal{N}_\Lambda$ and $r\mathcal{N}_\sigma^2$ many such monomials, respectively, each of the same form (with possibly different $\boldsymbol{\beta}, \boldsymbol{\beta}', \boldsymbol{\zeta}', \boldsymbol{\zeta}, \boldsymbol{\xi}$) as (2.3), with the respective $k_\mathbf{J}, k_\Lambda, k_\mathbf{h}$ or k_Δ increased by one.*

Proof Fixing k_J, k_Λ, k_h, k_Δ which sum up to k, we proceed by separately considering the effect each of $L_h\phi$, $L_J\phi$, $L_\Lambda\phi$ and $L_\Delta\phi$ has on the monomial ϕ. First,

$$(L_h\phi)(\mathbf{x}) = \prod_{\ell=1}^{s} J_{\alpha_\ell} \prod_{\ell=1}^{k_J} J_{\beta_\ell} \prod_{\ell=1}^{k_\Lambda} \Lambda_{\beta'_\ell} \prod_{\ell=1}^{k_h} h_{\zeta'_\ell} \prod_{\ell=1}^{2k_\Delta} \sigma_{\zeta_\ell} \sum_{j=1}^{N} h_j \partial_j \Big(\prod_{\ell=1}^{r} x_{\xi_\ell} \Big), \quad (2.5)$$

with non-zero contribution only from $j \in \boldsymbol{\xi}$, yielding at most r non-zero terms. To each of these corresponds a monomial of the form of (2.3), for $k_h \mapsto k_{h+1}$, $\boldsymbol{\zeta}' \mapsto \boldsymbol{\zeta}' \amalg (j)$ and $\boldsymbol{\xi} \mapsto (\boldsymbol{\xi} \setminus (j)) \amalg (0)$. Next,

$$(L_J\phi)(\mathbf{x}) = \prod_{\ell=1}^{s} J_{\alpha_\ell} \prod_{\ell=1}^{k_J} J_{\beta_\ell} \prod_{\ell=1}^{k_\Lambda} \Lambda_{\beta'_\ell} \prod_{\ell=1}^{k_h} h_{\zeta'_\ell} \prod_{\ell=1}^{2k_\Delta} \sigma_{\zeta_\ell} \sum_{i,j=1}^{N} J_{ij} x_i \partial_j \Big(\prod_{\ell=1}^{r} x_{\xi_\ell} \Big), \quad (2.6)$$

with non-zero contribution only when $j \in \boldsymbol{\xi}$. With $i \leq N$ the total number of resulting non-zero monomials is now at most rN, each having the stated form with $k_J \mapsto k_J+1$, $\boldsymbol{\beta} \mapsto \boldsymbol{\beta} \amalg (ij)$ and $\boldsymbol{\xi} \mapsto (\boldsymbol{\xi} \setminus (j)) \amalg (i)$. Likewise, we have that

$$(L_\Lambda\phi)(\mathbf{x}) = \prod_{\ell=1}^{s} J_{\alpha_\ell} \prod_{\ell=1}^{k_J} J_{\beta_\ell} \prod_{\ell=1}^{k_\Lambda} \Lambda_{\beta'_\ell} \prod_{\ell=1}^{k_h} h_{\zeta'_\ell} \prod_{\ell=1}^{2k_\Delta} \sigma_{\zeta_\ell} \sum_{i,j=1}^{N} \Lambda_{ij} x_i \partial_j \Big(\prod_{\ell=1}^{r} x_{\xi_\ell} \Big), \quad (2.7)$$

with non-zero contributions only for $j \in \boldsymbol{\xi}$. Enumerating over $i \leq N$, gives now at most $r\mathcal{N}_\Lambda$ non-zero monomials, of the stated form, with $k_\Lambda \mapsto k_\Lambda+1$, $\boldsymbol{\beta}' \mapsto \boldsymbol{\beta}' \amalg (ij)$ and $\boldsymbol{\xi} \mapsto (\boldsymbol{\xi} \setminus (j)) \amalg (i)$. Finally,

$$(L_\Delta\phi)(\mathbf{x}) = \prod_{\ell=1}^{s} J_{\alpha_\ell} \prod_{\ell=1}^{k_J} J_{\beta_\ell} \prod_{\ell=1}^{k_\Lambda} \Lambda_{\beta'_\ell} \prod_{\ell=1}^{k_h} h_{\zeta'_\ell} \prod_{\ell=1}^{2k_\Delta} \sigma_{\zeta_\ell} \sum_{j=1}^{N}$$
$$\Big(\sum_{i,i'=0}^{N} \sigma_{ij}\sigma_{i'j} x_i x_{i'} \Big) \partial_j \partial_j \Big(\prod_{\ell=1}^{r} x_{\xi_\ell} \Big), \quad (2.8)$$

is non-zero only for the summands in which $j \in \boldsymbol{\xi}$. Enumerating over $0 \leq i, i' \leq N$ (recalling the convention that $x_0 \equiv 1$), gives at most $r\mathcal{N}_\sigma^2$ non-zero monomials, of the stated form, with $k_\Delta \mapsto k_\Delta + 1$, $\boldsymbol{\zeta} \mapsto \boldsymbol{\zeta} \amalg (ij) \amalg (i'j)$ and $\boldsymbol{\xi} \mapsto (\boldsymbol{\xi} \setminus (j,j)) \amalg (i,i')$. $\square$

Fixing N, k, an s-tuple of pairs $\boldsymbol{\alpha}$, an r-tuple of indices $\boldsymbol{\gamma}$ and $W \in \{L_J, L_\Lambda, L_h, L_\Delta\}^k$, upon inductively applying Lemma 2.5, we are able to express Wf as the sum of at most

$$r^k N^{k_J} \mathcal{N}_\Lambda^{k_\Lambda} \mathcal{N}_\sigma^{2k_\Delta}, \quad (2.9)$$

many non-zero monomials of the form of (2.3). Recall that for a monomial ϕ, we use I_ζ for the number of $\zeta_\ell \notin (0j)_j$, I_α for the number of distinct pairs in $\boldsymbol{\alpha}$, $I_{\alpha \amalg \beta}$ for the

number of distinct pairs in $\boldsymbol{\alpha} \sqcup \boldsymbol{\beta}$, and introduce $I_\star = I_{\alpha \sqcup \beta} - I_\alpha$, which counts the number of distinct pairs in $\{\boldsymbol{\beta}\} \backslash \{\boldsymbol{\alpha}\}$. A careful examination of the proof of Lemma 2.5, yields the following significant refinement upon the crude bound of (2.9).

Proposition 2.6 *Fix $N, r, s, k \geq 0$, an s-tuple of pairs $\boldsymbol{\alpha}$, an r-tuple of indices $\boldsymbol{\gamma}$, and a word $W \in \{L_J, L_\Lambda, L_h, L_\triangle\}^k$. Then, of the monomials in such expansion of Wf, at most*

$$\binom{k_J}{I_\star, n_1, \ldots, n_{I_{\alpha \sqcup \beta}}} \binom{2k_\triangle}{I_\zeta} r^k \, N^{I_\star} \mathcal{N}_\Lambda^{k_\Lambda} \mathcal{N}_\sigma^{I_\zeta} \tag{2.10}$$

have I_ζ elements of $\boldsymbol{\zeta}$ with $\zeta_\ell \notin (0j)_j$, and the $I_{\alpha \sqcup \beta} = I_\alpha + I_\star$ distinct pairs in $\boldsymbol{\alpha} \sqcup \boldsymbol{\beta}$ appear in multiplicities $\{n_\ell + \mathbf{1}_{\{\ell > I_\alpha\}}\}_{\ell \leq I_{\alpha \sqcup \beta}}$ within the sequence $\boldsymbol{\beta}$ of length k_J. (N.b. we ordered the (n_ℓ) with multiplicities in $\boldsymbol{\beta}$ of the distinct pairs of $\boldsymbol{\alpha}$ appearing first, and the multiplicities in $\boldsymbol{\beta}$ of the remaining $I_\star$ distinct pairs next.)

Proof The first improvement in (2.10) over (2.9) is from observing that the growth factor $\mathcal{N}_\sigma$ applies only in those I_ζ of the $2k_\triangle$ applications of $L_\triangle$ within W which have led to an element $\zeta_\ell \notin (0j)_j$ (see (2.8)), and that there are at most $\binom{2k_\triangle}{I_\zeta}$ ways to choose which I_ζ elements of $\boldsymbol{\zeta}$ are not from the 0-th row of $\boldsymbol{\sigma}$.

Similarly, the growth factor N in counting the number of monomials after applying L_J is only relevant during the $I_\star$ applications of L_J within W in which a new pair (ij) is selected (see (2.6)). The left-most term in (2.10) counts the number of ways to select the locations of these $I_\star$ new elements within the k_J long sequence $\boldsymbol{\beta}$, and thereafter to partition the remaining $k_J - I_\star$ consistently with having the prescribed $n_\ell \geq 0$ repeats for each of the $I_{\alpha \sqcup \beta}$ distinct pairs in question. Putting all this together yields the stated bound (2.10) on the number of relevant monomials in the expansion of Wf. $\square$

Proof of Proposition 2.3. Combining Proposition 2.6 with the bound (2.4) we deduce that for any word W of length k and any $\boldsymbol{\alpha}$ whose I_α distinct terms appear in multiplicities $(c_\ell)_{\ell \leq I_\alpha}$,

$$\mathbb{E}[|Wf(\mathbf{X}_0)|] \leq C_\mu(r) \mathbf{C}_\star^s \, k_J! \frac{(4r\mathbf{C}_\star)^k}{N^{(k_J+s)/2}} \sum_{I_\star=0}^{k_J} \frac{N^{I_\star}}{I_\star!}$$

$$\sum_{(n_\ell)_{\ell \leq I_{\alpha \sqcup \beta}}} \prod_{\ell=1}^{I_\alpha} \frac{(n_\ell + c_\ell - 1)!}{n_\ell!}, \tag{2.11}$$

where the inner sum is over all partitions of $k_J - I_\star$ into $I_{\alpha \sqcup \beta}$ indistinguishable integers $n_\ell \geq 0$. Since $\sum_\ell c_\ell = s$ and $n_\ell + c_\ell \leq k_J + s$ for all ℓ, the right-most product is at most $(k_J + s)^s$. Further, the number of $(n_\ell)_\ell$ considered here is at most the number of integer partitions of k_J, which grows slower than e^{k_J} (c.f. the Hardy-Ramanujan asymptotic partition formula [24]). Thus, we find that for $C(r, s, C_\mu, \mathbf{C}_\star)$ finite and

any word W of length k,

$$\mathbb{E}[|Wf(\mathbf{X}_0)|] \le \frac{C}{N^{s/2}} (4re\mathbf{C}_\star)^{k-k_\mathbf{J}} k_\mathbf{J}! (k_\mathbf{J} + s)^s \left(\frac{4re\mathbf{C}_\star}{\sqrt{N}}\right)^{k_\mathbf{J}} \sum_{I_\star=0}^{k_\mathbf{J}} \frac{N^{I_\star}}{I_\star!}. \quad (2.12)$$

Since $k! \ge k_\mathbf{J}!(k - k_\mathbf{J})!$, the bounds (2.12) and (2.2) will yield the stated absolute convergence of the infinite series. Specifically, fixing $T < \infty$ and setting $\delta = 1/(16Tre\mathbf{C}_\star)$, we have that

$$\sum_{k=0}^{\infty} \frac{T^k}{k!} \mathbb{E}[|L^k f(\mathbf{X}_0)|] \le \frac{C}{N^{s/2}} \sum_{k=0}^{\infty} \sum_{k_\mathbf{J} \le k} \frac{(4T)^k}{k!} (4re\mathbf{C}_\star)^{k-k_\mathbf{J}} k_\mathbf{J}! (k_\mathbf{J} + s)^s$$

$$\left(\frac{4re\mathbf{C}_\star}{\sqrt{N}}\right)^{k_\mathbf{J}} \sum_{I_\star=0}^{k_\mathbf{J}} \frac{N^{I_\star}}{I_\star!}$$

$$\le \frac{C}{N^{s/2}} \sum_{k'=k-k_\mathbf{J}=0}^{\infty} \frac{\delta^{-k'}}{k'!} \sum_{k_\mathbf{J}=0}^{\infty} (k_\mathbf{J} + s)^s \left(\delta\sqrt{N}\right)^{-k_\mathbf{J}} \sum_{I_\star=0}^{k_\mathbf{J}} \frac{N^{I_\star}}{I_\star!},$$

$$(2.13)$$

which is finite for any fixed $N > \delta^{-2}$, thereby concluding the proof. $\square$

2.3 Controlling the differences of the k'th order Taylor coefficients

By Corollary 2.4, we have that

$$\sup_{t \in [0,T]} \left| \mathbb{E}[f(\mathbf{X}_t)] - \tilde{\mathbb{E}}[f(\tilde{\mathbf{X}}_t)] \right|$$

$$\le \sup_{t \in [0,T]} \sum_{k \ge 0} \frac{t^k}{k!} \left| \mathbb{E}[L^k f(\mathbf{X}_0)] - \tilde{\mathbb{E}}[L^k f(\mathbf{X}_0)] \right|$$

$$\le \sum_{k \ge 0} \frac{(4T)^k}{k!} \sup_{W \in \{L_\mathbf{J}, L_\mathbf{\Lambda}, L_\mathbf{h}, L_\Delta\}^k} \left| \mathbb{E}[Wf(\mathbf{X}_0)] - \tilde{\mathbb{E}}[Wf(\mathbf{X}_0)] \right|$$

$$\le \sum_{k \ge 0} \frac{(4T)^k}{k!} \sup_{W \in \{L_\mathbf{J}, L_\mathbf{\Lambda}, L_\mathbf{h}, L_\Delta\}^k} \sum_{\phi \in (Wf)(\mathbf{x})} \left| \mathbb{E}[\phi(\mathbf{X}_0)] - \tilde{\mathbb{E}}[\phi(\mathbf{X}_0)] \right|, \quad (2.14)$$

where the last sum is over ϕ appearing in the monomial decomposition of $Wf(\mathbf{x})$ per Claim 2.2. To bound the differences of expectations on the RHS of (2.14), we next control the type of monomials ϕ of the form (2.3) in the expansion of Wf, for which we may possibly have $\mathbb{E}[\phi(\mathbf{X}_0)] \ne \tilde{\mathbb{E}}[\phi(\mathbf{X}_0)]$.

Lemma 2.7 *For any $k, s \ge 0$, every s-tuple of pairs $\boldsymbol{\alpha}$, and every $W \in \{L_\mathbf{J}, L_\mathbf{\Lambda}, L_\mathbf{h}, L_\Delta\}^k$, the monomials ϕ in the expansion of Wf in Claim rm 2.2 may have $\mathbb{E}[\phi(\mathbf{X}_0)] \ne$*

$\tilde{\mathbb{E}}[\phi(\mathbf{X}_0)]$ *only if*

$$k_{\mathbf{J}} + s \geq 3 \quad and \quad k_{\mathbf{J}} \geq 2I_\star + I_{\alpha,1}^+ , \tag{2.15}$$

where as before, $I_\star = I_{\alpha \sqcup \beta} - I_\alpha$ *denotes the number of distinct elements in* $\{\beta\} \setminus \{\alpha\}$.

Proof By the independence of $\mathbf{J}, \tilde{\mathbf{J}}$ and μ, if $\mathbb{E}[\phi(\mathbf{X}_0)] \neq \tilde{\mathbb{E}}[\phi(\mathbf{X}_0)]$ for some $\phi = \phi_{\beta,\beta',\zeta,\zeta',\xi}$ as in (2.3), then

$$\mathbb{E}\Big[\prod_{i=1}^{s} J_{\alpha_i} \prod_{\ell=1}^{k_{\mathbf{J}}} J_{\beta_\ell}\Big] \neq \tilde{\mathbb{E}}\Big[\prod_{i=1}^{s} J_{\alpha_i} \prod_{\ell=1}^{k_{\mathbf{J}}} J_{\beta_\ell}\Big],$$

which for independent, zero-mean $(J_{ij})_{ij}$ of matching variances $\frac{1}{N}\mathbf{m} = \frac{1}{N}\tilde{\mathbf{m}}$, requires that simultaneously:

No pair $\alpha_\star$ appears exactly once in the concatenation $\alpha \sqcup \beta$.   (2.16)

Some $\alpha_\star$ appears more than twice in the concatenation $\alpha \sqcup \beta$.   (2.17)

The condition (2.16) implies that each of the $I_\star$ distinct elements in $\{\beta\} \setminus \{\alpha\}$ must appear at least twice in $\{\beta\}$, to which end we need at least $2I_\star$ applications of $L_{\mathbf{J}}$ to select those elements. In addition, some other $I_{\alpha,1}$ of the $k_{\mathbf{J}}$ applications of $L_{\mathbf{J}}$ must align exactly with the pairs (α_{ij}) appearing only once in α, so necessarily $k_{\mathbf{J}} \geq 2I_\star + I_{\alpha,1}$. Further, the condition (2.17) requires $k_{\mathbf{J}} + s \geq 3$ and when no pair appears more than twice in α, an extra application of $L_{\mathbf{J}}$ beyond the preceding $2I_\star + I_{\alpha,1}$ is needed for producing the third appearance of some $\alpha_\star$, as stated in (2.15). $\square$

We are now able to prove that the expectations of monomials of the form $f_{\alpha,\gamma}(\mathbf{X}_t)$ are universal.

Proof of Proposition 2.1. Fixing α, γ, in view of Lemma 2.7, it suffices when bounding the RHS of (2.14), to consider only words W and monomials ϕ for which (2.15) holds. By restricting attention only to monomials for which (2.15), holds, we find as in (2.11), that for any α whose I_α distinct terms appear in multiplicities $(c_\ell)_{\ell \leq I_\alpha}$, and every word W of length k such that $k_{\mathbf{J}} + s \geq 3$,

$$\big|\mathbb{E}[Wf(\mathbf{X}_0)] - \tilde{\mathbb{E}}[Wf(\mathbf{X}_0)]\big| \leq 2C_\mu(r)\mathbf{C}_\star^s k_{\mathbf{J}}! \frac{(4r\mathbf{C}_\star)^k}{N^{(k_{\mathbf{J}}+s)/2}} \sum_{\{I_\star : k_{\mathbf{J}} \geq 2I_\star + I_{\alpha,1}^+\}} \frac{N^{I_\star}}{I_\star!}$$

$$\sum_{(n_\ell)_{\ell \leq I_{\alpha \sqcup \beta}}} \prod_{\ell=1}^{I_\alpha} \frac{(n_\ell + c_\ell - 1)!}{n_\ell!},$$

where as in (2.11), the inner sum runs over all partitions of $k_{\mathbf{J}} - I_\star$ into $I_{\alpha \sqcup \beta}$ indistinguishable integers $n_\ell \geq 0$. Reasoning as we did leading up to (2.12), we find that

$$\big|\mathbb{E}[Wf(\mathbf{X}_0)] - \tilde{\mathbb{E}}[Wf(\mathbf{X}_0)]\big| \leq \frac{2C}{N^{s/2}} (4re\mathbf{C}_\star)^{k-k_{\mathbf{J}}} k_{\mathbf{J}}! (k_{\mathbf{J}} + s)^s \Big(\frac{4re\mathbf{C}_\star}{\sqrt{N}}\Big)^{k_{\mathbf{J}}}$$

$$\sum_{\{I_\star : k_\mathbf{J} \geq 2I_\star + I^+_{\alpha,1}\}} \frac{N^{I_\star}}{I_\star!} . \tag{2.18}$$

Plugging (2.18) into (2.14), as in the derivation of (2.13), we get for $\delta = 1/(16 Tr e\mathbf{C}_\star)$ and $N \geq \rho := (2/\delta)^2$,

$$\sup_{t \in [0,T]} \left| \mathbb{E}[f(\mathbf{X}_t)] - \tilde{\mathbb{E}}[f(\tilde{\mathbf{X}}_t)] \right|$$

$$\leq \frac{2C}{N^{s/2}} \sum_{k' \geq 0} \frac{\delta^{-k'}}{k'!} \sum_{I_\star \geq 0} \frac{1}{I_\star!} \sum_{k_\mathbf{J} \geq 2I_\star + I^+_{\alpha,1}} (k_\mathbf{J} + s)^s \, \delta^{-k_\mathbf{J}} N^{I_\star - k_\mathbf{J}/2}$$

$$\leq \bar{C} N^{-(s+I^+_{\alpha,1})/2} \sum_{I_\star \geq 0} \frac{\rho^{I_\star}}{I_\star!} \sum_{k_\mathbf{J} \geq 0} (k_\mathbf{J} + s)^s 2^{-k_\mathbf{J}} , \tag{2.19}$$

where $\bar{C} = 2Ce^{-1/\delta} \rho^{I^+_{\alpha,1}/2}$. This completes the proof, as both series on the RHS of (2.19) are finite and independent of N. □

Remark 2.8 In the case of *symmetric* random matrices $\mathbf{A}$, $\tilde{\mathbf{A}}$ (where only the upper triangular and diagonal elements are independent), we identify index pairs $\beta = ij$ and $\tilde{\beta} = ji$ as being the same. We do so whenever considering I_α, $I_{\alpha,1}$, $I^+_{\alpha,1}$, $I_{\alpha \sqcup \beta}$, $I_\star$, and the multiplicities $(n_\ell)_\ell$, as well as in the restrictions (2.16)–(2.17) imposed on the multiplicities within $\boldsymbol{\alpha} \sqcup \boldsymbol{\beta}$. Once this is done, the only difference in our proof is to replace in (2.10) the weight r^k by $(2r)^k$.

3 The extension to multi-time polynomial observables

In this section, we extend the results of Sect. 2 to more general observables, namely those that contain coefficients that depend on the driving martingale, and those that depend on the trajectory through multiple times, rather than just one. We then use those extensions to prove Theorem 1. To this end, fix any l, any $(\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(l)})$ each consisting of s_i pairs, any $(\boldsymbol{\gamma}^{(1)}, \ldots, \boldsymbol{\gamma}^{(l)})$ each consisting of r_i indices, and also fix m indices $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)$. Fix l times $0 \leq t_1 \leq \cdots \leq t_l \leq T$ and m times $0 \leq u_1 \leq \cdots \leq u_m \leq T$. For $f_{\boldsymbol{\alpha}^{(i)}, \boldsymbol{\gamma}^{(i)}}$ as in (2.1), consider observables of the form,

$$g_{(\boldsymbol{\alpha}^{(i)}), (\boldsymbol{\gamma}^{(i)}), \boldsymbol{\xi}}(\mathbf{t}, \mathbf{u}) = \left(\prod_{i=1}^{l} f_{\boldsymbol{\alpha}^{(i)}, \boldsymbol{\gamma}^{(i)}}(\mathbf{X}_{t_i}) \right) \left(\prod_{i=1}^{m} M_{\xi_i}(u_i) \right) . \tag{3.1}$$

Let $\bar{r} = \sum_i r_i + m$ and $\bar{\boldsymbol{\alpha}}$ denote the concatenation $\boldsymbol{\alpha}^{(1)} \sqcup \cdots \sqcup \boldsymbol{\alpha}^{(l)}$ of length $\bar{s} := \sum_i s_i$.

Proposition 3.1 *There exist finite* $C(\bar{r}, \bar{s}, m, l, T, \mathbf{C}_\star, C_\mu(\bar{r}))$ *such that for every* l, m, *every* $(\boldsymbol{\alpha}^{(i)})_{i \leq l}$, $(\boldsymbol{\gamma}^{(i)})_{i \leq l}$, $\boldsymbol{\xi}$, *every* $\mathbf{t} \in [0, T]^l$, $\mathbf{u} \in [0, T]^m$ *and* $g(\mathbf{t}, \mathbf{u}) =$

$g_{(\boldsymbol{\alpha}^{(i)}),(\boldsymbol{\gamma}^{(i)}),\boldsymbol{\xi}}(\mathbf{t}, \mathbf{u})$ *as in* (3.1),

$$\left|\mathbb{E}[g(\mathbf{t}, \mathbf{u})] - \tilde{\mathbb{E}}[g(\mathbf{t}, \mathbf{u})]\right| \leq C N^{-(\bar{s}+I^+_{\bar{\alpha},1})/2}.$$

We proceed to prove Proposition 3.1, which we thereafter combine with a short combinatorial estimate bounding the number of terms with specific values of $I^+_{\bar{\alpha},1}$ to establish Theorem 1.

3.1 Proof of Proposition 3.1

We start with the case of $m = 0$ to which we will reduce the case of $m > 0$.

Lemma 3.2 *Proposition* 3.1 *holds when* $m = 0$.

Proof Fixing l, $(\boldsymbol{\alpha}^{(i)})_{i \leq l}$ and $(\boldsymbol{\gamma}^{(i)})_{i \leq l}$, we set here $f^{(i)}(\mathbf{x}) = f_{\boldsymbol{\alpha}^{(i)}, \boldsymbol{\gamma}^{(i)}}(\mathbf{x})$ and

$$g(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(l)}) := \prod_{i=1}^{l} f^{(i)}(\mathbf{x}^{(i)}) = \prod_{i=1}^{l} J_{\alpha_1^{(i)}} \cdots J_{\alpha_{r_i}^{(i)}} x_{\gamma_1^{(i)}}^{(i)} \cdots x_{\gamma_{s_i}^{(i)}}^{(i)}, \qquad (3.2)$$

and for any l-tuple of times $\mathbf{t} = (t_1, \ldots, t_l) \in [0, T]^l$, evaluate (3.2) on the argument $(\mathbf{X}_{t_1}, \ldots, \mathbf{X}_{t_l})$: i.e., let

$$g(\mathbf{t}) = g_{(\boldsymbol{\alpha}^{(i)}),(\boldsymbol{\gamma}^{(i)})}(\mathbf{t}) = g(\mathbf{X}_{t_1}, \ldots, \mathbf{X}_{t_l}).$$

We express the expectation $\mathbb{E}_{\mathbf{B}}$ with respect to the Brownian motion of $g(\mathbf{t})$, in terms of the (diffusion) semi-group operator as

$$\mathbb{E}_{\mathbf{B}}[g(\mathbf{t})] = \left(P_{t_1} f^{(1)} P_{t_2-t_1} f^{(2)} \cdots P_{t_l-t_{l-1}} f^{(l)}\right)(\mathbf{X}_0).$$

Expanding each semi-group operator in terms of powers of the generator L, the above is precisely

$$\sum_{k_1 \geq 0} \frac{t_1^{k_1}}{k_1!} L^{k_1} \left[f^{(1)} \sum_{k_2 \geq 0} \frac{(t_2-t_1)^{k_2}}{k_2!} L^{k_2} \left[f^{(2)} \cdots \sum_{k_l \geq 0} \frac{(t_l-t_{l-1})^{k_l}}{k_l!} L^{k_l} f^{(l)} \right] \right](\mathbf{X}_0)$$

$$= \sum_{k_1, \ldots, k_l \geq 0} \left(\prod_{i=1}^{l} \frac{(t_i-t_{i-1})^{k_i}}{k_i!} \right) [L^{k_1} f^{(1)} L^{k_2} f^{(2)} \cdots L^{k_l} f^{(l)}](\mathbf{X}_0).$$

Taking the difference in expectations between $\mathbb{E}$ and $\tilde{\mathbb{E}}$, upon justifying swapping the expectation with the infinite sum (as done in Sect. 2.2), and using the fact that

$$\binom{k}{k_1, \ldots, k_l} l^{-k} \leq \sum_{\substack{k_1, \ldots, k_l \geq 0 \\ \sum k_i = k}} \binom{k}{k_1, \ldots, k_l} l^{-k} = 1, \qquad (3.3)$$

for every $k_1, k_2, \ldots, k_l$ such that $k_1 + \cdots + k_l = k$, we obtain that

$$\left|\mathbb{E}[g(\mathbf{t})] - \tilde{\mathbb{E}}[g(\mathbf{t})]\right| \leq \sum_{\substack{k \geq 0 \\ \sum_i k_i = k}} \sum_{\substack{k_1, \ldots, k_l}} \frac{l^k T^k}{k!} \sum_{\substack{W_1, \ldots, W_l \\ W_i \in \{L_{\mathbf{J}}, L_{\mathbf{\Lambda}}, L_{\mathbf{h}}, L_{\Delta}\}^{k_i}}} \left|\mathbb{E}[(W_1 f^{(1)} \cdots W_l f^{(l)})(\mathbf{X}_0)] - \tilde{\mathbb{E}}[(W_1 f^{(1)} \cdots W_l f^{(l)})(\mathbf{X}_0)]\right|.$$

The following structural property for words appearing in the above will allow us to reduce the analysis of multi-time observables to the combinatorial analysis of one-time observables $f_{\bar{\alpha}, \bar{\gamma}} = f^{(1)} f^{(2)} \cdots f^{(l)}$, for $\bar{\alpha} = \alpha^{(1)} \amalg \cdots \amalg \alpha^{(l)}$ and $\bar{\gamma} := \gamma^{(1)} \amalg \cdots \amalg \gamma^{(l)}$, which we have already completed. □

Claim Fix $k_1, \ldots, k_l \geq 0$ such that $\sum_i k_i = k$ and words $W_i \in \{L_{\mathbf{J}}, L_{\mathbf{\Lambda}}, L_{\mathbf{h}}, L_{\Delta}\}^{k_i}$, $i = 1, \ldots, l$, with $k_{\mathbf{J}}^i, k_{\mathbf{\Lambda}}^i, k_{\mathbf{h}}^i, k_{\Delta}^i$, of each appearing, respectively. Then, the function

$$(W_1 f^{(1)} W_2 f^{(2)} \cdots W_l f^{(l)})(\mathbf{x})$$

consists of a sum of (not necessarily distinct) monomials of the form

$$\phi(\mathbf{x}) = \prod_{i=1}^{s_1} J_{\alpha_i^{(1)}} \cdots \prod_{i=1}^{s_l} J_{\alpha_i^{(l)}} \prod_{\ell=1}^{\sum k_{\mathbf{J}}^i} J_{\beta_\ell} \prod_{\ell=1}^{\sum k_{\mathbf{\Lambda}}^i} \Lambda_{\beta_\ell'} \prod_{\ell=1}^{\sum k_{\mathbf{h}}^i} h_{\zeta_\ell'} \prod_{\ell=1}^{2 \sum k_\Delta} \sigma_{\zeta_\ell} \prod_{\ell=1}^{\sum r_i} x_{\xi_\ell}.$$

Moreover, each monomial $\phi(\mathbf{x})$ appearing in this expansion, must also appear in such monomial expansion of $W f_{\bar{\alpha}, \bar{\gamma}}$ for $W = W_1 \cdots W_l \in \{L_{\mathbf{J}}, L_{\mathbf{\Lambda}}, L_{\mathbf{h}}, L_{\Delta}\}^k$.

Proof The structure of the monomials is evident. Every such monomial in $W_1 f^{(1)} W_2 f^{(2)} \cdots W_l f^{(l)}$ must also appear in the monomial expansion of $[W_1 \cdots W_l] f_{\bar{\alpha}, \bar{\gamma}}$ because a subset of the terms in the latter are obtained by applying the letters in W_l to $f^{(l)}$, then the letters in W_{l-1} to $f^{(l-1)}(W_l f^{(l)})$, and so on. Finally, observe that $W_1 \cdots W_l$ is always a word in $\{L_{\mathbf{J}}, L_{\mathbf{\Lambda}}, L_{\mathbf{h}}, L_{\Delta}\}^k$. □

With Claim 3.1 in hand, we further get that

$$|\mathbb{E}[g(\mathbf{t})] - \tilde{\mathbb{E}}[g(\mathbf{t})]| \leq \sum_{\substack{k \geq 0 \\ \sum k_i = k}} \sum_{k_1, \ldots, k_l} \frac{4^k l^k T^k}{k!} \sup_{\substack{W_1, \ldots, W_l \\ W_i \in \{L_{\mathbf{J}}, L_{\mathbf{\Lambda}}, L_{\mathbf{h}}, L_{\Delta}\}^{k_i}}} \sum_{\phi \in (W_1 f^{(1)} \cdots W_l f^{(l)})(\mathbf{x})} \left|\mathbb{E}[\phi(\mathbf{X}_0)] - \tilde{\mathbb{E}}[\phi(\mathbf{X}_0)]\right|$$

$$\leq \sum_{\substack{k \geq 0 \\ \sum k_i = k}} \sum_{k_1, \ldots, k_l} \frac{(4lT)^k}{k!} \sup_{W \in \{L_{\mathbf{J}}, L_{\mathbf{\Lambda}}, L_{\mathbf{h}}, L_{\Delta}\}^k} \sum_{\phi \in (W f_{\bar{\alpha}, \bar{\gamma}})(\mathbf{x})} \left|\mathbb{E}[\phi(\mathbf{X}_0)] - \tilde{\mathbb{E}}[\phi(\mathbf{X}_0)]\right|$$

$$\leq \sum_{k \geq 0} \frac{(k+1)^l (4lT)^k}{k!} \sup_{W \in \{L_{\mathbf{J}}, L_{\mathbf{\Lambda}}, L_{\mathbf{h}}, L_{\Delta}\}^k} \sum_{\phi \in (W f_{\bar{\alpha}, \bar{\gamma}})(\mathbf{x})}$$

$$\left| \mathbb{E}[\phi(\mathbf{X}_0)] - \tilde{\mathbb{E}}[\phi(\mathbf{X}_0)] \right|,\tag{3.4}$$

where the sums are over the monomials ϕ in the decomposition of $W_1 f^{(1)} \cdots W_l f^{(l)}$ and that of $W f_{\bar{\alpha},\bar{\gamma}}$ per Claim 3.1. Note that each summand on the RHS of (3.4) is at most some $(k+1)^l l^k$ times the corresponding summand of (2.14) for the choice $f = f_{\bar{\alpha},\bar{\gamma}}$ for which we have deduced the bound of (2.18). Utilizing the latter and the elementary bound $k+1 \le (k_{\mathbf{J}}+1)(k+1-k_{\mathbf{J}})$, by proceeding as in the derivation of (2.19), we find that for $C = C(\bar{r}, \bar{s}, C_\mu(\bar{r}), \mathbf{C}_\star)$ finite, $\delta = 1/(16\, l\, T \bar{r}\, e\, \mathbf{C}_\star)$ positive and $N \ge (2/\delta)^2$,

$$\sup_{\mathbf{t} \in [0,T]^l} \left| \mathbb{E}[g(\mathbf{t})] - \tilde{\mathbb{E}}[g(\mathbf{t})] \right| \le \frac{2C}{N^{\bar{s}/2}} \sum_{k' \ge 0} \frac{\delta^{-k'}}{k'!} (k'+1)^l \sum_{I_\star \ge 0} \frac{1}{I_\star!}$$
$$\sum_{k_{\mathbf{J}} \ge 2I_\star + I_{\bar{\alpha},1}^+} (k_{\mathbf{J}} + \bar{s})^{\bar{s}+l} \delta^{-k_{\mathbf{J}}} N^{I_\star - k_{\mathbf{J}}/2}$$
$$\le \bar{C} N^{-(\bar{s}+I_{\bar{\alpha},1}^+)/2}$$

for some finite $\bar{C} = \bar{C}(l, \bar{r}, \bar{s}, T, \mathbf{C}_\star, C_\mu(\bar{r}))$. $\square$

We now add in the driving martingale observables (i.e., $m > 0$) and conclude the proof of Proposition 3.1.

Proof of Proposition 3.1. We reduce the situation $m > 0$ to the combinatorial calculations of Lemma 3.2 by utilizing the following expansion from Ito's lemma:

$$M_{\xi_i}(u) = X_{\xi_i}(u) - \int_0^u (L\, x_{\xi_i})(\mathbf{X}_\tau) d\tau.$$

When expanding (3.1) in this manner, the terms containing only products of $X_{\xi_i}(u_i)$ can be absorbed into $\boldsymbol{\gamma}$, in which case their difference in expectations has already been handled in Lemma 3.2, so by linearity it suffices for us to focus on handling terms of the form

$$h_{(\boldsymbol{\alpha}^{(i)}),(\boldsymbol{\gamma}^{(i)}),\boldsymbol{\xi}}(\mathbf{t}, \mathbf{u}) = \left(\prod_{i=1}^l f_{\boldsymbol{\alpha}^{(i)},\boldsymbol{\gamma}^{(i)}}(\mathbf{X}_{t_i}) \right) \left(\prod_{i=1}^m \int_0^{u_i} (Lx_{\xi_i})(\mathbf{X}_{\tau_i}) d\tau_i \right)$$
$$= \int_0^{u_1} \cdots \int_0^{u_m} \widehat{h}(\mathbf{t}, \boldsymbol{\tau}) d\tau_1 \cdots d\tau_m,$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m) \in [0, T]^m$ and where, setting $f^{(i)}(\mathbf{x}) = f_{\boldsymbol{\alpha}^{(i)},\boldsymbol{\gamma}^{(i)}}(\mathbf{x})$,

$$\widehat{h}(\mathbf{t}, \boldsymbol{\tau}) := \prod_{i=1}^l f^{(i)}(\mathbf{X}_{t_i}) \prod_{i=1}^m (Lx_{\xi_i})(\mathbf{X}_{\tau_i}).$$

Thus, fixing $l, m, (\boldsymbol{\alpha}^{(i)}), (\boldsymbol{\gamma}^{(i)}), \boldsymbol{\xi}$ and letting $h(\mathbf{t}, \mathbf{u}) = h_{(\boldsymbol{\alpha}^{(i)}), (\boldsymbol{\gamma}^{(i)}), \boldsymbol{\xi}}(\mathbf{t}, \mathbf{u})$ we obtain after swapping the expectation and integrals that

$$\mathbb{E}[h(\mathbf{t}, \mathbf{u})] = \int_0^{u_1} \cdots \int_0^{u_m} \mathbb{E}[\widehat{h}(\mathbf{t}, \boldsymbol{\tau})] d\tau_1 \cdots d\tau_m,$$

which thereby yields the following bound on the relevant difference in expectations

$$\left| \mathbb{E}[h(\mathbf{t}, \mathbf{u})] - \tilde{\mathbb{E}}[h(\mathbf{t}, \mathbf{u})] \right| \leq T^m \sup_{\boldsymbol{\tau} \in [0, T]^m} \left| \mathbb{E}[\widehat{h}(\mathbf{t}, \boldsymbol{\tau})] - \tilde{\mathbb{E}}[\widehat{h}(\mathbf{t}, \boldsymbol{\tau})] \right|.$$

Proceeding hereafter WLOG to bound the difference in expectations for $\widehat{h}(\mathbf{t}, \boldsymbol{\tau})$, we suppose for ease of exposition that $0 \leq t_l = \tau_0 \leq \tau_1 \leq \cdots \leq \tau_m$ (the situation where the two groups intertwine is similarly analyzed with the obvious modifications). As done in the proof of Lemma 3.2, first expressing $\mathbb{E}_\mathbf{B}$ in terms of the semi-group operator and then expanding that in powers of the generator L we find that

$$\mathbb{E}_\mathbf{B}[\widehat{h}(\mathbf{t}, \boldsymbol{\tau})]$$
$$= P_{t_1}\left[f^{(1)} P_{t_2-t_1}\left[f^{(2)} \cdots P_{t_l-t_{l-1}}\left[f^{(l)} P_{\tau_1-t_l}\left[Lx_{\xi_1} \cdots P_{\tau_m-\tau_{m-1}} Lx_{\xi_m} \right] \right] \right] \right](\mathbf{X}_0)$$
$$= \sum_{k \geq 0} \sum_{\substack{(k_i) \geq 0, (\ell_i) \geq 1 \\ \sum k_i + \sum \ell_i = k+m}} \prod_{i=1}^l \frac{(t_i - t_{i-1})^{k_i}}{k_i!} \prod_{i=1}^m \frac{(\tau_i - \tau_{i-1})^{\ell_i-1}}{(\ell_i - 1)!} L^{k_1}$$
$$\left[f^{(1)} \cdots L^{k_l}\left[f^{(l)} L^{\ell_1}\left[x_{\xi_1} \cdots L^{\ell_m} x_{\xi_m} \right] \right] \right](\mathbf{X}_0).$$

At this point, proceeding as in the derivation of (3.4), up to the transformations

$$k \mapsto k + m =: \bar{k}, \qquad l \mapsto l + m =: \bar{l}, \quad \text{and} \quad (f^{(l+1)}, \ldots, f^{(\bar{l})}) \mapsto (x_{\xi_1}, \ldots, x_{\xi_m}),$$

we first use (3.3) to get the bound

$$\left| \mathbb{E}[\widehat{h}(\mathbf{t}, \boldsymbol{\tau})] - \tilde{\mathbb{E}}[\widehat{h}(\mathbf{t}, \boldsymbol{\tau})] \right|$$
$$\leq \sum_{k \geq 0} \sum_{\substack{(k_i) \geq 0, (\ell_i) \geq 1 \\ \sum k_i + \sum \ell_i = \bar{k}}} \frac{4^{\bar{k}} (\bar{l}T)^k}{k!} \sup_{\substack{W_1, \ldots, W_l, W_1', \ldots, W_m' \\ W_i \in \{L_\mathbf{J}, L_\mathbf{\Lambda}, L_\mathbf{h}, L_\Delta\}^{k_i} \\ W_i' \in \{L_\mathbf{J}, L_\mathbf{\Lambda}, L_\mathbf{h}, L_\Delta\}^{\ell_i}}}$$
$$\sum_{\phi \in (W_1 f^{(1)} \cdots W_m' x_{\xi_m})(\mathbf{x})} \left| \mathbb{E}[\phi(\mathbf{X}_0)] - \tilde{\mathbb{E}}[\phi(\mathbf{X}_0)] \right|,$$

with the sum running over monomial decomposition of $(W_1 f^{(1)} \cdots W_l f^{(l)} W_1' x_{\xi_1} \cdots W_m' x_{\xi_m})(\mathbf{x})$. Then, utilizing again Claim 3.1, as well as the bound $k! \geq \bar{k}!/(\bar{k})^m$, we arrive at

$$\left| \mathbb{E}[\widehat{h}(\mathbf{t}, \boldsymbol{\tau})] - \tilde{\mathbb{E}}[\widehat{h}(\mathbf{t}, \boldsymbol{\tau})] \right|$$

$$\leq \sum_{\bar{k} \geq m} \left(\frac{\bar{k}}{\bar{l}T} \right)^m \frac{(\bar{k})^{\bar{l}} (4\bar{l}T)^{\bar{k}}}{\bar{k}!} \sup_{W \in \{L_{\mathbf{J}}, L_{\boldsymbol{\Lambda}}, L_{\mathbf{h}}, L_{\Delta}\}^{\bar{k}}} \sum_{\phi \in (W f_{\bar{\alpha}, \bar{\gamma}})(\mathbf{x})} \left| \mathbb{E}[\phi(\mathbf{X}_0)] - \tilde{\mathbb{E}}[\phi(\mathbf{X}_0)] \right|,$$

(3.5)

where as before $\bar{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^{(1)} \amalg \cdots \amalg \boldsymbol{\alpha}^{(l)}$ is of length $\bar{s} = \sum_i s_i$, while $\bar{\boldsymbol{\gamma}}$ of length $\bar{r} = \sum r_i + m$ has now the additional elements $(x_{\xi_i})_{i \leq m}$. Up to this update of $\bar{r}$ and the immaterial weight factor $(\bar{k}/(\bar{l}T))^m$ of its summands, the expression on the RHS of (3.5) is the same as that in (3.4). We thus conclude as in the proof of Lemma 3.2 that for some $C(l, m, \bar{r}, \bar{s}, T, \mathbf{C}_\star, C_\mu(\bar{r}))$ all $\mathbf{t} \in [0, T]^l$ and $\mathbf{u} \in [0, T]^m$,

$$\left| \mathbb{E}[h(\mathbf{t}, \mathbf{u})] - \tilde{\mathbb{E}}[h(\mathbf{t}, \mathbf{u})] \right| \leq C N^{-(\bar{s} + I_{\bar{\alpha},1}^+)/2}. \quad \square$$

3.2 Proof of Theorem 1.

Fix $T, m, p, C_{\mathbf{a}}, \mathbf{a} \in \mathbb{R}^{N^m}$ such that $\|\mathbf{a}\|_\infty \leq C_{\mathbf{a}}$, and $\mathbf{t} \in [0, T]^p$. For every $\ell \leq m$, fix observables $\mathcal{Y}^{(\ell,1)}, \ldots, \mathcal{Y}^{(\ell,p)} \in \mathfrak{F}$ and let $F(\mathbf{t})$ be as in (1.5) with those choices. By linearity of expectations and the uniform bound on $\|\mathbf{a}\|_\infty$, it suffices to show that uniformly over $i_1, \ldots, i_m$,

$$\sup_{\mathbf{t} \in [0,T]^p} \left| \mathbb{E}\Big[\prod_{\ell \leq m} \mathcal{Y}_{i_\ell}^{(\ell,1)}(t_1) \cdots \mathcal{Y}_{i_\ell}^{(\ell,p)}(t_p) \Big] - \tilde{\mathbb{E}}\Big[\prod_{\ell \leq m} \mathcal{Y}_{i_\ell}^{(\ell,1)}(t_1) \cdots \mathcal{Y}_{i_\ell}^{(\ell,p)}(t_p) \Big] \right|$$

$$\leq C N^{-1/2}$$

(3.6)

We denote by $\bar{s}$ the number of $\mathcal{Y}$ terms appearing in the preceding product which is a coordinate of $\mathbf{G}_t$. In case $\bar{s} = 0$, the bound (3.6) follows from considering Proposition 3.1 at $\bar{s} = 0$, in which case $I_{\bar{\alpha},1}^+ = 1$. Otherwise, we expand every term in that product which is a coordinate of $\mathbf{G}_t$ to obtain a sum of monomials of the form of (3.1). Each of these monomials has a sequence $\bar{\boldsymbol{\alpha}}$ of length $\bar{s}$, and as a result of such expansion there are at most $\bar{s}^{\bar{s}} N^{I_{\bar{\alpha}}}$ monomials with precisely $I_{\bar{\alpha}}$ distinct pairs in the sequence $\bar{\boldsymbol{\alpha}}$. Note that for any $\bar{\boldsymbol{\alpha}}$,

$$\bar{s} + I_{\bar{\alpha},1}^+ \geq 2I_{\bar{\alpha}} + 1.$$

Indeed, each pair which appears once in $\bar{\boldsymbol{\alpha}}$, is counted both in $\bar{s}$ and in $I_{\bar{\alpha},1}$, all other pairs are counted at least twice in $\bar{s}$, and for any $\bar{\boldsymbol{\alpha}}$ of maximal multiplicity two, we have added one to $I_{\bar{\alpha},1}^+$. Consequently, the bound of Proposition 3.1 on the difference in expectation for each of these $\bar{s}^{\bar{s}} N^{I_{\bar{\alpha}}}$ many monomials is at most $C N^{-I_{\bar{\alpha}} - 1/2}$ for some constant $C(T, m, p, \mathbf{C}_\star, C_\mu)$. From this, the bound (3.6) immediately follows upon enumerating over the at most $\bar{s}$ many choices for $I_{\bar{\alpha}}$. $\quad \square$

4 Concentration for quadratic observables: Proof of Theorem 2

Assuming henceforth that $\mathbf{M}_t$ is a scaled Brownian motion (i.e., that σ_{ij} are identically zero for $i \neq 0$), our goal is to prove Theorem 2 about the uniform over $\mathbf{t} \in [0, T]^2$ concentration property of the quadratic observable of (1.10),

$$F(\mathbf{t}) = F_{\mathcal{Y}, \mathcal{Y}', \mathbf{a}}(\mathbf{X}_{t_1}, \mathbf{X}_{t_2}) = \frac{1}{N} \sum_i a_i \mathcal{Y}_i(t_1) \mathcal{Y}'_i(t_2),$$

(for uniformly bounded non-random $\mathbf{a} = (a_i)_i$ and $\mathcal{Y}, \mathcal{Y}'$ in the collection $\mathfrak{F} = \{\mathbf{1}_t, \mathbf{X}_t, \mathbf{G}_t, \mathbf{M}_t\}$ of (1.4)). To this end, we introduce in Sect. 4.1 high probability localizing sets $\mathcal{L}_{N,R}$ on which various norms of $\mathbf{X}_t$ (and our observables $F(\mathbf{t})$), are uniformly bounded. Sect. 4.2 shows that on $\mathcal{L}_{N,R}$, such $F(\mathbf{t})$ are $O(N^{-1/2})$-Lipschitz in a mixed ℓ^2-norm. Combining these facts we prove Theorem 2 in Sect. 4.3.

4.1 Localizing the process

Denote the 2-to-2 matrix norm by

$$\|\mathbf{J}\|_{2 \to 2} := \sup_{\mathbf{x}: \|\mathbf{x}\|=1} \|\mathbf{J}\mathbf{x}\| = \|\mathbf{J}^T\|_{2 \to 2} = \sup_{\mathbf{x}: \|\mathbf{x}\|=1} \Big(\sum_{i \leq N} G_i(\mathbf{x})^2 \Big)^{1/2},$$

and for each constant R consider the following localization subset of $\mathcal{E}_N := \mathbb{R}^N \times \mathbb{R}^{N^2} \times C([0, T], \mathbb{R}^N)$,

$$\mathcal{L}_{N,R} := \Big\{ (\mathbf{X}_0, \mathbf{J}, \mathbf{M}) \in \mathcal{E}_N : \|\mathbf{X}_0\|^2 + N \|\mathbf{J}\|_{2 \to 2}^2 + \sup_{t \in [0,T]} \|\mathbf{M}_t\|^2 \leq RN \Big\}, \quad (4.1)$$

We begin by bounding the probability that $(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) \notin \mathcal{L}_{N,R}$.

Lemma 4.1 *There exists $C = C(T, C_\mu, C_\mathbf{A}, C_\sigma) > 0$ and $R_0(T, C_\mu, C_\mathbf{A}, C_\sigma) < \infty$, such that for every $R \geq R_0$ if $\mu, \mathbb{P}_\mathbf{A}$ satisfy Hypotheses 1–2, then*

$$\mathbb{P}\big(\mathcal{L}_{N,R}^c\big) \leq \exp(-\sqrt{RN}/C).$$

Proof We bound $\mathcal{L}_{N,R}^c$ by the union of the events where each of the three norms is greater than $\sqrt{RN/3}$. First, since $\mathbf{M}_t$ is a Brownian motion (scaled by $(\sigma_{0j})_j$), by Doob's maximal inequality for the sub-martingale $\exp(\delta \|\mathbf{M}_t\|^2)$, we have for some $C(C_\sigma) > 0$ any $R \geq T R_0(C_\sigma)$ and all N,

$$\mathbb{P}_\mathbf{B}\Big(\sup_{t \in [0,T]} \|\mathbf{M}_t\| > \sqrt{RN/3} \Big) \leq \exp(-RN/(CT)). \quad (4.2)$$

Next, since μ satisfies Hypotheses 1–2, the independent $X_i(0)$ have uniform (in i and N), second moments and exponential tails. Hence, applying [30, Theorem 3] for the

centered sum of i.i.d. variables that stochastically dominate $X_i^2(0)$, we have for some $C(C_\mu) > 0$, any $R \geq R_0(C_\mu)$ and all N,

$$\mu\left(\|\mathbf{X}_0\|^2 > RN/3\right) \leq \exp(-\sqrt{RN}/C).$$

It thus remains only to show that when $\mathbb{P}_\mathbf{A}$ satisfies Hypothesis 2, we have for some $C(C_\mathbf{A}) > 0$ any $R \geq R_0(C_\mathbf{A})$ and all N,

$$\mathbb{P}_\mathbf{A}(\|\mathbf{A}\|_{2\to 2} > \sqrt{RN/3}) \leq \exp(-\sqrt{RN}/C). \tag{4.3}$$

To this end, recall [28, Theorem 2] that there exists a universal constant C such that for any matrix $\mathbf{A}$ with independent, zero-mean entries of second moments m_{ij} and fourth moments b_{ij},

$$\mathbb{E}_\mathbf{A}[\|\mathbf{A}\|_{2\to 2}] \leq C\left(\max_{i\leq N}\left(\sum_{j\leq N} m_{ij}\right)^{1/2} + \max_{j\leq N}\left(\sum_{i\leq N} m_{ij}\right)^{1/2} + \left(\sum_{1\leq i,j\leq N} b_{ij}\right)^{1/4}\right).$$

For $\mathbb{P}_\mathbf{A}$ satisfying Hypothesis 1, b_{ij} and m_{ij} are bounded uniformly in i, j and N (see (1.8)). Hence, in the case where $\mathbf{A}$ is composed of independent entries, for some $C(C_\mathbf{A})$ finite and all N,

$$\mathbb{E}_\mathbf{A}[\|\mathbf{A}\|_{2\to 2}] \leq C\sqrt{N}. \tag{4.4}$$

Likewise, representing a symmetric $\mathbf{A}$ as $\mathbf{A} = \mathbf{A}^+ + \mathbf{A}^-$, with $\mathbf{A}^+$ the upper triangle (including the diagonal) part of $\mathbf{A}$ and $\mathbf{A}^-$ its lower triangle part, [28, Theorem 2] holds for the matrices $\mathbf{A}^-$ and $\mathbf{A}^+$ of zero-mean, independent entries (with uniformly bounded forth moments). Thus, (4.4) holds also in this case up to a factor of 2. Thanks to (4.4), if $\sqrt{R} \geq 4C$ then

$$\mathbb{P}_\mathbf{A}\left(\|\mathbf{A}\|_{2\to 2} > \sqrt{RN/3}\right) \leq \mathbb{P}_\mathbf{A}\left(|\|\mathbf{A}\|_{2\to 2} - \mathbb{E}_\mathbf{A}[\|\mathbf{A}\|_{2\to 2}]| > \sqrt{RN}/4\right).$$

Recall that $\|\mathbf{A}\|_{2\to 2}$, which is the largest singular value of $\mathbf{A}$, is 1-Lipschitz in its entries (endowed with the Euclidean norm, on $\mathbf{A}^+$ when $\mathbf{A}$ assumed symmetric). Indeed, this follows by combining the triangle inequality $|\|\mathbf{A}\|_{2\to 2} - \|\mathbf{B}\|_{2\to 2}| \leq \|\mathbf{A} - \mathbf{B}\|_{2\to 2}$ with the domination of the operator norm by the Frobenius norm, $\|\mathbf{A} - \mathbf{B}\|_{2\to 2} \leq \|\mathbf{A} - \mathbf{B}\|_F$. Hypothesis 2 for $\mathbb{P}_\mathbf{A}$ thus yields the bound (4.3). $\square$

We further have on the sets $\mathcal{L}_{N,R}$ the following localization for both $(\mathbf{X}_t)_{t\in[0,T]}$ and $(\mathbf{G}_t)_{t\in[0,T]}$.

Proposition 4.2 *There exists $R_0(T, \mathbf{C}_\star)$ and $C_0(\mathbf{C}_\star)$ such that if $R \geq R_0$, and $(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) \in \mathcal{L}_{N,R}$, then*

$$\frac{1}{\sqrt{N}}\sup_{t\in[0,T]}\{\|\mathbf{X}_t\|\} \leq e^{C_0\sqrt{R}T}, \qquad \frac{1}{\sqrt{N}}\sup_{t\in[0,T]}\{\|\mathbf{G}_t\|\} \leq e^{C_0\sqrt{R}T}. \tag{4.5}$$

In addition, for every **a** *such that* $\|\mathbf{a}\|_\infty \leq C_{\mathbf{a}}$ *(uniformly over N) and every* $\mathcal{Y}, \mathcal{Y}' \in \mathfrak{F}$, *if* $F(\mathbf{t})$ *is as in* (1.10)*, we have for all* $k \geq 1$,

$$\limsup_{N\to\infty} \mathbb{E}\Big[\Big(\frac{1}{\sqrt{N}} \sup_{t\in[0,T]} \|\mathbf{X}_t\|\Big)^k\Big] < \infty, \quad \limsup_{N\to\infty} \mathbb{E}\Big[\Big(\frac{1}{\sqrt{N}} \sup_{t\in[0,T]} \|\mathbf{G}_t\|\Big)^k\Big] < \infty,$$

(4.6)

$$\limsup_{N\to\infty} \mathbb{E}\Big[\sup_{\mathbf{t}\in[0,T]^2} |F(\mathbf{t})|^k\Big] < \infty.$$

(4.7)

Proof Setting $e_N(t) = \frac{1}{\sqrt{N}}\|\mathbf{X}_t\|$, we get upon expanding (1.2), that

$$(e_N(t))^2 \leq \frac{1}{N} \sum_{j\leq N} |X_j(t)|$$

$$\Big(|X_j(0)| + |M_j(t)| + \int_0^t |h_j|ds + \int_0^t |G_j(\mathbf{X}_s)|ds + \int_0^t |\Lambda_j(\mathbf{X}_s)|ds\Big)$$

$$=: I_1 + I_2 + I_3 + I_4 + I_5.$$

From the definition of the 2-to-2 norm, evidently

$$\|\mathbf{G}_s\| = \sqrt{\sum_{j\leq N} G_j(\mathbf{X}_s)^2} \leq \|\mathbf{J}\|_{2\to2}\|\mathbf{X}_s\|, \qquad \sqrt{\sum_{j\leq N} \Lambda_j(\mathbf{X}_s)^2} \leq \|\mathbf{\Lambda}\|_{2\to2}\|\mathbf{X}_s\|. \tag{4.8}$$

Hence, by Cauchy–Schwarz,

$$I_1 \leq e_N(t)\frac{1}{\sqrt{N}}\|\mathbf{X}_0\|, \qquad I_2 \leq e_N(t)\frac{1}{\sqrt{N}}\|\mathbf{M}_t\|, \qquad I_3 \leq e_N(t)\,C_{\mathbf{h}}\,T,$$

$$I_4 \leq e_N(t)\frac{1}{\sqrt{N}}\int_0^t \|\mathbf{G}_s\|ds \leq e_N(t)\|\mathbf{J}\|_{2\to2}\int_0^t e_N(s)ds,$$

$$I_5 \leq e_N(t)\frac{1}{\sqrt{N}}\int_0^t \Big(\sum_{j\leq N} |\Lambda_j(\mathbf{X}_s)|^2\Big)^{1/2}ds \leq e_N(t)C_{\mathbf{\Lambda}}\int_0^t e_N(s)ds,$$

where in the last inequality we rely on our assumption that $\|\mathbf{\Lambda}\|_{1\to1} \leq C_{\mathbf{\Lambda}}$ and $\|\mathbf{\Lambda}\|_{\infty\to\infty} \leq C_{\mathbf{\Lambda}}$, to deduce that $\|\mathbf{\Lambda}\|_{2\to2} \leq C_{\mathbf{\Lambda}}$. Combining these bounds on $(I_i)_{i\leq5}$, and dividing out by $e_N(t)$, we see that

$$e_N(t) \leq \frac{1}{\sqrt{N}}\Big[\|\mathbf{X}_0\| + \|\mathbf{M}_t\|\Big] + C_{\mathbf{h}}T + (\|\mathbf{J}\|_{2\to2} + C_{\mathbf{\Lambda}})\int_0^t e_N(s)ds.$$

By Gronwall's inequality, using the localization to $\mathcal{L}_{N,R}$, it then follows that for any $t \in [0, T]$,

$$e_N(t) \leq (\sqrt{R} + C_{\mathbf{h}}T)\exp\big((\sqrt{R} + C_{\mathbf{\Lambda}})t\big),$$

yielding the LHS of (4.5) as soon as $R \geq R_0(T, \mathbf{C}_\star) \geq 1$. From the LHS of (4.8) we know that $\|\mathbf{G}_t\| \leq \sqrt{R} \|\mathbf{X}_t\|$ throughout $\mathcal{L}_{N,R}$, hence after suitably increasing C_0 and R_0, the RHS of (4.5) holds as well.

To deduce the uniformly bounded moment estimate of (4.6) for $\mathbf{X}_t$, recall first from the LHS of (4.5) that

$$Z_{N,\mathbf{X}}^k := \big(\sup_{t \leq T} e_N(t)\big)^k \leq e^{C_0 \sqrt{R}Tk} =: f(R), \quad \forall R \geq R_0, \ (\mathbf{X}_0, \mathbf{J}, \mathbf{M}) \in \mathcal{L}_{N,R}.$$

Combining the latter bound with that of Lemma 4.1, we arrive at

$$\begin{aligned}
\mathbb{E}[Z_{N,\mathbf{X}}^k] &= \int_0^\infty f'(R)\mathbb{P}(Z_{N,\mathbf{X}}^k > f(R))dR \\
&\leq f(R_0) + \int_{R_0}^\infty f'(R)\mathbb{P}(\mathcal{L}_{N,R}^c)dR \leq f(R_0) + \int_{R_0}^\infty f'(R)e^{-\sqrt{RN}/C}dR.
\end{aligned}$$
(4.9)

The RHS decreases in N and as $f'(R) = (C_0 Tk)/(2\sqrt{R})f(R)$, it is finite for $\sqrt{N}/C > C_0 Tk$, yielding the LHS of (4.6). The RHS of (4.6) follows by applying the same reasoning to $Z_{N,\mathbf{G}}^k = \big(N^{-1/2}\sup_{t \in [0,T]}\|\mathbf{G}_t\|\big)^k$ while utilizing the RHS of (4.5).

Turning to (4.7), note that for any $k \geq 1$ and $F(\mathbf{t})$ of (1.10) with $\|\mathbf{a}\|_\infty \leq C_{\mathbf{a}}$, by Cauchy–Schwarz,

$$|F(\mathbf{t})|^k \leq C_{\mathbf{a}}^k \sqrt{Z_{N,\mathcal{Y}}^{2k}}\sqrt{Z_{N,\mathcal{Y}'}^{2k}}, \quad \text{where} \quad Z_{N,\mathcal{Y}}^{2k} := \big(\frac{1}{\sqrt{N}}\sup_{t \in [0,T]}\{\|\mathcal{Y}(t)\|\}\big)^{2k}.$$

Thus, yet another application of Cauchy–Schwarz results with

$$\mathbb{E}\big[\sup_{\mathbf{t} \in [0,T]^2}|F(\mathbf{t})|^k\big] \leq C_{\mathbf{a}}^k \sqrt{\mathbb{E}[Z_{N,\mathcal{Y}}^{2k}]}\sqrt{\mathbb{E}[Z_{N,\mathcal{Y}'}^{2k}]} \leq C_{\mathbf{a}}^k \max_{\mathcal{Y} \in \mathfrak{F}} \mathbb{E}[Z_{N,\mathcal{Y}}^{2k}].$$

If $\mathcal{Y}$ is $\mathbf{1}$, this latter expectation is simply 1. If $\mathcal{Y}$ is $\mathbf{M}$, using the tail bound of (4.2) in combination with (4.9) (now for $f(R) = (R/3)^k$), the latter expectation is uniformly bounded in N. Lastly if $\mathcal{Y}$ is from $\{\mathbf{X}, \mathbf{G}\}$, the expectation above is uniformly bounded in N by (4.6). Combining these yields the desired (4.7). □

4.2 A Lipschitz estimate on quadratic observables

Our next proposition shows that on $\mathcal{L}_{N,R}$ all $F(\mathbf{t})$ of the form (1.10) are $O(N^{-1/2})$-Lipschitz in the $(\mathbf{X}_0, \mathbf{J}, \mathbf{M})$ endowed with the following mixed 2-norm on $\mathcal{E}_N$,

$$\|(\mathbf{X}_0, \mathbf{J}, \mathbf{M})\|_{\text{MIX}}^2 := \|\mathbf{X}_0\|^2 + N\sum_{1 \leq i,j \leq N} J_{ij}^2 + \sup_{t \in [0,T]} \|\mathbf{M}_t\|^2$$
(4.10)

(which is taken from [3, Hypothesis 1.1]).

Proposition 4.3 *Fixing* **a** *such that* $\|\mathbf{a}\|_\infty \leq C_{\mathbf{a}}$ *and* $\mathcal{Y}, \mathcal{Y}' \in \mathfrak{F}$, *denote by* $F(\mathbf{t}; (\mathbf{X}_0, \mathbf{J}, \mathbf{M}))$ *the observable in* (1.10) *evaluated on the trajectory* $\mathbf{X}_t$ *constructed out of the triplet* $(\mathbf{X}_0, \mathbf{J}, \mathbf{M})$. *There exist* $R_0(T, C_{\mathbf{a}}, \mathbf{C}_\star)$ *and* $C(T, C_{\mathbf{a}}, \mathbf{C}_\star)$ *such that for any* $R \geq R_0$ *all* N *and* $(\mathbf{X}_0, \mathbf{J}, \mathbf{M}), (\mathbf{X}_0', \mathbf{J}', \mathbf{M}')$ *in* $\mathcal{L}_{N,R}$

$$\sup_{\mathbf{t}\in[0,T]^2} |F(\mathbf{t}; (\mathbf{X}_0, \mathbf{J}, \mathbf{M})) - F(\mathbf{t}; (\mathbf{X}_0', \mathbf{J}', \mathbf{M}'))|$$

$$\leq \frac{Ce^{C\sqrt{R}}}{\sqrt{N}} \|(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) - (\mathbf{X}_0', \mathbf{J}', \mathbf{M}')\|_{\mathrm{MIX}}.$$

The key to Proposition 4.3 is to show that $\mathbf{X}_t$ is $O(1)$-Lipschitz on $\mathcal{L}_{N,R}$ endowed with $\|\cdot\|_{\mathrm{MIX}}$. Specifically, denoting by $\mathbf{X}_t(\mathbf{X}_0, \mathbf{J}, \mathbf{M})$ the solution to (1.2), constructed from the triplet $(\mathbf{X}_0, \mathbf{J}, \mathbf{M})$ and $\mathbf{X}_t'(\mathbf{X}_0, \mathbf{J}, \mathbf{M})$ the solution constructed from the triplet $(\mathbf{X}_0', \mathbf{J}', \mathbf{M}')$, our next lemma establishes a uniform over $\mathcal{L}_{N,R}$ Lipschitz bound on $\|\mathbf{X}_t - \mathbf{X}_t'\|$.

Lemma 4.4 *There exist* $R_0(T, \mathbf{C}_\star), C(T, \mathbf{C}_\star)$ *such that for all* $R \geq R_0$ *and* $(\mathbf{X}_0, \mathbf{J}, \mathbf{M}), (\mathbf{X}_0', \mathbf{J}', \mathbf{M}') \in \mathcal{L}_{N,R}$,

$$\sup_{t\in[0,T]} \|\mathbf{X}_t(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) - \mathbf{X}_t'(\mathbf{X}_0', \mathbf{J}', \mathbf{M}')\| \leq e^{C\sqrt{R}} \|(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) - (\mathbf{X}_0', \mathbf{J}', \mathbf{M}')\|_{\mathrm{MIX}}.$$

Proof Following the strategy of proof of [3, Lemma 2.6], we let

$$e_N(t) := \frac{1}{\sqrt{N}} \|\mathbf{X}_t(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) - \mathbf{X}_t'(\mathbf{X}_0', \mathbf{J}', \mathbf{M}')\|,$$

and expanding over $j \leq N$, we have by the definition of the solution $\mathbf{X}_t$ for the SDS (1.2)–(1.3), that

$$e_N(t)^2 \leq \frac{1}{N} \sum_{j\leq N} |X_j(t) - X_j'(t)|$$

$$\left(|X_j(0) - X_j'(0)| + |M_j(t) - M_j'(t)| + \int_0^t |\Lambda_j(\mathbf{X}_s) - \Lambda_j(\mathbf{X}_s')| ds \right.$$

$$\left. + \int_0^t |G_j(\mathbf{X}_s) - G_j(\mathbf{X}_s')| ds + \int_0^t |G_j(\mathbf{X}_s') - G_j'(\mathbf{X}_s')| ds \right)$$

$$=: I_1 + I_2 + I_3 + I_4 + I_5,$$

where $\mathbf{G}'(\cdot)$ is defined as $\mathbf{G}(\cdot)$ but constructed using $\mathbf{J}'$ instead of $\mathbf{J}$. By Cauchy–Schwarz,

$$I_1 \leq e_N(t) \frac{1}{\sqrt{N}} \|\mathbf{X}_0 - \mathbf{X}_0'\|, \quad \text{and} \quad I_2 \leq e_N(t) \frac{1}{\sqrt{N}} \|\mathbf{M}_t - \mathbf{M}_t'\|.$$

Recalling (4.8), we similarly find that

$$I_3 \leq e_N(t) \frac{1}{\sqrt{N}} \int_0^t \Big(\sum_{j \leq N} |\Lambda_j (\mathbf{X}_s - \mathbf{X}'_s)|^2 \Big)^{1/2} ds \leq e_N(t) C_\Lambda \int_0^t e_N(s) ds \,.$$

Turning to the terms involving $\mathbf{G}(\cdot)$ or $\mathbf{G}'(\cdot)$, observe first that

$$\begin{aligned}
\|\mathbf{G}(\mathbf{X}_t) - \mathbf{G}(\mathbf{X}'_t)\| &\leq \|\mathbf{J}\|_{2 \to 2} \|\mathbf{X}_t - \mathbf{X}'_t\|] \,, \qquad \text{and} \\
\|\mathbf{G}(\mathbf{X}_t) - \mathbf{G}'(\mathbf{X}_t)\| &\leq \|\mathbf{J} - \mathbf{J}'\|_{2 \to 2} \|\mathbf{X}_t\| \,.
\end{aligned} \tag{4.11}$$

Using the localization to $\mathcal{L}_{N,R}$, we thus find that

$$\begin{aligned}
I_4 &\leq e_N(t) \frac{1}{\sqrt{N}} \int_0^t \|\mathbf{G}(\mathbf{X}_s) - \mathbf{G}(\mathbf{X}'_s)\| ds \leq e_N(t) \|\mathbf{J}\|_{2 \to 2} \int_0^t e_N(s) ds \\
&\leq e_N(t) \sqrt{R} \int_0^t e_N(s) \,, \\
I_5 &\leq e_N(t) \frac{1}{\sqrt{N}} \int_0^t \|\mathbf{G}(\mathbf{X}'_s) - \mathbf{G}'(\mathbf{X}'_s)\| ds \leq e_N(t) \|\mathbf{J} - \mathbf{J}'\|_{2 \to 2} \frac{1}{\sqrt{N}} \int_0^t \|\mathbf{X}'_s\| ds \\
&\leq e_N(t) \|\mathbf{J} - \mathbf{J}'\|_{2 \to 2} T e^{C_0 \sqrt{R} T} \,,
\end{aligned}$$

where in the last inequality we further assumed $R \geq R_0(T, \mathbf{C}_\star)$, utilizing the LHS of (4.5). Further increasing R_0 such that $T e^{C_0 \sqrt{R_0} T} \geq 1$, upon combining the bounds on $(I_i)_{i \leq 5}$, and dividing out by $e_N(t)$, we see that

$$\begin{aligned}
e_N(t) &\leq \frac{T e^{C_0 \sqrt{R} T}}{\sqrt{N}} \Big[\|\mathbf{X}_0 - \mathbf{X}'_0\| + \sqrt{N} \|\mathbf{J} - \mathbf{J}'\|_{2 \to 2} + \sup_{t \leq T} \|\mathbf{M}_t - \mathbf{M}'_t\| \Big] \\
&\quad + \Big[C_\Lambda + \sqrt{R} \Big] \int_0^t e_N(s) ds \,.
\end{aligned}$$

Recall that $\|\mathbf{J}\|_{2 \to 2}^2 \leq \sum_{ij} J_{ij}^2$, so by Gronwall's inequality, there exist $C(T, \mathbf{C}_\star)$, such that

$$e_N(t) \leq \frac{e^{C \sqrt{R}}}{\sqrt{N}} \|(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) - (\mathbf{X}'_0, \mathbf{J}', \mathbf{M}')\|_{\text{MIX}} \,,$$

for any $R \geq R_0$, every N and all $t \in [0, T]$, as claimed. $\qquad \square$

Proof of Proposition 4.3. Fix $\mathcal{Y}^1, \mathcal{Y}^2 \in \mathfrak{F}$, $\mathbf{a}$ such that $\|\mathbf{a}\|_\infty \leq C_{\mathbf{a}}$ and $\mathbf{t} = (t_1, t_2) \in [0, T]^2$. Equipped with Lemma 4.4 and (4.11) it remains to establish a Lipschitz control on differences of $F(\mathbf{t}; (\mathbf{X}_0, \mathbf{J}, \mathbf{M}))$ in terms of differences of $\|\mathbf{G}_t\|$, $\|\mathbf{X}_t\|$ and $\|\mathbf{M}_t\|$ corresponding to any pair of triplets $(\mathbf{X}_0, \mathbf{J}, \mathbf{M})$ and $(\mathbf{X}'_0, \mathbf{J}', \mathbf{M}')$ in $\mathcal{L}_{N,R}$. To this end,

we start with the following bound on differences of $F(\mathbf{t}; \cdot)$:

$$
\begin{aligned}
\left| F(\mathbf{t}; (\mathbf{X}_0, \mathbf{J}, \mathbf{M})) - F(\mathbf{t}; (\mathbf{X}'_0, \mathbf{J}', \mathbf{M}')) \right| &\leq \frac{C_{\mathbf{a}}}{N} \sum_{i \leq N} \left| \mathcal{Y}_i^1(\mathbf{X}_{t_1}) - \mathcal{Y}_i^1(\mathbf{X}'_{t_1}) \right| \left| \mathcal{Y}_i^2(\mathbf{X}_{t_2}) \right| \\
&\quad + \frac{C_{\mathbf{a}}}{N} \sum_{i \leq N} \left| \mathcal{Y}_i^1(\mathbf{X}'_{t_1}) \right| \left| \mathcal{Y}_i^2(\mathbf{X}_{t_2}) - \mathcal{Y}_i^2(\mathbf{X}'_{t_2}) \right|.
\end{aligned}
$$

Since the two terms on the RHS can be bounded symmetrically, WLOG we focus on the first one, which by Cauchy–Schwarz, is at most

$$
C_{\mathbf{a}} \sup_{\mathcal{Y} \in \mathfrak{F}, t \in [0,T]} \left\{ \frac{1}{\sqrt{N}} \| \mathcal{Y}(\mathbf{X}_t) - \mathcal{Y}(\mathbf{X}'_t) \| \right\} \sup_{\mathcal{Y} \in \mathfrak{F}, t \in [0,T]} \left\{ \frac{1}{\sqrt{N}} \| \mathcal{Y}(\mathbf{X}_t) \| \right\}, \tag{4.12}
$$

where as before, $\mathbf{X}'_t$ is constructed out of the triplet $(\mathbf{X}'_0, \mathbf{J}', \mathbf{M}')$. Now recall from $(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) \in \mathcal{L}_{N,R}$ and Proposition 4.2, that the right-most term in (4.12) is at most $\exp(C_0 \sqrt{RT})$ for all $R \geq R_0$, in which case by the preceding

$$
\begin{aligned}
\sup_{\mathbf{t} \in [0,T]^2} &\left| F(\mathbf{t}; (\mathbf{X}_0, \mathbf{J}, \mathbf{M})) - F(\mathbf{t}; (\mathbf{X}'_0, \mathbf{J}', \mathbf{M}')) \right| \\
&\leq \frac{2 C_{\mathbf{a}} e^{C_0 \sqrt{RT}}}{\sqrt{N}} \sup_{\mathcal{Y} \in \mathfrak{F}, t \in [0,T]} \| \mathcal{Y}(\mathbf{X}_t) - \mathcal{Y}(\mathbf{X}'_t) \|. \tag{4.13}
\end{aligned}
$$

Recall Lemma 4.4 and (4.11), to deduce that for some $C(T, \mathbf{C}_\star) > 0$, every $R \geq R_0$, and all $(\mathbf{X}_0, \mathbf{J}, \mathbf{M})$, we have $(\mathbf{X}'_0, \mathbf{J}', \mathbf{M}') \in \mathcal{L}_{N,R}$,

$$
\sup_{\mathcal{Y} \in \mathfrak{F}, t \in [0,T]} \| \mathcal{Y}(\mathbf{X}_t) - \mathcal{Y}(\mathbf{X}'_t) \| \leq \sqrt{R} e^{C \sqrt{R}} \| (\mathbf{X}_0, \mathbf{J}, \mathbf{M}) - (\mathbf{X}'_0, \mathbf{J}', \mathbf{M}') \|_{\text{MIX}}.
$$

Putting these all together, we deduce that there exists some other $R_0(T, \mathbf{C}_\star)$ and $C(T, C_{\mathbf{a}}, \mathbf{C}_\star)$, such that for all $R \geq R_0(T, \mathbf{C}_\star)$,

$$
\begin{aligned}
\sup_{\substack{(\mathbf{X}_0, \mathbf{J}, \mathbf{M}), (\mathbf{X}'_0, \mathbf{J}', \mathbf{M}') \in \mathcal{L}_{N,R} \\ \mathbf{t} \in [0,T]^2}} &\left| F(\mathbf{t}; (\mathbf{X}_0, \mathbf{J}, \mathbf{M})) - F(\mathbf{t}; (\mathbf{X}'_0, \mathbf{J}', \mathbf{M}')) \right| \\
&\leq \frac{C e^{C \sqrt{R}}}{\sqrt{N}} \| (\mathbf{X}_0, \mathbf{J}, \mathbf{M}) - (\mathbf{X}'_0, \mathbf{J}', \mathbf{M}') \|_{\text{MIX}}. \qquad \square
\end{aligned}
$$

We conclude this subsection by combining the respective exponential concentrations of Lipschitz functions due to μ, $\mathbb{P}_{\mathbf{A}}$ and $\mathbb{P}_{\mathbf{B}}$.

Lemma 4.5 *Suppose that* μ, $\mathbb{P}_{\mathbf{A}}$ *satisfy* Hypothesis 2. *Then* $\mathbb{P} = \mu \otimes \mathbb{P}_{\mathbf{A}} \otimes \mathbb{P}_{\mathbf{B}}$ *satisfies exponential concentration of Lipschitz functions with respect to* $(\mathcal{E}_N, \| \cdot \|_{\text{MIX}})$.

Proof Fix any function f that is 1-Lipschitz on $(\mathcal{E}_N, \|\cdot\|_{\text{MIX}})$. Let us expand

$$f(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) - \mathbb{E}[f(\mathbf{X}_0, \mathbf{J}, \mathbf{M})] = (f(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) - \mathbb{E}_{\mathbf{B}}[f(\mathbf{X}_0, \mathbf{J}, \mathbf{M})])$$
$$+ (\mathbb{E}_{\mathbf{B}}[f(\mathbf{X}_0, \mathbf{J}, \mathbf{M})] - \mathbb{E}_{\mathbf{J},\mathbf{B}}[f(\mathbf{X}_0, \mathbf{J}, \mathbf{M})])$$
$$+ (\mathbb{E}_{\mathbf{J},\mathbf{B}}[f(\mathbf{X}_0, \mathbf{J}, \mathbf{M})] - \mathbb{E}[f(\mathbf{X}_0, \mathbf{J}, \mathbf{M})]),$$

where the subscripts of the expectations indicate which random variables the expectation is taken over. Call the above three differences $I_{\mathbf{M}}$, $I_{\mathbf{J}}$ and $I_{\mathbf{X}_0}$ say. For every $\mathbf{X}_0, \mathbf{J}$ fixed, $f(\mathbf{X}_0, \mathbf{J}, \mathbf{M})$ is 1-Lipschitz in $\mathbf{M} \in C([0, T], \mathbb{R}^N)$ endowed with the norm $\sup_{t \leq T} \|\cdot\|$. As such, from the exponential concentration of Lipschitz functions satisfied by $\mathbb{P}_{\mathbf{B}}$ with respect to $C([0, T], \mathbb{R}^N)$ endowed with $\sup_{t \leq T} \|\cdot\|$ (see e.g., the discussion around [3, Hypothesis 1.1]), there exists $C = C(C_{\boldsymbol{\sigma}}) > 0$ such that for every $r > 0$,

$$\sup_{\mathbf{X}_0, \mathbf{J}} \mathbb{P}_{\mathbf{B}}(|I_{\mathbf{M}}| > r/3) \leq Ce^{-r/C}.$$

Similarly, we have that for every fixed $\mathbf{X}_0$, $\mathbb{E}_{\mathbf{B}}[f(\mathbf{X}_0, \mathbf{J}, \mathbf{M})]$ is 1-Lipschitz in $\mathbf{J}$ endowed with its rescaled Frobenius norm $\sum_{i,j} (\sqrt{N} J_{ij})^2$, and finally, $\mathbb{E}_{\mathbf{J},\mathbf{B}}[f(\mathbf{X}_0, \mathbf{J}, \mathbf{M})]$ is 1-Lipschitz in $\mathbf{X}_0$ endowed with its ℓ^2 norm. Altogether, expanding

$$\mathbb{P}(|f(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) - \mathbb{E}[f(\mathbf{X}_0, \mathbf{J}, \mathbf{M})]| > r)$$
$$\leq \mathbb{E}[\mathbb{P}_{\mathbf{B}}(|I_{\mathbf{M}}| > r/3 \mid \mathbf{X}_0, \mathbf{J})] + \mathbb{E}[\mathbb{P}_{\mathbf{J}}(|I_{\mathbf{J}}| > r/3 \mid \mathbf{X}_0)] + \mu(|I_{\mathbf{X}_0}| > r/3)$$

we see that the exponential concentrations for 1-Lipschitz functions of μ, $\mathbb{P}_{\mathbf{A}}$ and $\mathbb{P}_{\mathbf{B}}$ lift to exponential concentration of $\mathbb{P}$ for functions that are 1-Lipschitz in the triplet $(\mathbf{X}_0, \mathbf{J}, \mathbf{M})$ on $(\mathcal{E}_N, \|\cdot\|_{\text{MIX}})$. □

4.3 Proof of Theorem 2

We first prove a concentration estimate for F at a fixed pair of times $\mathbf{t} \in [0, T]^2$, before extending this to the full trajectory $(F(\mathbf{t}))_{\mathbf{t} \in [0,T]^2}$ by bounding the modulus of continuity of F.

Proposition 4.6 *Suppose* μ, $\mathbb{P}_{\mathbf{A}}$ *satisfy* Hypotheses 1–2. *There exist* $C(T, C_{\mathbf{a}}, \mathbf{C}_{\star}, C_{\mu})$ *large, such that for any F as in* (1.10) *with* $\|\mathbf{a}\|_{\infty} \leq C_{\mathbf{a}}$, $\mathcal{Y}, \mathcal{Y}' \in \mathfrak{F}$, *all* $\mathbf{t} \in [0, T]^2$, $\lambda > 0$ *and* $N \geq N_0(T, C_{\mathbf{a}}, \mathbf{C}_{\star}, C_{\mu})$,

$$\mathbb{P}(|F(\mathbf{t}) - \mathbb{E}[F(\mathbf{t})]| > \lambda) \leq q_N(\lambda) := \begin{cases} Ce^{-\lambda\sqrt{N}/C} + \lambda^{-1}e^{-\sqrt{N}/C}, & \lambda \leq C \\ e^{-(\log \lambda)\sqrt{N}/C}, & \lambda > C \end{cases}.$$
$$(4.14)$$

Proof In proving [3, Lemma 2.5] it is shown, using a Lipschitz extension, that if $\mathbb{P}$ satisfies exponential concentration for Lipschitz functions as in (1.11) and V is an

A-Lipschitz function on a set $\mathcal{L}$ on which $|V|$ is uniformly bounded by K, then for some universal constant $C > 0$ and every $\lambda > 0$,

$$\mathbb{P}(|V - \mathbb{E}[V]| \geq \lambda) \leq Ce^{-\lambda/(2AC)} + \mathbb{P}(\mathcal{L}^c) + \frac{2}{\lambda}(\sqrt{\mathbb{E}[V^2]} + K)\sqrt{\mathbb{P}(\mathcal{L}^c)}.\quad(4.15)$$

Recall from Lemma 4.5 that $\mathbb{P} = \mu \otimes \mathbb{P}_{\mathbf{A}} \otimes \mathbb{P}_{\mathbf{B}}$ satisfies exponential concentration for Lipschitz functions in $(\mathcal{E}_N, \|\cdot\|_{\mathrm{MIX}})$ and Proposition 4.3 that $V = F(\mathbf{t}; \cdot)$ is $\frac{D(R)}{\sqrt{N}}$-Lipschitz on $\mathcal{L} = \mathcal{L}_{N,R}$ for $D(R) = C_1 e^{C_1\sqrt{R}}$, for some $C_1(T, C_{\mathbf{a}}, \mathbf{C}_\star)$ for every $R \geq R_0(T, C_{\mathbf{a}}, \mathbf{C}_\star)$, all N, and every $F, \mathbf{t}$ as in Theorem 2.

Further, increasing R_0 as needed for Lemma 4.1 and Proposition 4.2, yields

$$\sup_{(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) \in \mathcal{L}_{N,R}} |F(\mathbf{t}; (\mathbf{X}_0, \mathbf{J}, \mathbf{M}))| \leq K(R) \quad \text{where}$$

$$K(R) := C_{\mathbf{a}} \max(R, e^{2C_0\sqrt{R}T}),$$

as well as guaranteeing that $C_2^2 := \sup_{N,\mathbf{t}}\{\mathbb{E}[F(\mathbf{t})^2]\}$ is finite and that $\mathbb{P}(\mathcal{L}_{N,R}^c) \leq \exp(-\sqrt{RN}/C_3)$ for some $C_3(T, C_\mu, C_{\mathbf{A}}, C_\sigma)$. Plugging all this into (4.15) gives us a family of upper bounds for $R \geq R_0$,

$$q_N(\lambda; R) = Ce^{-\lambda\sqrt{N}/(2D(R)C)} + e^{-\sqrt{RN}/C_3} + \frac{2}{\lambda}(C_2 + K(R))e^{-\sqrt{RN}/(2C_3)}.$$

For $R = R_0$ we can embed the constant factor $2D(R_0)$ into C and further adjust C_3 to bound the pre-exponent $2(C_2 + K(R_0))$ within the factor $\exp(-\sqrt{R_0N}/(2C_3))$ multiplying it, resulting with $q_N(\lambda; R_0)$ as in the top line on the RHS of (4.14). For a better tail decay, consider $R_\lambda = (\eta \log \lambda)^2 \geq R_0$, with $\eta = 1/(2C_1)$ so $D(R_\lambda) = C_1 e^{C_1\eta \log \lambda} \leq C_1\lambda/\log \lambda$ for all $\lambda \geq 4$. In addition, once $\sqrt{N}/(2C_3) \geq 4C_0T$ we can again embed the pre-exponent $2(C_2 + K(R_\lambda))/\lambda$ within the factor $\exp(-\sqrt{R_\lambda N}/(2C_3))$ multiplying it . Thus, upon adjusting the various constants we end up with $q_N(\lambda; R_\lambda)$ as in the bottom line on the RHS of (4.14). □

Setting hereafter R for the larger of R_0 and R_λ values from the preceding proof of Proposition 4.6, recall that the event $\mathcal{L}_{N,R}^c$ was already ruled out as part of the derivation of (4.14). Thus, proceeding to prove Theorem 2, we fix $\varepsilon = N^{-k}$, $k > 1$, and apply Proposition 4.6 at the $M_N = \lceil TN^k \rceil^2$ grid points $\mathbf{t}_{i,j} = (i\varepsilon, j\varepsilon)$ within $[0, T]^2$, to deduce by the union bound that

$$\mathbb{P}(\mathcal{L}_{N,R}^c) + \mathbb{P}\Big(\sup_{i,j}\big|F(\mathbf{t}_{i,j}) - \mathbb{E}[F(\mathbf{t}_{i,j})]\big| > \lambda, \mathcal{L}_{N,R}\Big) \leq M_N\, q_N(\lambda).$$

It is easy to check that $2M_N q_N(\lambda)$ is further bounded by $p_N(3\lambda)$ of (1.12) once we suitably enlarge the constant C on the RHS of (1.12) relative to that of (4.14). In addition, since the right-most term in (4.15) exceeds one whenever $\mathbb{E}[|V|\mathbf{1}_{\mathcal{L}_{N,R}^c}] = $

$\mathbb{E}[|F(\mathbf{t})|\mathbf{1}_{\mathcal{L}_{N,R}^c}] \geq \lambda/2$, if that inequality holds for any $\mathbf{t} \in [0,T]^2$, then $q_N(\lambda)$ and in turn $p_N(3\lambda)$ of (1.12) would exceed one. Thus, we may assume WLOG that

$$\sup_{\mathbf{t},\mathbf{s}:\mathbf{t}+\mathbf{s}\in[0,T]^2} \{\mathbb{E}[|F(\mathbf{t}+\mathbf{s}) - F(\mathbf{t})|\,\mathbf{1}_{\mathcal{L}_{N,R}^c}]\} \leq 2 \sup_{\mathbf{t}\in[0,T]^2} \mathbb{E}[|F(\mathbf{t})|\mathbf{1}_{\mathcal{L}_{N,R}^c}] \leq \lambda \quad (4.16)$$

We can then expand

$$\mathbb{P}\Big(\sup_{\mathbf{t}\in[0,T]^2} \big|F(\mathbf{t}) - \mathbb{E}[F(\mathbf{t})]\big| > 3\lambda \Big)$$

$$\leq \mathbb{P}(\mathcal{L}_{N,R}^c) + \mathbb{P}\Big(\sup_{i,j} |F(\mathbf{t}_{i,j}) - \mathbb{E}[F(\mathbf{t}_{i,j})]| > \lambda, \mathcal{L}_{N,R} \Big)$$

$$+ M_N \sup_{i,j} \mathbb{P}\Big(\sup_{\mathbf{s}\in[0,\varepsilon]^2} |F(\mathbf{t}_{i,j} + \mathbf{s}) - F(\mathbf{t}_{i,j})| > \lambda, \mathcal{L}_{N,R} \Big)$$

$$+ \mathbf{1}\Big\{ \sup_{\mathbf{t},\mathbf{s}} \mathbb{E}[|F(\mathbf{t}+\mathbf{s}) - F(\mathbf{t})|\mathbf{1}_{\mathcal{L}_{N,R}}] > \lambda \Big\} .$$

Restricting to $\lambda > 1/\sqrt{N}$ (as otherwise $p_N(3\lambda) \geq 1$), and using $p_N(3\lambda) \gg M_N \exp(-(\lambda^2 \wedge \lambda)N^k/C')$ (as $k > 1$) with the above, the stated bound of Theorem 2, follows from the following short-time estimates.

Lemma 4.7 *There exists $C'(C_\sigma)$, such that for every $\varepsilon \leq 1$, $\lambda \geq C'\varepsilon$, and F as in Theorem 2,*

$$\sup_{\mathbf{t}\in[0,T-\varepsilon]^2} \mathbb{P}\Big(\sup_{\mathbf{s}\in[0,\varepsilon]^2} |F(\mathbf{t}+\mathbf{s}) - F(\mathbf{t})| > \lambda, \mathcal{L}_{N,R} \Big) \leq 2e^{-(\lambda^2 \wedge \lambda)/(C'\varepsilon)} . \quad (4.17)$$

In particular, for any $N \geq N_0(T, C_\mathbf{a}, C_\mu, \mathbf{C}_\star)$ and $\lambda \geq N^{-1/2} = \varepsilon^{1/(2k)}$, $k > 1$,

$$\sup_{\mathbf{t}\in[0,T-\varepsilon]^2,\mathbf{s}\in[0,\varepsilon]^2} \mathbb{E}\big[|F(\mathbf{t}+\mathbf{s}) - F(\mathbf{t})|\mathbf{1}_{\mathcal{L}_{N,R}}\big] \leq \lambda . \quad (4.18)$$

Proof Similarly to the computation leading to (4.13), we find that for any $\mathbf{t}+\mathbf{s} \in [0,T]^2$ and F as in Theorem 2, evaluated on the solution $\mathbf{X}_t(\mathbf{X}_0, \mathbf{J}, \mathbf{M})$ that corresponds to some $(\mathbf{X}_0, \mathbf{J}, \mathbf{M}) \in \mathcal{L}_{N,R}$

$$|F(\mathbf{t}+\mathbf{s}) - F(\mathbf{t})| \leq \frac{2C_\mathbf{a}e^{C_0\sqrt{R}T}}{\sqrt{N}} \max_{\mathcal{Y}\in\mathfrak{F}} \max_{i=1,2} \{\|\mathcal{Y}(t_i + s_i) - \mathcal{Y}(t_i)\|\} .$$

When $\mathcal{Y} = \mathbf{1}$ this difference is zero, whereas in case $\mathcal{Y} = \mathbf{X}$ and $s_i \leq \varepsilon$, assuming WLOG that $R_0, \mathbf{C}_\star \geq 1$, we have on $\mathcal{L}_{N,R}$, by (4.5) and the RHS of (4.8), that

$$\|\mathbf{X}_{t_i+s_i} - \mathbf{X}_{t_i}\| \leq \|\mathbf{M}_{t_i+s_i} - \mathbf{M}_{t_i}\| + \int_{t_i}^{t_i+s_i} [\|\mathbf{G}_u\| + C_\Lambda\|\mathbf{X}_u\| + \sqrt{N}C_\mathbf{h}]\,du$$

$$\leq \|\mathbf{M}_{t_i+s_i} - \mathbf{M}_{t_i}\| + 3\varepsilon\sqrt{N}\mathbf{C}_\star e^{C_0\sqrt{R}T} . \quad (4.19)$$

Further, similarly to the LHS of (4.11), on $\mathcal{L}_{N,R}$,

$$\|\mathbf{G}_{t_i+s_i} - \mathbf{G}_{t_i}\| \leq \|\mathbf{J}\|_{2\to 2}\|\mathbf{X}_{t_i+s_i} - \mathbf{X}_{t_i}\| \leq \sqrt{R}\|\mathbf{X}_{t_i+s_i} - \mathbf{X}_{t_i}\|,$$

so up to extra factor $\sqrt{R}$ the bound (4.19) applies for $\mathcal{Y} = \mathbf{G}$, and considering all cases we get for $\mathbf{s} \in [0,\varepsilon]^2$,

$$|F(\mathbf{t}+\mathbf{s}) - F(\mathbf{t})| \leq \frac{2C_{\mathbf{a}}\sqrt{R}e^{C_0\sqrt{R}T}}{\sqrt{N}} \max_{i=1,2} \|\mathbf{M}_{t_i+s_i} - \mathbf{M}_{t_i}\|$$
$$+ 6\varepsilon C_{\mathbf{a}}\mathbf{C}_\star \sqrt{R}e^{2C_0\sqrt{R}T}. \tag{4.20}$$

For some $C' > 0$, when $R = R_0$ and $\lambda \geq C'\varepsilon$, the right most term in (4.20) can not exceed $\lambda/2$. The same applies for $R = R_\lambda = (\eta \log \lambda)^2$ provided $\eta \leq 1/(3C_0T)$. By the same reasoning, for such η and some $C_4(T, C_{\mathbf{a}}, R_0) > 0$, the factor multiplying $\|\mathbf{M}_{t_i+s_i} - \mathbf{M}_{t_i}\|$ in (4.20), is in both cases at most $(\sqrt{\lambda}\vee 1)/(2C_4\sqrt{N})$. Recall from (4.2) and the stationarity of Brownian increments, that there exists $C(C_{\boldsymbol{\sigma}})$ such that for every $L \geq \varepsilon^2 L_0(C_{\boldsymbol{\sigma}})$, every N,

$$\sup_{t\in[0,T-\varepsilon]} \mathbb{P}_{\mathbf{B}}\Big(\sup_{s\in[0,\varepsilon]}\{\|\mathbf{M}_{t+s} - \mathbf{M}_t\|\} > L\sqrt{N}\Big) \leq e^{-3L^2/(C\varepsilon)}. \tag{4.21}$$

Combining (4.20) and (4.21), we thus get that for some $C'(C_{\boldsymbol{\sigma}})$, for every $\lambda \geq C'\varepsilon$, and every N, $\mathbf{t} = (t_1, t_2)$,

$$\mathbb{P}\Big(\sup_{\mathbf{s}\in[0,\varepsilon]^2} |F(\mathbf{t}+\mathbf{s}) - F(\mathbf{t})| > \lambda, \ \mathcal{L}_{N,R}\Big)$$
$$\leq 2\max_{i=1,2} \mathbb{P}\Big(\sup_{s\in[0,\varepsilon]} \|\mathbf{M}_{t_i+s} - \mathbf{M}_{t_i}\| > C_4(\lambda \wedge \sqrt{\lambda})\sqrt{N}\Big)$$
$$\leq 2e^{-(\lambda^2\wedge\lambda)/(C'\varepsilon)},$$

as claimed in (4.17). Next, by Cauchy-Schwarz, (4.7) and (4.17), there exists $C(T, C_{\mathbf{a}}, C_\mu, \mathbf{C}_\star)$ such that for every $N \geq N_0(T, C_{\mathbf{a}}, C_\mu, \mathbf{C}_\star)$, every $\lambda \geq 2C'\varepsilon$, every $\mathbf{t}, \mathbf{s}$ and all F,

$$\mathbb{E}\big[|F(\mathbf{t}+\mathbf{s}) - F(\mathbf{t})|\mathbf{1}_{\mathcal{L}_{N,R}}\big] \leq \frac{\lambda}{2} + 2\mathbb{P}\big(|F(\mathbf{t}+\mathbf{s}) - F(\mathbf{t})| > \frac{\lambda}{2}, \ \mathcal{L}_{N,R}\big)^{1/2}$$
$$\sup_{\mathbf{t}\in[0,T]^2}\{\sqrt{\mathbb{E}[F(\mathbf{t})^2]}\}$$
$$\leq \frac{\lambda}{2} + Ce^{-(\lambda^2\wedge\lambda)/(4C'\varepsilon)}.$$

Our assumption that $\lambda \geq \varepsilon^{1/(2k)}$ for some $k > 1$ guarantees that the right most term is at most $\lambda/2$ (as soon as $N \geq N_0$), thereby establishing (4.18). $\qquad\square$

References

1. Bayati, M., Lelarge, M., Montanari, A.: Universality in polytope phase transitions and message passing algorithms. Ann. Appl. Probab. **25**(2), 753–822 (2015). https://doi.org/10.1214/14-AAP1010
2. Ben Arous, G., Dembo, A., Guionnet, A.: Aging of spherical spin glasses. Probab. Theory Related Fields **120**(1), 1–67 (2001). https://doi.org/10.1007/PL00008774
3. Ben Arous, G., Dembo, A., Guionnet, A.: Cugliandolo-Kurchan equations for dynamics of spin-glasses. Probab. Theory Related Fields **136**(4), 619–660 (2006). https://doi.org/10.1007/s00440-005-0491-y
4. Ben Arous, G., Gheissari, R., Jagannath, A.: Algorithmic thresholds for tensor PCA. Ann. Probab. **48**(4), 2052–2087 (2020). https://doi.org/10.1214/19-AOP1415
5. Ben Arous, G., Gheissari, R., Jagannath, A.: Bounding flows for spherical spin glass dynamics. Commun. Math. Phys. **373**(3), 1011–1048 (2020). https://doi.org/10.1007/s00220-019-03649-4
6. Ben Arous, G., Guionnet, A.: Large deviations for Langevin spin glass dynamics. Probab. Theory Related Fields **102**(4), 455–509 (1995). https://doi.org/10.1007/BF01198846
7. Ben Arous, G., Guionnet, A.: Symmetric Langevin spin glass dynamics. Ann. Probab. **25**(3), 1367–1422 (1997). https://doi.org/10.1214/aop/1024404517
8. Chatterjee, S.: A simple invariance theorem Available at arXiv:math/0508213 (2005)
9. Chatterjee, S.: A generalization of the Lindeberg principle. Ann. Probab. **34**(6), 2061–2076 (2006). https://doi.org/10.1214/009117906000000575
10. Chen, W.K., Lam, W.K.: Universality of approximate message passing algorithms. arXiv:2003.10431 (2020)
11. Crisanti, A., Horner, H., Sommers, H.J.: The spherical p-spin interaction spin-glass model. Zeitschrift für Physik B Condensed Matter **92**(2), 257–271 (1993). https://doi.org/10.1007/BF01312184
12. Crisanti, A., Sommers, H.J.: The spherical *p*-spin interaction spin glass model: the statics. Zeitschrift für Physik B Condensed Matter **87**(3), 341–354 (1992). https://doi.org/10.1007/BF01309287
13. Crisanti, A., Sompolinsky, H.: Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model. Phys. Rev. A **36**, 4922–4939 (1987). https://doi.org/10.1103/PhysRevA.36.4922
14. Cugliandolo, L.F., Dean, D.S.: Full dynamical solution for a spherical spin-glass model. J. Phys. A: Math. Gen. **28**(15), 4213–4234 (1995). https://doi.org/10.1088/0305-4470/28/15/003
15. Cugliandolo, L.F., Kurchan, J.: Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. Phys. Rev. Lett. **71**, 173–176 (1993). https://doi.org/10.1103/PhysRevLett.71.173
16. Dembo, A., Guionnet, A., Mazza, C.: Limiting dynamics for spherical models of spin glasses at high temperature. J. Stat. Phys. **128**(4), 847–881 (2007). https://doi.org/10.1007/s10955-006-9239-z
17. Dembo, A., Lubetzky, E., Zeitouni, O.: Universality for Langevin-like spin glass dynamics. Ann. Appl. Probab. (2021)
18. Dembo, A., Subag, E.: Dynamics for spherical spin glasses: disorder dependent initial conditions. J. Stat. Phys. **181**(2), 465–514 (2020). https://doi.org/10.1007/s10955-020-02587-z
19. Erdős, L., Schlein, B., Yau, H.T.: Universality of random matrices and local relaxation flow. Inventiones mathematicae **185**(1), 75–119 (2011). https://doi.org/10.1007/s00222-010-0302-7

20. Erdős, L., Yau, H.: A Dynamical Approach to Random Matrix Theory. Courant Lecture Notes. Courant Institute of Mathematical Sciences, New York University (2017). https://books.google.com/books?id=6kUzDwAAQBAJ
21. Gromov, M., Milman, V.D.: A topological application of the isoperimetric inequality. Am. J. Math. **105**(4), 843–854 (1983)
22. Guerra, F., Toninelli, F.L.: The thermodynamic limit in mean field spin glass models. Commun. Math. Phys. **230**(1), 71–79 (2002). https://doi.org/10.1007/s00220-002-0699-y
23. Guionnet, A.: Dynamics for spherical models of spin-glass and aging. In: Spin glasses, Lecture Notes in Math., vol. 1900, pp. 117–144. Springer, Berlin (2007). https://doi.org/10.1007/978-3-540-40908-3_5
24. Hardy, G.H., Ramanujan, S.: Asymptotic formulaæ in combinatory analysis. Proc. Lond. Math. Soc. **s2–17**(1), 75–115 (1918). https://doi.org/10.1112/plms/s2-17.1.75
25. Hertz, J.A., Grinstein, G., Solla, S.A.: Irreversible spin glasses and neural networks. In: van Hemmen, J.L., Morgenstern, I. (eds.) Heidelberg Colloquium on Glassy Dynamics, pp. 538–546. Springer, Berlin (1987)
26. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA **79**(8), 2554–2558 (1982). https://doi.org/10.1073/pnas.79.8.2554
27. Kinzel, W.: Neural networks with asymmetric bonds. In: van Hemmen, J.L., Morgenstern, I. (eds.) Heidelberg Colloquium on Glassy Dynamics, pp. 529–537. Springer, Berlin (1987)
28. Latala, R.: Some estimates of norms of random matrices. Proc. Am. Math. Soc. **133**(5), 1273–1282 (2005)
29. Mossel, E., O'Donnell, R., Oleszkiewicz, K.: Noise stability of functions with low influences: Invariance and optimality. In: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science, FOCS '05, p. 21–30. IEEE Computer Society, USA (2005). https://doi.org/10.1109/SFCS.2005.53
30. Nagaev, A.V.: Integral limit theorems taking large deviations into account when Cramer's condition does not hold. i. Theory Probab. Appl. **14**(1), 51–64 (1969). https://doi.org/10.1137/1114006
31. Øksendal, B.: Stochastic Differential Equations: An Introduction with Applications. Hochschultext / Universitext. Springer (2003)
32. Panchenko, D.: The Parisi ultrametricity conjecture. Ann. Math. **177**(1), 383–393 (2013). https://doi.org/10.4007/annals.2013.177.1.8
33. Panchenko, D.: The Sherrington-Kirkpatrick Model. Springer, Berlin (2013)
34. Rotar', V.I.: Limit theorems for polylinear forms. J. Multivar. Anal. **9**(4), 511–530 (1979)
35. Talagrand, M.: Gaussian averages, Bernoulli averages, and Gibbs' measures. Random Struct. Algorithms **21**(3), 197–204 (2002). https://doi.org/10.1002/rsa.10059
36. Talagrand, M.: The Parisi formula. Ann. Math. **163**(1), 221–263 (2006)
37. Tao, T., Vu, V.: Random matrices: Universality of local eigenvalue statistics. Acta Math. **206**(1), 127–204 (2011). https://doi.org/10.1007/s11511-011-0061-3
38. Tao, T., Vu, V.: Random matrices: The Universality phenomenon for Wigner ensembles Available at arXiv:1202.0068 (2012). https://cds.cern.ch/record/1421029
39. Vershynin, R.: High–Dimensional Probability. Cambridge University Press (to appear) (2018)
40. Wigner, E.P.: On the distribution of the roots of certain symmetric matrices. Ann. Math. **67**(2), 325–327 (1958)
41. Xu, Z.B., Hu, G.Q., Kwong, C.P.: Asymmetric Hopfield-type networks: theory and applications. Neural Netw. **9**(3), 483–501 (1996). https://doi.org/10.1016/0893-6080(95)00114-X