Bounding Flows for Spherical Spin Glass Dynamics

Gérard Ben Arous¹,², Reza Gheissari¹,³, Aukosh Jagannath⁴,⁵

¹ Courant Institute, New York University, New York, USA
² NYU Shanghai, Pudong, China. E-mail: benarous@cims.nyu.edu
³ University of California, Berkeley, Berkeley, USA. E-mail: gheissari@berkeley.edu
⁴ Harvard University, Cambridge, USA
⁵ University of Waterloo, Waterloo, Canada. E-mail: a.jagannath@uwaterloo.ca

Received: 8 August 2018 / Accepted: 19 October 2019
Published online: 6 December 2019 – © Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract: We introduce a new approach to studying spherical spin glass dynamics based on differential inequalities for one-time observables. Using this approach, we obtain an approximate phase diagram for the evolution of the energy $H$ and its gradient under Langevin dynamics for spherical $p$-spin models. We then derive several consequences of this phase diagram. For example, at any temperature, uniformly over all starting points, the process must reach and remain in an absorbing region of large negative values of $H$ and large (in norm) gradients in order 1 time. Furthermore, if the process starts in a neighborhood of a critical point of $H$ with negative energy, then both the gradient and energy must increase macroscopically under this evolution, even if this critical point is a saddle with index of order $N$. As a key technical tool, we estimate Sobolev norms of spin glass Hamiltonians, which are of independent interest.

1. Introduction

We introduce here a simple new way to study the dynamics of spherical spin glasses that is relevant on short time scales. We apply this method to the Langevin dynamics for the spherical $p$-spin model. In a companion work, we use this approach to study Langevin dynamics for Tensor PCA and the related signal recovery problem [12].

Let $S^N_S = S^{N-1}(\sqrt{N}) = \{x \in \mathbb{R}^N : \sum x_i^2 = N\}$, denote the $(N - 1)$-sphere in dimension $N$ of radius $\sqrt{N}$, and consider the $p$-spin Hamiltonian

$$H(x) = \frac{1}{N(p-1)/2} \sum_{i_1,\ldots,i_p=1}^{N} J_{i_1,\ldots,i_p} x_{i_1} \cdots x_{i_p},$$

where the coefficients, $J_{i_1,\ldots,i_p}$, are i.i.d. standard Gaussian random variables. Observe that this function is typically of order $\sqrt{N}$, while its extreme values are of order $N$. Recall that these models are prototypical examples of complex energy landscapes: they have
exponentially many critical points of every index \([1,2,41]\). We study here the Langevin dynamics corresponding to this Hamiltonian, namely the Markov process \(X_t\) which is the solution to the stochastic differential equation

\[
\begin{cases}
  dX_t = \sqrt{2} dB_t - \beta \nabla H(X_t) dt \\
  X_0 = x_0
\end{cases},
\]

where \(B_t\) is spherical Brownian motion, \(\nabla H\) is the spherical gradient of \(H\), and \(\beta > 0\) is the inverse temperature. This process is reversible and ergodic with stationary measure given by the Gibbs measure, \(d\pi_{\beta,N}(x) \propto e^{-\beta H(x)} dx\), where \(dx\) is the normalized volume measure on \(S^N\).

The dynamics of mean-field models of spin glasses has received a tremendous amount of attention in both the physics and mathematics literatures. This history is too rich to be summarized here and we refer the reader to the general surveys \([15,19,26]\) in the physics literature and \([5,32]\) in the mathematics literature. One regime of interest in the study of spin glass dynamics are exponential timescales at low temperatures, as this is the timescale to equilibrium \([14,30]\). For an overview of related work on spin-glass dynamics on such “activated” timescales, see Sect. 1.3 below.

In this paper, we are interested in the dynamics on much shorter timescales, namely timescales that are order 1 in \(N\). The classical approach to studying Langevin dynamics of spherical spin glasses on these timescales is via the “Cugliandolo–Kurchan equations”. These govern the evolution of certain natural two-time observables via a system of integro-differential equations \([10,11,25,27,28]\). At this time, however, it seems challenging to use these techniques to answer the following kinds of natural questions.

**Question 1.** (Going down quickly). Does the dynamics reach and remain at a macroscopic fraction of the global minimum energy, i.e., energies of order \(N\), in short time?

**Question 2.** (Visiting critical points). Does the dynamics visit and get slowed by critical points?

**Question 3.** (Escaping critical points). When started in a neighborhood of a critical point, can the dynamics escape in finite time, and, if so, in which way?

**Question 4.** (Varying initial data). How do the answers to these questions change as we vary the initial energy and gradient?

Our goal in this paper is to provide an elementary approach to obtain precise answers to these questions. As an application of our approach, we obtain the following answers to the above.

1. There is a constant \(T_0\) such that for all initial data, the energy must reach the extreme, order \(N\), scale by time \(T_0\), and remain there. Moreover, we provide explicit bounds on the fraction of the ground state reached depending on \(p\) and \(\beta\).
2. This absorbing region is a region of macroscopically large gradients (norm of order \(\sqrt{N}\)). In particular, the dynamics does not come close to critical points in order 1 times.
3. When started from any critical point of \(H\), the dynamics must increase instantaneously in gradient. When that critical point has negative energy, the dynamics also increases in energy at the order \(N\) scale. Perhaps surprisingly, this happens even when the critical point has diverging index. Put simply, the dynamics must climb high dimensional saddles of negative energy for an order 1 time.
One can obtain sharper results for any fixed energy and norm of gradient of the initial data.

The core of our approach is what we call bounding flows. We analyze the evolution of

$$U_N(t) = (u_N(t), v_N(t)) = \left( -\frac{H(X_t)}{N}, \frac{|\nabla H(X_t)|^2}{N} \right) ,$$

from any initial point, for any finite $T$. We do this by studying all possible weak limit points of the laws of $U_N$ in the space of probability measures, $\mathcal{M}_1(C([0, T])^2)$. To this end, we develop an elementary, abstract approach to show that these limit points exist and are $C^1$, by reducing this tightness, to estimating a Sobolev norm of $H$, which we call the $G$-norm (see Definition 4.1). This norm is the natural norm in which to study high dimensional problems of this type and plays an important role in the accompanying [12]. We estimate it using tools from Gaussian comparison theory and differential geometry in Sect. 4. We then show that these observables form what we call a quasi-autonomous family. Roughly, this is a system of observables, which is closed but for an additional set of “auxiliary observables” which themselves have limits. (For a precise definition see Sect. 3.) In the case of the variables $u_N, v_N$, there is a single auxiliary observable which is related to the alignment of the gradient of $H$ with the eigenvectors of its Euclidean Hessian along the flow. We then uniformly bound the auxiliary observable by “truncating” the Euclidean Hessian, yielding a pair of differential inequalities for $\dot{U}$ only in terms of the pair $(u, v)$. This allows us to confine our dynamics for any choice of initial energy and gradient by the flowlines for the corresponding autonomous systems. In this paper, our truncation is elementary and one can imagine continuing this program by treating the Hessian as a part of our “quasi-autonomous” family. This would yield new auxiliary variables and thereby sharpen the results.

1.1. Bounding flows. We begin by proving the following differential inequalities for the evolution of $U$ in Sect. 5 (see Sect. 2.1 for a sketch of the derivation). In the following we equip $C([0, T])^k$ with the norm topology.

Theorem 1.1. For every $p, \beta, T$ and every sequence of initial points $x_N \in \mathcal{S}^N$, the laws of $U_N(t)$ from (1.3) are tight in $\mathcal{M}_1(C([0, T])^2)$. Moreover, every limit $U(t) = (u(t), v(t))$ is almost surely continuously differentiable and satisfies

$$\dot{u} = \mathcal{F}_1(u, v) \quad \text{and} \quad \mathcal{F}_2^L(u, v) \leq \dot{v} \leq \mathcal{F}_2^U(u, v),$$

where for some $\Lambda_p < \infty$ (see Remark 1.3 for explicit bounds),

$$\mathcal{F}_1(u, v) = -pu + \beta v,$$
$$\mathcal{F}_2^L(u, v) = 2p(p - 1) - 2(p - 1)v + 2pu(pu - \beta v) - 2\beta \Lambda_p v,$$
$$\mathcal{F}_2^U(u, v) = 2p(p - 1) - 2(p - 1)v + 2pu(pu - \beta v) + 2\beta \Lambda_p v.$$  

More broadly, this enables us to bound the trajectory of $U_N(t)$ as we vary the initial data. The differential inequalities above translate to precise inequalities for the integral curves of the corresponding dynamical systems. Indeed, for any initial data $U(0)$, the integral curves of $U(t)$ are almost surely confined to the region in between the corresponding integral curves of the lower and upper bounding differential systems of
1.2. Answers to the four questions. To understand the consequences of Theorem 1.1, let us now state more precisely the answers to the four questions from Sect. 1. We present the full phase diagram from Fig. 1 in Sect. 2 and prove it in Sect. 7.
Throughout the paper, let \( P \) denote the law of \( H \) and \( Q_x \) denote the law of \( X_t \) started from \( X_0 = x \). As an answer to Questions 1 and 2, we obtain the following.

**Theorem 1.2** (Going down quickly and avoiding critical points). For every \( \beta > 0 \) and \( p > 2 \), there exists an explicit \( u_c > 0 \) such that the following holds. For every \( \epsilon > 0 \), there exists \( T_0(\epsilon) > 0 \) such that \( P \)-almost surely,

\[
\lim_{N \to \infty} \inf_{x \in \mathbb{S}^N} Q_x \left( \left\{ H(X_t) \leq - (u_c - \epsilon) N \text{ and } |\nabla H(X_t)|^2 > \left( \frac{pu_c}{\beta} - \epsilon \right) N \text{ for all } t \in [T_0, T] \right\} \right) = 1, \tag{1.6}
\]

for every \( T > T_0 \).

Before discussing the remaining questions, we pause to make the following important remark on the constants \( \Lambda_p, u_c \) and the dependence of \( u_c \) on \( \beta \).

**Remark 1.3.** Let \( G \) be the restriction of the Euclidean Hessian of \( H \) to the tangent space, \( T_x \mathbb{S}^N \) [see (4.10)]; then \( \Lambda_p \) in (1.5) can be taken to be any constant satisfying

\[
\Lambda_p > \limsup_{N \to \infty} \mathbb{E} \left[ \sup_{x \in \mathbb{S}^N} |G(x)|_{op} \right], \tag{1.7}
\]

where \( |G|_{op} \) denotes the operator norm for \( G \) when viewed as a bilinear form on \( T_x \mathbb{S}^N \). An explicit, but suboptimal, admissible choice of \( \Lambda_p \) is

\[
\Lambda_p = \sqrt{p(p-1)(\sqrt{2} + E_{0,p-2})}, \quad \text{where} \quad E_{0,p} := \lim_{N \to \infty} \left\| \frac{H_{N,p}}{N} \right\|_{L^\infty}, \tag{1.8}
\]

seen to satisfy (1.7) via standard Gaussian comparison inequalities [30]. (For \( p = 1, 2 \) we have \( E_{0,1} = 1 \) and \( E_{0,2} = \sqrt{2} \); for \( p \geq 3 \), see [1,24,34] for explicit formulae for the ground state energy).

For any choice of \( \Lambda_p \) and \( \beta > 0 \), the minimal energy reached is bounded above by \(-u_c N\) where

\[
u_c = u_c(\beta) = \beta \left( 1 + \frac{\beta \Lambda_p}{p-1} \right)^{-1}. \tag{1.9}
\]

Using the sub-optimal choice of \( \Lambda_p \) from (1.8), we can plug in for various values of \( p \), and obtain explicit upper bounds on the energy the process reaches and remains at for all order one times. In particular, if we compare the low-temperature values of \(-u_c N\) to the conjectured threshold energy \(-E_{\infty,p} N\) (the energy at which low-temperature Langevin dynamics is believed to get stuck on order one times when initialized uniformly at random) the choice of \( \Lambda_p \) from (1.8) ensures that uniformly over all initializations, the energy reaches the following fraction of the threshold in order 1 times:

\[
\sup_\beta \frac{u_c(\beta)}{E_{\infty,p}} = \frac{1}{2(\sqrt{2} + E_{0,p-2})}. \tag{1.10}
\]

We plot the resulting lower bounds on \( u_c/E_{\infty,p} \) as a function of \( \beta \) for various choices of \( p \) in Fig. 3.
Fig. 3. Lower bounds on the fraction of the ground state (left) and threshold (right) energies, that we guarantee $X_t$ reaches in $O(1)$ time, uniformly over all $X_0$.

One can in fact obtain a more precise absorbing region for the dynamics which holds on the same time scale $T_0$ as that described in Theorem 1.2. This absorbing region is related to the fixed points of the bounding flow and the transit between these fixed points by the flows. This is illustrated in Fig. 1 and described in detail in Sect. 2.2. We note here, importantly, that the bounding “ellipsoid” from Fig. 1, shrinks to a point as $\beta \to 0$, whereas when $\beta \to \infty$, the upper curve converges to the vertical line, $u = u_c$, defined by the event in Theorem 1.2 (that is, the fixed point for the “upper bounding flow” diverges at low temperature).

Let us now turn to the behavior of the dynamics near critical points. As a second consequence of the bounding flows approach, we prove that critical points of negative energy—including those with indexes of order $N$—repel $X_t$ to regions of larger gradients as well as larger energies. The fact that the distance to the critical point is repulsive is explained by the fact that diffusions do not hit small balls in high dimensions; but the fact that the energy increases rapidly even when the Hessian has $(\frac{1}{2} - \epsilon)N$ negative eigenvalues—and thus directions of lower energy—is surprising.

To be precise, for $(u_0, v_0) \in \mathbb{R}_+^2$, let

$$\Omega_N(u_0, v_0) = \{ x \in S^N : H(x) < -u_0 N, \ |\nabla H(x)|^2 < v_0 N \}. \quad (1.10)$$

We obtain the following answer to the third question.

**Theorem 1.4 (Climbing saddles and wells).** Fix any $\beta > 0$, $p > 2$. For every $\eta > 0$ and every $\delta < p n^p \beta^{-1}$, there exists $c_1, c_2, \rho > 0$, such that $\mathbb{P}$-almost surely,

$$\lim_{N \to \infty} \inf_{x \in \Omega_N(\eta, \delta)} Q_x \left( \left\{ \frac{H(X_t) - H(X_0)}{|\nabla H(X_t)|^2 - |\nabla H(X_0)|^2} > c_2 t N \text{ for all } t \in [0, \rho] \right\} \right) = 1.$$

In fact, it is clear from the above that, not only does the dynamics leave critical regions quickly, it cannot even get close to them. For more on this, see Sect. 7.

As an answer to the fourth question, we can obtain trajectory-wise and pointwise comparisons as consequences of the differential inequalities allowing us to confine $U(t)$ started from $U(0)$ by the bounding flows started from the same initial data (see Sect. 6).

We end this section with the following important remarks.

**Remark 1.5.** Recall that $-E_{\infty, p} N$ is the threshold energy as in [2], i.e., the energy level below which the expected number of local minima diverges exponentially in $N$. It is believed in the physics literature (see e.g., [16]) that in order 1 times, starting from a uniform point on $S^N$, the complexity threshold $-E_{\infty, p} N$ is where $H(X_t)$ gets stuck...
and where short-time aging is exhibited, as the critical points there have Hessians with very few negative eigenvalues, each of small magnitude.

Theorem 1.2 and Theorem 1.4, do not precisely identify where (or if) $H(X_t)$ gets stuck on order 1 time scales. However, they indicate that if the slowdown and aging at $-E_{\infty,p}N$ are due to the few flat directions that “point down” near critical points, $X_t$ needs to maintain a large modulus of its gradient throughout its descent, and needs to stay macroscopically far from the critical points themselves.

Remark 1.6. We note here the following interesting observation about the differential system in Theorem 1.1. Under the ansatz that $G(\nabla H, \nabla H)(X_t) = O(\sqrt{N})$ for all order one times, $\Lambda_p$ disappears, the differential inequalities become equalities, and $(u_N, v_N)$ becomes fully autonomous. In this case, the resulting fixed point is at $(\beta, p)$. Surprisingly, this corresponds to $\langle (-H, (\nabla H)^2_N) \rangle$ in the entire replica symmetric regime $(\beta < \beta_c)$, where $\langle \cdot \rangle$ denotes expectation under $\pi$. In particular, under this ansatz, this would occur even in a window $\beta \in (\beta_d, \beta_c)$ where there is slow mixing.

Given that $\nabla H(x)$ and $G(x)$ can be shown to be independent, one can show that if $X_0$ is chosen uniformly over $S^N$, the ansatz holds for $t = 0$; moreover the ansatz is true under the Gibbs measure $\pi$ whenever $\beta < \beta_c$. One might thus hope to show that when $\beta$ is small, from a start chosen uniformly at random, $G(\nabla H, \nabla H)(X_t)$ remains microscopic throughout the trajectory of $X_t$ until equilibration, and the system $(u_N(t), v_N(t))$ closes. We note that this cannot be the case when $\beta > \beta_c$, where the Gibbs expectation of $G(\nabla H, \nabla H)(x)$ must in fact be order $N$ to guarantee that $\langle (-H, (\nabla H)^2_N) \rangle$ is still a fixpoint of the system when $\langle -H_N \rangle \neq \beta$.

If desired, this could be extended further by starting from a point chosen uniformly at random amongst points of a given energy and modulus of gradient. Again, the independence of $G$ from both $H$ and $\nabla H$ would imply that the ansatz above holds with high probability at $t = 0$. For such initializations, including the uniform at random initialization, one could then improve the bounding flows and narrow the corridors to which the dynamics are confined in short times, by using the bound $|G(\nabla H, \nabla H)(X_t)| \leq \Lambda_p^t \nu$ instead of the sup-norm bound of $\Lambda_p \nu$ used in (1.5).

1.3. Previous results. The dynamics of mean field spin glass models are expected to exhibit a deep and rich structure on both short and long timescales. At the level of convergence to equilibrium, when $\beta$ is small, the measure $\pi$ admits a logarithmic Sobolev inequality with constant $c > 0$ which is order 1 in $N$ [30], so the dynamics converges to $\pi$ in order 1 time. For large $\beta$, if $L$ is the infinitesimal generator of $X_t$, the spectral gap of $-L$ decays exponentially in $N$ [14,30] so that the process $X_t$ reaches equilibrium only after a time that is exponential in $N$. For a physics approach to the analysis of the spectrum of $-L$, see [17]. For studies of the spectrum and spectral gap in related spin glass models see, [4,21,38].

At low temperatures, there has been extensive research on the behavior of glassy dynamics on timescales that are exponentially large but shorter than the time to equilibrium, sometimes referred to as activated dynamics. On such timescales, glassy dynamics have traditionally been found to exhibit the important feature of aging. Following the influential works of [18,20] in the physics literature, this question was studied for Random Hopping time dynamics of the Random Energy Model (REM) in [7–9], and for the $p$-spin model in [6,13,22]. More recently, aging has been established for Metropolis dynamics of the REM in the recent important works [23,29]. For an analysis of Glauber dynamics of the REM in the physics literature, see [3].
These dynamics are also interesting in timescales much shorter than the equilibration time, where the dynamics is expected to exhibit the phenomenon of short-time aging at low temperatures. This picture was first suggested in the physics literature in [39]. For spherical \( p \)-spin models, the classical way to study dynamics in this regime is to introduce the two-time autocorrelation

\[
C_N(t, s) = \frac{1}{N} \sum_i X_t^i X_s^i,
\]

corresponding to the overlap between \( X_t \) and \( X_s \). It is possible to find a second two-time function, the response function \( R_N(t, s) \), such that the pair \((C_N, R_N)\) converges, as \( N \to \infty \), to the solution of a system of coupled integro-differential equations. This was proposed in [25,27] and is now commonly referred to as the Cugliandolo–Kurchan equations. The main focus in the physics literature has been to understand what these equations imply about \( C_N(t, s) \) when \( t \) and \( s \) tend to infinity together, after the thermodynamical \( N \to \infty \) limit has been taken; in the low temperature regime, aging can be seen in the behavior of \( C_N \) in this limit. This picture was established rigorously in the simpler case of \( p = 2 \) in [10] where the equations for \( C \) and \( R \) decouple. These equations have also been proven to hold for \( p \geq 3 \) in [11]; however, here the decoupling is not expected and it remains a deep and challenging question to prove aging.

One can also study the evolution of the energy, \( H(X_t) \), via the Cugliandolo–Kurchan equations. However, due to the difficulty of the equations for \((C_N, R_N)\) when \( p > 2 \), this expression is not particularly amenable to rigorous analysis without an ansatz on the form of \( C \) and \( R \). Moreover, the study of these equations is restricted to certain random initial conditions, typically the cases where the initial distribution of \( X_0 \) is well-concentrated and independent of the field \( H \). Observe that the approach through the Cugliandolo–Kurchan equations is exact, whereas the bounding flows approach pays the price of inequalities. In exchange, this allows us to analyze one-time observables, from which we can glean meaningful information about the behavior of \( X_t \). Furthermore, this information can depend in a precise way on the initial data, which we are now free to choose as we wish. Finally, we observe that the analysis of the Hessian as it effects the dynamics has also been developed using the Cugliandolo–Kurchan approach in the physics literature in [36].

2. Outline and Ideas of Proofs

In this section, we sketch out a derivation of the differential inequality of Theorem 1.1, then formalize the phase diagram depicted in Fig. 1 in Theorem 2.6, from which we will directly conclude Theorems 1.2 and 1.4. The rest of the paper will then develop the necessary regularity theory and make the derivation of the differential inequality rigorous, before turning to comparison theory of bounded differential systems and proving Theorem 2.6.

2.1. Derivation of Eq. (1.4). We will sketch the derivation of the differential inequality of Theorem 1.1. Making this rigorous will be the project of Sects. 3–5. Let us begin first by considering the evolution of \( H(X_t) \). Using that \( p \)-spin Hamiltonian is a homogeneous function of order \( p \), one can see that is an approximate spherical harmonic, \( \Delta H = \frac{1}{N} \sum_i X_t^i X_s^i \)
\[-pH + o(N),\] where the error term comes from the trace of the Euclidean Hessian [see (4.10) and (4.3)]. Consequently, by Ito’s lemma,

\[u_N(t) \approx u_N(0) + \int_0^t \left( p \frac{H(X_s)}{N} + \beta \frac{\nabla H(X_s)}{N} \right) ds + M_u^t.\]

where \(M_u^t\) is a martingale term. Given the above integral equation, it is then natural to consider the evolution of \(|\nabla H(X_t)|^2\). Again using Ito’s lemma,

\[v_N(t) = v_N(0) + \frac{1}{N} \int_0^t \left( \Delta |\nabla H|^2 (X_s) - 2\beta \nabla^2 H(X_s)(\nabla H(X_s), \nabla H(X_s)) \right) ds + M_v^t,\]

where \(M_v^t\) is again a martingale term. Examining the drift terms above, we will find that

\[\Delta |\nabla H(X_s)|^2 \approx 2p(p - 1)N + 2p^2 \frac{1}{N} (H(X_s))^2 - 2(p - 1) |\nabla H(X_s)|^2,\]

by elementary geometric considerations on the sphere and Bochner’s formula [see (4.8)]. At the same time, if we let \(G\) be the Euclidean Hessian restricted to \(T S^N\) as in Remark 1.3, then

\[\nabla^2 H(X_s)(\nabla H(X_s), \nabla H(X_s)) \approx G(\nabla H, \nabla H)(X_s) - p \frac{1}{N} H(X_s) |\nabla H(X_s)|^2.\]

At this point, bounding the operator norm of \(G\) by \(\Lambda_p\) per (1.7), we then find that

\[v_N(0) + \int_0^t \mathcal{F}^L_2(u_N(s), v_N(s)) ds - o(1) \leq v_N(t) \leq v_N(0) + \int_0^t \mathcal{F}^U_2(u_N(s), v_N(s)) ds + o(1).\]

The Sobolev estimates for \(H\) developed in Sect. 4 will establish that the approximations above hold \(L^p\)-eventually almost surely, uniformly over all choices of initial data, and will imply that both martingale terms are of order \(O(\sqrt{t/N})\). Sending \(N \to \infty\), we deduce (1.4).

### 2.2. Phase portrait of Fig. 1 via bounding flows

We now describe the precise phase diagram we can deduce from the differential inequalities above for all the subsequential limits of \((- \frac{H(X_t)}{N}, \frac{\nabla H(X_t)}{N})\). Let us begin by decomposing the phase space into a set of regions in terms of which our phase diagram depicted in Fig. 1 is defined. (As \(v_N\) is deterministically nonnegative, we are always restricting our flows to \(\mathbb{R} \times \mathbb{R}_+\).) These regions will be described by certain zero-sets of the lower and upper bounding flows governed by the differential inequality in Theorem 1.1. To that end, recall the definitions of \(\mathcal{F}_1\) and \(\mathcal{F}_2^{L/U}\) from (1.5) for any \((x_0, y_0)\) in \(\mathbb{R}^2\), let \(A_L(t)\) and \(A_U(t)\) denote the solutions to the initial value problems

\[
\begin{align*}
\dot{A}_L(t) &= (\mathcal{F}_1(A_L(t)), \mathcal{F}_2^L(A_L(t))) \\
A_L(0) &= (x_0, y_0)
\end{align*}
\quad , \quad 
\begin{align*}
\dot{A}_U(t) &= (\mathcal{F}_1(A_U(t)), \mathcal{F}_2^U(A_U(t))) \\
A_L(0) &= (x_0, y_0)
\end{align*}
\]

(2.1)
Fig. 4. The curves $\ell_1(u)$ (purple), $f_L(u)$ (green), and $f_U(u)$ (orange). The two black curves are $\gamma_U(u; z_c)$ and $\gamma_L(u; \bar{z}_c)$ above and below $\ell_1(u)$ respectively.

(As both $\Phi_L$ and $\Phi_U$ are locally Lipschitz, existence and uniqueness are ensured.) Call $A_L(t)$ the lower bounding flow and $A_U(t)$ the upper bounding flow. Now, define (refer to Fig. 4)

$$f_L(u) = \frac{p(p-1) + p^2 u^2}{p - 1 + p\beta u + \beta \Lambda_p},$$

$$f_U(u) = \frac{p(p-1) + p^2 u^2}{p - 1 + p\beta u - \beta \Lambda_p},$$

$$\ell_1(u) = pu\beta^{-1}$$

where we impose domains $\{u : 0 \leq f_L(u) < \infty\}$ and $\{u : 0 \leq f_U(u) < \infty\}$ on the first two. We observe the following important facts about these functions.

**Remark 2.1.** For each $u$ in the domain of $f_L(u)$, the map $v \mapsto F_2^L(u, v)$ has a unique zero. In particular, $f_L(u)$ is the unique function such that $F_2^L(u, f_L(u)) = 0$. Likewise for $f_U(u)$ with respect to $F_2^U$. It follows from this that the unique (attractive) fixpoint of $A_L$ is given when $f_L(u) = \ell_1(u)$ and by explicit calculation, we see this happens at $z_c = (u_c, f_L(u_c))$ for $u_c$ defined in (1.9). Similarly, for $\bar{u}_c$ defined as

$$\bar{u}_c = \begin{cases} \beta \left(1 - \frac{\beta \Lambda_p}{p-1}\right)^{-1} & \text{if } \beta \Lambda_p < p-1, \\ \infty & \text{if } \beta \Lambda_p \geq p-1, \end{cases}$$

when $\bar{u}_c < \infty$, $\bar{z}_c = (\bar{u}_c, f_U(\bar{u}_c))$ is the unique (attractive) fixpoint of $A_U$. Finally, we remark that by explicit calculation, one can see that $f'_L(u) > 0$ for all $u \geq u_c$ and likewise, $f'_U(u) > 0$ for all $u \geq \bar{u}_c$ when $\bar{u}_c < \infty$.

Let us also use the following hitting time notation.

**Definition 2.2.** For fixed initial data $(x_0, y_0)$ and a set $B \subset \mathbb{R}^2$, let

$$\tau_B^L = \inf\{t \geq 0 : A_L(t) \in B\}, \quad \text{and} \quad \tau_B^U = \inf\{t \geq 0 : A_U(t) \in B\}$$

and for a dynamical system $U : [0, T] \to \mathbb{R} \times \mathbb{R}_+$ with $U(0) = (x_0, y_0)$, let

$$\tau_B = \tau_B(U(t)) = \inf\{t \geq 0 : U(t) \in B\}.$$
We omit the dependence of the above on $U(t)$ and $(x_0, y_0)$ when unambiguous. We use the convention that whenever hitting times are infinite, corresponding events are vacuously satisfied.

The bounding dynamics depend crucially on whether $(x_0, y_0)$ is above or below $\ell_1(u)$. Let

$$W_+ = \{(x, y) \in \mathbb{R} \times \mathbb{R}_+ : f_1(x, y) > 0\} \quad \text{and} \quad W_- = \{(x, y) \in \mathbb{R} \times \mathbb{R}_+ : f_1(x, y) < 0\}. \quad (2.4)$$

Observe that for $(x_0, y_0) \in W_+$, we have $\tau_{W^+_L}^L > 0$ and $(A_L(t))_{t\leq \tau_{W^+_L}^L}$ is the graph of a function, call it $\gamma_L(u; (x_0, y_0))$ with domain $\text{Dom}(\gamma_L, (x_0, y_0))$. The same holds for $(x_0, y_0) \in W_-$. Define $\gamma_U(\cdot; (x_0, y_0))$ and $\text{Dom}(\gamma_L; (x_0, y_0))$ analogously for $A_U(t)$.

These allow us to define the subsets of $\mathbb{R} \times \mathbb{R}_+$ with respect to which we prove a phase diagram. The absorbing set for our dynamics will approximately be $A_0$, which is defined as follows:

$$A_0 = \begin{cases} \{(u, v) : v \in [\gamma_L(u; A_U(\tau_{W^+_L}^U; z_u)), \gamma_U(u; z_u)]\} & \text{if } \bar{u}_c < \infty, \\ \{(u, v) : v \in [f_L(u), \gamma_U(u; z_u)]\} & \text{otherwise}. \end{cases} \quad (2.5)$$

Notice that the $u$–values here are implicitly constrained by the domains of of $\gamma_L$ and $\gamma_U$. As we wish to prove certain hitting times are finite, it will be helpful to enlarge $A_0$ by $\delta$ near the fixpoints of $A_L$ and $A_U$. To this end, let

$$A_{0,\delta} = A_0 \cup B_{\ell_1}(z_u, \delta) \cup B_{\ell_1}(\bar{z}_c, \delta), \quad (2.6)$$

where if $\bar{u}_c = \infty$, the second ball is empty. Define the following other regions:

$$A_4 = \{(u, v) : u < 0\},$$
$$A_3 = \{(u, v) : v \leq \ell_1(u) \land f_L(u)\},$$
$$A_2 = \{(u, v) : v > pu\beta^{-1}\} \setminus (A_0 \cup A_4),$$
$$A_1 = \{(u, v) : v \in [f_L(u), \ell_1(u)]\} \setminus A_0. \quad (2.7)$$

Note that if $\beta \Lambda_p$ is sufficiently large and $\bar{u}_c = \infty$, we have $A_1 = \emptyset$. We invite the reader to refer to and use Fig. 5 as a guide throughout this section and Sect. 7.
Remark 2.3. Since $\mathcal{F}_2^I(u, \ell_1(u)) > 0$ for all $u < \bar{u}_c$ while $\mathcal{F}_2^I(u, \ell_1(u)) < 0$ for all $u > u_c$ the set $A_0$ is well-defined and contains the entire segment $\{(u, v) : u \in [u_c, \bar{u}_c], v = \ell_1(u)\}$. (In particular this implies that the union $\bigcup_{i=0}^4 A_i = \mathbb{R} \times \mathbb{R}_+$.) If we also consider the fact that $\mathcal{F}_2^L(u, v) \geq 0$ for all $v \leq f_L(u)$ we deduce that $A_0 \subset \{(u, v) : u \geq u_c, v \geq puc^β - 1\}$.

We are now in position to describe the phase portrait for $U_N(t)$. It will be useful to restrict our attention to bounded subsets of $\mathbb{R} \times \mathbb{R}_+$. The phase portrait will describe how dynamical systems satisfying the differential inequality of Theorem 1.1 and a boundedness assumption move between the sets $(A_i)_{i=0, \ldots, 4}$ in finite time. For two vectors $v, w \in \mathbb{R}^2$ we say $v \leq_2 w$ if $v_1 = w_1$ and $v_2 \leq w_2$.

**Definition 2.4.** Let $U : [0, T] \to \mathbb{R}^2$ be $C^1$. We say that $U$ satisfies **Condition I** if

$$\Phi_L(U(t)) \leq_2 \dot{U}(t) \leq_2 \Phi_U(U(t)). \quad (2.8)$$

**Definition 2.5.** We say that $U : [0, T] \to \mathbb{R}^2$ satisfies **Condition B** if there exists a compact (nonempty) $W \subset \mathbb{R} \times \mathbb{R}_+$ such that if $U(0) \in W$ then $U(t) \in W$ for all $t \leq T$.

For ease of notation, we let $A_i$ denote the sets from (2.6)–(2.7) restricted to the compact set $W$ for which $U$ satisfies **Condition B**. Recall that $τ_A$ is the hitting time of $A$ for $U$ started from $U(0) = (x_0, y_0)$. The phase diagram depicted in Fig. 1 can be described formally as follows.

**Theorem 2.6.** Let $U : [0, T] \to \mathbb{R}^2$ satisfy **Condition I** and **Condition B**. For every $ε > 0$, there exists $δ > 0$ and $T_0$ such that the following holds:

1. When $U$ starts in $A_4$, it exits into $A_2$ in time at most $T_0$.
2. When $U$ starts in $A_3 \setminus A_{0,δ}$, it exits into $A_{0,δ} \cup A_1 \cup A_2$ in time at most $T_0$.
3. When $U$ starts in $A_2 \setminus A_{0,δ}$, it exits into $A_{0,δ} \cup A_1$ in time at most $T_0$.
4. When $U$ starts in $A_1 \setminus A_{0,δ}$, it exits into $A_{0,δ}$ in time at most $T_0$.
5. There exists a set $A_ε$ having $A_{0,δ} \subset A_ε \subset [u_c - ε, \infty) \times [v_c - ε, \infty)$ such that when $U$ starts in $A_ε$, it stays in $A_ε$ for all times.

**Remark 2.7.** The set $A_ε$ will be a non-uniform enlargement of the set $A_0$ by some $O(ε)$ distance; in particular, the enlargement is different around the corners of the set $A_0$, and is defined precisely in (7.1) of Sect. 7, while proving Theorem 2.6.

We defer the proof of Theorem 2.6 to Sect. 7 and its proof requires trajectory-wise and point-wise comparisons that are developed in Sect. 6 and may be independently interesting. With Theorem 2.6, however, we are in position to easily complete the proofs of Theorems 1.2 and 1.4.

2.3. **Proofs of Theorems 1.2–1.4.** We now read off our main results from Theorem 2.6, up to a regularity estimate ensuring that **Condition B** is satisfied. Since we study properties of all possible limit points for a sequence of initial data, we introduce the following notation for convenience.

**Definition 2.8.** For every sequence $x_N \in S^N$, let $\mathcal{W}(\{x_N\})$ denote the set of all possible limiting laws of $(u_N(t), v_N(t))_{t \in [0, T]}$ started from $X(0) = x_N$. Notice the implicit dependence on $T$ here.
Proof of Theorem 1.2. First observe by Theorem 1.1 that $\mathbb{P}$-eventually almost surely, for every $T$, every sequence $(x_N) \in S^N$, and every limit law $\mathcal{L} \in \mathcal{U}((x_N))$, $\mathcal{L}$-almost surely, $U \sim \mathcal{L}$ satisfies Condition I. It will follow from standard $L^\infty$ bounds on $H/N$ and $|\nabla H|^2/N$ (special cases of Theorem 4.3) that there exists $W \subset \mathbb{R} \times \mathbb{R}_+$ such that $\mathbb{P}$-eventually almost surely, for every $T$, every sequence $(x_N) \in S^N$, and every limit law $\mathcal{L} \in \mathcal{U}((x_N))$, $\mathcal{L}$-almost surely, $U(0) \in W$ and Condition B holds.

Now fix any $\epsilon > 0$ and let $\delta > 0$ be that given by Theorem 2.6. Let $(x_N)$ denote a sequence of near minimizers of (1.6) and let $\mathcal{L} \in \mathcal{U}((x_N))$. As noted above, $\mathcal{L}$-a.s., $U \sim \mathcal{L}$ satisfies the conditions of Theorem 2.6. Consequently, by stitching together items (1)–(4) in Theorem 2.6, there is a $\tau_{A_{0,\delta}}$ such that $\tau_{A_{0,\delta}} \leq 4T_0$. By item (5) of Theorem 2.6, $\tau_{A_c} \leq \tau_{A_{0,\delta}}$ and thus, for every $T > 4T_0$,

$$\mathcal{L}\left(\left\{\inf_{t \in [4T_0,T]} u(t) > u_c - \epsilon\right\} \cap \left\{\inf_{t \in [4T_0,T]} v(t) \geq v_c - \epsilon\right\}\right) = 1.$$ 

As this event is strongly open in $C([0, T])^2$, it follows by the Portmanteau Lemma that

$$\lim_{N \to \infty} Q_{x_N} \left(\left\{\inf_{t \in [4T_0,T]} u(t) > u_c - \epsilon\right\} \cap \left\{\inf_{t \in [4T_0,T]} v(t) \geq v_c - \epsilon\right\}\right) = 1.$$ 

The result follows as our $(x_N)$ was such that the probability in (1.6) got arbitrarily close to its infimum over all $x \in S^N$ as $N$ went to $\infty$. □

Proof of Theorem 1.4. For each $\eta > 0$, let $\delta_0(\eta) = p\eta/\beta = \ell_1(\eta)$. Fix any $\delta < \delta_0$ and let $\Omega = \{(u, v): u > \delta, v < \delta\}$. There exists a $c > 0$ such that $\mathcal{F}_1$ and $\mathcal{F}_2^L$ from (1.5) satisfy $\mathcal{F}_1 < -c$ and $\mathcal{F}_2 > c$ on $\Omega$. Fix any sequence $(x_N) \in S^N$ such that $(-H(x_N)/N, |\nabla H(x_N)|^2/N) \in \Omega$. Then for every $\mathcal{L} \in \mathcal{U}((x_N))$, by continuity of $\mathcal{F}_1, \mathcal{F}_2^L$, there exist $c_1, \rho > 0$ such that $\mathcal{L}$-a.s.,

$$U^1(t) - U^1(0) < c_1 t \quad \text{and} \quad U^2(t) - U^2(0) > c_1 t$$ 

for every $t \leq \rho$. The result then follows by the Portmanteau Lemma. □

3. Quasi-autonomy for Langevin-Type Systems on Spheres in Large Dimensions

We seek to study the properties of scaling limits of observables of Langevin dynamics. The key observation is that certain observables are quasi-autonomous. Informally, this will mean that they asymptotically (in $N$) satisfy autonomous differential inequalities.

In this section, we introduce the notion of a quasi-autonomous family of observables. We then present an elementary result proving quasi-autonomy for a sufficiently regular family of observables evolving with respect to Langevin dynamics on spheres in large dimensions.

Notation. Throughout, we say that $f_N \preceq_g g_N$ if there is a $C(a) > 0$ such that $f_N \preceq C(a)g_N$ for every $N$. We write $f_N = O(g_N)$ if $f_N \preceq g_N$ and $f_N = o(g_N)$ if $f_N/g_N \to 0$ as $N \to \infty$.

Let $(V_N)$ be a sequence of smooth functions with $V_N \in C^\infty(S^N)$. Let $X_t$ be the stochastic process on $S^N$ with generator

$$L = \Delta - \beta \langle \nabla V_N, \nabla \cdot \rangle.$$ 

(3.1)
Here and in the following \(\langle \cdot, \cdot \rangle\) will refer to the inner product on \(TS^N\) given by the metric \(g\) and \(\Delta\), \(\nabla\) will be the covariant Laplacian and derivative respectively. Finally, \(d(x, y)\) will be the distance on \(S^N\). Notice that \(X_t\) can be seen as the solution to the Langevin equation with Hamiltonian \(V_N\):

\[
\begin{align*}
\frac{dX_t}{dt} &= \sqrt{2}d_B - \beta \nabla V_N(X_t)dt, \\
X_0 &= x,
\end{align*}
\]

where \(B_t\) is Brownian motion on \(S^N\). Since \(V_N\) is smooth, one can solve this SDE in the strong sense (see e.g., [33]). We call \(X_t\) the Langevin dynamics corresponding to \(V_N\).

Let \(Q_x\) denote the law of \(X_t\) started from \(X_0 = x\) and let \(E_x\) denote the corresponding expectation.

As \(B_t\) moves on order \(\sqrt{N}\) distances in order 1 time, it is natural to consider Hamiltonians whose gradients are on this scale. To this end, we say that a sequence of functions \((V_N(x))\) with \(V_N \in C^\infty(S^N)\) is \(C\text{-regular}\) for \(C > 0\) if they satisfy the following gradient estimate: if, for every \(N\),

\[
\| |\nabla V_N|\|_{L^\infty(S^N)} \leq C\sqrt{N}. \tag{GE}
\]

The following lemma shows that if the sequence of Hamiltonians, \((V_N)\), is \(C\text{-regular}\) then the “gradient descent” and “diffusive” natures of \(X_t\) are on the same scale.

**Lemma 3.1.** There is a universal constant \(K > 0\) such that the following holds.

1. For all \(N \geq 1\), every \(x \in S^N\), and every \(t \geq 0\),

\[
E_x\left[ d(B_t, x)^4 \right] \leq KN^2t^2.
\]

2. If \((V_N)\) is \(C\text{-regular}\), then for every \(N \geq 1\), \(x \in S^N\) and every \(T \geq 0\) and \(t \in [0, T]\),

\[
E_x\left[ d(X_t, x)^4 \right] \leq \left( K + \beta^4C^4T^2 \right) N^2t^2.
\]

**Proof.** Recall Stroock’s representation [33, Exam. 3.3.2] for \(B_t\) which, in Itô form, is given by

\[
B_t - B_0 = \int_0^t \left( Id - \frac{B_s \otimes B_s}{N} \right) dW_s - \int_0^t \frac{B_s}{N} ds,
\]

where \(W_t\) is Brownian motion on \(\mathbb{R}^N\), \(Id\) is the identity on \(\mathbb{R}^N\). Since \(B_t \in S^N\), we have that

\[
\left| \frac{B_s}{N} \right| = \frac{1}{\sqrt{N}} \quad \text{and} \quad \left| Id - \frac{B_s \otimes B_s}{N} \right|_F^2 = N - 1,
\]

where \(|\cdot|_F\) is the Frobenius norm. Observe that by the Burkholder–Davis–Gundy inequality and Itô’s isometry,

\[
E_x\left[ \left| \int_0^t \left( Id - \frac{B_s \otimes B_s}{N} \right) dW_s \right|^4 \right] \lesssim (N - 1)^2 t^2.
\]
Recall further that the metric on the sphere and Euclidean space are comparable:

\[ d(x, y) \leq \pi |x - y| \quad \forall x, y \in S^N. \]

Consequently, for every \( x \in S^N \),

\[
E_x \left[ d(B_t, x)^4 \right] \lesssim E_x \left[ \left| \int_0^t \left( I_d - \frac{B_s \otimes B_s}{N} \right) dB_s \right|^4 \right] + E_x \left[ \left| \int \frac{B_s}{N} ds \right|^4 \right] \lesssim N^2 t^2.
\]

This yields the first item; the second item is then an immediate consequence of the fact that \( X_t \) solves the SDE (3.2), the estimate on \( B_t \), and the \( C^1 \)-regularity of \( V \). (GE)

For \( n \geq 1 \), let \( ((F^k_N)_{k \in [n]})_N \) be a sequence of \( n \)-tuples of smooth functions, i.e., for each \( N \), \( (F^k_N)_{k \in [n]} \in C^\infty(S^N)^n \). We call such a sequence a family of observables. We seek to study families of observables whose fluctuations are of order 1 in time.

**Definition 3.2.** A family of observables \( ((F^k_N)_{k \in [n]})_N \) is mild if for every \( k \in [n] \), there is a \( K_k \geq 0 \) such that \( F^k_N \) satisfies the Sobolev estimate,

\[
\limsup_{N \to \infty} \left( \| F^k_N \|_{L^\infty(S^N)} + \sqrt{N} \| \nabla F^k_N \|_{L^\infty(S^N)} \right) \leq K_k. \tag{SE}
\]

With this notion in hand, we present the main tightness lemma, which is an elementary application of the Kolmogorov criterion for tightness. We are concerned with tightness in the spaces \( C([0, T]) \) endowed with the strong topology and \( C([0, T])^n \) endowed with the product topology.

**Lemma 3.3.** Let \( X_t \) be Langevin dynamics corresponding to a \( C \)-regular sequence \( (V_N) \). Let \( n \geq 1 \) and let \( (F^k_N)_{k \leq n} \subset C^\infty(S^N)^n \) be a mild family of observables. Then for all \( T \geq 0 \) and for every sequence \( (x_N) \) with \( x_N \in S^N \), the family \( \{(F^k_N(X_t))_N\} \) is tight in \( C([0, T])^n \).

**Proof.** As \( C([0, T])^n \) is equipped with the product topology, it suffices, by a union bound, to show that for each \( k \) and \( \epsilon > 0 \), there is a compact \( E^k_\epsilon \subset C([0, T]) \) such that for every \( N \),

\[
Q (F^k_N(X_t), \in E^k_\epsilon) \geq 1 - \epsilon.
\]

For this, it suffices to check the following Kolmogorov-type criterion (see, e.g., [40, Exer. 2.4.2]): there exists \( L_k \) such that for every \( x \in S^N \), every \( t, s \leq T \) and every \( N \),

\[
E_x \left[ |F^k_N(X_t) - F^k_N(X_s)|^4 \right] \leq L_k(t-s)^2, \quad \| F^k_N \|_{L^\infty(S^N)} \leq L_k.
\]

This follows by the assumption \( (SE) \) and Lemma 3.1 combined with the Markov property.

With the above in hand, we can now introduce the main notion of this section, namely, that of a quasi-autonomous family of observables.
Definition 3.4. Let $V$ be C-regular. A family of observables $(F_N^k)_{k \leq \ell}$ is $V$-quasi-autonomous if there exists $n \geq \ell$ and auxiliary observables $(F_N^{k,n})_{k=\ell+1}$ such that the following hold.

1. The augmented family, $(F_N^k)_{k \in [n]}$, is mild.
2. If $X_t$ is the Langevin dynamics with Hamiltonian $V_N$, then for every $k \leq \ell$, there is a smooth function $F_k : \mathbb{R}^n \to \mathbb{R}$ and a bounded stochastic process $g_N^k(t)$ adapted to the filtration of $(B_s)_{s \leq t}$ such that for every $N$,

$$LF_N^k(X_t) = F_k(F_N^1(X_t), \ldots, F_N^n(X_t)) + g_N^k(t),$$

(3.1)

where we recall $L$ is given by (3.1)] and moreover, the sequence $g_N^k$ satisfies

$$\lim_{N \to \infty} \sup_{x \in S_N} \mathbb{E}_x \left[ \|g_N^k\|_{L^\infty([0,T])} \right] = 0.$$  

(C2)

We call the functions, $(F_k)_{k \leq \ell}$, the dynamical functions of the family.

We then have the following theorem which is the main result of this section.

Theorem 3.5. Let $V$ be C-regular. Let $(F_N^k)_{k \leq \ell}$ be a $V$-quasiautonomous family of observables with auxiliary observables $(F_N^{k,n})_{k=\ell+1}$ and dynamical functions $(F_k)_{k \leq \ell}$. Finally, let $u_N^k(t) = F_N^k(X_t)$ denote the evolution of the augmented family. Then for every $T \geq 0$, $(u_N^k(t))_{i=1}^n$ is tight in $C([0,T])^n$. Furthermore, for any weak limit point of this sequence, $(u^k(t))_{k \leq \ell}$, we have that for every $m \leq \ell$, $u^m$ is continuously differentiable and satisfies the integral equation

$$u^m(t) = u^m(0) + \int_0^t F_m(u^1(s), \ldots, u^n(s)) ds.$$  

(3.3)

Proof. The tightness follows by the assumption that the augmented family is mild and Lemma 3.3. We now turn to the proof of the integral equations. By Itô’s lemma and the assumptions, for every $k \leq \ell$, we can write

$$u_N^k(t) = u_N^k(0) + \int_0^t F_k(u_N^1(s), \ldots, u_N^n(s)) ds + \mathcal{E}_N^k(t),$$

$$\mathcal{E}_N^k(t) = \int_0^t g_N^k(s) ds + M_t^k,$$

where $M_t^k$ is a martingale with quadratic variation

$$[M_t^k]_t = \int_0^t \left| \nabla F_N^k(X_s) \right|^2 ds.$$

Fix any convergent subsequence $(N_i)_i$,

$$\left( u_N^k \right)_{(N_i)} \overset{(d)}{\to} \left( u^k \right),$$

where $(u^k)$ is a weak limit point. If we let

$$\tilde{u}_N^k = u_N^k - \mathcal{E}_N^k,$$
then

\[ \bar{u}_{N_i}^k(t) \xrightarrow{(d)} \bar{u}^k(t) = u(0) + \int_0^t \mathcal{F}_k(u^1(s), \ldots, u^k(s)) ds. \]

On the other hand, for every \( k \), there exists \( K_k \) such that

\[ E_x \left[ \| \mathcal{M}_{\ell}^k \|_{C([0,T])} \right] \leq E_x \sqrt{\int_0^T \left| \nabla F_N^k(X_s) \right|^2 ds} \leq \sqrt{\frac{K_k T}{N}}, \]

by Burkholder–Davis–Gundy and the mildness criterion (SE). Thus, by assumption on \( g_N^k \),

\[ \limsup_{N \to \infty} \sup_{x \in S^N} E_x \left[ \| \bar{u}_N^k - u_N^k \|_{C([0,T])} \right] = \limsup_{N \to \infty} \sup_{x \in S^N} E_x \left[ \| \mathcal{G}_N^k \|_{C([0,T])} \right] = 0. \]

It then follows, by a standard approximation argument [35, Thm. 3.29], that

\[ (u_{N_i}^k) \xrightarrow{(d)} \bar{u}^k, \]

implying the claimed integral inequality. Since \((\mathcal{F}_k)_{k \leq \ell}\) are smooth and \((u^k)^n_{k=1}\) are continuous, \( u^k \) is continuously differentiable for each \( k \leq \ell \). Thus the corresponding \( u^k \) are as well. \( \square \)

4. Regularity Theory for \( p \)-spin Models and the \( \mathcal{G}^k \) Norm

In order to apply the preceding, we need to control the regularity of the \( p \)-spin Hamiltonian in the appropriate Sobolev norm. Throughout our entire discussion of the \( p \)-spin models, we fix \( p > 2 \) and \( H \) will always refer to the \( p \)-spin Hamiltonian, \( H = H_{N,p} \) from (1.1).

Recall the homogeneous Sobolev norm \( \dot{W}^{k,\infty}(S^N) \),

\[ \| f \|_{\dot{W}^{k,\infty}} := \left\| \sqrt{\langle \nabla^k f, \nabla^k f \rangle} \right\|_{L^\infty(S^N)}. \]

It will turn out to be easier to work with the following, more controlled, function space.

**Definition 4.1.** We say that a function is in the space \( \mathcal{G}^k_K(S^N) \), if

\[ \| f \|_{\mathcal{G}^k_K} := \sum_{0 \leq \ell \leq k} K^{\ell/2} \left\| \left| \nabla^\ell f \right|_{op} \right\|_{L^\infty(S^N)} < \infty. \]

Here, \( \left| \nabla^k f \right|_{op} (x) \) denotes the natural operator norm when \( \nabla^k f \) is viewed as a \( k \)-form acting on the \( k \)-fold product of the tangent space \( T_x S^N \). Throughout the paper, unless otherwise specified \( \left| \nabla^k f \right| \) will denote this norm. We let \( \mathcal{G}^k(S^N) = \mathcal{G}^k_N(S^N) \) denote the special case \( K = N \).
Remark 4.2. By equivalence of norms in finite dimensional vector spaces, $G^k$ is the classical $W^{k,\infty}$ space with an equivalent norm. We use this norm as opposed to the usual one, based on Frobenius norms, for the following reason. For $k \geq 2$, we study operator norms of random tensors. It is well known that for operators of this type there is a marked difference in the scaling of the Frobenius and operator norms in the dimension (see, e.g., [44]). The choice of $K = N$ compensates for this difference in scaling at every $k$. To avoid ambiguity about the choice of Frobenius norm as opposed to operator norm, we denote this space here by $G^k$ as opposed to $W^{k,\infty}$.

The main goal of this section is the following estimate. In the following we say that a function $f(x)$ is of at most polynomial growth if there are $n, C, c > 0$ such that $|f(x)| \leq C |x|^p + c$.

Theorem 4.3 (Regularity). For every $k$, there exist $K_1(p, k), K_2(p, k), c_p > 0$ such that $K_1$ is of at most polynomial growth in $p$, and

(1) $H$ is in $G^k$ uniformly in $N$ with high probability:

$$
\limsup_{N \to \infty} \frac{1}{N} \log \mathbb{P}(\|H\|_{G^k} \geq K_1 N) \leq -c, \quad (4.1)
$$

(2) $H$ satisfies $\dot{W}^{k,\infty}$ bounds for $k = 1, 2$ uniformly in $N$ with high probability:

$$
\limsup_{N \to \infty} \frac{1}{N} \log \mathbb{P}(\|H\|_{\dot{W}^{k,\infty}} \geq K_2 \sqrt{N}) \leq -c. \quad (4.2)
$$

In the present paper, Theorem 4.3 is only applied with $k \leq 4$; however, we prove it in this greater generality as it can be applied to deduce tightness of a much greater family of observables, and proves relevant in other settings, e.g., [12].

Remark 4.4. We note here that by an application of Borell’s inequality, and the preceding estimates, the same bound also holds for the so called “mixed $p$-spin models”, i.e., for Hamiltonians of the form $H = \sum a_p H_p(x)$ where $\sum a_p^2 2^p < \infty$, except the constants now depend on the sequence $(a_p)$ instead of $p$ (and, of course, the polynomial dependence is lost).

It will also be essential to control certain spectral properties of the Hessian. Recall that the restriction of the Euclidean Hessian to $TS^N$, denoted $G = G_{N,p}$, is distributed as $C_p \cdot M$ where $M$ is drawn from the Gaussian orthogonal ensemble (GOE) normalized such that its limiting spectrum is supported on $[-\sqrt{2}, \sqrt{2}]$ (see, e.g., [2]). We then have the following uniform estimates on $G$.

Theorem 4.5. There exist $f(p, c), g(p, c), h(p, c) > 0$ such that for every $\delta > 0$, and every $c > 0$,

$$
\mathbb{P}\left( \sup_{x \in S^N} |\text{tr} G(x)| \geq c N^{\frac{1}{2} + \delta} \right) \lesssim \exp(-f(c) N^{1+2\delta}), \quad (4.3)
$$

$$
\mathbb{P}\left( \sup_{x \in S^N} |\text{tr} G^2(x) - C_p^2 N E_{\mu SC}[\lambda^2]| \geq c N^{\frac{1}{2} + \delta} \right) \lesssim \exp(-g(c) N^{1+2\delta}), \quad (4.4)
$$

$$
\mathbb{P}\left( \sup_{x \in S^N} |\nabla \text{tr}(G(x))| \geq c N^\delta \right) \lesssim \exp(-h(c) N^{1+2\delta}). \quad (4.5)
$$
Observe that in (4.4), using that \(C_p^2 = 2p(p - 1)\) and the moments of \(\mu_{SC}\), we have
\[
\mathbb{E}[\text{tr} \, G^2(n)] = C_p^2 N E_{\mu_{SC}}[\lambda^2] = p(p - 1)N.
\]

As the proofs of Theorem 4.3 and 4.5 are technical and independent of the rest of the paper, readers interested primarily in the main implications of the paper can freely skip to Sect. 5.

4.1. Some facts regarding the geometry of \(S^N\). In the rest of this section, we will heavily use certain elementary facts regarding the geometry of \(S^N\). We recount them here.

We begin by observing the following estimate on the covering number of \(S^N\). For a set \(A \subset \mathbb{R}^N\) equipped with the Euclidean metric, let \(N(\mathcal{A}, r)\) be the minimum number of Euclidean balls of radius \(r\) needed to cover \(A\). Recall (see, e.g., [44, Lemma 5.1]) that
\[
N(S^N - 1, 1, r) \leq 4r^{-N}.
\]

In particular, observe that it takes \(C N \delta^{-N/2}_N\) balls of the form \(B_x = \{y : \frac{1}{N} \sum_i x_i y_i \geq 1 - \delta_N\}\) to cover the sphere \(S^N\) for some universal \(C > 0\).

We now remind the reader of the following results regarding the differential geometry of \(S^N\). Recall that the Ricci tensor of \(S^N\) satisfies
\[
\text{Ric}_{S^N} = (1 - \frac{1}{N}) \text{Id},
\]
where \(\text{Id}\) is the identity on \(T S^N\). As a consequence, Bochner’s formula in this setting reads: for any \(u \in C^\infty(S^N)\),
\[
\frac{1}{2} \Delta |\nabla u|^2 = \left|\nabla^2 u\right|_F^2 + \langle \nabla \Delta u, \nabla u \rangle + (1 - \frac{1}{N}) \langle \nabla u, \nabla u \rangle.
\]

The principle curvatures of \(S^N\) are all equal, \(\kappa_i = \kappa\), with \(\kappa = \pm 1/\sqrt{N}\), where the choice of sign depends on the choice of normal vector. Let us work with the convention that \(\kappa\) is negative (an “outward” facing normal). Then the second fundamental form satisfies
\[
II(X, Y) = \frac{1}{\sqrt{N}} \langle X, Y \rangle P,
\]
where \(P\) is the “outward” normal direction. As \(H\) is a degree \(p\) homogeneous function, \(PH = pH/\sqrt{N}\). Consequently, the Euclidean Hessian of \(H\), restricted to \(T S^N\), can be expressed as
\[
G(x) = \nabla^2 H(x) + II(\cdot, \cdot) H(x),
\]
\[
= \nabla^2 H(x) + p \frac{H(x)}{N} \text{Id},
\]
where \(\text{Id}\) is the identity (i.e., the metric tensor). Going further, we note the following representation of higher covariant derivatives of \(H\). Let \(\text{Sym}_k\) denote the symmetric group of the set \([k]\) and let \(\hat{\nabla}\) be the Euclidean derivative.
Lemma 4.6. Let \( f \in C^\infty(S^N) \), \( x \in S^N \) and let \( \{X_i\}_{i=1}^k \subset T_xS^N \). For each \( k \geq 2 \) and \( \ell \leq k \), there exist functions \( c_{\ell,k} : \text{Sym}_{2k-\ell} \rightarrow \mathbb{R} \) such that

\[
\nabla^k f(X_1, \ldots, X_k) = \sum_{\ell=1}^k \sum_{\sigma \in \text{Sym}_{k-\ell}} c_{\ell,k}(\sigma) A_{\ell,k}(\sigma),
\]

where

\[
A_{\ell,k}(\sigma) = \frac{1}{N^{(k-\ell)/2}} \hat{\nabla}^\ell f(W_{\sigma(1)}, \ldots, W_{\sigma(\ell)}) \cdot \left\{ W_{\sigma(\ell+1)}, W_{\sigma(\ell+2)}, \ldots, W_{\sigma(2k-\ell)}, W_{\sigma(2k-\ell)} \right\},
\]

and \( W_{i,k} = X_i \) for \( i \leq k \) and \( W_{i,k} = P \) for \( k+1 \leq i \leq 2k-\ell \).

The proof of Lemma 4.6 goes by induction, using the above facts—as it is cumbersome and not relevant to the rest of the paper, we defer the proof of this representation to Appendix A.

4.2. Proof of Item 1 of Theorem 4.3. Before turning to the proof of Item 1 of Theorem 4.3, note the following estimate regarding injective tensor norms of i.i.d. Gaussian tensors; here \( g_{i_1,\ldots,i_k} \) are each i.i.d. standard Gaussian random variables. (In fact, Lemma 4.7 only requires that \( g_{i_1,\ldots,i_k} \) have sub-Gaussian tails.)

Lemma 4.7. Let \( A = (g_{i_1,\ldots,i_k}/N^{k-1/2})_{i_1,\ldots,i_k} \) be an i.i.d. Gaussian \( k \)-tensor on \( \mathbb{R}^N \). For every \( k \geq 1 \), there exists \( C(k) > 0 \) such that the injective tensor norm of \( A \) satisfies

\[
\mathbb{E} \left[ \sup_{X_1,\ldots,X_k \in \mathbb{S}^{N-1}(1)} A(X_1, \ldots, X_k) \right] \leq CN^{1-(k/2)}.
\]

Proof. Let \( \Sigma = \Sigma_\epsilon \) be an \( \epsilon \)-net of \( \mathbb{S}^{N-1}(1) \) for an \( \epsilon \) to be determined and let \( \Sigma^k \) be its \( k \)-fold Cartesian product. By multi-linearity of \( A \) and the triangle inequality,

\[
|A| \leq \sup_{\Sigma^k} A(X_1, \ldots, X_k) + \epsilon k |A|,
\]

where we recall that \( |A| \) denotes the operator norm of \( A \). Then if \( \epsilon \leq 1/(2k) \), we have

\[
\mathbb{P}(|A| \geq \lambda) \leq \mathbb{P} \left( \bigcup_{\Sigma^k} \{ |A(X_1, \ldots, X_k)| \geq \lambda/2 \} \right).
\]

For any \( X_1, \ldots, X_k \) unit vectors,

\[
\mathbb{E}[A(X_1, \ldots, X_k)^2] = \frac{1}{N^{k-1}},
\]

by definition of \( A \), so that by Gaussian concentration,

\[
\mathbb{P} \left( |A(X_1, \ldots, X_k)| \geq \lambda N^{1-(k/2)} \right) \leq \exp(-N^{2-k}\lambda^2/(2N^{1-k})) \leq \exp(-\lambda^2 N/2).
\]

By isotropy and a union bound over \( \Sigma^k \) (bounding \( |\Sigma^k| \) by (4.6)), we see that

\[
\mathbb{P}(|A| \geq \lambda N^{1-(k/2)}) \leq k(4\epsilon^{-1})^N e^{-\lambda^2 N/8}.
\]

The upper bound on \( \mathbb{E}[|A|] \) then follows by integration. \( \square \)
With these preliminaries in hand, we are in position to prove Theorem 4.3. Observe that $H$ is almost surely smooth and for $x, y \in S^N$,

$$
\mathbb{E}[H(x)H(y)] = NR^p(x, y)
$$

where $R(x, y) = \frac{1}{N} \sum_i x_i y_i$ is the normalized overlap of $x$ and $y$.

**Proof of Theorem 4.3.** Since the Frobenius norm and operator norm of a matrix always satisfy $|A|_F \leq \sqrt{N} |A|$ for any $A$, the second item is an immediate consequence of the first.

The proof of the first item is in two steps. First, we bound the operator norm of the Euclidean derivatives (denoted $\hat{\nabla}^k$) viewed as operators on the bundle $T S^N \oplus N S^N$ (where $N S^N$ is the normal bundle), via standard Gaussian comparison inequalities. We then prove the bounds on the $G^k$ norms for all $k$, inductively using the corresponding bounds on the Euclidean derivatives.

**Step 1.** For every $k$, there exists $K_0(p, k) > 0$ and $c(p, k, K) > 0$ such that for every $K > K_0$,

$$
P\left( \sup_{x \in S^N} \sup_{X_1, \ldots, X_k} \hat{\nabla}^k H(x)(X_1, \ldots, X_k) \geq KN^{1-k/2} \right) \lesssim \exp(-cN). \quad (4.12)
$$

In particular, $K_0(p, k)$ has at most polynomial growth in $p$.

**Proof.** Clearly it suffices to consider $k \leq p$. Let $E = S^N \times (S^{N-1}(1))^k$ and consider the field $\psi : E \to \mathbb{R}$ defined by

$$
\psi(x, X_1, \ldots, X_k) = \frac{1}{N^{(p-1)/2}} \sum_{i_1, \ldots, i_p} J_{i_1, \ldots, i_p} X_{i_1}^{i_1} \cdots X_{i_k}^{i_k} \cdot x_{i_{k+1}} \cdots x_{i_p}.
$$

Observe that as Gaussian processes on $E$,

$$
\hat{\nabla}^k H(x)(X_1, \ldots, X_k) \overset{(d)}{=} C_{p, k} \psi(x, X_1, \ldots, X_k),
$$

for some constants $C_{p, k}$ which are of at most polynomial growth in $p$. Thus it suffices to estimate $\sup_E \psi$, which we do as follows.

The covariance of $\psi$ satisfies for every $(x, X_1, \ldots, X_k), (x', X'_1, \ldots, X'_k) \in E$,

$$
\text{Cov}(\psi(x, X_1, \ldots, X_k), \psi(x', X'_1, \ldots, X'_k)) = \mathbb{E} \left[ \psi(x, X_1, \ldots, X_k) \psi(x', X'_1, \ldots, X'_k) \right]
= \frac{1}{N^{k-1}} R(x, x')^{p-k} \prod_i X_i \cdot X'_i, \quad (4.13)
$$

where $X \cdot Y$ is the usual Euclidean inner product. Evidently, this is uniformly bounded on all of $E^2$ by $N^{-(k-1)}$. Consequently, Borell’s inequality implies that

$$
\frac{1}{N} \log P\left( \sup_E \psi \geq \mathbb{E} [\sup_E \psi] + u \right) \leq -u^2/(2N^{-(k-1)}). \quad (4.14)
$$

It remains to bound the expectation $\mathbb{E} [\sup_E \psi]$. 
This will follow by Gaussian comparison. Define \( \phi_1 : S^N \to \mathbb{R} \) and \( \phi_2 : (S^{N-1}(1))^k \to \mathbb{R} \), by
\[
\phi_1(x) = \frac{1}{N^{k/2}} H_{p-k}(x),
\]
\[
\phi_2(X_1, \ldots, X_k) = \frac{1}{N(k-1)/2} \sum_{i_1, \ldots, i_k} g_{i_1, \ldots, i_k} X_{1}^{i_1} \cdots X_{k}^{i_k},
\]
where \( H_{p-k} \) is a \((p-k)\)-spin Hamiltonian and \((g_{i_1, \ldots, i_k})_{i_1, \ldots, i_k}\) are i.i.d. standard Gaussians. Observe that, if we let \( X = (X_1, \ldots, X_k) \) and \( X' = (X'_1, \ldots, X'_k) \) in \((S^{N-1}(1))^k\),
\[
\mathbb{E}[(\psi(x, X) - \psi(x', X'))^2] = \frac{1}{N^{k-1}} \left[ R(x, x)^{p-k} + R(x', x')^{p-k} - 2R(x, x')^{p-k} \prod X_i \cdot X'_i \right]
\]
\[
\leq \mathbb{E}[(\phi_1(x) - \phi_1(x'))^2] + \mathbb{E}[(\phi_2(X) - \phi_2(X'))^2]
\]
\[
= \mathbb{E}[(\phi_1(x) + \phi_2(X) - (\phi_1(x') + \phi_2(X'))^2].
\]
Thus by the Sudakov-Fernique inequality [37],
\[
\mathbb{E} \left[ \sup_{(x, X) \in E} \psi(x, X) \right] \leq \mathbb{E} \left[ \sup_{(x, X) \in E} \phi_1(x) + \phi_2(X) \right] \leq \mathbb{E} \left[ \sup_{x \in S^N} \phi_1(x) \right] + \mathbb{E} \left[ \sup_{x \in (S^{N-1}(1))^k} \phi_2(x) \right].
\]
It thus suffices to separately bound these two expectations. By Dudley’s entropy bound [37] and (4.6), the ground state of the \((p-k)\)-spin glass is order \( N \), so \( \phi_1 \) satisfies
\[
\mathbb{E}[^{\psi_1}] \leq C(p) N^{1-(k/2)},
\]
where \( C(p) \) is of at most polynomial growth in \( p \) (for a more precise bound see, e.g., [2]). The estimate on \( \mathbb{E}[^{\psi_2}] \) follows from Lemma 4.7. The result then follows from (4.14).

We now turn to the proof of the main bounds we desire on the sphere.

**Step 2.** It suffices to show the following inductively. For every \( k \), there exists a \( K_0(p, k) > 0 \) of at most polynomial growth in \( p \) and a \( c(p, k, K) > 0 \) such that for every \( K > K_0 \),
\[
\Pr \left( \sup_{x \in S^N} \sup_{X_1, \ldots, X_k \in (S^{N-1}(1))^k} \nabla^k H(X_1, \ldots, X_k) \geq KN^{1-k/2} \right) \lesssim \exp(-cN). \tag{4.15}
\]

**Proof.** To bound the case \( k = 1 \), it suffices to bound the operator norm of \( \nabla H \) as an operator on \( T^N \). Thus that case is complete by (4.12) with \( k = 1 \). For every \( k \geq 2 \), by Lemma 4.6, it suffices to show that for each \( \ell \leq k \), we have the bound
\[
\sup_{X_1, \ldots, X_k : \|X_i\| = 1} \sup_{\sigma \in \text{Sym}_{2k-\ell}} A_{\ell, k}(\sigma) \leq K(p, k) N^{1-k/2}.
\]
For some \( K \) of at most polynomial growth in \( p \). This follows from the fact that each of the inner products in \( A_{\ell, k}(\sigma) \) are bounded by 1, and the \( \ell \)’th Euclidean derivative part of \( A_{\ell, k}(\sigma) \) is bounded by \( K \cdot N^{1-\ell/2} \) as per (4.12); Consequently, each summand in the expansion (4.11) is at most some \( C_{\ell, k, p} N^{1-k/2} \), implying the desired. \( \Box \)
4.3. **Proof of Theorem 4.5.** The proof of Theorem 4.5 relies on the following general lemma that boosts a pointwise concentration estimate to a uniform bound on the sphere via a minimal modulus of continuity bound. Since we consider isotropic fields, we let \( n = (\sqrt{N}, 0, \ldots, 0) \) be the north pole and stand in as a fixed point on \( S^N \).

**Lemma 4.8.** Suppose that \( F \) is an isotropic random field on \( S^N \) that is \( \mathbb{P} \)-a.s. continuous and satisfies the following:

1. There exists \( c_1(c) > 0 \) such that for every \( c, \delta > 0 \),
   \[
   \mathbb{P}(|F(n) - \mathbb{E}[F(n)]| \geq cN^{\frac{1}{2} + \delta}) \lesssim \exp(-c_1N^{1+2\delta}).
   \]

2. There exists an \( \alpha > 0 \) and \( c_2(c) > 0 \) such that for every \( c, \delta > 0 \),
   \[
   \mathbb{P}\left( \sup_{x, y : \|x - y\| \geq 1 - N^{-\alpha}} |F(x) - F(y)| \geq cN^{\frac{1}{2} + \delta} \right) \lesssim \exp(-c_2N^{1+2\delta}).
   \]

Then there exists \( f(c) > 0 \) depending on \( c_1, c_2 \) such that for every \( c, \delta > 0 \)
   \[
   \mathbb{P}\left( \sup_{x \in S^N} |F(x) - \mathbb{E}[F(n)]| \geq cN^{\frac{1}{2} + \delta} \right) \lesssim \exp(-f(c)N^{1+2\delta}).
   \]

Similarly, if \( F \) is almost surely nonnegative and \( \mathbb{E}[F(n)] = O(\sqrt{N}) \), if we replace condition 2 with

(2') There exists an \( \alpha > 0 \) and \( c_2(c) > 0 \) such that for every \( c, \delta > 0 \),
   \[
   \mathbb{P}\left( \sup_{x, y : \|x - y\| \geq 1 - \delta_N} |F^2(x) - F^2(y)| \geq cN^{1+2\delta} \right) \lesssim \exp(-c_2N^{1+2\delta}),
   \]

then there exists \( f(c) > 0 \) depending on \( c_1, c_2 \) such that for every \( c, \delta > 0 \),
   \[
   \mathbb{P}\left( \sup_{x \in S^N} F(x) \geq cN^{\frac{1}{2} + \delta} \right) \lesssim \exp(-f(c)N^{1+2\delta}).
   \]

**Proof.** Let \( \eta_N = N^{-\alpha} \) for the \( \alpha \) given by condition 2. Then by (4.6), there exists a set of points \( x_1, \ldots, x_K \in S^N \) such that the balls \( B_i = \{ y \in S^N : R(x_i, y) \geq 1 - \eta_N \} \) cover \( S^N \) and for some universal \( C, c_\star > 0 \),
   \[ K \leq C^N N^{\alpha N/2} \leq \exp(c_\star N \log N). \]

For each \( i \), condition (1) holds with \( n \) replaced by \( x_i \) by isotropy. Fix any \( c > 0 \) and \( \delta > 0 \); for every \( 1 \leq i \leq K \), define the events
   \[ \Gamma_1^i = \left\{ \sup_{x, y : \|x - y\| \geq 1 - \eta_N} |F(x) - F(y)| < \frac{1}{2} cN^{\frac{1}{2} + \delta} \right\}, \]
   \[ \Gamma_2^i = \left\{ |F(x_i) - \mathbb{E}[F(n)]| < \frac{1}{2} cN^{\frac{1}{2} + \delta} \right\}. \]

By the choice of \( \{x_i\} \),
   \[ (\Gamma_1^1 \cap \bigcap_{i \leq K} \Gamma_2^i) \subset \left\{ \sup_{x \in S^N} |F(x) - \mathbb{E}[F(n)]| < cN^{\frac{1}{2} + \delta} \right\}. \]
As a result, by isotropy, a union bound over $1 \leq i \leq K$, and conditions 1–2, for $c_1 = c_1(c/2)$ and $c_2 = c_2(c/2)$, we obtain for every $\delta > 0$,

$$\mathbb{P}\left( \sup_{x \in S^N} |F(x) - \mathbb{E}[F(n)]| \geq cN^{1+\delta} \right) \leq \mathbb{P}\left( (\Gamma^1)^c \right) + K \mathbb{P}\left( (\Gamma_2^2)^c \right) \lesssim \exp\left( -\frac{1}{2}[c_1 \wedge c_2]N^{1+2\delta} \right),$$

yielding the first inequality. When condition 2 is replaced with condition 2′, we let

$$\Gamma^1 = \left\{ \sup_{x,y: R(x,y) \geq 1 - \eta_N} |F^2(x) - F^2(y)| < \frac{1}{\sqrt{2}} c^2 N^{1+2\delta} \right\},$$

$$\Gamma_i^2 = \left\{ F(x_i) < \frac{1}{\sqrt{2}} c^2 N^{1+\delta} \right\} = \left\{ F^2(x_i) < \frac{1}{2} c^2 N^{1+2\delta} \right\},$$

and obtain

$$\mathbb{P}\left( \sup_{x \in S^N} F^2(x) \geq c^2 N^{1+2\delta} \right) \leq \mathbb{P}\left( (\Gamma^1)^c \right) + K \mathbb{P}\left( (\Gamma_2^2)^c \right).$$

The first probability on the right-hand side is bounded by condition 2′. The second probability is bounded by condition 1 combined with the bound on $\mathbb{E}[F(n)]$. The conclusion then follows.

Recall the Euclidean Hessian $G$ given in (4.10). We now wish to verify that condition (1) in Lemma 4.8 holds for $\text{tr } G$, $\text{tr } G^2$ and $|\nabla \text{tr } G|$.

**Lemma 4.9.** For $F$ given by each of $\text{tr } G$, $\text{tr } G^2$, and $\sqrt{N} |\nabla \text{tr } G|$, there exists $c_1(c) > 0$ such that for every $c$, $\delta > 0$,

$$\mathbb{P}( |F(n) - \mathbb{E}[F(n)]| \geq cN^{1+\delta} ) \lesssim \exp(-c_1 N^{1+2\delta}).$$

**Proof.** As $G(n)$ is a GOE, up to a constant factor, $G^2(n)$ is a Wishart matrix. The results for $\text{tr } G(n)$ and $\text{tr } G^2(n)$, then follow by standard concentration estimates for spectral statistics [31, Corr. 1.6, Corr 1.8] or sums of i.i.d. random variables [37].

Now consider $F(n) = \sqrt{N} |\nabla \text{tr } G(n)|$. Since $G$ is the restriction of the Euclidean Hessian of $H$ to $TS^N$, an explicit calculation yields

$$\mathbb{E} \left[ E_i \text{tr } G(n)E_j \text{tr } G(n) \right] = c \rho \delta_{ij},$$

for some $c_p > 0$ independent of $N$, where $\{E_i\}$ are an orthonormal frame for $T_nS^N$. (See, e.g., [42] for similar calculations.) In particular, $\nabla \text{tr } G(n)$ is a standard Gaussian vector with independent entries, up to a constant factor depending at most on $p$. The result then follows by concentration of norms of Gaussian vectors [37].

We now prove the uniform continuity required to obtain conditions 2–2′ in Lemma 4.8.

**Lemma 4.10.** Let $F$ be given by $\text{tr } G$, $\text{tr } G^2$ and $N |\nabla \text{tr } G|^2$. There exists $\alpha > 0$ such that for $\eta_N \asymp N^{-\alpha}$, there exists $c_2(c) > 0$ such that for every $c$, $\delta > 0$, we have

$$\mathbb{P}\left( \sup_{x,y: R(x,y) \geq 1 - \eta_N} |F(x) - F(y)| \geq cN^{1+\delta} \right) \lesssim \exp(-c_2 N^{1+2\delta}).$$
Proof. We prove the desired for $\sqrt{N} |\nabla \text{tr } G|$. The cases $\text{tr } G$ and $\text{tr } G^2$ follow by the same argument (and are in fact implicitly proved in what follows). For every $x \in \mathcal{S}^{N}$, $X \in T_{x}\mathcal{S}^{N}$ with $|X| = 1$,

$$\nabla X |\nabla \text{tr } G|^2 = 2\nabla^2 (\text{tr } G)(\nabla \text{tr } G, X) \leq 2 |\nabla^2 \text{tr } G| |\nabla \text{tr } G|.$$  

Since contractions commute with covariant derivatives it follows, by (4.10), that

$$\|\nabla^k \text{tr } G\|_{L^\infty} \leq \frac{1}{N^{1+2\delta}} \mathbb{P}\left(\|\nabla^k \text{tr } G\|_{L^\infty} \geq N^{\frac{2-k}{2}+\delta}, k = 1, 2\right) < -c,$$

for some $c > 0$ for all $N$ sufficiently large. This implies that

$$\frac{1}{N^{1+2\delta}} \log \mathbb{P}\left(\|\nabla |\nabla \text{tr } G|^2\|_{L^\infty} \geq N^{\frac{1}{2}+2\delta}\right) < -c,$$

holds for all $N$ sufficiently large. Choosing $\alpha = 1 + \delta$ yields the result. \qed

Proof of Theorem 4.5. By Lemma 4.8, it first suffices to prove that conditions 1–2 hold for $\text{tr } G$ and $\text{tr } G^2$ as these are both isotropic, almost surely continuous fields on $\mathcal{S}^N$. For these two fields, condition 1 follows from Lemma 4.9 and condition 2 follows from Lemma 4.10. For $\sqrt{N} |\nabla \text{tr } G(x)|$, it is again an isotropic, almost surely continuous field on $\mathcal{S}^N$ and we observed in the proof of Lemma 4.9 earlier that $\mathbb{E}[\sqrt{N} |\nabla \text{tr } G(n)|] = O(\sqrt{N})$. Condition 1 for this field follows from Lemma 4.9, and condition 2′ follows from Lemma 4.10. \qed

5. Tightness and Differentiability of Observables

We begin our analysis of dynamics for $p$-spin models in this section. Here $X_t$ will be Langevin dynamics with Hamiltonian $H = H_{N,p}$. Our main goal will be to conclude the tightness and differentiability of weak limit points for the family of observables

$$u_N(t) = -\frac{H(X_t)}{N}, \quad v_N(t) = \frac{|\nabla H(X_t)|^2}{N},$$

and the auxiliary observable

$$w_N(t) = \frac{1}{N} G(\nabla H, \nabla H)(X_t).$$

Let $\mathcal{L}^T_{xN}$, a probability measure on $C([0,T]^3$, be the law of the augmented family $(u_N, v_N, w_N)$ started from $X_0 = x_N$, and let

$$\mathcal{F}_1(x, y) = -px + \beta y, \quad (5.1)$$
$$\mathcal{F}_2(x, y, z) = 2p(p - 1) + 2p^2x^2 - 2(p - 1)y - 2p\beta xy - 2\beta z. \quad (5.2)$$

We then have the following theorem.
Theorem 5.1. \( \mathbb{P} \)-almost surely, the following holds. For every sequence \( x_N \in S^N \), the family \( (u_N, v_N, w_N) \) is tight. That is \( \mathcal{L}_{x_N,N}^T \) is precompact in the narrow topology. Furthermore, for any limit point \( \mathcal{L} \), the family \( (u, v, w) \sim \mathcal{L} \) satisfies the integral equations

\[
\begin{align*}
    u(t) &= u(0) + \int_0^t \mathcal{F}_1(u(s), v(s)) ds, \\
    v(t) &= v(0) + \int_0^t \mathcal{F}_2(u(s), v(s), w(s)) ds
\end{align*}
\]

and the family \( (u(s), v(s)) \) are continuously differentiable.

Proof. We first need the following estimates on \( \Delta H_N \) and \( \Delta |\nabla H_N|^2 \) to estimate \( Lu_N \) and \( Lv_N \).

Lemma 5.2. For every \( p > 2 \), there exists \( f(\cdot) > 0 \) such that for every \( c, \delta > 0 \),

\[
\begin{align*}
    \Pr \left( \| \Delta H + p H \|_{L^{\infty}(S^N)} \geq c N^{1/2+\delta} \right) &\lesssim \exp(-f(c) N^{1+2\delta}), \\
    \Pr \left( \| \Delta |\nabla H|^2 - 2A \|_{L^{\infty}(S^N)} \geq c N^{1/2+\delta} \right) &\lesssim \exp(-f(c) N^{1+2\delta}),
\end{align*}
\]

where

\[
A = Np(p - 1) + p^2 \frac{H^2}{N} - (p - 1) |\nabla H|^2.
\]

Proof. We begin with the first estimate. Taking the trace of (4.10), we see that

\[
\Delta H = -p(1 - \frac{1}{N}) H + tr G(x).
\]

The result then follows by (4.3) and (4.1).

Now for the second estimate, by the Ricci bound (4.7) and Bochner’s formula (4.8),

\[
\frac{1}{2} \Delta |\nabla H|^2 = tr(G^2) + (1 - \frac{1}{N})p^2 \frac{H^2}{N} - 2p tr(G) \frac{H}{N} - (p - 1)(1 - \frac{1}{N}) |\nabla H|^2 + (\nabla tr G, \nabla H).
\]

The result then follows upon applying (4.4) to the first term, (4.3) to the third term, and Cauchy–Schwarz along with (4.5) to the fifth term. \( \square \)

Proof of Theorem 5.1. The result will follow by an application of Theorem 3.5. We begin by observing that \( H \) satisfies the gradient estimate (GE) by Theorem 4.3. Thus \( H \) is \( \mathbb{P} \)–eventually almost surely \( K_1(p, k) \)-regular for any fixed \( k \).

We now check that these observables are mild. Let

\[
F^1_N = -\frac{H}{N}, \quad F^2_N = \frac{|\nabla H|^2}{N}, \quad F^3_N = \frac{1}{N} G(\nabla H, \nabla H).
\]

Observe that

\[
\nabla X G(\nabla H, \nabla H) = \nabla G(X, \nabla H, \nabla H) + 2 G(\nabla^2 H(\cdot, X), \nabla H),
\]

so that

\[
|\nabla G(\nabla H, \nabla H)| \leq |\nabla G| |\nabla H|^2 + 2 \left| \nabla^2 H \right| |G| |\nabla H|.
\]
Thus by the $G^3$-norm bound from Theorem 4.3, and the definition of $G$, there is a $K_p$ such that eventually $\mathbb{P}$-almost surely,

$$
\left\| F_N^k \right\|_{G^1} \leq K_p \quad \forall k \leq 3.
$$

Thus the Sobolev estimates (SE) hold. Consequently, this family is mild.

To show that $\{F_N^k\}_{k=1}^2$ are quasiautonomous, let

$$
g_1^1(t) = \frac{1}{N} (\Delta H(X_t) + p H(X_t)),
$$

$$
g_2^2(t) = \frac{1}{N} \left( \frac{1}{2} \Delta |\nabla H|^2 - A \right)(X_t),
$$

where $A$ is as in (5.7). The processes $g_1, g_2$ satisfy (C2) by Lemma 5.2. Recalling the generator $L$ of the Langevin dynamics and writing out $LF_N^k$ using the identity,

$$
\langle \nabla H, \nabla |\nabla H|^2 \rangle = 2\nabla^2 H(\nabla H, \nabla H) = 2G(\nabla H, \nabla H) - 2\frac{H}{N} |\nabla H|^2,
$$

we see we have the splitting (C1). In particular, the family $\{F_N^k\}_{k=1}^2$ is $H$-quasiautonomous with dynamical functions $\{F_k^2\}_{k=1}^2$ and auxiliary function $F_3^N$. Thus the conditions of Theorem 3.5 are satisfied. Consequently the integral equations hold and any weak limit point is such that $(u, v)$ are continuously differentiable.

5.1. A differential inequality for the evolution of $(u(t), v(t))$. We end this section by observing the following consequence of Theorem 5.1. Recall $\Lambda_p > 0$ from (1.7), and recall from (1.5),

$$
\mathcal{F}_1(u, v) = -pu + \beta v,
$$

$$
\mathcal{F}_2^L(u, v) = 2p(p-1) - 2(p-1)v + 2pu(pu - \beta v) - 2\beta \Lambda_p v,
$$

$$
\mathcal{F}_2^U(u, v) = 2p(p-1) - 2(p-1)v + 2pu(pu - \beta v) + 2\beta \Lambda_p v.
$$

Moreover, define the maps

$$
\Phi_L(u, v) = \begin{pmatrix} \mathcal{F}_1 \\ \mathcal{F}_2^L \end{pmatrix}, \quad \Phi_U(u, v) = \begin{pmatrix} \mathcal{F}_1 \\ \mathcal{F}_2^U \end{pmatrix},
$$

so that for $(x_0, y_0) \in \mathbb{R} \times \mathbb{R}_+$, $A_L(t)$ is the solution to the initial value problem $\dot{A}_L(t) = \Phi_L(A_L(t))$ and $A_L(0) = (x_0, y_0)$, and $A_U(t)$ is similarly defined.

Also, recall that for any sequence $x_N \in S^N$, we let $\mathcal{U}((x_N))$ denote the set of all possible limiting laws of $(u_N(t), v_N(t))_{t \in [0, T]}$ started from $X_0 = x_N$, and suppress the dependence on $T$ in the notation. We are then able to conclude the proof of the differential inequality. Here and in the following, inequalities for vectors are to be interpreted as holding in both coordinates simultaneously.
Proof of Theorem 1.1. For every $T$ and every sequence of initial data $(x_N)$, the tightness of the laws of $(U_N(t))_{t \in [0,T]}$ was established in Theorem 5.1. Let $\mathcal{U}((x_N))$ be as in Definition 2.8, which by the tightness must be non-empty, fix any $\mathcal{L} \in \mathcal{U}((x_N))$, and let $(U(t))_{t \in [0,T]} = (u(t), v(t))_{t \in [0,T]}$ be distributed according to $\mathcal{L}$. By Theorem 5.1, $U$ is $\mathcal{L}$-almost surely differentiable.

By (5.3), $\mathbb{P}$-almost surely, $\dot{u} = F_1(u, v)$ $\mathcal{L}$-a.s. By (1.7),

$$-\Lambda_p v \leq w \leq \Lambda_p v,$$

$\mathcal{L}$-a.s. Consequently, $\mathcal{L}$-a.s.,

$$F_2^L (u, v) \leq F_2(u, v, w) \leq F_2^U (u, v),$$

for all $t \geq 0$. This combined with (5.4), implies that $\mathbb{P}$–almost surely, for every $T$, every sequence $x_N \in \mathcal{S}^N$ and every $\mathcal{L} \in \mathcal{U}((x_N))$, if $U(t) \sim \mathcal{L}$,

$$\mathcal{L} \left( \Phi_L(U) \leq \dot{U} \leq \Phi_U(U) \right) = 1.$$

6. Comparison Theory for Limiting Dynamics

Before turning to the proof of Theorem 2.6, we will first understand some general consequences of Theorem 1.1 for trajectories of the limiting dynamics $(u(t), v(t))$ (see Fig. 6).

6.1. A trajectory-wise comparison. Our main result in this section is a means to use the integral curves of two “bounding flows” to bound that of our dynamics. Informally, the goal is to confine integral curves of dynamical systems satisfying the differential inequality of Theorem 1.1 by the integral curves of the lower and upper bounding dynamical systems.

We wish to compare any weak limit, $U \sim \mathcal{L}$, of $(u_N(t), v_N(t))$ to these systems. To show this comparison result, we first observe the following basic fact from calculus. Recall the notation $v \preceq_2 w$ if $v_1 = w_1$ and $v_2 \leq w_2$ and define $\succeq_2$ similarly.

![Fig. 6. Theorem 6.3 shows that $U(t) = (u(t), v(t))$ is $\mathcal{L}$-a.s. contained in the rectangle drawn out by $A_L(t)$ (red) and $A_U(t)$ (black) for all $t$ such that the rectangle is contained in $V_-$ (below the gold line).]
Lemma 6.1. Let $X, Y : [0, T] \to \mathbb{R}^2$ be $C^1$. Let $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$ be $C^1$ be such that $\dot{Y} = \Psi(Y)$. Let $E_+ = \{ \Psi_1 > 0 \}$ and $E_- = \{ \Psi_1 < 0 \}$. We then have the following. Suppose that

- $X(0) = Y(0) \in E_+ \cup E_-$,
- For $t > 0$, if $X(0) \in E_+$, then $\Psi_1(X), \Psi_1(Y) > 0$, and if $X(0) \in E_-$ then $\Psi_1(X), \Psi_1(Y) < 0$,
- $X$ satisfies the differential inequality $\dot{X} \leq_2 \Psi(X)$ (resp. $\dot{X} \geq_2 \Psi(X)$).

Then the functions $f_X(u) = X_2 \circ X_1^{-1}(u)$ and $f_Y(u) = Y_2 \circ Y_1^{-1}(u)$ are well-defined, differentiable for $u \neq X_1(0)$, and $f_X(u) \leq f_Y(u)$ (resp. $f_X(u) \geq f_Y(u)$) on the intersection of their domains.

Proof. Without loss of generality, take $X(0) = Y(0) = (0, 0)$. Furthermore, it suffices to consider the case where $(0, 0) \in E_+$ and $\dot{X} \leq_2 \Psi(X)$, as the others are identical. By the positivity assumption for $\Psi_1(X)$ and $\Psi_1(Y)$, $X_1(t)$ and $Y_1(t)$ are invertible so that the functions $f_X$ and $f_Y$ are well-defined. Furthermore, the inverse functions are differentiable for $u > 0$. It remains to show the inequality between $f_X$ and $f_Y$.

Suppose that there is some $w$ for which $f_X(w) > f_Y(w)$. Observe that for every $u > 0$ for which $f_X(u) = f_Y(u)$, we have

$$
 f_X'(u) \leq \frac{\Psi_2(u, f_X(u))}{\Psi_1(u, f_X(u))} = f_Y'(u),
$$

by the third assumption. Thus $w > 0$. On the other hand, let $v = \sup \{ u \leq w : f_X(u) = f_Y(u) \}$. Then by continuity of $f_X, f_Y$ and maximality of $v$, we have that $f_X(u) > f_Y(u)$ for all $u \in (v, w)$. On the other hand, by continuity of $f_X, f_Y$ and (6.1), we have that $f_X'(v) \leq f_Y'(v)$. This is a contradiction. \qed

In the following, we use the convention that $x \to \infty$ means $x = \infty$. We can deduce the following corollary of the preceding. Recall from Sect. 2.2, the definitions of $A_L(t), A_U(t)$, the graphs of their trajectories, $\gamma_L(u; (x_0, y_0))$ and $\gamma_U(u; (x_0, y_0))$, and the sets $W_-$ and $W_+$ defined in (2.4).

Corollary 6.2. Suppose that $X : [0, T] \to \mathbb{R}^2$ is $C^1$ and satisfies Condition 1. Then,

1. If $X(0) \in W_\pm$, then $(X(t))_{t \in W_\pm}$ is the graph of a function $\gamma(\cdot; X(0))$.
2. The domains of $\gamma, \gamma_L$, and $\gamma_U$ satisfy

$$
 \text{Dom}(\gamma_L(\cdot; X(0))) \subseteq \text{Dom}(\gamma(\cdot; X(0))) \subseteq \text{Dom}(\gamma_U(\cdot; X(0))),
$$

when $X(0) \in W_+$, and the reverse inclusions are satisfied when $X(0) \in W_-$.
3. If $X(0) \in W_+ \cup W_-$, then

$$
 \begin{cases}
 \gamma_L(u; X(0)) \leq \gamma(u; X(0)) & \forall u \in \text{Dom}(\gamma_L(\cdot; X(0))) \cap \text{Dom}(\gamma(\cdot; X(0))) \\
 \gamma(U; X(0)) \leq \gamma_U(u; X(0)) & \forall u \in \text{Dom}(\gamma_U(\cdot; X(0))) \cap \text{Dom}(\gamma(\cdot; X(0)))
 \end{cases}
$$

4. If $X(0) \in \{ F_1 = 0 \}$ and $X_1(0) < u_c$ (resp. $X_1(0) > u_c$), then $\gamma, \gamma_L$, and $\gamma_U$ are still well-defined and items 2 and 3 still hold.
Proof. By Condition I and the assumptions of items (1)–(3), we may apply Lemma 6.1 by truncating $X$ at $\tau_{W^c}$, from which items (1)–(3) follow. For item (4) we consider the case $X(0) < u_c$ and when $\tilde{u}_c$ is finite, the case $X(0) > \tilde{u}_c$ is analogous. In this case, by (2.8), $F_1(X(0)) = 0$ and

$$0 < F_2^L(X(0)) \leq F_2(X(0)) \leq F_2^U(X(0)).$$

Thus by continuity and the definitions of $F_i^{L/U}$ it follows that $F_1(X(t)), F_1(A_L(t)), \text{ and } F_1(A_U(t))$ are all strictly positive for $t > 0$. Thus the conditions of Lemma 6.1 still hold. \hfill \Box

6.2. A pointwise comparison. One might also be interested in obtaining pointwise-in-time comparisons between the flows as in usual comparison theory. To this end we observe the following which is not used in the proof of the phase diagram, but we believe is illustrative. Define the region

$$V_- = \{(u, v) \in \mathbb{R}^2 : v \leq 2pu/\beta\}.$$

Theorem 6.3. Suppose that $X$ satisfies Condition I. If $X(0) \in V_-^0$ then:

1. $X(t) \geq A_L(t)$ for every $t \leq \tau_{V_-}^c \land \tau_{V_-}$, and $X(t) \leq A_U(t)$ for every $t \leq \tau_{V_-}^U \land \tau_{V_-}$.
2. In particular, if $\tau_{B}^\square$ is the hitting time of $B$ for $(A_L^1(t), A_U^2(t))$, then for every $t \leq \tau_{V_-}^\square$

$$A_L(t) \leq X(t) \leq A_U(t).$$

As a result, if $X(0)$ is such that $\tau_{V_-}^\square = \infty$, then $A_L(t) \leq X(t) \leq A_U(t)$ for all $t \geq 0$.

Observe the following, whose proof is immediate from the definition of $A_L$ and $A_U$.

Observation 6.4. For every $(x_0, y_0) \in W^c_+$, if $X(0) = (x_0, y_0)$, we have that $\tau_{V_-}^\square > 0$.

The proof of the Theorem 6.3 uses the classical comparison theorem of Chaplygin (see, e.g., [43]).

Lemma 6.5 (Chaplygin’s Lemma). Let $\Omega \subset \mathbb{R}^2$ and let $A, B : [0, T] \rightarrow \Omega$ be continuously differentiable. Suppose that the following hold:

1. $B(0) \leq A(0)$ (resp. $B(0) \geq A(0)$)
2. There is a locally Lipschitz $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ such that: $\dot{B} = \Psi(B)$ and $\dot{A} \geq \Psi(A(t))$ for every $t \leq T$ (resp. $\dot{A} \leq \Psi(A(t))$)
3. $\Psi$ is quasi-increasing: that is, for each $x$, the map $\Psi_1(x, \cdot)$ is monotone increasing, and for every $y$, the map $\Psi_2(\cdot, y)$ is monotone increasing.

Then $A(t) \geq B(t)$ (resp. $B(t) \geq A(t)$) for all $t \leq T$.

Proof of Theorem 6.3. We check that $X(t)$ satisfies the conditions of Chaplygin’s lemma. The first condition is satisfied by choice of initial data. The second follows by Condition I. Finally, by an explicit computation, it is easy to see that the region $V_-$ is the domain in which condition (3) holds for both $\Phi_L$ and $\Phi_U$. The result then follows by Lemma 6.5. \hfill \Box
7. Proof of Full Phase Diagram

In this section, we turn to proving Theorem 2.6. One can of course obtain sharper results for a given choice of initial data, however, we do not pursue this direction.

7.1. Properties of the bounding flows. We begin this section by first studying the phase diagram for the two “bounding flows”; we will then use the results we glean to obtain corresponding results for the flow itself.

As in previous sections, let \( A_L \) and \( A_U \) be the lower and upper bounding flows respectively, and recall the definition of \( W_+ \) and \( W_- \) from (2.4) as well as \( \bar{u}_c \) from (2.3) and the decomposition of \( \mathbb{R} \times \mathbb{R}_+ \) in (2.6)–(2.7).

Lemma 7.1. We have the following:

1. If \( (x_0, y_0) \in W_+ \) then \( A_L^1(\tau^L_{W_+}) \geq u_c \) and likewise, if \( \bar{u}_c < \infty \), then \( A_U^1(\tau^U_{W_+}) \geq \bar{u}_c \).
2. Let \( B \) be a bounded set. For every \( \delta > 0 \) there exists \( T_0 \) such that for every \( (x_0, y_0) \in B \cap W_+ \), we have \( \tau^L_{A_0,\delta \cup A_1} \leq T_0 \) and if \( \bar{u}_c < \infty \) then also \( \tau^U_{A_0,\delta \cup A_1} \leq T_0 \).

Proof. The fact that \( A_L^1(\tau^L_{W_+}) \geq u_c \) is shown as follows. We first claim that \( \tau^L_{[u < u_c] \cap W_-} \geq \tau^L_{\{u = u_c\}} \). This follows from the fact that \( F_2^L(u, \ell_1(u)) > 0 \) and \( F_1(u, \ell_1(u)) = 0 \) for all \( u < u_c \), implying that \( A_L^1(t) \) cannot hit \( \{(u, \ell_1(u)) : u < u_c\} \) for any \( t \in (0, \tau^L_{[u \geq u_c]}) \).

Then since \( F_1(u, v) \geq 0 \) when \( u \geq u_c \) and \( v \geq f_L(u) \), \( A_L^1(t) \geq u_c \) for all \( t \in [\tau^L_{[u \geq u_c]} \cap W_+, \tau^L_{W_+}] \), implying the desired. Whenever \( \bar{u}_c < \infty \) the analogous claim for \( A_U^1(\tau^U_{W_+}) \) follows similarly.

It remains to prove the upper bounds on \( \tau^L_{A_0,\delta} \) and \( \tau^U_{A_0,\delta} \). We prove the bound for \( A_L(t) \) as the bound for \( A_U(t) \) (when \( \bar{u}_c < \infty \)) follows analogously. Fix \( \delta > 0 \). By continuity, there exists an \( \eta > 0 \) such that \( \ell_1(u) + \eta < f_L(u) \) for every \( u \leq u_c - \delta \). Split up \( A_4 \cup A_2 \) into the following sets:

\[
E_1 = \{(u, v) : (W_+ \cap B) \setminus A_{0,\delta} \ : \ u \leq u_c - \delta, v \leq \ell_1(u) + \eta\},
\]

\[
E_2 = \{(u, v) : (W_+ \cap B) \setminus A_{0,\delta} \ : \ u \leq u_c + \delta\} \setminus E_1,
\]

\[
E_3 = (W_+ \cap B) \setminus (E_1 \cup E_2).
\]

Observe that there exists a \( c > 0 \) such that \( \inf_{E_1} F^L_{E_1} \geq c \). As a result, since \( E_1 \) is bounded from above by \( E_2 \cup A_{0,\delta} \), there exists \( T_1 \) such that for every \( (x_0, y_0) \in E_1 \), \( \tau^L_{E_2 \cup A_{0,\delta}} \leq T_1 \). Observe that furthermore, \( \inf_{E_2} F^L_{E_2} \geq c \), for some other \( c > 0 \). The continuity of \( A_L(t) \) then implies that, for every \( (x_0, y_0) \in E_2 \), \( A_L \) cannot enter \( E_1 \) from \( E_2 \), which is to say that \( \tau^L_{E_2} > \tau^L_{W_+} \) for \( (x_0, y_0) \in E_2 \). Moreover, since \( E_2 \) is a bounded set and bounded to the right by \( A_{0,\delta} \cup E_3 \), there exists a \( T_2 \) such that for every \( (x_0, y_0) \in E_2 \), \( \tau^L_{A_{0,\delta} \cup E_3} \leq T_2 \), and, by continuity of \( A_L(t) \), the flow cannot enter \( E_2 \) from \( A_{0,\delta} \cup E_3 \). Since \( \sup_{E_3} F^L_{E_3} \leq -c \), and \( E_3 \) is bounded and bounded below by \( A_{0,\delta} \cup A_1 \), there exists \( T_3 \) such that for every \( (x_0, y_0) \in E_3 \), \( A_L \) satisfies \( \tau^L_{A_{0,\delta} \cup A_1} \leq T_3 \). Setting \( T_0 = T_1 + T_2 + T_3 \) yields the desired estimate. \( \square \)

Lemma 7.2. If \( \bar{u}_c < \infty \), \( A_1 \) is nonempty, and we have the following:

1. If \( (x_0, y_0) \in A_1 \), then \( A_L(\tau^L_{A_1}) \in A_0 \) and \( A_U(\tau^U_{A_1}) \in A_0 \).
(2) Let $B$ be a bounded set. For every $\delta > 0$, there exists $T_0$ such that for every $(x_0, y_0) \in B \cap A_1$, we have that $\tau^L_{A_0, \delta} \leq T_0$ and that $\tau^U_{A_0, \delta} \leq T_0$.

Proof. Fix any $\delta > 0$. First of all, since $f_{U}(u)$ intersects $\ell_1(u)$ at $\bar{z}_c \in A_0$, there exists $c > 0$ such that for every $(u, \ell_1(u))$ with $u \geq \sup \{x : (x, y) \in A_0, \delta \}$ we have that $F_2(u, \ell_1(u)) \leq -c$. As a consequence, we have both $A_L(\tau^L_{A_1}) \notin A_2$ and $A_U(\tau^U_{A_1}) \notin A_2$.

Moreover, since $f_L(u) < f_U(u)$ on the domain of $f_U(u)$, there is some $c' > 0$ such that $F_2(u, f_L(u)) \geq c'$ for every $u \geq u_c + \delta$. Thus $A_U(\tau^U_{A_1}) \notin A_3$ so that $A_U(\tau^U_{A_1}) \in A_0$.

To deduce item 1 above, it remains to show that $A_L(\tau^L_{A_1}) \notin A_3$. By Lemma 6.1 and the fact that $A_L(\tau^L_{A_1}) \notin A_2$, for every $(x_0, y_0) \in A_1$, the set $(A_L(t))_{t \leq \tau^L_{A_1}}$ is the graph of some $\gamma_L(u; (x_0, y_0))$ with

$$\gamma'(u) = -\frac{F_2(u, \gamma_L(u))}{F_1(u, \gamma_L(u))}.$$ 

This implies that if $A_L(t_0) = (u_0, f_L(u_0))$ for some $u_0 > u_c$, then $\gamma'_L(u_0) = 0$ while $f'_L(u_0) > 0$. As a result, $\gamma_L(u) \in A_1$ for all $u \leq u_0$ sufficiently close to $u_0$. Consequently, $A_L(t) \in A_1$ for all $t > t_0$ sufficiently close to $t_0$. By continuity of $A_L(t)$, it then follows that $A_L(\tau^L_{A_1}) \notin A_3$ as desired.

To see item 2, notice that $\inf_{A_1 \setminus A_0, \delta} |(F_1, F_2)| \geq c$ for some $c > 0$ and, moreover, $F_1(u, v) \leq 0$ and $F_2(u, v) \leq 0$ for every $(u, v) \in A_1$. Since $A_1$ is bounded to the left, there exists a $T_0$ such that for every $(x_0, y_0) \in A_1$, we have that $\tau^L_{A_1 \setminus A_0, \delta} \leq T_0$.

By item 1, this implies that $\tau^L_{A_0, \delta} \leq T_0$. Item 2 for the upper bounding flow $A_U(t)$ follows similarly, after splitting up $A_1$ into those points above $\{u, f_U(u)\}$ and those below $\{u, f_U(u)\}$ and noting that $A_U(t)$ cannot cross $f_U(u)$ from above. \hfill \square

7.2. Proof of Theorem 2.6. In the following, we will be interested in phase portraits for systems that satisfy Condition I and Condition B as defined in Definitions 2.4–2.5. Thus, for the rest of this section, we assume that $U$ satisfies both Condition I and Condition B with respect to some $T$ and $W$. In particular, we will take this $T$ to be sufficiently large such that all finite hitting times we consider are less than $T$. As before, let $A_j$ denote the sets from (2.6)–(2.7) restricted to the compact set $W$ for which $U$ satisfies Condition B. Recall that for any set $A \subset W$, we let $\tau_A$ denote the hitting time for $U$ started from $U(0) = (x_0, y_0)$.

Proof of item (1) of Theorem 2.6. We wish to show that for every $(x_0, y_0) \in A_4$, we have $U(\tau^L_{A_1}) \in A_2$ and $\tau^L_{A_1} \leq T_0$. The fact that $U(\tau^L_{A_1}) \in A_2$ is trivial by definition of $A_4$ and $A_2$ (as $\partial A_4 \subset A_2$). By item (2) of Corollary 6.2 and Lemma 7.1, for every $x_0 < 0$ and every $u \in [x_0, 0]$, for $T$ large enough, $u$ is in the domain of $\gamma_L(\cdot; (x_0, y_0))$ and

$$\ell_1(u) = \gamma_L(u; (x_0, y_0)) \leq \gamma(u; (x_0, y_0)).$$

By Lemma 7.1, there exists $T_0$ such that for every $(x_0, y_0) \in A_4 \subset W_+$, we have $\tau_{A_2} \leq T_0$. This implies by Corollary 6.2 that $(U(t))_{t \leq \tau_{A_2}}$ is the graph of a function $\gamma(u; (x_0, y_0))$ whose domain contains $[x_0, 0]$. For every $t \leq \tau_{A_2}$, we have $|\dot{U}(t)| \leq c > 0$ for $c(p, \beta; (x_0, y_0))$ depending continuously on $(x_0, y_0)$. Together these imply that
for every \((x_0, y_0) \in A_4\), there exists \(T_{x_0, y_0}(p, \beta)\) depending continuously on \((x_0, y_0)\) such that \(\tau_{A_4}^{\ell} \leq T_{x_0, y_0}\). As \(T_{x_0, y_0}\) is continuous and \(W\) is compact, the supremum over all \((x_0, y_0) \in A_4\) of \(T_{x_0, y_0}\) is some finite \(T_0\) as desired. □

**Proof of item (2) of Theorem 2.6.** Tautologically, for every \((x_0, y_0) \in A_3\), we have \(U(\tau_{A_3}^\ell) \in A_0 \cup A_1 \cup A_2\); it remains to show that \(\tau_{A_0,\delta \cup A_1 \cup A_2} \leq T_0\). Fix any \(\delta\) small and observe that for every \((u, v) \in A_3 \setminus A_{0,\delta},\)

\[
\mathcal{F}_1(u, v) \leq 0 \quad \text{and} \quad \mathcal{F}_2(u, v) \geq 0.
\]

Moreover, there is a \(c(\delta) > 0\) such that for every \((u, v) \in A_3 \setminus A_{0,\delta}\), we have \(|\Phi_L| > c\) and, for every \((x_0, y_0) \in A_3 \setminus A_{0,\delta}\), for every \(t \leq \tau_{A_3 \setminus A_{0,\delta}}\), we have \(|\bar{U}(t)| > c\). As a result, by compactness of \(A_3 \setminus A_{0,\delta}\) and compactness of \(W\), there exists \(T_0\) such that for every \((x_0, y_0) \in A_3 \setminus A_{0,\delta},\)

\[
\tau_{A_3 \setminus A_{0,\delta}} = \tau_{A_0,\delta \cup A_1 \cup A_2} \leq c^{-1} \sup_{w \in A_3, w' \in \delta(A_3 \setminus A_{0,\delta})} d(w, w') \leq T_0. \quad \Box
\]

**Proof of item (3) of Theorem 2.6.** We first show that \(U(\tau_{A_3}^\ell) \in A_0 \cup A_1\). The fact that \(U(\tau_{A_3}^\ell) \notin A_4\) follows because \(\mathcal{F}_1(u, v) \geq 0\) for all \((u, v) \in W^+\). By item (2) of Corollary 6.2 combined with Lemma 7.1, we see that for every \(u \in [x_0, u_c)\), if \(T\) is large enough, \(u\) is in the domain of \(\gamma_L(\cdot; (x_0, y_0))\) and

\[
\ell_1(u) < \gamma_L(u; (x_0, y_0)) \leq \gamma(u; (x_0, y_0));
\]

this implies that \(U(\tau_{A_3}^\ell) \notin A_3\) implying \(U(\tau_{A_3}^\ell) \in A_0 \cup A_1\).

It remains to show that for every \(\delta > 0\), there is a \(T_0\) such that \(\tau_{A_0,\delta \cup A_1} \leq T_0\). Recall that by Lemma 7.1, for every \(\delta > 0\), there exists \(T_0(\delta) < \infty\) such that for every \((x_0, y_0)\), \(\tau_{A_0,\delta \cup A_1} \leq T_0\). This implies by items (1) and (2) of Corollary 6.2, \((U(t))_{t \leq \tau_{A_0,\delta \cup A_1}}\) is the graph of a function \(\gamma(u; (x_0, y_0))\) with domain containing \([x_0, U(\tau_{A_0,\delta \cup A_1}))\].

Now notice that for every \(\delta > 0\) and every \(t \leq \tau_{A_0,\delta \cup A_1}\), we have \(|\bar{U}(t)| \geq c > 0\) for some \(c(p, \beta; (x_0, y_0))\) that depends continuously on \((x_0, y_0)\). As the boundary of \(A_2\) relative to \([u \geq x_0, v \geq \gamma_L(u)] \in A_0, A_1\), we have that for every \((x_0, y_0) \in A_2\), there exists \(T_{x_0, y_0}(p, \beta)\) depending continuously on \((x_0, y_0)\) such that \(\tau_{A_0,\delta \cup A_1} \leq T_{x_0, y_0}\). As \(T_{x_0, y_0}\) is continuous and \(W\) is compact, we may bound \(T_{x_0, y_0}\) uniformly over all such \((x_0, y_0)\) by some finite \(T_0\), as desired. □

**Proof of item (4) of Theorem 2.6.** By Lemma 7.2, \(A_L(\tau_{A_1}^\ell), A_U(\tau_{A_1}^U) \in A_0\) and, for every \(\delta > 0\), there is a \(T_1\) such that \(\tau_{A_0,\delta}^L \leq T_1\) and \(\tau_{A_0,\delta}^U \leq T_1\). Thus by Corollary 6.2, \(U(\tau_{A_3}^\ell) \in A_0\).

We now show that for every \(\delta > 0\) there is a \(T_0\) such that \(\tau_{A_0,\delta} \leq T_0\). We begin with the following velocity estimates. By (2.8), there exists \(c > 0\) such that \(|\bar{U}(t)| > c\) for every \(t \leq \tau_{(A_1 \setminus A_{0,\delta}^\ell)}\). We now split the region \(A_1 \setminus A_{0,\delta}^\ell\) into two regions as follows. For starting points \((x_0, y_0)\) satisfying \(x_0 > \tilde{u}_c + \delta\) and \(y_0 = \ell_1(x_0) - \eta\), as \(\eta \to 0\), we have that \(\gamma_L(u; (x_0, y_0)) \to \infty\). Since \(\gamma_L(u; (x_0, y_0))\) depends continuously on \((x_0, y_0)\), there exists an \(\eta(\delta) > 0\) small such that for every \((u, v)\) with \(u \geq \tilde{u}_c + \delta\) and \(v = \ell_1(u) - \eta\), we have \(\gamma_L(u; (u, v)) > p / \beta\). With respect to this \(\eta\), define the sets

\[
E_+ = \{(u, v) \in A_1 \setminus A_{0,\delta} : y \geq \ell_1(u) - \eta\}, \quad \text{and} \quad E_- = (A_1 \setminus A_{0,\delta}) \setminus E_+.
\]
For every \((u, v) \in E_+\), we have that \(F_1(u, v) \leq 0\) and \(F_2^U(u, v) \leq 0\), so that by (2.8),
\[
\tau_{E- \cup A_{0, \delta}} = \tau_{E^*_1} \leq 2c^{-1} \sup_{w \in E_+, w' \in \partial A_1} d_{\ell_1}(w, w') = T_2,
\]
for some finite \(T_2(\delta)\), uniformly over \((x_0, y_0) \in E_+\). On the other hand, since integral curves of locally smooth flows cannot cross and since the flow \(A_U\) is outward along the boundary of \(E_+\), relative to \(A_1 \setminus A_{0, \delta}\) (as \(\gamma_1^U(u; (u, v)) > \beta/p\) and \(F_1 \leq 0\)), it follows that \(\tau_{E^*_c} = \tau_{A_{0, \delta}}^U\) for every \((x_0, y_0) \in E_-\). As a result, by Corollary 6.2 and we also see that \(\tau_{E^*_c} = \tau_{A_{0, \delta}}\). Moreover, there is a \(c > 0\) such that \(\inf_{E_-} F_1 \leq -c\), so by compactness of \(W\) and the fact that \(E_+\) is bounded to the left by \(A_{0, \delta}\), there is a \(T_3(\delta)\) such that for every \((x_0, y_0) \in E_-\), we have \(\tau_{A_{0, \delta}} = \tau_{E^*_c} \leq T_3\). Thus for every \((x_0, y_0) \in A_1\) we have that \(\tau_{A_{0, \delta}} \leq T_2 + T_3\) as desired.

**Proof of item (5) of Theorem 2.6.** It remains, for each choice of \(\epsilon > 0\), to construct an absorbing set \(A_{\epsilon}\) containing \(A_{0, \delta}\) for some \(\delta(\epsilon) > 0\). Notice that while the above indicates that the hitting time to \(A_{0, \delta}\) is finite, \(\tau_{A_{0, \delta}}^U\) started from \((x_0, y_0) \in A_{0, \delta}\) is not necessarily \(\infty\). Let \(z(\epsilon) = (u_c - \epsilon, \ell_1(u_c - \epsilon))\) and \(\tilde{z}(\epsilon) = (\bar{u}_c + \epsilon, \ell_1(\bar{u}_c + \epsilon))\) so that \(z_c = z(0)\) and \(\tilde{z}_c = \tilde{z}(0)\):
\[
A_{\epsilon} = \begin{cases} 
\{(u, v) \in W : v \in [\ell_1(u) \vee \gamma_L(u; \tilde{z}(\epsilon)) \vee A_U(\tau_{W_+}^U; \tilde{z}(\epsilon))], \gamma_U(u; z(\epsilon))] \} & \text{if } \bar{u}_c < \infty \\
\{(u, v) \in W : v \in [f_L(u), \gamma_U(u; z(\epsilon))] \} & \text{otherwise}
\end{cases}
\]
where again, the \(u\) values are implicitly constrained by the domain of \(\gamma_U(u; z(\epsilon))\). Let
\[
B_{\epsilon}(z_c) = [(u_c - \epsilon, u_c + \epsilon] \times [\ell_1(u_c - \epsilon), \ell_1(u_c + \epsilon)] \cap W_-
\]
and define \(B_{\epsilon}(\tilde{z}_c)\) analogously but intersected with \(W_+\) (again if \(\bar{u}_c = \infty\), this is empty). Let
\[
A_{\epsilon} = A_{\epsilon} \cup B_{\epsilon}(z_c) \cup B_{\epsilon}(\tilde{z}_c). \tag{7.1}
\]
It follows by the definitions and the continuity of \(\Phi_L\) and \(\Phi_U\) that for every \(\epsilon > 0\), there is a \(\delta > 0\) such that \(A_{0, \delta} \subset A_{\epsilon}\). We now turn to showing that for each \(\epsilon > 0\), \(A_{\epsilon}\) is an absorbing set.

By continuity of \(U\), it suffices to show that if for any \(t_0 \geq 0\), \(U(t_0) \in \partial A_{\epsilon}\) then for all \(t > t_0\) sufficiently close to \(t_0\), \(U(t) \in A_{\epsilon}\). We decompose the boundary into its constituent parts and analyze each part. We consider the case \(\bar{u}_c < \infty\). Notice that by definition,
\[
\partial A_{\epsilon} \subset \Gamma(\gamma_L) \cup \Gamma(\gamma_U) \cup \partial B_{\epsilon}(z_c) \cup \partial B_{\epsilon}(\tilde{z}_c),
\]
where \(\Gamma(\gamma_L)\) is the graph of \(\gamma_L(u; \tilde{z}(\epsilon)) \vee A_U(\tau_{W_+}^U; \tilde{z}(\epsilon))\) and \(\Gamma(\gamma_U)\) is the graph of \(\gamma_U(u; z(\epsilon))\).

First, suppose that \(U(t_0) \in \Gamma(\gamma_U)\). Since this is in the region \(W_+\),
\[
\gamma(w; U(t_0)) \leq \gamma_U(w; U(t_0)),
\]
for \(w > u\) sufficiently close to \(u\) by Corollary 6.2. For such \(w\), \(\gamma_U(w; z(\epsilon)) = \gamma_U(w; U(t_0))\) as these are integral curves, so it follows from Corollary 6.2 that for \(t\)
sufficiently close to \( t_0 \), \( U(t) \in \mathcal{A}_\varepsilon \) as desired. The case when \( U(t_0) \in \Gamma(\gamma_L) \) of course follows analogously.

Now suppose that \( U(t_0) \in \partial B_\varepsilon(z_c) \). Observe that if \( \hat{n}_z \) denotes the outward pointing normal at \( z \in \partial B_\varepsilon(z_c) \), then for \( z \) in the right face and the bottom face we have that \( U(t_0) \cdot \hat{n}_U(t_0) < 0 \), which yields the desired statement. The remaining face is in the interior of \( \mathcal{A}_\varepsilon \) so we can ignore this case. The case when \( U(t_0) \in \partial B_\varepsilon(\tilde{z}_c) \) also follows analogously.

In the situation when \( \tilde{u}_c = \infty \), \( \Gamma(\gamma_L) \) is replaced by \( \Gamma(f_L) = \{(u, v) : u > u_c, v = f_L(u)\} \) and the corresponding estimate follows by recalling from the proof of item (4) of Theorem 2.6 that if \((x_0, y_0) \in \Gamma(f_L)\), then \( A_L(t) \in A_0 \cup A_1 \) for all sufficiently small \( t \). The desired then again follows by Corollary 6.2.  

Acknowledgements. The authors thank the anonymous referees for their helpful comments and suggestions. The authors thank Giulio Biroli and Chiara Cammarota for interesting discussions. This research was conducted while G.B.A. was supported by NSF DMS1209165 and BSF 2014019, and A.J. was supported by NSF OISE-1604232. R.G. thanks NYU Shanghai for its hospitality during the time some of this work was completed.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix A. Proof of Lemma 4.6

In this section, we provide a proof of Lemma 4.6 which we had deferred.

Proof of Lemma 4.6. The proof is by induction. As a base case, observe that

\[
\nabla^2 f(X_1, X_2) = \hat{\nabla}^2 f(X_1, X_2) - \frac{1}{\sqrt{N}} \hat{\nabla} f(P) \langle X_1, X_2 \rangle
\]

\[
= \hat{\nabla}^2 f(X_1, X_2) + \frac{1}{\sqrt{N}} \sum_{\sigma \in S_3} c_{1,2}(\sigma) \hat{\nabla} f(W_{\sigma(1)}^{1,2}, W_{\sigma(2)}^{1,2}, W_{\sigma(3)}^{1,2}),
\]

for some suitably defined \( c_{1,2} \). Suppose now that (4.11) holds for some \( k \geq 2 \). Then, by the inductive hypothesis, to compute \( \nabla^{k+1} f \) it suffices to compute \( \nabla_{Y} A_{\ell,k}(\sigma) \) where \( Y = X_{k+1} \).

First notice that \( A_{\ell,k} \) is a product of the form

\[
\frac{1}{N^{(k-\ell)/2}} \hat{\nabla}^\ell f \otimes Id \otimes \ldots \otimes Id.
\]

As the covariant derivative of the metric tensor is zero, we see by the product rule, it suffices to compute \( V = \nabla_{Y} \hat{\nabla}^\ell f(W_{\sigma(1)}^{\ell,k}, \ldots, W_{\sigma(\ell)}^{\ell,k}) \). From now on we suppress the dependence of \( W \) on \( k, \ell \) for readability. Let \( I = \{i : \sigma(i) \leq k\} \). Then \( V \) is the covariant derivative of an \(|I|\)-form. In particular,

\[
V = Y \hat{\nabla}^\ell f(W_{\sigma(1)}, \ldots, W_{\sigma(\ell)}) - \sum_{s \in I} \hat{\nabla}^\ell f(W_{\sigma(1)}, \ldots, \nabla_Y W_{\sigma(s)}, \ldots, W_{\sigma(\ell)})
\]

\[
= \hat{\nabla}^{\ell+1} f(W_{\sigma(1)}, \ldots, W_{\sigma(\ell)}, Y) + \sum_{s \in I} \hat{\nabla}^\ell f(W_{\sigma(1)}, \ldots, II(Y, W_{\sigma(s)}), \ldots, W_{\sigma(\ell)})
\]

\[
+ \sum_{s \in I^C} \hat{\nabla}^\ell f(W_{\sigma(1)}, \ldots, \hat{\nabla}_Y W_{\sigma(s)}, \ldots, W_{\sigma(\ell)}).
\]
where the second equality comes from the definition of the covariant derivative and the Gauss formula. Since the shape operator of the sphere is multiplication by \((-1/\sqrt{N})\), we see for \(s \in I^\ell\),

\[
\hat{\nabla}_Y W_{\sigma(s)} = \hat{\nabla}_Y P = \frac{1}{\sqrt{N}} Y = \frac{1}{\sqrt{N}} Y \langle P, P \rangle.
\]

by the Weingarten equation. Plugging this, \(Y = X_{k+1}\), and (4.9) in the above, we see that

\[
V = \hat{\nabla}^{\ell+1} f(W_{\rho(1)}^{\ell+1,k+1}, \ldots, W_{\rho(\ell+1)}^{\ell+1,k+1})
+ \frac{1}{\sqrt{N}} \sum_{\rho' \in E} \hat{\nabla}^\ell f(W_{\rho'(1)}^{\ell,k+1}, \ldots, W_{\rho'([^\ell])}^{\ell,k+1}) \left\{ W_{\rho(\ell+1)^{\ell+1}}, W_{\rho'(\ell+1)^{\ell+1}} \right\},
\]

where \(\rho \in \text{Sym}_{2(k+1)-([\ell]+1)}\) is such that \(\rho(\ell+1) = k+1\) and \(E \subset \text{Sym}_{2(k+1)-[\ell]}\). Observe that in the first term, we have increased the indices \(k\) and \(\ell\) and in the second term we have gained a factor of \(N^{-1/2}\) and an inner-product. Thus

\[
\nabla X_{k+1} A_{\ell,k}(\sigma) = A_{\ell+1,k+1}(\rho) + \sum_{\rho' \in E} A_{\ell,k+1}(\rho'),
\]

for some \(\rho\) and where the sum is again over some subset of \(\text{Sym}_{2(k+1)-[\ell]}\). Note that \(\rho\) and this subset \(E\) depend only on \(\sigma, \ell,\) and \(k\). Thus, by the inductive hypothesis,

\[
\nabla^{k+1} f(X_1, \ldots, X_{k+1}) = (\nabla X_{k+1} \nabla^k f)(X_1, \ldots, X_k) = \sum_{\ell=1}^k \sum_{\sigma \in \text{Sym}_{2k-\ell}} c_{\ell,k}(\sigma) \nabla X_{k+1} A_{\ell,k}(\sigma)
= \sum_{\ell=1}^{k+1} \sum_{\sigma \in \text{Sym}_{2k-\ell}} c_{\ell,k+1}(\sigma) A_{\ell,k+1}(\rho(\sigma))
+ \sum_{\rho' \in E(\sigma)} A_{\ell,k+1}(\rho'),
\]

for some \(c_{\ell,k+1}\), concluding the proof. \(\square\)

References

1. Auffinger, A., Ben Arous, G.: Complexity of random smooth functions on the high-dimensional sphere. Ann. Probab. 41(6), 4214–4247 (2013)
2. Auffinger, A., Ben Arous, G., Černý, J.: Random matrices and complexity of spin glasses. Commun. Pure Appl. Math. 66(2), 165–201 (2013)
3. Baity-Jesi, M., Biroli, G., Cammarota, C.: Activated aging dynamics and effective trap model description in the random energy model. J. Stat. Mech. Theory Exp. 2018(1), 013301 (2018)
4. Bauerschmidt, R., Bodineau, T.: A very simple proof of the LSI for high temperature spin systems. arXiv preprint arXiv:1712.03676 (2017)
5. Ben Arous, G.: Aging and spin-glass dynamics. In: Proceedings of the International Congress of Mathematicians, Vol. III, pp. 3–14. Higher Ed. Press, Beijing (2002)
6. Ben Arous, G., Bovier, A., Černý, J.: Universality of the REM for dynamics of mean-field spin glasses. Commun. Math. Phys. 282(3), 663–695 (2008)
7. Ben Arous, G., Bovier, A., Gayrard, V.: Aging in the random energy model. Phys. Rev. Lett. 88(8), 087201 (2002)
8. Ben Arous, G., Bovier, A., Gayrard, V.: Glauber dynamics of the random energy model. I. Metastable motion on the extreme states. Commun. Math. Phys. 235(3), 379–425 (2003)
9. Ben Arous, G., Bovier, A., Gayrard, V.: Glauber dynamics of the random energy model. II. Aging below the critical temperature. Commun. Math. Phys. 236(1), 1–54 (2003)
10. Ben Arous, G., Dembo, A., Guionnet, A.: Aging of spherical spin glasses. Probab. Theory Related Fields 120(1), 1–67 (2001)
11. Ben Arous, G., Dembo, A., Guionnet, A.: Cugliandolo–Kurchan equations for dynamics of spin-glasses. Probab. Theory Related Fields 136(4), 619–660 (2006)
12. Ben Arous, G., Gheissari, R., Jagannath, A.: Algorithmic thresholds for tensor PCA. Preprint available on arXiv (2018)
13. Ben Arous, G., Gün, O.: Universality and extremal aging for dynamics of spin glasses on subexponential time scales. Commun. Pure Appl. Math. 65(1), 77–127 (2012)
14. Ben Arous, G., Jagannath, A.: Spectral gap estimates in mean field spin glasses. Commun. Math. Phys. arXiv:1705.04243 (to appear)
15. Berthier, L., Biroli, G.: Theoretical perspective on the glass transition and amorphous materials. Rev Modern Phys 83(2), 587 (2011)
16. Bouchaud, J.-P.: Weak ergodicity breaking and aging in disordered systems. J. Phys. I 2(9), 1705–1713 (1992)
17. Bouchaud, J.-P., Cugliandolo, L.F., Kurchan, J., Mézard, M.: Out of equilibrium dynamics in spin-glasses and other glassy systems. Spin glasses and random fields, pp. 161–223 (1998)
18. Bouchaud, J.-P., Dean, D.S.: Aging on parisi’s tree. J Phys I 5(3), 265–286 (1995)
19. Bovier, A., Faggionato, A.: Spectral characterization of aging: the REM-like trap model. Ann. Appl. Probab. 15(3), 1997–2037 (2005)
20. Bovier, A., Gayrard, V.: Convergence of clock processes in random environments and ageing in the $p$-spin SK model. Ann. Probab. 41(2), 817–847 (2013)
21. Černý, J., Wassmer, T.: Aging of the metropolis dynamics on the random energy model. Probab. Theory Related Fields 167(1–2), 253–303 (2017)
22. Chen, W.-K., Sen, A.: Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed $p$-spin models. Commun. Math. Phys. 350(1), 129–173 (2017)
23. Cugliandolo, L.F.: Course 7: Dynamics of glassy systems. In: Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter, pp. 367–521. Springer (2003)
24. 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)
25. 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)
26. Gayrard, V.: Aging in metropolis dynamics of the REM: a proof. ArXiv e-prints (2016)
27. Gheissari, R., Jagannath, A.: On the spectral gap of spherical spin glass dynamics. Ann. Inst. H. Poincaré Probab. Stat. 55(2), 756–776 (2019)
28. Guionnet, A., Zeitouni, O.: Concentration of the spectral measure for large matrices. Electron. Commun. Probab. 5, 119–136 (2000)
29. Guionnet, A.: Dynamics for spherical models of spin-glass and aging. In: Spin Glasses, Volume 1900 of Lecture Notes in Mathematics, pp. 117–144. Springer, Berlin (2007)
30. Hsu, E.P.: Stochastic analysis on manifolds, Volume 38 of Graduate Studies in Mathematics. American Mathematical Society, Providence (2002)
31. Jagannath, A., Tobasco, I.: Low temperature asymptotics of spherical mean field spin glasses. Commun. Math. Phys. 352(3), 979–1017 (2017)
32. Kallenberg, O.: Foundations of modern probability. In: Probability and Its Applications, 2nd edn. Springer, New York (2002)
33. Kurshan, J., Laloux, L.: Phase space geometry and slow dynamics. J. Phys. A Math. Gen. 29(9), 1929 (1996)
34. Ledoux, M., Talagrand, M.: Probability in Banach spaces. In: Classics in Mathematics. Springer, Berlin (2011) (Isoperimetry and processes, reprint of the 1991 edition)
35. Mathieu, P.: Convergence to equilibrium for spin glasses. Commun. Math. Phys. 215(1), 57–68 (2000)
36. Sompolinsky, H., Zippelius, A.: Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses. Phys. Rev. B 25(11), 6860 (1982)
37. Stroock, D.W., Varadhan, S.R.S.: Multidimensional Diffusion Processes, vol. 233. Springer, Berlin (1979)
41. Subag, E.: The complexity of spherical p-spin models—a second moment approach. The Annals of Probability 45, 3385–3450 (2015)
42. Subag, E.: The geometry of the Gibbs measure of pure spherical spin glasses. Invent. Math. 210, 1–75 (2017)
43. Szarski, J.: Differential inequalities. Monografie Matematyczne, Tom 43. Państwowe Wydawnictwo Naukowe, Warsaw (1965)
44. Vershynin, R.: High-Dimensional Probability. Cambridge University Press, Cambridge (2018)

Communicated by H. Spohn