POTENTIAL HESSIAN ASCENT IV: SAMPLING THE SHERRINGTON–KIRKPATRICK MODEL AT $\beta < 1$

HOLDEN LEE, JUSPREET SINGH SANDHU, AND JONATHAN SHI

Abstract. We give a polynomial-time algorithm to sample from the Gibbs measure of the Sherrington–Kirkpatrick (SK) model with $o_n(1)$ error in total-variation distance (TVD) at any inverse-temperature $\beta < 1$. The algorithm combines algorithmic stochastic localization (ASL) with rejection sampling over path-space via Jarzynski's equality (JE).

The analysis extends the authors' prior $\beta < 1/2$ result [DLSS26a] by replacing all global regularity requirements in the stochastic differential equation error analysis with local regularity around likely trajectories. The relaxed regularity is established using Celentano's proof of the local strong convexity of the TAP free energy [Cel24]. The analysis utilizes the cavity interpolation theory and free probability toolkit developed in the authors' previous result, where the former applies nearly verbatim and the latter applies supplemented with Lipschitz and C^2 extensions of various functions.

We give a list of *desiderata* encapsulating the approximation and regularity properties required of the TAP free energy, relaxing those of [DLSS26a, §2.5] to only require local regularity. These generic desiderata are potentially applicable in other settings where a free energy surrogate exists, giving algorithmic sampling guarantees while bypassing the usual functional inequality–based approach.

Contents

1. Introduction 1
2. Sampling via ASL-JE under local regularity 4
3. Quantitative SDE analysis 6
4. Uniform TAP Hessian positivity over the stochastic localization process 19
5. Resolvent analysis: Regularization, uniform concentration and interpolation 24
6. Properties of the algorithmic stochastic process 36
7. Sampling from the time-T localized distribution 42
A. Resolvent product regularity and subordination at the TAP value 45
B. Self-stability for the cavity interpolation of [DLSS26a, §5] at $\beta < 1$ 47
C. Algorithmic error analysis 49

Department of Applied Mathematics and Statistics, Johns Hopkins University, USA
Department of Computer Science, Colorado State University, USA
No affiliation
E-mail addresses: hlee283@jhu.edu, js.sandhu@colostate.edu, jshi@cs.cornell.edu.

1. Introduction

The Gibbs measure for the Sherrington–Kirkpatrick (SK) model of system-size n is a (random) probability measure on $\{-1,1\}^n$ given by

$$\mu_{\beta A}(\sigma) = \frac{1}{Z(\beta A)} \exp\left(\frac{1}{2}\langle \sigma, \beta A \sigma\rangle\right),$$

where $A \sim \mathsf{GOE}(n)$ and the constant $Z(\beta A) := \sum_{\sigma\in\{-1,1\}^n} \exp\left(\frac{1}{2}\langle \sigma, \beta A\sigma\rangle\right)$ is the partition function.

Building on the cavity interpolation theory and free-probability toolkit developed by the authors in their previous result [DLSS26a], we give a polynomial time algorithm to sample from the Gibbs measure of the SK model at every fixed inverse temperature $\beta < 1$ with $o_n(1)$ total-variation distance (TVD) error. In the absence of an external field, this characterizes the entire high-temperature regime [dAT78, Lop26b] and $\beta = 1$ is believed to be the onset of computational hardness for sampling [AMS22, Sel25] since the system exhibits full replica-symmetry breaking (fRSB) at $\beta > 1$ [JSS25, Lop26a]. This is another step towards fully resolving the conjecture that Glauber dynamics exhibits fast-mixing for the SK model below the dAT line [BKMR25, Open-Problem 15]. In forthcoming work [LSS26a], we make further progress towards the conjecture by generalizing the conductance arguments and Wiener analysis techniques developed in another prior result of the authors [DLSS26b] – we prove functional inequalities with access to only "local" regularity properties of the algorithmic stochastic localization (ASL) process.[1]

Theorem 1.1 (Main theorem)**.**
Let $0 < \beta < 1$. Given any fixed $\delta > 0$, for all $n \geqslant n_0(\beta,\delta)$, with probability at least $1-\delta$ over $A \sim \mathsf{GOE}(n)$, Algorithm 2 gives a sample from a distribution that is $o_n(1)$ close in TVD error to $\mu_{\beta A}$ in polynomial time.

We use (nearly) the same notation and technical toolkit as in [DLSS26a]. Therefore, we refer the reader to [DLSS26a, §1.4] for notation and preliminaries, and [DLSS26a, §B.4] for the relevant free probability tools.

1.1. Technical overview.

The prior result of the authors gives a sampling algorithm for $\beta < 1/2$ through a set of "desiderata" [DLSS26a, §2] which include pointwise regularity of certain matricial functions of operators that drive the algorithmic process. There are known counter-examples to such pointwise regularity for $0.798 < \beta < 1$ — these are a consequence of the fact that the TAP free energy for the SK model [FMM21] is known not to be *globally* strongly convex in $(-1,1)^n$ when $0.798 < \beta < 1$ [GSS23, Theorem 1.2]. However, a result of Celentano [Cel24] uses control of the empirical distribution of approximate-message passing (AMP) iterates [Cel24, Proposition 3.8] in conjunction with the Sudakov–Fernique inequality to demonstrate local strong convexity for the TAP free energy at $\beta < 1$. This hints that a "relaxation" of the desiderata in [DLSS26a, §2] to require regularity only over a "safe" set of points $S_A(c) \subseteq (-1,1)^n$, those where the TAP free energy continues to be locally strongly convex, likely allows the analysis template developed in [DLSS26a, §3–6] to cover the entire high-temperature regime.

In this work, we accomplish this goal through two important contributions:

(1) We develop the precise formulation for local regularity of various key matricial functionals of the ideal and algorithmic processes, updating the desiderata to reflect this (§2.2).
(2) We systematically extend the SDE error analysis (§3), control over the diagonal of the algorithmic covariance (§5), and proofs for the regularity properties (§6) developed in [DLSS26a, §3, §6 & §B] to hold over non-convex sets under *local* regularity, and demonstrate uniform local strong convexity of the TAP free energy along the stochastic localization (SL) process (§4), thereby proving the requisite "relaxed" desiderata.

Algorithm. The only change to the algorithm of [DLSS26a, Algorithm 3] is to reject any trajectories which leave the safe set, and use a modified sampler to sample from the localized distribution that is known to work for any value of β. The sampler for the localized distribution is directly sourced from Kumar, Sarkar, Tian

[1] A consequence of the generalization of [DLSS26b, §3, §4 & §6] in the forthcoming work [LSS26a] is a weak Poincaré inequality for the SK model up to $\beta < 1$, followed by a proof that a warm start can be algorithmically obtained.

and Zhu [KSTZ26]. Complete descriptions of the main algorithm and the modified ASL-TAP-JE sampler are provided in §2.3.

Algorithm 1 Informal description of our algorithm

repeat
  Initialize $y_0 = 0$.
  $\widetilde{y}_T \leftarrow \mathsf{approximate}(dy_t = m_t dt + dB_t)$ for time $T = O(1)$, while computing Jarzynski weights $\widetilde{w}_T$, restarting if trajectory exits safe set.
  Accept $\widetilde{y}_T$ with some probability computed from $\widetilde{w}_T$.
until accept ▷ Now $\widetilde{y}_T$ is approximately from the time-T distribution of $dy_t = m_t dt + dB_t$.
Initialize $\sigma_0 = \mathrm{sign}(\widetilde{y}_T)$.
Run the sampler for the localized distribution (Theorem 7.1).

Analysis. The cavity interpolation theory built to bound the error in approximating the Gibbs covariance by the inverse of the TAP Hessian already works up to $\beta < 1$ (see §B) and the free probability analysis template remains roughly the same[2] (see §5).

The first step in the analysis comes with a proof that, with high probability over the input and the SL process, the TAP Hessian continues to be uniformly locally strongly convex at any time $t \in [0, T(\beta)]$ (§4). This proof combines Celentano's argument for local strong convexity of the TAP Hessian [Cel24, Theorem 2] in small balls around late-stage AMP iterates with the characterization of "stability" of the SL process for the SK model established in [AMS22, Lemma 4.9] (§4.2). From this, it is a straightforward corollary to obtain a Löwner order "sandwich" for the TAP Hessian on the safe set (§4.3).

The second step in the analysis deals with regularity properties of the inverse of the TAP Hessian, which we interpret as a resolvent. In [DLSS26a, §B], we obtained precise control over the diagonal elements of the squared resolvent, whose importance arises from the fact that those diagonal elements give a measure of how close the TAP Hessian is to being a covariance on the discrete hypercube (the diagonal elements of the covariance matrix are the variances of individual coordinates, which for binary variables are fully determined by their means). Using that diagonal control, [DLSS26a, §6] established regularity and SL-like properties for the coefficients of the algorithmic SDE.

In the present work, we first introduce a C^2 extension of the resolvent from the safe set $S_A(c)$ onto the entirety of $(-1, 1)^n$ (§5.1). This C^2 extension is swapped into both parts of the diagonal control argument where regularity is needed to bound fluctuations: the chaining argument obtaining uniform concentration of the resolvent diagonal (§5.2), as well as a Gaussian interpolation argument comparing the expected resolvent to its freely independent limit (§5.3).

The regularity and SL-likeness of the coefficients then follow (§6.1 – §6.5) almost exactly as in [DLSS26a, §6.1–§6.5], only with Lipschitz bounds proved directly through the resolvent identity instead of through derivative bounds on straight lines that now are no longer guaranteed to exist within the safe set. For SDE well-posedness and transportation inequalities, we swap in a simple Lipschitz extension of the coefficients, and analyze the modified SDE which is guaranteed to agree with the original up until exit from the safe set (§3.3). Finally, the Jarzynski weights are the only Lipschitz bound that can't be obtained only with the resolvent identity: for these, we design short paths within the safe set and integrate bounded derivatives along these short paths to obtain the Lipschitz bound (§6.6).

The third step in the analysis is straightforward given existing work – a simple choice of $T \geqslant T(\beta)$ gives a localized distribution where the external field is strong enough to ensure concentration on a small Hamming ball with high probability over the field, and fast mixing on any Hamming ball of that size [KSTZ26].

The final step in the analysis is to bound the error between the SDE's of the ideal and algorithmic process, when the latter is designed to stop the moment the process leaves the safe set. This requires redoing the

[2] The inverse of the TAP Hessian (a resolvent) needs to be regularized so that it extends smoothly outside the safe set. This leads to certain analytic technicalities, but does not alter the overall structure of the argument in [DLSS26a, §B]. The authors do, however, give a simpler version of the free interpolation argument.

approach taken in [DLSS26a, §3] with delicate error handling for the stopped process to obtain $O(1)$ KL divergence bounds, followed by a "survival"-based Jarzynski's equality (JE) analysis that takes into account the stopped process for the analysis of the rejection sampling step.

1.2. **Related work.** There are two main approaches to sampling from spin glass models. One is to prove functional inequalities (e.g. Poincaré or modified log-Sobolev inequalities), typically via stochastic localization, and deduce rapid mixing of Glauber or Langevin dynamics. It can be sufficient to show weak functional inequalities, which essentially establish the efficacy of these dynamics given a warm start. The other is to algorithmize stochastic localization itself. Our method combines these perspectives: we use ASL and Jarzynski reweighing up to a large finite time, and then use a fast-mixing result for the localized distribution. Since our previous result [DLSS26a], a few notable results have made progress on understanding sampling from the SK model in various settings – we review only these and a few auxiliary results, and refer the reader to [DLSS26a, §1.4] for a more complete overview.

Functional inequalities and fast mixing. Recently, the authors proved a weak Poincaré inequality for the SK model when $\beta < 1/2$ by transferring regularity of approximate stochastic localization to the target measure, yielding fast-mixing for Glauber dynamics from a warm start [DLSS26b]. Shortly thereafter, Wang [Wan26] proved fast-mixing of single-site Glauber dynamics for the SK model at $\beta < 1/2$, from every initial configuration and uniformly over all external fields. Even more recently, Boban, Li, and Oveis Gharan [BLG26] introduced rank-1-perturbed trickledown theorems and obtained polynomial mixing of Glauber dynamics for the SK model at $\beta \leqslant 1/2 + \varepsilon$ for some explicit absolute constant $0 < \varepsilon < 10^{-4}$. For $\beta \leqslant 1/2 + \varepsilon$, these results imply fast-mixing for Glauber dynamics.

Bandeira, El Alaoui, and Rödder [BAR26] prove polynomial mixing of Glauber dynamics for high-overlap Gibbs measures, including the planted SK model with sufficiently strong Gaussian external field. Kumar, Sarkar, Tian, and Zhu [KSTZ26] give polynomial-time samplers for the SK model on sufficiently highly-magnetized slices at every fixed inverse temperature, and combine these with annealing to obtain samplers under strong external fields. We use this result (nearly verbatim) to sample from the localized distribution.

The regime $\beta < 1$ is sharp for our approach and believed to be the onset of hardness for efficient sampling. Specifically, it is known that no stable algorithm (a class containing ASL) can sample with small W_2-distance error for $\beta > 1$ [AMS22], and worst-case Glauber dynamics mixes in exponential time at sufficiently low temperature [Sel25].

Interpolations, overlaps and covariances. Various mathematical results and techniques provide the foundation for analyzing the Gibbs overlap array and covariance for "planted" spin-glass models (cf. [DLSS26a, §1.4]). Most relevant to this paper, the prior result of the authors [DLSS26a, §4–5] develops the precise cavity moment estimates for the planted SK model with SL field needed for the covariance approximation. While stated in [DLSS26a, §5] as a result that is bottlenecked at $\beta < 1/2$, the result extends easily to $\beta < 1$ (see §B). Inspired by the ability to Guerra interpolate the planted SK model with SL tilt, Li and the last two authors prove sub-Gaussian overlap concentration for the Curie–Weiss random-field model at $\beta < 1$ with an external field satisfying the Nishimori fixed-point condition [LSS26b].

PHA, free probability and local TAP convexity. Subag pioneered Hessian Ascent—following the top eigenvector of the Hessian of the objective function—to optimize spherical spin glass models [Sub21], which inspired both Montanari's AMP-based optimization algorithm [Mon19] and the later Potential Hessian Ascent (PHA) framework for the SK model [JSS25, JSS]. For spin glasses on the hypercube, PHA adds a coordinate-wise potential to the objective function, recovering the generalized TAP free energy [CPS23]. Resolvent bounds and techniques from free-probability control the diagonal entries of matrix functions of the Hessian, which determine how individual coordinates evolve and give access to regularity properties. In the prior work of the authors, these techniques were used to provide global regularity properties [DLSS26a, §6]. The present work establishes diagonal control in the region of local TAP convexity even when global control is impossible. A simplification of the previous argument is enabled by comparisons of matricial functions of Gaussian random matrices with their free limits developed by Bandeira, Boedihardjo, and van Handel [BBv23].

Work	Method	Guarantee	Regime ($\beta < \cdots$)
Functional inequalities / fast mixing			
Eldan, Koehler, and Zeitouni [EKZ22]; Anari et al. [AJK+22]	PI / MLSI via SL	ε-TVD	1/4
Anari, Koehler, and Vuong [AKV24]	MLSI via SL	ε-TVD	≈ 0.295
Davies et al. [DLSS26b]	weak PI; warm-start Glauber	$o(1)$-TVD	1/2
Wang [Wan26]	PI / MLSI; Glauber	ε-TVD	1/2 (any ext. field)
Boban, Li, and Oveis Gharan [BLG26]	rank-1 trickledown; Glauber	ε-TVD	$\leqslant 1/2 + \varepsilon$
Algorithmic stochastic localization			
El Alaoui, Montanari, and Selke [AMS22]; Celentano [Cel24]	ASL + AMP	$o(n)$-W_2^2	1/2, then 1
Davies et al. [DLSS26a]	ASL-TAP / PHD + JE	$o(1)$-TVD	1/2
This work	ASL-TAP / PHD + JE	$o(1)$-TVD	1

TABLE 1.1. Sampling guarantees for the SK model. "ε-TVD" denotes error ε achievable for every fixed $\varepsilon > 0$ with running time polynomial in $1/\varepsilon$; "$o(1)$-TVD" denotes error vanishing as $n \to \infty$.

2. Sampling via ASL-JE under local regularity

Our sampling algorithm closely follows the implementation of *algorithmic stochastic localization* described in [DLSS26a, §2], so we defer to that section for the full explanation, and will describe only the differences here. In multiple places, the original approach used *global strong convexity* of the TAP free energy and the consequent regularity of various random processes defined over the domain $(-1,1)^n$ of the magnetization. Whereas this global strong convexity has a particularly simple proof at $\beta < 1/2$, it is known to fail for $\beta > 0.798$ [GSS23]. Thus, in the current regime of interest $1/2 \leqslant \beta < 1$, many of the TAP-based quantities used in the argument of [DLSS26a] actually fail to exist.

We circumvent this issue by giving an analysis that only requires *local convexity* of the TAP free energy around likely SL paths, which can be proved up to $\beta < 1$ by building upon an argument of Celentano [Cel24]. We define a *safe subset* of $(-1,1)^n$ characterized by this local convexity property, and relax our SDE and JE analyses to require regularity only up to (the unlikely event of an) escape from this safe set. Different parts of the argument require different treatments of the complement of the safe set, including Lipschitz or C^2 extensions of functions to the complement of the safe set, the construction of short paths that stay inside the safe set, and the use of the resolvent identity to prove Lipschitz bounds without derivative bounds on straight-line paths.

2.1. TAP free energy and the safe set.

First, we recall the definition of the TAP free energy, then use that to define our notion of safe sets.

Definition 2.1 (TAP free energy for the planted SK model, [TAP77, FMM21])**.** Denote the TAP free energy for the Hamiltonian $\langle m, Am\rangle$ at inverse temperature $\beta < 1$ as

$$\mathcal{F}_{\mathrm{TAP}}(m,y) := -\frac{\beta}{2}\langle m, Am\rangle - \langle y, m\rangle - \sum_{i=1}^{n} h(m_i) - \frac{1}{4}n\beta^2\left(1 - \frac{1}{n}\|m\|_2^2\right)^2, \tag{2.1}$$

where $h(x) := -\frac{1+x}{2}\log\left(\frac{1+x}{2}\right) - \frac{1-x}{2}\log\left(\frac{1-x}{2}\right)$ and $\widehat{Q}(m) := \nabla_m^2 \mathcal{F}_{\mathrm{TAP}}(m)^{-1}$ where $m \in (-1,1)^n$.

Note that $\nabla_m^2 \mathcal{F}_{\mathrm{TAP}}(m,y)$ does not depend on y, so we will write this simply as $\nabla_m^2 \mathcal{F}_{\mathrm{TAP}}(m)$. Note

$$\nabla_m^2 \mathcal{F}_{\mathrm{TAP}} = -\beta A + \mathsf{diag}\left(\frac{1}{1-m^2}\right) + \beta^2\left(1 - \frac{\|m\|_2^2}{n}\right) I_n - \frac{2\beta^2}{n} mm^{\mathsf{T}} =: \widehat{Q}(m)^{-1}, \tag{2.2}$$

where the definition is when $\nabla_m^2 \mathcal{F}_{\mathrm{TAP}}(m) \succ 0$. We define the set of safe points in the solid cube as those near which $\mathcal{F}_{\mathrm{TAP}}(m,\cdot)$ is sufficiently strongly convex in m.

Definition 2.2 (Safe set in the hypercube)**.** Fix $A \in \mathbb{R}^{n\times n}$ and $D(m) := \mathsf{diag}\left(\frac{1}{1-m^2}\right)$ for any $m \in (-1,1)^n$. Let

$$\mathcal{S}_{\beta,A}(c) := \left\{ m \in (-1,1)^n \mid c\, D(m) \prec \nabla^2 \mathcal{F}_{\mathrm{TAP}}(m) \right\}.$$

Given $m \in (-1,1)^n$, there is a unique $y \in \mathbb{R}^n$ such that (m,y) solve the TAP equation $\nabla_m \mathcal{F}_{\mathrm{TAP}}(m,y) = 0$. Define $\widehat{y} : (-1,1)^n \to \mathbb{R}^n$ to be this value of y:

$$\widehat{y}(m) := -\frac{\beta}{2}\langle m, Am\rangle - \sum_{i=1}^n h(m_i) - \frac{1}{4} n\beta^2 \left(1 - \frac{1}{n}\|m\|_2^2\right)^2. \tag{2.3}$$

Note that $\widehat{y}$ may not be injective even on $\mathcal{S}_{\beta,A}(c)$, but it is locally; see Lemma 3.4.

2.2. **Desiderata.** We now provide the updated list of desiderata, modifying those provided in [DLSS26a, §2] to work within the safe set and to present a more minimal interface. These desiderata formalize the requirement of the TAP free energy to be a good approximation of the actual free energy while having good regularity properties, to prove efficient sampling with low error measured in total variation distance.

The following desiderata are written over SL paths $\{m_t\}_{t\geqslant 0}$ which come as a solution to HD.

(d1) **($O(1)$-error for covariance estimate, Theorem B.3)** The following covariance estimate holds for $0 < \beta < 1$:

$$\mathop{\mathbb{E}}_{A,B_t}\left[\left\|\nabla_m^2 \mathcal{F}_{\mathrm{TAP}}(m_t) Q_t - I_n\right\|_F^2\right] \leqslant \varepsilon_{\mathrm{cov}}(\beta,t)^2,$$

where $\sup_{t\in[0,T]} \varepsilon_{\mathrm{cov}}(\beta,t) < \infty$.

(d2a) **(Uniform local strong convexity of TAP Hessian over SL process, Theorem 4.2)** For every T, there are constants $\rho := \rho(\beta,T) > 0$, $c_{\mathrm{TAP}} := c_{\mathrm{TAP}}(\beta,T) > 0$ such that

$$\mathop{\mathbb{P}}_{A,B_t}\left\{\forall t \in [0,T],\ B_\square(m_t, \rho\sqrt{n}) \subseteq \mathcal{S}_{\beta,A}(c_{\mathrm{TAP}})\right\} \geqslant 1 - o_n(1),$$

where $B_\square(x,r) := \{u \in (-1,1)^n : \|u - x\|_\infty < r\}$ is the x-centered ball of radius r in the ℓ_2 distance within $(-1,1)^n$.

The following desiderata are written over all $m \in \mathcal{S}_{\beta,A}(c_{\mathrm{TAP}})$, on a high-probability event on A:

(d2b) **(Local regularity of the TAP Hessian, Lemma 4.3)** There exist finite $c_{\mathrm{TAP}}(\beta,T), C_{\mathrm{TAP}}(\beta) > 0$, such that

$$C_{\mathrm{TAP}}(\beta)^{-1} D^{-1}(m) \preceq \widehat{Q}(m) \preceq c_{\mathrm{TAP}}(\beta,T)^{-1} D^{-1}(m).$$

(d3) **(Drift error in PHD-TAP, Corollary 6.10)**

$$\|f(m) - m\|_2^2 \leqslant \varepsilon_{\mathrm{drift}}(\beta)^2,$$

where $\varepsilon_{\mathrm{drift}}(\beta)$ is a constant depending only on β and

$$f(m) = \left(E_{\mathcal{D}_n}\left[\widehat{Q}(m)^2\right] D(m)^2 - \frac{\beta^2}{n}\left(\mathsf{Tr}\left[\widehat{Q}^2(m)\right] I_n + 2\widehat{Q}^2(m)\right)\right) m \tag{2.4}$$

(d4) **(Lipschitzness of diffusion and drift terms, Corollary 6.10)**

(a) There exists a constant L such that for all $m, w \in \mathcal{S}_{\beta,A}(c_{\mathrm{TAP}})$,

$$\left\|\widehat{Q}(m) - \widehat{Q}(w)\right\|_F \leqslant L\|m - w\|_2.$$

(b) There exists a constant L_{drift} such that for all $m, w \in \mathcal{S}_{\beta,A}(c_{\mathrm{TAP}})$, with f defined as in [DLSS26a, (2.3)],
$$\left\|\widehat{Q}(m)(m - f(m)) - \widehat{Q}(w)(w - f(w))\right\|_2 \leqslant L_{\mathrm{drift}} \|m - w\|_2 .$$

(d5) **(Lipschitzness of JE weight integrand, Corollary 6.10)** Let
$$\omega(m) := \frac{1}{2}\big(\mathsf{Tr}[\widehat{Q}(m)] + \|m\|_2^2\big).$$
Then there is an $L_\omega < \infty$ such that
$$|\omega(m) - \omega(w)| \leqslant L_\omega \|m - w\|_2 \qquad (\forall m, w \in \mathcal{S}_{\beta,A}(c)).$$

Note that as a consequence of these desiderata, we will also be able to prove (d2a) for the algorithmic $\widehat{m}_t$. This is done via a SDE-based argument in Lemma 3.2.

Finally, we need an efficient sampling algorithm for the localized distribution. For large enough time, the measure concentrates on a small Hamming ball, so the following statement suffices.

(d6) **(Efficient sampling on small enough wedges, Theorem 7.1)** There exists small enough $\varepsilon(\beta) > 0$ such that with high probability over A, there is a $\mathsf{poly}(n, \log(1/\varepsilon))$-time algorithm to sample from $\mu_{\beta' A, y}|_{B(\mathrm{sign}(y), \varepsilon(\beta)n)}$ with TV distance ε, for all $y \in \mathbb{R}^n$ and all $0 \leqslant \beta' \leqslant \beta$, where $B(\sigma, k) = \{\sigma' \in \{\pm 1\}^n : d(\sigma, \sigma') \leqslant k\}$ is the Hamming ball around σ with radius k.

One of the high-probability events we will need is simply an operator norm bound on A.

Proposition 2.3 (GOE operator norm concentration, [DS01, Theorem II.1])**.**
Let $A \sim \mathsf{GOE}(n)$. Then,
$$\mathbb{P}\{\|A\|_{\mathsf{op}} > 3\} \leqslant 2e^{-n/4}. \tag{2.5}$$

Proof. This is given by the Gaussian eigenvalue tail bound of [DS01, Theorem II.11] applied to A and $-A$. □

2.3. **Algorithm.** We display the full algorithm in Algorithm 2. First, we run ASL-TAP-JE up to stopping time τ_c as derived in [DLSS26a, §2.2–2.3]; we separate this out into Algorithm 3. As in [DLSS26a, Algorithm 3] we conduct rejection sampling as described in [DLSS26a, §2.3] to turn our weighted sample for ρ_T into an approximate sample for μ_T. Finally, we sample from the localized distribution μ_{y_T} using our approximate sample as a warm-start.

Algorithm 2 SK Sampler

1: **Input:** matrix A, inverse temperature β, total time T which is a multiple of step size η.
2: **Output:** Approximate sample from Gibbs measure μ_A.
3: **Part 1:** Approximately sample from μ_T.
4: Run Algorithm 4 (Approximate Rejection Sampler) with subroutine for sampling and density ratio estimation given by Algorithm 3 (ASL-TAP-JE with density ratio estimation) to obtain $\widetilde{y}_T$.
5: **Part 2:** Approximately sample from localized distribution $\mu_{\beta A, \widetilde{y}_T}$.
6: Run the sampler for $\mu_{\beta A, \sigma_0}$ restricted to the Hamming ball $B(\sigma_0, c(\beta)n)$, given by Theorem 7.1 ([KSTZ26]).

3. Quantitative SDE analysis

We first recall the various processes which appear in our algorithm and analysis. For a full derivation, see [DLSS26a, §3]. First, the ideal process we would like to approximate is stochastic localization in y-space or

Algorithm 3 (ASL-TAP-JE) with density ratio estimation

Input: matrix A, inverse temperature β, total time T, step size η, number of steps S for mirror descent, error parameter ε
Let $\widetilde{y}_0 = 0$, $\widetilde{m}_0 = 0$, $\widetilde{x}_0 = 0$, and $\widetilde{w}_0 = 0$.
for $t \in \{0, \eta, \dots, T - \eta\}$ **do**
  Draw $\xi_t \sim \mathcal{N}(0, I_n)$.
  $\widetilde{y}_{t+\eta} = \widetilde{y}_t + \eta \widetilde{m}_t + \sqrt{\eta}\xi_t$.
  $\widetilde{x}^{(0)}_{t+\eta} = \widetilde{x}_t + \eta\left[D(\widetilde{m}_t)\widehat{Q}(\widetilde{m}_t)(\widetilde{m}_t - f(\widetilde{m}_t)) + E_{\mathcal{D}_n}\left(\widehat{Q}(\widetilde{m}_t)^2\right) D(\widetilde{m}_t)^2\widetilde{m}_t\right] + \sqrt{\eta}D(\widetilde{m}_t)\widehat{Q}(\widetilde{m}_t)\xi_t$.
  for $s \in \{0, \dots, S-1\}$ **do** ▷ Solve $\nabla_m \mathcal{F}_{\mathrm{TAP}}(m, y) = 0$ for m when $y = \widetilde{y}_{t+\eta}$ using mirror descent
    $\widetilde{x}^{(s+1)}_{t+\eta} = \widetilde{x}^{(s)}_{t+\eta} - \lambda \nabla_m \mathcal{F}_{\mathrm{TAP}}(\tanh(\widetilde{x}^{(s)}_{t+\eta}), \widetilde{y}_{t+\eta})$ where $\nabla_m \mathcal{F}_{\mathrm{TAP}}$ is given in (3.2) and $\lambda = C_{\mathrm{TAP}}(\beta)^{-1}$ in (d2b).
  end for
  $\widetilde{x}_{t+\eta} = \widetilde{x}^{(S)}_{t+\eta}$.
  $\widetilde{m}_{t+\eta} = \tanh(\widetilde{x}_{t+\eta})$.
  $\widetilde{w}_{t+\eta} = \widetilde{w}_t + \frac{\eta}{2}\left[\mathsf{Tr}(\widehat{Q}(\widetilde{m}_t)) + \|\widetilde{m}_t\|^2\right]$.
  If $\nabla^2 \mathcal{F}_{\mathrm{TAP}}(\widetilde{m}_{t+\eta}) \not\succ cD(m)$, then restart.
end for
Estimate $\widehat{Z}_{\beta A, \widetilde{y}_T}$ using simulated annealing ([DLSS26a, Algorithm 5]) with a sampler given by Theorem 7.1 ([KSTZ26]).
Output: $\left(\widetilde{y}_T, e^{\widetilde{w}_T}\widehat{Z}_{\beta A, \widetilde{y}_T} e^{\mathcal{F}_{\mathrm{TAP}}(\widetilde{m}_T, \widetilde{y}_T)}\right)$

Hessian dynamics in m-space:

$$dy_t = m_t dt + dB_t, \qquad y_0 = 0, \tag{SL}$$

$$dm_t = Q_t dB_t, \qquad m_0 = \mathop{\mathbb{E}}_{x_0 \sim \mu}[\sigma] = 0, \tag{HD}$$

$$m_t = \langle \sigma \rangle_{y_t} = \mathop{\mathbb{E}}_{\mu_{y_t}} \sigma, \qquad Q_t = \mathbf{Cov}(\mu_{y_t}), \tag{3.1}$$

where $\mu_y(\sigma) \propto \mu(\sigma) e^{\langle y, \sigma \rangle}$ and $\mu = \mu_{\beta A}$ is the desired Gibbs measure. Two natural algorithmic processes arise from using an estimate of m_t in (SL) or an estimate of Q_t in (HD), and then enforcing the TAP equation

$$0 = \nabla \mathcal{F}_{\mathrm{TAP}}(m, y) = -\beta A m - y + \underbrace{\frac{1}{2}\log\left(\frac{1+m}{1-m}\right)}_{\tanh^{-1}(m)} + \beta^2\left(1 - \frac{1}{n}\|m\|_2^2\right) m \tag{TAP}$$

to solve for the other variable (y or m). This constraint is given by differentiating (TAP), which gives

$$d\widehat{y}_t = \widehat{Q}(\widehat{m}_t)^{-1} d\widehat{m}_t + f(\widehat{m}_t) dt. \tag{3.2}$$

Define $\widehat{Q}(m)$ as in (2.2). Then the two options give

$$\begin{cases} d\widehat{y}_t = \widehat{m}_t dt + dB_t \\ d\widehat{m}_t = \widehat{Q}(\widehat{m}_t)\left(\widehat{m}_t - f(\widehat{m}_t)\right) dt + \widehat{Q}(\widehat{m}_t) dB_t \end{cases} \tag{ASL-TAP}$$

$$\begin{cases} d\widehat{m}_t = \widehat{Q}(\widehat{m}_t) dB_t \\ d\widehat{y}_t = f(\widehat{m}_t) dt + dB_t \end{cases} \tag{PHD-TAP}$$

These are defined if $\widehat{m}_t$ stays inside a safe set $\mathcal{S}_{\beta,A}(c)$, and we can extend them outside by modifying the drift and diffusion terms appropriately—see §3.3 and §3.4. For (ASL-TAP), we define Jarzynski weights by

$$dw_t = \omega(\widehat{m}_t)\, dt, \qquad \omega(m) := \frac{1}{2}\left[\mathsf{Tr}(\widehat{Q}(m)) + \|m\|^2\right]. \tag{JE}$$

Finally, for analysis purposes, we also need to consider (SL) and keep track of the estimated magnetization for tilt y_t according to (TAP), which we denote as m_t^y (i.e., we consider y under the law of the SL process, but find what the m would be if it were the algorithmic process). We derive this by solving (3.2) with $d\widehat{y}_t$

replaced by dy_t in (SL).

$$\begin{cases} dy_t = m_t\,dt + dB_t \\ dm_t^y = \widehat{Q}(m_t^y)(dy_t - f(m_t^y)\,dt) = \widehat{Q}(m_t^y)((m_t - f(m_t^y))\,dt + dB_t). \end{cases} \tag{SL-TAP}$$

3.1. $O(1)$ **Wasserstein and KL bounds.** We first consider a general setup and give a Wasserstein bound for general SDE's with different drift and diffusion terms, when Lipschitzness holds locally. Let $\mathcal{S} \subseteq \Omega \subseteq \mathbb{R}^n$, where $\mathcal{S}$ and Ω are open and $\mathcal{S}$ is considered the "safe set". Consider 2 Itô processes

$$\begin{aligned} dX_t &= f_t\,dt + \Sigma_t\,dB_t \\ d\widehat{X}_t &= \widehat{f}_t\,dt + \widehat{\Sigma}_t\,dB_t \end{aligned} \tag{3.3}$$

and assume that pathwise unique strong solutions exist for X_t, and for $\widehat{X}_t$ up to the first exit time

$$\tau = \inf\left\{t \geqslant 0 : \widehat{X}_t \notin \mathcal{S}\right\}.$$

Suppose that $X_0 = \widehat{X}_0$. We will think of X_t as an ideal process and $\widehat{X}_t$ as an approximate process. Define the expanded safe set and safe pair set by

$$\begin{aligned} \mathcal{S}(r) &= \{x \in \Omega : B(x,r) \cap \Omega \subseteq \mathcal{S}\}, \\ \mathcal{S}^2(r_1,r_2) &= \left\{(x_1,x_2) \in \Omega^2 : x_1 \in \mathcal{S}(r_1),\ \|x_1 - x_2\| < r_2\right\}, \quad r_1 \geqslant r_2 \\ \mathcal{S}^2(r) &= \mathcal{S}^2(r,r), \end{aligned}$$

and define the path-space analogues by

$$\begin{aligned} \mathfrak{S}_t &= \{\gamma : \forall s \in [0,t],\ \gamma_s \in \mathcal{S}\} \\ \mathfrak{S}_t(r) &= \{\gamma : \forall s \in [0,t],\ \gamma_s \in \mathcal{S}(r)\} \\ \mathfrak{S}_t^2(r_1,r_2) &= \left\{(\gamma_1,\gamma_2) : \forall s \in [0,t],\ (\gamma_{1,s},\gamma_{1,t}) \in \mathcal{S}^2(r_1,r_2)\right\}, \quad r_1 \geqslant r_2 \\ \mathfrak{S}_t^2(r) &= \mathfrak{S}_t^2(r,r). \end{aligned}$$

Note that if $(\gamma_1,\gamma_2) \in \mathfrak{S}_t^2(r_1,r_2)$, then $\gamma_1,\gamma_2 \in \mathfrak{S}_t$. Note that if we apply this to the safe set $\mathcal{S} = \mathcal{S}_{\beta,A}(c)$, then for the expanded safe set of (d2a),

$$\left\{(y_{[0,t]}, m_{[0,t]}) : \forall t \in [0,t],\ B_\square(m_t,r) \subseteq \mathcal{S}_{\beta,A}(c)\right\} \subseteq \mathcal{S}(r),$$

so the $1 - o_n(1)$ probability transfers to $\mathcal{S}(r)$.

Lemma 3.1 (Grönwall bound for Wasserstein error of SDE's, under local Lipschitzness)**.** *Consider the setup of (3.3) where X_t and $\widehat{X}_t$ are diffusion processes, that is, $f_t = f_t(X_t)$, $\Sigma_t = \Sigma_t(X_t)$, $\widehat{f}_t = \widehat{f}_t(X_t)$, $\widehat{\Sigma}_t = \widehat{\Sigma}(\widehat{X}_t)$. Suppose also that for $t \in [0,T]$,*

(1) $\mathbb{E}\left[\mathbb{1}_{\tau(\widehat{X}_t)>t}\left\|f_t(\widehat{X}_t) - \widehat{f}_t(\widehat{X}_t)\right\|^2\right] \leqslant \varepsilon_f(t)^2$ and $\mathbb{E}\left[\mathbb{1}_{\tau(X_t)>t}\left\|\Sigma_t(X_t) - \widehat{\Sigma}_t(X_t)\right\|_F^2\right] \leqslant \varepsilon_\Sigma(t)^2$, where $X_t, \widehat{X}_t$ evolve according to the SDE (3.3).
(2) f_t is L_f-segmentwise Lipschitz and $\widehat{\Sigma}_t$ is L_Σ-segmentwise Lipschitz on $\mathcal{S}$ (with respect to the $\|\cdot\|_2$ and $\|\cdot\|_F$ norms), for some $L_f, L_\Sigma > 0$.
(3) $\mathbb{P}(X_{[0,T]} \in \mathfrak{S}_T(r)) \geqslant 1 - \delta_1$.

Then for $t \in [0,T]$, with X_t and $\widehat{X}_t$ driven by the same Brownian motion in (3.3) (i.e. synchronous coupling), we have

$$\mathbb{E}\left[\mathbb{1}_{\mathfrak{S}_t^2(r)}[X_{[0,t]}, \widehat{X}_{[0,t]}]\left\|X_t - \widehat{X}_t\right\|^2\right] \leqslant \int_0^t \left(\frac{\sqrt{2}}{L_f}\varepsilon_f(s)^2 + 2\varepsilon_\Sigma(s)^2\right) e^{(2\sqrt{2}L_f + 2L_\Sigma^2)(t-s)} ds.$$

Moreover, if $r_2 \leqslant r$ and $t \leqslant \frac{1}{2L_1}\log r_2$,

$$\mathbb{P}[(X_{[0,t]}, \widehat{X}_{[0,t]}) \notin \mathfrak{S}^2(r,r_2)] \leqslant \delta_1 + \frac{32tL_\Sigma^2 + \int_0^t 32\varepsilon_\Sigma(s)^2 + 8\varepsilon_f(s)\,ds}{\log r_2} + \frac{1}{r_2}.$$

Note in particular that if $\varepsilon_f(t), \varepsilon_\Sigma(t), L_f, L_\Sigma = O(1)$ (uniformly over $t \in [0,T]$), $\delta_1 = o_n(1)$, and $r = \omega_n(1)$, then the expectation is $O(1)$ and the bad probability is $o_n(1)$.

Proof. Let $\mathcal{F}_t$ be the filtration for the Brownian motion. Let $\Delta_t := X_t - \widehat{X}_t$. Note

$$d\Delta_t = (f_t(X_t) - \widehat{f}_t(\widehat{X}_t))dt + (\Sigma_t(X_t) - \widehat{\Sigma}_t(\widehat{X}_t))dB_t.$$

By Itô's formula, for $t \geqslant u$,

$$\begin{aligned} d\left\|\Delta_t\right\|^2 &= 2\left\langle \Delta_t, (f_t(X_t) - \widehat{f}_t(\widehat{X}_t))dt + (\Sigma_t(X_t) - \widehat{\Sigma}_t(\widehat{X}_t))dB_t \right\rangle + \left\|\Sigma_t(X_t) - \widehat{\Sigma}_t(\widehat{X}_t)\right\|_F^2 dt \\ \implies \frac{d}{dt}\mathbb{E}\left[\left\|\Delta_t\right\|^2 \mid \mathcal{F}_u\right] &= 2\,\mathbb{E}\left[\left\langle \Delta_t, f_t(X_t) - \widehat{f}_t(\widehat{X}_t)\right\rangle \mid \mathcal{F}_u\right] + \mathbb{E}\left[\left\|\Sigma_t(X_t) - \widehat{\Sigma}_t(\widehat{X}_t)\right\|_F^2 \mid \mathcal{F}_u\right] \\ &\leqslant c\,\mathbb{E}\left[\left\|\Delta_t\right\|^2 \mid \mathcal{F}_u\right] + \frac{1}{c}\,\mathbb{E}\left[\left\|f_t(X_t) - \widehat{f}_t(\widehat{X}_t)\right\|^2 \mid \mathcal{F}_u\right] + \mathbb{E}\left[\left\|\Sigma_t(X_t) - \widehat{\Sigma}_t(\widehat{X}_t)\right\|_F^2 \mid \mathcal{F}_u\right] \end{aligned}$$

for any $c > 0$. Now for $x \in \mathfrak{S}(r)$ and $\widehat{x} \in B(x, r)$, by the definition of $\mathfrak{S}(r)$ and segmentwise Lipschitzness on $\mathcal{S}$,

$$\begin{aligned} \left\|\widehat{f}_t(\widehat{x}) - f_t(x)\right\|^2 &\leqslant 2\left\|f_t(\widehat{x}) - f_t(x)\right\|^2 + 2\left\|\widehat{f}_t(\widehat{x}) - f_t(\widehat{x})\right\|^2 \leqslant 2L_f^2\left\|\widehat{x} - x\right\|^2 + 2\left\|\widehat{f}_t(\widehat{x}) - f_t(\widehat{x})\right\|^2 \\ \left\|\widehat{\Sigma}_t(\widehat{x}) - \Sigma_t(x)\right\|_F^2 &\leqslant 2\left\|\widehat{\Sigma}_t(\widehat{x}) - \widehat{\Sigma}_t(x)\right\|_F^2 + 2\left\|\widehat{\Sigma}_t(x) - \Sigma_t(x)\right\|_F^2 \leqslant 2L_\Sigma^2\left\|\widehat{x} - x\right\|^2 + 2\left\|\widehat{\Sigma}_t(x) - \Sigma_t(x)\right\|_F^2. \end{aligned}$$

Plugging in gives that for $t = u$, on the event $\mathfrak{S}^2(r)$,

$$\begin{aligned} \frac{d}{dt}\mathbb{E}\left[\left\|\Delta_t\right\|^2 \mid \mathcal{F}_u\right] &\leqslant \left(c + \frac{2L_f^2}{c} + 2L_\Sigma^2\right)\mathbb{E}\left[\left\|\Delta_t\right\|^2 \mid \mathcal{F}_u\right] + \frac{2}{c}\,\mathbb{E}\left[\left\|\widehat{f}_t(\widehat{X}_t) - f_t(\widehat{X}_t)\right\|^2 \mid \mathcal{F}_u\right] + 2\,\mathbb{E}\left[\left\|\widehat{\Sigma}_t(X_t) - \Sigma_t(X_t)\right\|_F^2 \mid \mathcal{F}_u\right] \\ &= \left(2\sqrt{2}L_f + 2L_\Sigma^2\right)\mathbb{E}\left[\left\|\Delta_t\right\|^2 \mid \mathcal{F}_u\right] + \frac{\sqrt{2}}{L_f}\,\mathbb{E}\left[\left\|\widehat{f}_t(\widehat{X}_t) - f_t(\widehat{X}_t)\right\|^2 \mid \mathcal{F}_u\right] + 2\,\mathbb{E}\left[\left\|\widehat{\Sigma}_t(X_t) - \Sigma_t(X_t)\right\|_F^2 \mid \mathcal{F}_u\right] \end{aligned}$$

by taking $c = \sqrt{2}L_f$. Now take $L = 2\sqrt{2}L_f + 2L_\Sigma^2$ and $E_t(x, \widehat{x}) = \frac{\sqrt{2}}{L_f}\left\|\widehat{f}_t(\widehat{x}) - f_t(\widehat{x})\right\|^2 + 2\left\|\widehat{\Sigma}_t(x) - \Sigma_t(x)\right\|_F^2$, and $\varepsilon(t) = \frac{\sqrt{2}}{L_f}\varepsilon_f(t)^2 + 2\varepsilon_\Sigma(t)^2$. This shows that for $t = u$, on the event $\mathfrak{S}^2(r)$,

$$\frac{d}{dt}\mathbb{E}[M_t \mid \mathcal{F}_u] \leqslant 0 \text{ where } M_t = e^{-Lt}\left\|\Delta_t\right\|^2 - \int_0^t E_t(X_t, \widehat{X}_t)e^{-Ls}\,ds$$

Define the stopping time

$$\tau_2 = \inf\left\{t \in [0, T] : (X_{[0,t]}, \widehat{X}_{[0,t]}) \notin \mathfrak{S}_t^2(r)\right\}.$$

On the event that $\tau_2 \leqslant u$, we have $\mathbb{E}[\left\|\Delta_{t\wedge\tau_2}\right\|^2 \mid \mathcal{F}_u] = \left\|\Delta_{\tau_2}\right\|^2$, so the above chain of inequalities also holds. Hence

$$\frac{d}{dt}\mathbb{E}\left[M_{t\wedge\tau_2}\right] \leqslant 0.$$

Hence

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{\mathfrak{S}_t^2(r)}[X_{[0,t]}, \widehat{X}_{[0,t]}]\left\|X_t - \widehat{X}_t\right\|^2] \leqslant \mathbb{E}\left[\left\|\Delta_{t\wedge\tau_2}\right\|^2\right] &\leqslant \int_0^t \mathbb{E}\left[\mathbb{1}_{\mathfrak{S}^2(r)}[X_{[0,t]}, \widehat{X}_{[0,t]}]E_t(X_t, \widehat{X}_t)\right]e^{L(t-s)}ds \\ &\leqslant \int_0^t \varepsilon(s)e^{L(t-s)}ds. \end{aligned}$$

The second part follows from the next lemma, Lemma 3.2 with $L_b = L_f$, $\varepsilon_b(t) = \varepsilon_f(t)$, $L_C = L_\Sigma$, $\varepsilon_C(t) = \varepsilon_\Sigma(t)$. □

The following lemma shows that, given that an "ideal" process stays within the expanded safe set with high probability, given Lipschitz and expected error bounds for the drift and diffusion terms for the difference, the approximate process also stays close with high probability.

Lemma 3.2.

Consider the setup of (3.3), and suppose that $\Delta_t = X_t - \widehat{X}_t$ satisfies

$$d\Delta_t = b_t\,dt + C_t\,dB_t.$$

Define the stopping time $\tau_2 = \inf\left\{t \geqslant 0 : (X_t, \widehat{X}_t) \notin \mathcal{S}^2(r)\right\}$. Suppose that we can write $b_t = b_t^{\mathrm{Lip}} + b_t^{\mathrm{err}}$ and $C_t = C_t^{\mathrm{Lip}} + C_t^{\mathrm{err}}$ such that when $(X_t, \widehat{X}_t) \in \mathcal{S}^2(r)$,

$$\left\|b_t^{\mathrm{Lip}}\right\| \leqslant L_b\left\|\Delta_t\right\|, \qquad \mathbb{E}\left[\mathbb{1}_{t<\tau_2}\left\|b_t^{\mathrm{err}}\right\|^2\right] \leqslant \varepsilon_b(t)^2$$

$$\left\|C_t^{\mathrm{Lip}}\right\|_F \leqslant L_C \left\|\Delta_t\right\|, \qquad \mathbb{E}\left[\mathbb{1}_{t<\tau_2} \left\|C_t^{\mathrm{err}}\right\|_F^2\right] \leqslant \varepsilon_C(t)^2.$$

Suppose that $\mathbb{P}(X_{[0,t]} \notin \mathfrak{S}_t(r)) \leqslant \delta_1$. *Then if* $r \leqslant r_2$ *and* $t \leqslant \frac{1}{2L_1} \log r_2$,

$$\mathbb{P}[(X_{[0,t]}, \widehat{X}_{[0,t]}) \notin \mathfrak{S}^2(r, r_2)] \leqslant \delta_1 + \frac{32tL_C^2 + \int_0^t 32\varepsilon_C(s)^2 + 8\varepsilon_b(s)\, ds}{\log r_2} + \frac{1}{r_2}.$$

Proof. We will use a (super)martingale concentration inequality. Let $r_t = \sqrt{\left\|\Delta_t\right\|^2 + 1}$. By Itô's formula,

$$d\log(\left\|\Delta_t\right\|^2 + 1) = \frac{2\left\langle \Delta_t, b_t\, dt + C_t\, dB_t\right\rangle}{r_t^2} + \frac{2\left\|C_t\right\|_F^2}{r_t^2}\, dt - \frac{\left\|C_t^\intercal \Delta_t\right\|^2}{r_t^4} dt.$$

On the event $\mathfrak{S}_t^2(r)$,

$$\frac{\left\|b_t\right\|}{r_t} \leqslant \frac{\left\|b_t^{\mathrm{Lip}}\right\|}{r_t} + \frac{\left\|b_t^{\mathrm{err}}\right\|}{r_t} \leqslant L_b + \left\|b_t^{\mathrm{err}}\right\|$$
$$\frac{\left\|C_t\right\|_F^2}{r_t^2} \leqslant \frac{2\left\|C_t^{\mathrm{Lip}}\right\|_F^2}{r_t^2} + \frac{2\left\|C_t^{\mathrm{err}}\right\|_F^2}{r_t^2} \leqslant 2L_C^2 + 2\left\|C_t^{\mathrm{err}}\right\|_F^2.$$

Hence, for $t < \tau_2$,

$$d\log(\left\|\Delta_t\right\|^2 + 1) \leqslant \left(2L_b + 4L_C^2 + 2\left\|b_t^{\mathrm{err}}\right\| + 4\left\|C_t^{\mathrm{err}}\right\|_F^2\right) dt + \frac{2\left\langle \Delta_t, C_t\, dB_t\right\rangle}{r_t^2}.$$

Hence, letting $L_1 = 2L_b + 4L_C^2$, $F_t = 2\left\|b_t^{\mathrm{err}}\right\| + 4\left\|C_t^{\mathrm{err}}\right\|_F^2$, and $N_t = \log(\left\|\Delta_t\right\|^2 + 1) - L_1 t - \int_0^t F_s\, ds$, for $t < \tau_2$,

$$dN_t \leqslant \frac{2\left\langle \Delta_t, C_t\, dB_t\right\rangle}{r_t^2}.$$

The quadratic variation is bounded as

$$[N]_{t\wedge\tau_2} \leqslant \int_0^{t\wedge\tau_2} \frac{4\left\|\Delta_s\right\|^2 \left\|C_s\right\|_{\mathsf{op}}^2}{r_s^4}\, ds \leqslant \int_0^{t\wedge\tau_2} \frac{4\left\|C_s\right\|_F^2}{r_s^2}\, ds \leqslant \int_0^{t\wedge\tau_2} 8L_C^2 + 8\left\|C_s^{\mathrm{err}}\right\|_F^2\, ds \leqslant 8tL_C^2 + 8\int_0^{t\wedge\tau_2} \left\|C_s^{\mathrm{err}}\right\|_F^2\, ds.$$

We have the martingale concentration bound

$$\mathbb{P}\left(\sup_{s\leqslant t\wedge\tau_2} N_s \geqslant x,\, [N]_{t\wedge\tau_2} \leqslant v\right) \leqslant e^{-\frac{x^2}{2v}}.$$

We now take expectations for $[N]_{t\wedge\tau_2}$ and the random part of $N_{t\wedge\tau_2}$. We have

$$\mathbb{E}[N]_{t\wedge\tau_2} \leqslant 8tL_C^2 + 8\int_0^t \mathbb{E}\left[\mathbb{1}_{s\leqslant\tau_2} \left\|C_s^{\mathrm{err}}\right\|_F^2\right] ds \leqslant 8tL_C^2 + 8\int_0^t \varepsilon_C(s)^2\, ds =: A$$
$$\mathbb{E}\int_0^{t\wedge\tau_2} F_s\, ds \leqslant \int_0^t 2\varepsilon_b(s) + 4\varepsilon_C(s)^2\, ds =: B.$$

By Markov's inequality, with probability $1 - \delta$,

$$[N]_{t\wedge\tau_2} \leqslant \frac{2}{\delta}A, \qquad \int_0^{t\wedge\tau_2} F_s\, ds \leqslant \frac{2}{\delta}B.$$

Let $\mathcal{A}$ be the event when both these hold. Then

$$\mathbb{P}\left(\sup_{s\leqslant t\wedge\tau_2} N_s \geqslant x,\, \mathcal{A}\right) \leqslant \exp\left(-\frac{x^2}{4A/\delta}\right).$$

Note that if $\sup_{s\leqslant t\wedge\tau_2} \left\|\Delta_s\right\|^2 \geqslant e^{L_1 t + \frac{2}{\delta}B + x}$, under $\mathcal{A}$, then the above event holds. We have, for $e^{L_1 s + \frac{2}{\delta}B + x} \leqslant r$ and $x \geqslant 0$,

$$\mathbb{P}\left(\{\forall t \in [0,T],\, X_t \in \mathcal{S}\}^c \cup \left\{\sup_{s\leqslant t} \left\|\Delta_s\right\|^2 \geqslant e^{L_1 t + \frac{2}{\delta}B + x}\right\}\right)$$
$$\leqslant \mathbb{P}\left(\{\forall t \in [0,T],\, X_t \in \mathcal{S}\}^c\right) + \mathbb{P}\left(\mathcal{A}^c\right) + \mathbb{P}\left(\{\forall t \in [0,T],\, X_t \in \mathcal{S}\} \cap \mathcal{A} \cap \left\{\sup_{s\leqslant t\wedge\tau_2} \left\|\Delta_s\right\|^2 \geqslant e^{L_1 t + \frac{2}{\delta}B + x}\right\}\right)$$
$$\leqslant \delta_1 + \delta + \exp\left(-\frac{x^2}{4A/\delta}\right).$$

The first inequality is because, if $\|\Delta_t\|^2 \geqslant e^{L_1 s + \frac{2}{\delta}B + x}$, then either $t < \tau_2$ and hence $\sup_{s \leqslant t \wedge \tau_2} \|\Delta_s\|^2 \geqslant e^{L_1 s + \frac{2}{\delta}B + x}$, or $t \geqslant \tau_2$. In the latter case,

$$\sup_{s \leqslant t \wedge \tau_2} \|\Delta_s\|^2 \geqslant \|\Delta_{\tau_2}\|^2 \geqslant r \geqslant e^{L_1 s + \frac{2}{\delta}B + x}$$

as well. Choosing x such that $e^{L_1 s + \frac{2}{\delta}B + x} = r_2^2 \leqslant r^2$ gives

$$\mathbb{P}\left[(X_{[0,t]}, \widehat{X}_{[0,t]}) \notin \mathfrak{S}^2(r, r_2)\right] \leqslant \delta_1 + \delta + \exp\left(-\frac{\delta(2\log r_2 - L_1 t - \frac{2B}{\delta})^2}{4A}\right).$$

If $L_1 t \leqslant \frac{1}{2}\log r_2$, then taking $\delta = \frac{4\max\{A,B\}}{\log r_2}$ gives that this is $\leqslant \delta_1 + \frac{4\max\{A,B\}}{\log r_2} + \frac{1}{r_2}$. □

Given a path $(y_t)_{t \geqslant 0}$, define m_t^y to be the magnetization process if (m_t^y, y_t) solves (ASL-TAP), up to the exit time from $\mathcal{S}_{\beta,A}(c)$. Note that because we only have unique solutions to the TAP equation locally but not globally, m_t^y cannot simply be defined as a function $\widehat{m}(y_t)$ as in [DLSS26a], but can depend on the path. If $\widehat{y}(\mathcal{S}_{\beta,A}(c))$ is not simply connected, then a loop in y-space may lift to a path with different endpoints in m-space.

Lemma 3.3 (Error in estimated magnetization)**.**
Suppose that (d1), (d2a), (d2b), (d3), and (d4) hold to time T. Let y_t denote the SL process, and $m_t = \langle\sigma\rangle_{y_t}$ be the true magnetization for tilt y_t. Then for each $t \in [0,T]$ there is a constant $C(t)$ such that

$$\mathbb{E}\left[\mathbb{1}_{\mathcal{G}_t} \|m_t^y - m_t\|^2\right] \leqslant C(t),$$

where $\mathcal{G}_t = \{((m,y)_{[0,t]}, (\widehat{m},\widehat{y})_{[0,t]}) \in \mathfrak{S}_t^2(\rho\sqrt{n}, c\rho\sqrt{n}/2)\}$ and $\mathbb{P}(\mathcal{G}_t) = 1 - o_n(1)$. Here, $c = c_{\mathrm{TAP}}(\beta,T)$ and $\rho = \rho(\beta,T)$ as in (d2a).

Proof. Well-posedness of the SDE's is proven by Lemma 3.9. Consider the synchronously coupled SDE's

$$\text{(PHD-TAP)} \quad \begin{cases} d\widehat{m}_t = \widehat{Q}(\widehat{m}_t) dB_t \\ d\widehat{y}_t = f(\widehat{m}_t)dt + dB_t \end{cases}$$

$$\text{(Ideal)} \quad \begin{cases} dm_t = Q(m_t)dB_t = \mathbf{Cov}(\mu_{y_t})dB_t \\ dy_t = m_t dt + dB_t \end{cases}$$

over $\Omega := (-1,1)^n \times \mathbb{R}^n$. Then

$$\begin{aligned}
\mathop{\mathbb{E}}_{A,B_t}\left[\mathbb{1}_{\mathcal{S}_{\beta,A}(c)}[m_t] \left\|Q(m_t) - \widehat{Q}(m_t)\right\|_F^2\right] &\leqslant \mathop{\mathbb{E}}_{A,B_t}\left[\mathbb{1}_{\mathcal{S}_{\beta,A}(c)}[m_t] \left\|\widehat{Q}(m_t)\right\|_{\mathsf{op}}^2 \left\|Q(m_t) - \widehat{Q}(m_t)\right\|_F^2\right] \\
&\leqslant_{\text{(d2a)}} c^{-2} \mathop{\mathbb{E}}_{A,B_t}\left[\mathbb{1}_{\mathcal{S}_{\beta,A}(c)}[m_t] \left\|\widehat{Q}(m_t)^{-1}Q(m_t) - I_n\right\|_F^2\right] \\
&= c^{-2} \mathop{\mathbb{E}}_{A,B_t}\left[\mathbb{1}_{\mathcal{S}_{\beta,A}(c)}[m_t] \left\|\nabla_m^2 \mathcal{F}_{\mathrm{TAP}}(m_t)Q_t - I_n\right\|_F^2\right] \\
&\leqslant_{\text{(d1)}} c^{-2}\varepsilon_{\mathrm{cov}}(\beta,t)^2 .
\end{aligned}$$

Therefore, by Markov's inequality, for any δ, with probability $\geqslant 1-\delta$ over A,

$$\mathop{\mathbb{E}}_{B_t}\left[\mathbb{1}_{\mathcal{S}_{\beta,A}(c)}[m_t] \left\|Q(m_t) - \widehat{Q}(m_t)\right\|_F^2\right] = O(1)$$

where we consider β, T, δ fixed. We also have, with probability $\geqslant 1 - o_n(1)$ over A,

$$\mathop{\mathbb{E}}_{B_t}\left[\mathbb{1}_{\mathcal{S}_{\beta,A}(c)}[\widehat{m}_t] \left\|f(\widehat{m}_t) - \widehat{m}_t\right\|^2\right] \leqslant_{\text{(d3)}} \varepsilon_{\mathrm{drift}}^2 = O(1).$$

By (d4), we have that $\widehat{Q}$ is $O(1)$-Lipschitz with respect to $\|\cdot\|_\Omega$, and the identity map is clearly 1-Lipschitz. (d2a) supplies $1 - o_n(1)$ probability of (m_t, y_t) staying inside $\mathcal{S}_{\beta,A}(c)$. Define $\mathfrak{S}_t^2(r_1, r_2)$ using $\mathcal{S}_{\beta,A}(c)$ as the safe set. Then Lemma 3.1 applies to give

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_{\mathfrak{S}_t^2(\rho\sqrt{n})}[(m,y)_{[0,t]}, (\widehat{m},\widehat{y})_{[0,t]}] \left\|(m_t, y_t) - (\widehat{m}_t, \widehat{y}_t)\right\|^2\right] &= O(1) \\
\mathbb{P}[((m,y)_{[0,t]}, (\widehat{m},\widehat{y})_{[0,t]}) \notin \mathfrak{S}_t^2(\rho\sqrt{n}, \rho\sqrt{n}/2)] &= o_n(1).
\end{aligned}$$

Consider the map $\widehat{y}$ given by (2.3). We aim to apply Lemma 3.4, but because the ball is truncated by the cube, we work in dual space, $x = \tanh^{-1}(m)$. First, we claim that for $m \in (-1,1)^n$, $B(\tanh^{-1}(m), r) \subseteq$

$\tanh^{-1}(B(m,r)\cap(-1,1)^n)$. Since $\tanh:\mathbb{R}^n\to(-1,1)^n$ is a homeomorphism, this is equivalent to $\tanh(B(\tanh^{-1}(m),r))\subseteq B(m,r)\cap(-1,1)^n$. This is clear since tanh is contractive.

Now consider $\widehat{y}\circ\tanh:\tanh^{-1}(\mathcal{S}_{\beta,A}(c))\to\mathbb{R}^n$. Note $D(\widehat{y}\circ\tanh)(\tanh^{-1}(m))\succeq\widehat{Q}(m)^{-1}D(m)^{-1}\succeq cI_n$. When $((m,y)_{[0,t]},(\widehat{m},\widehat{y})_{[0,t]})\in\mathfrak{S}_t^2(\rho\sqrt{n},\rho\sqrt{n}/2)$, $B(\tanh^{-1}(\widehat{m}_t),\rho\sqrt{n}/2)\subseteq\tanh^{-1}(B(\widehat{m}_t,\rho\sqrt{n}/2)\cap(-1,1)^n)\subseteq\tanh^{-1}(\mathcal{S}_{\beta,A}(c))$, so by Lemma 3.4,

$$B(\widehat{y}(\widehat{m}_t)=\widehat{y}_t,c\rho\sqrt{n}/2)\subseteq\widehat{y}(B(\widehat{m}_t,\rho\sqrt{n}/2)\cap(-1,1)^n)\subseteq\widehat{y}(\mathcal{S}_{\beta,A}(c)).$$

Now let

$$\mathcal{G}_t=\{((m,y)_{[0,t]},(\widehat{m},\widehat{y})_{[0,t]})\in\mathfrak{S}_t^2(\rho\sqrt{n},c\rho\sqrt{n}/2)\}.$$

Under $\mathcal{G}_t$, for all $s\in[0,t]$, $y_s\in B(\widehat{y}_s,c\rho\sqrt{n}/2)$, so considering the local inverse map $\widehat{m}_{\widehat{y}_s}:B(\widehat{y}_s,c\rho\sqrt{n}/2)\to B(\widehat{m}_s,\rho\sqrt{n}/2)$, by continuity of m_s^y, we have that $\widehat{m}_{\widehat{y}_s}(y_s)=m_s^y$ and

$$\|m_s^y-\widehat{m}_s\|=\|\widehat{m}_{\widehat{y}_s}(y_s)-\widehat{m}_{\widehat{y}_s}(\widehat{y}_s)\|\leqslant c^{-1}\|y_s-\widehat{y}_s\|.$$

since $D\widehat{m}_{\widehat{y}_s}(y)=\widehat{Q}(\widehat{m}_{\widehat{y}_s}(y))$ has operator norm at most c^{-1}. Hence under $\mathcal{G}_t$,

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{\mathcal{G}_t}\|m_t-m_t^y\|^2\right]&\lesssim\mathbb{E}[\mathbb{1}_{\mathcal{G}_t}\|m_t-\widehat{m}_t\|^2]+\mathbb{E}[\mathbb{1}_{\mathcal{G}_t}\|\widehat{m}_t-m_t^y\|^2]\\&\lesssim\mathbb{E}[\mathbb{1}_{\mathcal{G}_t}\|m_t-\widehat{m}_t\|^2]+\mathbb{E}[\mathbb{1}_{\mathcal{G}_t}\|\widehat{y}_t-y_t\|^2]=O(1).\end{aligned}$$

□

Lemma 3.4 (Local Hadamard Theorem, [OR00, 5.3.11])**.**
Let $f:\Omega\to\mathbb{R}^n$ be a continuously differentiable map where $\Omega\subseteq\mathbb{R}^n$ is open. Suppose $B(m_0,r)\subseteq\Omega$ and $Df\succeq cI_n$ on Ω, $c>0$. Then $B(f(m_0),cr)\subseteq f(\Omega)$. Moreover, f is injective on any convex subset of Ω.

A standard Girsanov argument (see e.g. [Che24, §4.4]) then bounds the KL divergence between the SL and ASL-TAP.

Corollary 3.5 ($O(1)$ KL error for ASL-TAP)**.**
Suppose that (d1), (d2a), (d2b), (d3), and (d4) hold to time T. Let y_t and $\widehat{y}_t$ follow (SL) and (ASL-TAP) (with any extension outside of $\mathcal{S}_{\beta,A}(c)$). Then there is a distribution Q_T such that

$$\mathrm{TV}(\mathcal{L}(y_T),Q_T)=o_n(1),\quad \mathrm{KL}(Q_T\|\,\mathcal{L}(\widehat{y}_T))=O(1).$$

Proof. Let $\tau=\inf\left\{t\geqslant0:((m_t,y_t),(\widehat{m}_t,\widehat{y}_t))\notin\mathcal{S}_t^2(\rho\sqrt{n},c\rho\sqrt{n}/2)\right\}$. Let $\mathbf{P}_T$ and $\widehat{\mathbf{P}}_T$ denote the path measures of $(y_{t\wedge\tau})_{t=0}^T$ and $(\widehat{y}_{t\wedge\tau})_{t=0}^T$. Since m_t^y,m_t are bounded, Novikov's condition holds: $\mathbb{E}\exp\left(\frac{1}{2}\int_0^T\|m_t^y-m_t\|^2\,dt\right)<\infty$. By the data processing inequality and Girsanov's Theorem,

$$\begin{aligned}\mathrm{KL}(\mathcal{L}(y_{T\wedge\tau})\|\,\mathcal{L}(\widehat{y}_{T\wedge\tau}))\leqslant\mathrm{KL}(\mathbf{P}_T\|\widehat{\mathbf{P}}_T)&\leqslant\mathop{\mathbb{E}}_{\mathbf{P}_T}\log\frac{d\mathbf{P}_T}{d\widehat{\mathbf{P}}_T}\\&=-\mathbb{E}\int_0^{T\wedge\tau}\langle m_t-m_t^y,dB_t\rangle+\frac{1}{2}\mathbb{E}\int_0^{T\wedge\tau}\|m_t-m_t^y\|^2\,dt\\&=\frac{1}{2}\int_0^{T\wedge\tau}\mathbb{E}\|m_t-m_t^y\|^2\,dt=O(1),\end{aligned}$$

where the last equality follows from Lemma 3.3. Noting that $\mathrm{TV}(\mathcal{L}(y_{T\wedge\tau}),\mathcal{L}(y_T)),\mathrm{TV}(\mathcal{L}(\widehat{y}_{T\wedge\tau}),\mathcal{L}(\widehat{y}_T))\leqslant\mathbb{P}[\tau<T]=o_n(1)$ finishes the proof. □

3.2. **Jarzynski's equality with survival.** In the following, we derive a specific Jarzynski's equality formula to compare the Radon-Nikodym derivative between the SL-TAP and ASL-TAP process on path space for m, as well as the final-time marginal m_T, restricted to survival of the path. Note that we consider the final-time marginal m_T^y rather than y_T, because the map $\widehat{y}(m)$ may not be injective. As long as the survival event has probability $1-o(1)$, we can then do rejection sampling with the provided ratio, with an error of $o(1)$.

Lemma 3.6 (Jarzynski's equality with survival for SL-to-ASL ratio)**.**
Let $\tau=\inf\{t\geqslant0:m_t\notin\mathcal{S}_{\beta,A}(c)\}$ and $S_T=\{\tau>T\}$. For a measure $\mathbf{Q}_T$ on path space $m:[0,T]\to\mathbb{R}^n$, let $\mathbf{Q}_T^s(\cdot)=\mathbf{Q}_T(\cdot\cap S_T)$, and let $q_T^s(m)=\mathbb{E}_{m_{[0,T]}\sim\mathbf{Q}_T}[\mathbb{1}_{S_T}\mid m_T=m]$, i.e., the marginal distribution of m_T under $\mathbf{Q}_T^s$. Let $\mathbf{P}_T$ denote the path measure $m_{[0,T]}$ in (SL-TAP) and $\widehat{\mathbf{P}}_T$ denote the path measure of $\widehat{m}_{[0,T]}$ in the

(ASL-TAP). *Then we have that*

$$\frac{d\mathbf{P}_T^s}{d\widehat{\mathbf{P}}_T^s}(m_{[0,T]}) = e^{w_T} \cdot \frac{e^{-\mathcal{F}_{\mathrm{TAP}}(0,0)}}{e^{-\mathcal{F}_{\mathrm{TAP}}(m_T,\widehat{y}(m_T))}} \cdot e^{-\frac{nT}{2}} \cdot \frac{Z_{\beta A,\widehat{y}(m_T)}}{Z_{\beta A,0}}$$
$$\frac{dp_T^s}{d\widehat{p}_T}(m) = \mathbb{E}[e^{w_T}\mathbb{1}_{S_T} \mid m_T = m] \cdot \frac{e^{-\mathcal{F}_{\mathrm{TAP}}(0,0)}}{e^{-\mathcal{F}_{\mathrm{TAP}}(m,\widehat{y}(m))}} \cdot e^{-\frac{nT}{2}} \cdot \frac{Z_{\beta A,\widehat{y}(m)}}{Z_{\beta A,0}}$$

where w_T is defined by

$$dw_t = \omega(m_t)\,dt, \qquad \omega(m) := \frac{1}{2}\left[\mathsf{Tr}(\widehat{Q}(m)) + \|m\|^2\right].$$

Proof. We rederive the calculations from [DLSS26a, §3.2], but taking $m = m_t^y$ instead of assuming there is a unique TAP solution $m = \widehat{m}(y_t)$. Given that (m_t, y_t) is an Itô process satisfying (TAP) for all t, by Itô's formula,

$$d\mathcal{F}_{\mathrm{TAP}}(m_t,y_t) = \underbrace{\langle \nabla_m \mathcal{F}_{\mathrm{TAP}}(m_t,y_t), dm_t\rangle}_{0} - \langle m_t, dy_t\rangle + \frac{1}{2}\mathsf{Tr}\left(\nabla_m^2 \mathcal{F}_{\mathrm{TAP}}(m_t,y_t) d[y]_t\right) - \mathsf{Tr}(d[m,y]_t).$$

When y_t satisfies (SL), letting m_t^y be the solution to (TAP),

$$\begin{aligned} d\mathcal{F}_{\mathrm{TAP}}(m_t^y,y_t) &= -\langle m_t^y, dy_t\rangle + \frac{1}{2}\mathsf{Tr}\left(\nabla_m^2 \mathcal{F}_{\mathrm{TAP}}(m_t,y_t)\widehat{Q}(m_t^y)\widehat{Q}(m_t^y)^\intercal\right) dt - \mathsf{Tr}(\widehat{Q}(m_t^y)) \\ &= -\langle \widehat{m}_t^y, dy_t\rangle - \frac{1}{2}\mathsf{Tr}(\widehat{Q}(m_t^y))\,dt. \end{aligned}$$

Let $\mathbf{W}_T$ denote the Wiener measure. We consider the stopped ASL process defined by

$$d\widehat{y}_t = \mathbb{1}_{t<\tau}\widehat{m}_t\,dt + dB_t,$$

and denote the path measure by $\widehat{\mathbf{P}}_T^y$. By Girsanov's Theorem,

$$\frac{d\widehat{\mathbf{P}}_T^y}{d\mathbf{W}_T}(y_{[0,T]}) = \exp\left(\int_0^{T\wedge\tau} \langle m_t^y, dy_t\rangle - \frac{1}{2}\int_0^{T\wedge\tau} \|m_t^y\|^2\,dt\right).$$

Then

$$\begin{aligned} \mathbb{1}_{S_T}[m_{[0,T]}^y] e^{\widehat{w}_T(m_{[0,T]}^y)} \frac{d\widehat{\mathbf{P}}_T^y}{d\mathbf{W}_T}(y_{[0,T]}) &= \mathbb{1}_{S_T}\exp\left(\int_0^T \frac{1}{2}\left(\mathsf{Tr}(\widehat{Q}(m_t^y)) + \|m_t^y\|^2\right)dt\right) \\ &= \mathbb{1}_{S_T}\exp\left(-\int_0^T d\mathcal{F}_{\mathrm{TAP}}(m_t^y,y_t)\right) = \exp(\mathcal{F}_{\mathrm{TAP}}(0,0) - \mathcal{F}_{\mathrm{TAP}}(m_T^y,y_T)). \end{aligned}$$

Define the survival event on paths $y_{[0,T]}$ to be that $m_t^y \in \mathcal{S}_{\beta,A}(c)$ for $t \leqslant T$. Therefore,

$$\frac{d\widehat{\mathbf{P}}_T^{y,s}}{d\mathbf{W}_T^s}(y_{[0,T]}) = e^{-\widehat{w}_T(m_{[0,T]}^y)} \frac{e^{-\mathcal{F}_{\mathrm{TAP}}(m_T^y,y_T)}}{e^{-\mathcal{F}_{\mathrm{TAP}}(0,0)}}.$$

The SL process is given by $y_t = t\sigma + B_t$, i.e., $dy_t = \sigma\,dt + dB_t$, where $\sigma \sim \mu_{\beta A}$. By Girsanov's Theorem, conditioned on σ,

$$\frac{d\mathbf{P}_T^y(\cdot \mid \sigma)}{d\mathbf{W}_T}(y_{[0,T]}) = \exp\left(\int_0^T \langle \sigma, dy_t\rangle - \frac{1}{2}\int_0^T \|\sigma\|^2\,dt\right) = \exp\left(\langle \sigma, y_T\rangle - \frac{nT}{2}\right).$$

Averaging over $\sigma \sim \mu_{\beta A}$ then gives

$$\frac{d\mathbf{P}_T^y}{d\mathbf{W}_T}(y_{[0,T]}) = \sum_{\sigma\in\{\pm 1\}^n} \frac{\exp\left(\frac{\beta}{2}\langle \sigma, A\sigma\rangle\right)}{Z_{\beta A,0}} \exp\left(\langle \sigma, y_T\rangle - \frac{nT}{2}\right) = e^{-\frac{nT}{2}}\frac{Z_{\beta A,y_T}}{Z_{\beta A,0}}.$$

Note this only depends on the endpoint of the path. We have that $\frac{d\mathbf{P}_T^s}{d\mathbf{W}_T^s}$ equals the same expression. Then

$$\frac{d\mathbf{P}_T^{y,s}}{d\widehat{\mathbf{P}}_T^{y,s}} = \frac{d\mathbf{P}_T^{y,s}}{d\mathbf{W}_T^s} \Big/ \frac{d\widehat{\mathbf{P}}_T^{y,s}}{d\mathbf{W}_T^s} = e^{\widehat{w}_T(m_{[0,T]}^y)} \cdot \frac{e^{-\mathcal{F}_{\mathrm{TAP}}(0,0)}}{e^{-\mathcal{F}_{\mathrm{TAP}}(m_T^y,y_T)}} \cdot e^{-\frac{nT}{2}} \cdot \frac{Z_{\beta A,y_T}}{Z_{\beta A,0}}.$$

Pushing forward to m-space via $y_{[0,T]} \mapsto m_{[0,T]}^y$, multiplying by $\mathbb{1}_{S_T}$, and taking a conditional expectation with respect to the endpoint m_T^y then gives the second result. □

We need to show that with high probability, for the SL process y_t, considering the m_t^y chosen to satisfy the TAP equation, survival occurs with high probability.

Lemma 3.7.
Suppose that (d1), (d2a), (d2b), (d3), and (d4) hold to time T. Let y_t solve (SL). Define $\mathfrak{S}_t$ on the magnetization path space corresponding to $\mathcal{S}_{\beta,A}(c)$. Then for all $t \in [0,T]$,

$$\mathbb{P}[m_t^y \notin \mathfrak{S}_t] = o_n(1).$$

Proof. Abbreviate $\widetilde{m}_t = m_t^y$. We consider the mismatch between the true magnetization and the magnetization estimated from following the TAP equation,

$$\begin{aligned} dy_t &= m_t\,dt + dB_t \\ dm_t &= Q(m_t)\,dB_t \\ d\widetilde{m}_t &= \widehat{Q}(\widetilde{m}_t)(dy_t - f(\widetilde{m}_t)\,dt) = \widehat{Q}(\widetilde{m}_t)((m_t - f(\widetilde{m}_t))\,dt + dB_t). \end{aligned}$$

Let $\widetilde{\Delta}_t = m_t - \widetilde{m}_t$. We have

$$d\widetilde{\Delta}_t = [\underbrace{\widehat{Q}(\widetilde{m}_t)(m_t - \widetilde{m}_t)}_{b^{\mathrm{Lip}}} + \underbrace{\widehat{Q}(\widetilde{m}_t)(\widetilde{m}_t - f(\widetilde{m}_t))}_{b^{\mathrm{err}}}]\,dt + [\underbrace{(\widehat{Q}(m_t) - \widehat{Q}(\widetilde{m}_t))}_{C^{\mathrm{Lip}}} + \underbrace{(Q(m_t) - \widehat{Q}(m_t))}_{C^{\mathrm{err}}}]\,dB_t$$

Note

$$\begin{aligned} \left\|b_t^{\mathrm{Lip}}\right\| &\leqslant c^{-1}\left\|\Delta_t\right\| & \mathbb{E}\left[\mathbb{1}_{t<\tau}\left\|b_t^{\mathrm{err}}\right\|^2\right] &\leqslant c^{-2}\varepsilon_{\mathrm{drift}}^2 \\ \left\|C_t^{\mathrm{Lip}}\right\|_F &\leqslant L\left\|\Delta_t\right\| & \mathbb{E}\left[\mathbb{1}_{t<\tau}\left\|C_t^{\mathrm{err}}\right\|_F^2\right] &\leqslant \varepsilon_{\mathrm{cov}}^2. \end{aligned}$$

Then Lemma 3.2 gives that (defining the safe sets on just the magnetization path space)

$$\mathbb{P}[\widetilde{m}_{[0,t]} \notin \mathfrak{S}_t] \leqslant \mathbb{P}[(m_{[0,t]}, \widetilde{m}_{[0,t]}) \notin \mathfrak{S}_t^2(\rho\sqrt{n})] = o_n(1).$$

□

3.3. **Well-posedness.** We consider the initial condition $\widehat{m}_0 = 0$ and the equations

$$d\widehat{m}_t = \widehat{Q}(\widehat{m}_t)(\widehat{m}_t - f(\widehat{m}_t))\,dt + \widehat{Q}(\widehat{m}_t)\,dB_t, \tag{3.4}$$

$$d\widehat{m}_t = \widehat{Q}(\widehat{m}_t)\,dB_t, \tag{3.5}$$

for ASL-TAP and PHD, respectively. Each equation is used up to its first exit from $\mathcal{S}_{\beta,A}(c)$, with its own exit time denoted τ_c. As in [DLSS26a, §6.6], we introduce the dual coordinates $\widehat{x}_t = \tanh^{-1}(\widehat{m}_t)$ and write

$$\widehat{\Sigma}(x) := D(m)\widehat{Q}(m), \qquad \widehat{f}^*(x) := E_{\mathcal{D}_n}[\widehat{Q}(m)^2]D(m)^2 m, \qquad m = \tanh(x).$$

Here $\widehat{f}^*$ is not a convex conjugate. The two dual equations are

$$d\widehat{x}_t = \widehat{f}^*(\widehat{x}_t)\,dt + \widehat{\Sigma}(\widehat{x}_t)\,dB_t, \tag{d-PHD}$$

$$d\widehat{x}_t = \left[D(\widehat{m}_t)\widehat{Q}(\widehat{m}_t)(\widehat{m}_t - f(\widehat{m}_t)) + \widehat{f}^*(\widehat{x}_t)\right]dt + \widehat{\Sigma}(\widehat{x}_t)\,dB_t. \tag{d-ASL-TAP}$$

Lemma 3.8 (Boundedness of dual SDE coefficients)**.**
Assume (d3). Then the following hold within the domain $\{x \in \mathbb{R}^n : \tanh(x) \in \mathcal{S}_{\beta,A}(c)\}$:

(1) $\left\|\widehat{\Sigma}\right\|_{\mathsf{op}} \leqslant c^{-1}$.
(2) $\left\|\widehat{f}^*\right\|_2 \leqslant c^{-2}\sqrt{n}$.
(3) With $m = \tanh(x)$, $\left\|D(m)\widehat{Q}(m)(m - f(m))\right\|_2 \leqslant c^{-1}\varepsilon_{\mathrm{drift}}$

Proof. These follow directly from the bound on $D(m)\widehat{Q}(m)$ given by Definition 2.2, the bound $\|m\|_2 \leqslant \sqrt{n}$ implied by $m \in (-1,1)^n$, and the bound on $\|m - f(m)\|_2$ given by (d3). □

Lemma 3.9 (ASL-TAP and PHD stay in the cube up to safe-set exit)**.**
Assume $\|A\|_{\mathsf{op}} \leqslant 3$, (d3), and $0^n \in \mathcal{S}_{\beta,A}(c)$ for $c > 0$. Then the primal and dual equations (3.4)–(3.5) and (d-PHD)–(d-ASL-TAP) have pathwise unique strong solutions up to their respective first exits from the good

set $\mathcal{S}_{\beta,A}(c)$ or its dual-coordinate image $\tanh^{-1}(\mathcal{S}_{\beta,A}(c))$. Their stopped solutions are defined for all $t \geqslant 0$, and almost surely

$$\widehat{m}_{t\wedge\tau_c} \in (-1,1)^n \qquad (t < \infty).$$

The ASL tilt and JE weight also have finite limits at a finite exit and can be stopped there.

Proof. Write either dual equation as $d\widehat{x}_t = b(\widehat{x}_t)\,dt + \widehat{\Sigma}(\widehat{x}_t)\,dB_t$ on $U := \tanh^{-1}(\mathcal{S}_{\beta,A}(c))$, with $\widehat{x}_0 = 0$. Its coefficients are smooth on U. Then Lemma 3.8 gives

$$\sup_{x\in U} \|b(x)\|_2 \leqslant \infty, \qquad \sup_{x\in U} \left\|\widehat{\Sigma}(x)\right\|_{\mathsf{op}} \leqslant \infty.$$

With dist being the distance between a point and a set, define

$$U_k := \{x \in U : \|x\|_2 < k,\ \operatorname{dist}(x, U^c) > k^{-1}\}.$$

For k large enough that $0 \in U_k$, use $\overline{U}_k \subseteq U_{k+1}$ to choose $\chi_k \in C_c^\infty(U_{k+1})$ with $0 \leqslant \chi_k \leqslant 1$ and $\chi_k = 1$ on $\overline{U}_k$. The coefficient pairs $(\chi_k b, \chi_k \widehat{\Sigma})$, extended by zero outside U, are globally Lipschitz, so [Øks03, Theorem 5.2.1] gives global strong solutions $\widehat{x}^{(k)}$ with these coefficients, the same Brownian motion, and initial value 0. Set $\tau_k := \inf\{t : \widehat{x}_t^{(k)} \notin U_k\}$. The stopped uniqueness argument in that theorem gives, for $\ell \geqslant k$,

$$\tau_k \leqslant \tau_\ell, \qquad \widehat{x}^{(\ell)}_{t\wedge\tau_k} = \widehat{x}^{(k)}_{t\wedge\tau_k}.$$

Thus $\tau_c := \lim_k \tau_k$ and $\widehat{x}_t := \widehat{x}_t^{(k)}$ whenever $t < \tau_k$ define the pathwise unique solution before τ_c.

Taking the integrands to be zero for $s \geqslant \tau_c$, we have $\mathbb{E}_B \int_0^T \mathbf{1}_{\{s<\tau_c\}} \|\widehat{\Sigma}(\widehat{x}_s)\|_F^2\,ds \leqslant C_{\beta,c}^2 nT$. Hence Itô isometry, [Øks03, Theorem 3.2.5], and the drift bound give the continuous extension

$$\widehat{x}_{t\wedge\tau_c} := \int_0^t \mathbf{1}_{\{s<\tau_c\}} b(\widehat{x}_s)\,ds + \int_0^t \mathbf{1}_{\{s<\tau_c\}} \widehat{\Sigma}(\widehat{x}_s)\,dB_s, \qquad t < \infty.$$

On $\{\tau_c < \infty\}$, continuity gives $\widehat{x}_{\tau_k} \in \partial U_k$, so

$$\|\widehat{x}_{\tau_k}\|_2 = k \quad \text{or} \quad \operatorname{dist}(\widehat{x}_{\tau_k}, U^c) = k^{-1}.$$

Since $\widehat{x}_{\tau_k} \to \widehat{x}_{\tau_c} \in \mathbb{R}^n$, the first alternative is impossible for large k. Thus $\widehat{x}_{\tau_c} \in \partial U$, and τ_c is the first exit from U.

Itô's formula applied to tanh and locally to $\tanh^{-1}$ as in [DLSS26a, Lemma 6.9], gives the primal equations, pathwise uniqueness, and $\widehat{m}_{t\wedge\tau_c} = \tanh(\widehat{x}_{t\wedge\tau_c}) \in (-1,1)^n$. Finally, for $t < \tau_c$, recover

$$\widehat{y}_t = \int_0^t \widehat{m}_s\,ds + B_t, \qquad \widehat{w}_t = \frac{1}{2}\int_0^t \left(\mathsf{Tr}\widehat{Q}(\widehat{m}_s) + \|\widehat{m}_s\|_2^2\right) ds.$$

Since $\|\widehat{m}_s\|_2 \leqslant \sqrt{n}$ and $0 \leqslant \mathsf{Tr}\widehat{Q}(\widehat{m}_s) \leqslant n/c$ before exit, both have finite limits at a finite τ_c and can be stopped there. □

3.4. **Pathwise concentration.** We use the same transportation argument as [DLSS26a, Definition 6.10 and Theorems 6.11–6.12], applying it with globally Lipschitz extensions of the SDE's. We briefly restate the definition and theorems here.

Definition 3.10. A probability measure P on a metric space satisfies a T_p transportation inequality with constant C if $d_{W,p}(P,\mu)^2 \leqslant 2CH(\mu \mid P)$ for every probability measure μ, where $d_{W,p}$ denotes Wasserstein distance and H relative entropy.

Theorem 3.11 (T_1 inequality and sub-Gaussian concentration [DGW04, Theorem 1.1, Bobkov & Götze])**.** *If P satisfies T_1 with constant C, then every integrable K_F-Lipschitz function F satisfies*

$$\mathop{\mathbb{E}}_P[e^{\lambda(F-\mathbb{E}_P[F])}] \leqslant e^{\lambda^2 CK_F^2/2} \quad (\forall \lambda \in \mathbb{R}), \qquad P\{|F - \mathop{\mathbb{E}}_P[F]| > r\} \leqslant 2e^{-r^2/(2CK_F^2)} \quad (\forall r > 0).$$

Theorem 3.12 (T_2 transport inequality for regular Itô processes [Üst12, modification of Theorem 1])**.** *Let $C, K, T > 0$. Consider the global solution of $dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t$, $X_0 = 0$, on $[0,T]$. Suppose that $\|\sigma(x)\|_{\mathsf{op}} \leqslant C$ for all $x \in \mathbb{R}^n$, that b is K-Lipschitz in Euclidean norm, and that σ is $K/\sqrt{n}$-Lipschitz in $\|\cdot\|_{L^2}$. Then its path law on $C^0([0,T],\mathbb{R}^n)$ with uniform Euclidean distance satisfies T_2 with constant $3TC^2\exp(3T(T+4)K^2)$.*

Proof. This is the time-homogeneous special case of [DLSS26a, Theorem 6.12]. □

We now apply the transportation argument to the SDE's with Lipschitz-extended coefficients.

Theorem 3.13 (Transport and sub-Gaussian concentration for an ASL-TAP extension)**.** *Assume $\|A\|_{\mathsf{op}} \leqslant 3$, (d3), (d4), and $0^n \in \mathcal{S}_{\beta,A}(c)$ for $c > 0$. Let $0 < T < \infty$. Then there is a global strong solution*

$$d\widehat{m}_t^{\mathrm{ext}} = b_{\mathrm{ext}}(\widehat{m}_t^{\mathrm{ext}})\,dt + \sigma_{\mathrm{ext}}(\widehat{m}_t^{\mathrm{ext}})\,dB_t, \qquad \widehat{m}_0^{\mathrm{ext}} = 0, \tag{3.6}$$

whose coefficients agree on $\mathcal{S}_{\beta,A}(c)$ with $\widehat{Q}(m)(m - f(m))$ and $\widehat{Q}(m)$, respectively. Driven by the same Brownian motion, it agrees with the original ASL-TAP process up to their common first exit from $\mathcal{S}_{\beta,A}(c)$. Let P be its path law on $\mathcal{W} = C^0([0,T],\mathbb{R}^n)$ with the uniform norm $\|x\|_\infty = \sup_{0\leqslant t\leqslant T}\|x_t\|_2$. For a constant K depending on L and L_{drift}, set

$$c_{T_2} := 3Tc^{-2}\exp\bigl(3T(T+4)K^2\bigr). \tag{3.7}$$

Then:

(i) *For every probability measure μ on $\mathcal{W}$, $d_{W,2}(P,\mu)^2 \leqslant 2c_{T_2}H(\mu \mid P)$.*

(ii) *For every P-integrable K_F-Lipschitz functional F,*

$$\mathop{\mathbb{E}}_{P}[e^{\lambda(F-\mathbb{E}_P[F])}] \leqslant \exp\left(\frac{\lambda^2 c_{T_2}K_F^2}{2}\right) \qquad (\forall \lambda \in \mathbb{R}). \tag{3.8}$$

In sub-Gaussian tail-bound form, $P\{|F - \mathbb{E}_P[F]| > r\} \leqslant 2\exp(-r^2/(2c_{T_2}K_F^2))$ for $r > 0$.

Proof. Follow [DLSS26a, Theorem 6.13], extending from $\mathcal{S}_{\beta,A}(c)$ rather than from the whole cube. By (d4), Kirszbraun's theorem gives extensions satisfying

$$\|\sigma_{\mathrm{ext}}(x) - \sigma_{\mathrm{ext}}(y)\|_F \leqslant K\|x-y\|_2, \qquad \|b_{\mathrm{ext}}(x) - b_{\mathrm{ext}}(y)\|_2 \leqslant K\|x-y\|_2. \tag{3.9}$$

Project the diffusion extension in Frobenius norm onto the closed convex set $\{Q \in \mathrm{Sym}_n : 0 \preceq Q \preceq c^{-1}I_n\}$. This preserves its Lipschitz constant and its values on the good set. Global strong existence follows, and pathwise uniqueness identifies the paths until exit. Apply Theorem 3.12 with $C = c^{-1}$ and the above K. As in [DLSS26a, Theorem 6.13], $d_{W,1} \leqslant d_{W,2}$ then implies T_1, and Theorem 3.11 gives (ii). □

Lemma 3.14 (Indicator-weighted centered exponential moments)**.** *Let F be a real random variable with $\mathbb{E}[e^{s(F-\mathbb{E}[F])}] \leqslant e^{vs^2/2}$ for all $s \in \mathbb{R}$, for some $v > 0$. Then for any event S and every $\lambda \in \mathbb{R}$,*

$$\mathbb{E}\bigl[\mathbf{1}_S e^{\lambda(F-\mathbb{E}[F|S])}\bigr] \leqslant e^{2v\lambda^2}.$$

If $\mathbb{P}(S) = 0$, then the left-hand side is defined to be zero.

Proof. Let $q = \mathbb{P}(S) > 0$ since the $q = 0$ case holds by convention. Conditional Jensen's inequality and the MGF hypothesis at $s = -2\lambda$ give

$$q\,e^{-2\lambda(\mathbb{E}[F|S]-\mathbb{E}[F])} \leqslant q\,\mathbb{E}[e^{-2\lambda(F-\mathbb{E}[F])} \mid S] = \mathbb{E}\bigl[\mathbf{1}_S e^{-2\lambda(F-\mathbb{E}[F])}\bigr] \leqslant e^{2v\lambda^2}.$$

Taking square roots and then applying Cauchy–Schwarz, we obtain

$$\begin{aligned}\mathbb{E}\bigl[\mathbf{1}_S e^{\lambda(F-\mathbb{E}[F|S])}\bigr] &= e^{\lambda(\mathbb{E}[F]-\mathbb{E}[F|S])}\,\mathbb{E}\bigl[\mathbf{1}_S e^{\lambda(F-\mathbb{E}[F])}\bigr] \\ &\leqslant \bigl(q\,e^{-2\lambda(\mathbb{E}[F|S]-\mathbb{E}[F])}\bigr)^{1/2}\bigl(\mathbb{E}[e^{2\lambda(F-\mathbb{E}[F])}]\bigr)^{1/2} \\ &\leqslant e^{v\lambda^2}e^{v\lambda^2} = e^{2v\lambda^2},\end{aligned}$$

where the last step also uses the MGF hypothesis at $s = 2\lambda$. □

Corollary 3.15 (Sub-gaussian concentration of JE weights)**.** *Suppose $0^n \in \mathcal{S}_{\beta,A}(c)$ and assume (d3), (d4), and (d5). Stop the original ASL-TAP-JE process at τ_c, and set $S_T = \{\tau_c > T\}$. Then for every $0 < T < \infty$ and $\lambda \in \mathbb{R}$,*

$$\mathop{\mathbb{E}}_{B}\bigl[\mathbf{1}_{S_T} e^{\lambda(\widehat{w}_T - \mathbb{E}_B[\widehat{w}_T|S_T])}\bigr] \leqslant \exp\bigl(2\lambda^2 c_{T_2}T^2L_\omega^2\bigr), \tag{3.10}$$

where c_{T_2} is as in Theorem 3.13 and L_ω as in (d5). The left-hand side is zero by convention if $\mathbb{P}_B(S_T) = 0$.

Proof. Follow the path-functional argument in [DLSS26a, Corollary 6.14(d5)], using (d5) instead of integrating a gradient along the whole cube, and extend ω to all of $\mathbb{R}^n$, for example by

$$\omega_{\mathrm{ext}}(x) := \inf_{m \in \mathcal{S}_{\beta,A}(c)} \{\omega(m) + L_\omega \|x - m\|_2\}.$$

This agrees with ω on the good set and is L_ω-Lipschitz. The path functional $F(x) := \int_0^T \omega_{\mathrm{ext}}(x_t)\,dt$ is TL_ω-Lipschitz and integrable for the global law in Theorem 3.13, so (3.8) gives a sub-Gaussian MGF with variance proxy $v = c_{T_2} T^2 L_\omega^2$ for $F(\widehat{m}^{\mathrm{ext}})$ on the extended SDE. On S_T, the original and extended paths agree, hence $F(\widehat{m}^{\mathrm{ext}}) = \widehat{w}_T$ and their surviving-path means coincide, so finally, apply Lemma 3.14 to remove paths outside of S_T from the bound. □

3.5. **Algorithmic error analysis and proof of main theorem.** The following lemma will allow us to give a final error bound for the algorithm, given all the sources of error.

Lemma 3.16 (Jarzynski's equality with survival and rejection sampling)**.**
Suppose that f_t and $\widehat{f}_t$ are such that pathwise unique strong solutions to the following SDE's exist:

$$dx_t = f_t\,dt + dB_t \qquad x_0 \sim p_0 \tag{3.11}$$

$$d\widehat{x}_t = \widehat{f}_t\,dt + dB_t \qquad \widehat{x}_0 \sim p_0. \tag{3.12}$$

Let $\mathcal{S} \subseteq \mathbb{R}^n$ and for a process γ adapted to B_t, define the stopping time $\tau(\gamma) := \inf_{t \geqslant 0}\{t : \gamma_t \notin \mathcal{S}\}$. Let $\mathbf{P}_T, \widehat{\mathbf{P}}_T$ be the path measures for $x_{[0,T]}, \widehat{x}_{[0,T]}$ and $\mathbf{P}_T^s, \widehat{\mathbf{P}}_T^s$ be the path measures for $x_{[0,T]}, \widehat{x}_{[0,T]}$ conditioned on survival, $S_T = \{\tau > T\}$. Let $p_t, \widehat{p}_t$ be the density of $x_t, \widehat{x}_t$, respectively and p_t^s, $\widehat{p}_t^s$ be the densities for x_t, $\widehat{x}_t$, restricted to surviving paths S_t. Suppose that $\widehat{w}_t(\widehat{x}_{[0,T]})$ is defined so that

$$\frac{d\mathbf{P}_T^s}{d\widehat{\mathbf{P}}_T}(\gamma) = \mathbb{1}_{S_T}(\gamma)\frac{e^{\widehat{w}_T(\gamma)}}{\widehat{Z}_T}\frac{R_2(\gamma_T)}{Z_2}$$

where $\widehat{Z}_t = \mathbb{E}[e^{\widehat{w}_t} \mid S_t]$ and so $\frac{dp_T^s}{d\widehat{p}_T}(x) = \mathbb{E}[\mathbb{1}_{S_T}(\widehat{x}_{[0,T]})e^{\widehat{w}_T}/\widehat{Z}_T \mid \widehat{x}_T = x]\frac{R_2(x)}{Z_2}$. Here, $\widehat{w}_t = \widehat{w}_t(\widehat{x}_{[0,t]})$ unless specified otherwise. Suppose with $L = \exp\left(\frac{3}{\varepsilon}\left(\frac{1}{2}\int_0^T \varepsilon(t)^2 dt + 1\right)\right)$,

- *(0) (Survival probability) $\mathbf{P}_T(S_T) \geqslant 1 - \delta$ and $\widehat{\mathbf{P}}_T(S_T) \geqslant 1 - \widehat{\delta}$.*
- *(1) (Drift error) For a stopping time τ' such that $\tau' \leqslant \tau$, we have $\mathbb{E}_{x_{[0,T]}}\left[\mathbb{1}_{\{t<\tau'\}}\left\|\widehat{f}_t - f_t\right\|^2\right] \leqslant \varepsilon(t)^2$ and $\mathbf{P}_T[\tau' > T] \leqslant \delta'$.*
- *(2) (Exponential upper tail for log-weights) $\mathbb{E}\left[\mathbb{1}_{S_T}(\widehat{x}_{[0,T]})e^{\widehat{w}_T}\mathbb{1}_{e^{\widehat{w}_T}/\widehat{Z}_T \geqslant C_1}\right] \leqslant \varepsilon_1 \widehat{Z}_T$, for $\varepsilon_1 = \frac{c_3\varepsilon}{3L}$.*
- *(3) (Lower tail for log-weights) $\mathbb{P}\left[\{\widehat{x}_{[0,T]} \in S_T\} \cap \left\{\frac{e^{\widehat{w}_T}}{\widehat{Z}_T} \leqslant c_3\right\}\right] \leqslant \varepsilon_3$, for $\varepsilon_3 = \frac{\varepsilon}{3L}$.*

Then $\mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T}\frac{dp_T}{d\rho_T}(\widehat{x}_T) \wedge \frac{2C_1 L}{c_3}\right] \geqslant 1 - \varepsilon - \delta$, and given query access to $R_2(x)$, we can sample from p_T with TV error $2\varepsilon + \delta + \delta'$ by $O\left(\frac{C_1^2 L^2 \log(\frac{1}{\varepsilon})}{c_3^2}\right)$ simulations of (3.12), provided that $2\varepsilon + \delta < \frac{1}{2}$, by choosing the parameters $C_1' = C_2' = \frac{C_1 L}{c_3}$ and using the estimate $e^{\widehat{w}_T}R_2(\widehat{x}_T)$ in Algorithm 4. Suppose moreover that $\widetilde{x}_{[0,T]}$ is a random process and $\widetilde{w}_T(x_{[0,T]})$ are weights such that

- *(4) (Path error) $\mathrm{TV}(\widehat{\mathbf{P}}_T^s, \widetilde{\mathbf{P}}_T^s) \leqslant \varepsilon_4$ for some $\widetilde{\mathbf{P}}_T^s$ such that $\frac{d\widetilde{\mathbf{P}}_T^s}{d\widehat{\mathbf{P}}_T} \leqslant 1$, where $\varepsilon_4 = \frac{\varepsilon}{384C_1'^2}$.*[3]
- *(5) (Weight error) $\widetilde{\mathbf{P}}_T^s\left(|\widetilde{w}_T - \widehat{w}_T| > \frac{\varepsilon_4}{2}\right) \leqslant \frac{\varepsilon_4}{2}$.*
- *(6) (Ratio error) $\widetilde{R}_2(x)$ is a (randomized) estimate such that*

$$(a) \qquad \forall x, \quad \mathbb{P}\left(\frac{\widetilde{R}_2(x_T)}{R_2(x_T)} \notin [0, e^{\varepsilon_4/2}]\middle| x_T = x\right) \leqslant \frac{\varepsilon_4}{2}$$

$$(b) \qquad \mathbb{P}\left(\frac{\widetilde{R}_2(x_T)}{R_2(x_T)} \notin [e^{-\varepsilon_4/2}, e^{\varepsilon_4/2}]\right) \leqslant \varepsilon$$

[3]We define TV error for measures which may not be probability measures in the same way, $\mathrm{TV}(P, Q) = \sup_A |P(A) - Q(A)|$.

Then we can sample from p_T with TV error $6\varepsilon + \delta + \delta' + 768C_1'^2\widehat{\delta}$ by $O\left(\frac{C_1^2 L^2 \log(\frac{1}{\varepsilon})}{c_3^2}\right)$ simulations of (3.12) using the estimate $e^{\widetilde{w}_T(\widetilde{x}_{[0,T]})}\widetilde{R}_2(\widetilde{x}_T)$ in Algorithm 4, provided that this error is $< \frac{1}{2}$.

As the proof is very similar to [DLSS26a, Lemma 3.9], but with restriction to survival events in the right places, we defer it to §C. Note that we directly use the Radon-Nikodym derivative on path space rather than on the time-T marginal. We also allow an earlier stopping time in bounding the drift error: τ' can be more complicated (as in Lemma 3.3), so we only require the probability of $\tau' \leqslant T$ to be bounded under $\mathbf{P}_T$, whereas we need the probability of $\tau \leqslant T$ to be bounded under both $\mathbf{P}_T$ and $\widehat{\mathbf{P}}_T$.

In particular, note that if all the parameters are constant in n and $\delta, \delta', \widehat{\delta} = o_n(1)$, then there is a choice of $\varepsilon = o_n(1)$ such that the total error is still $o_n(1)$.

We can now prove the main theorem. As many of the arguments carry over from [DLSS26a, §8] in a straightforward manner, we focus on the necessary modifications.

Proof (Proof of Theorem 1.1). We check the conditions of Lemma 3.16 where the law $\mathbf{P}_T$ of the ideal process m_t is given by (HD), and the law of the approximate process $\widehat{\mathbf{P}}_T$ is given by (ASL-TAP) with weights (JE). The algorithm computes the Euler-Maruyama discretization with some step size h to be determined:

$$\widetilde{y}_{t+h} = \widetilde{y}_t + h\widetilde{m}_t + \sqrt{h}\xi_t, \quad \xi_t \sim \mathcal{N}(0, I_n), \qquad \widetilde{y}_0 = 0$$
$$\widetilde{w}_{t+h} = \widetilde{w}_t + \frac{h}{2}\left[\mathsf{Tr}(\widehat{Q}(\widetilde{m}_t)) + \|\widetilde{m}_t\|^2\right].$$

We first consider the case where $\widetilde{m}_t$ is computed exactly from $\widetilde{y}_t$ from (TAP). We take $\mathcal{S} = \mathcal{S}_{\beta,A}(c)$, so that $\tau = \inf\{t \geqslant 0 : m_t \notin \mathcal{S}_{\beta,A}(c)\}$ and $S_T = \{\forall t \in [0,T],\, m_t \in \mathcal{S}_{\beta,A}(c)\}$. We take

$$\tau' = \inf\left\{t \geqslant 0 : ((m_t, y_t), (\widehat{m}_t, \widehat{y}_t)) \notin \mathcal{S}^2(\rho\sqrt{n}, c\rho\sqrt{n}/2)\right\},$$

where the expanded pair safe set is defined with $\mathcal{S} = \{(m, y) : m \in \mathcal{S}_{\beta,A}(c)\}$. The Radon-Nikodym derivative is in the form specified by the lemma, by Lemma 3.6.

Part 0 (survival probability). Note that $\mathbf{P}_T(S_T^c) = o_n(1)$ directly from (d2a), while $\mathbf{P}_T(S_T^c) = o_n(1)$ follows from Lemma 3.7.

Part 1 (drift error). This holds with $\varepsilon(t) = O(1)$ by Lemma 3.3.

Part 2 & 3 (tails for log-weights). This follows from the subgaussianity of Jarzynski weights, Corollary 3.15 by [DLSS26a, Lemma 3.10], applied to the measure conditional on survival.

Part 4 (path error). Discretization error in [DLSS26a] is analyzed in path space, using Lipschitzness of a uniquely defined $\widehat{m}(y)$ solving (TAP). Instead, we can conduct the analysis on a lift of y-space $\widehat{y}(\mathcal{S}(c))$ based on homotopy equivalence in $\widehat{y}(\mathcal{S}(c))$; then $\widehat{m}$ is well-defined and Lipschitz on the lifted space. With $O\left(\frac{1}{n}\right)$ step size, Brownian excursions are $o(1)$ with high probability, so when $\widehat{m}$ lies within the expanded safe set, all arguments can be done within a single branch. We find that it suffices to take $O\left(\frac{\varepsilon_4^2}{n}\right)$ step size with an appropriate constant to obtain TV error ε_4.

Part 5 (weight error). Following the discretization arguments in [DLSS26a] with the above remarks, to obtain the needed error bound, it suffices to take $h = O\left(\frac{\varepsilon_4^2}{n^2}\right)$ with an appropriate constant.

Part 6 (ratio error). We need to check the estimation of $\widehat{Z}_{\beta A,y}$. With high probability over A, we can estimate the partition function on any small enough Hamming ball $B(\sigma, c(\beta)n) \in \{\pm 1\}^n$, by using the sampling guarantee for all temperatures $\beta' \leqslant \beta$ (7.1) along with simulated annealing, as in [DLSS26a, §7.5]; this gives an underestimate with high probability not depending on σ. When $\sigma = \mathrm{sign}(y_T)$ where y_T is the SL tilt, for large enough $T = T(\beta)$, with probability $1 - e^{-\Omega(n)}$, this wedge contains all but $e^{-\Omega(n)}$ of the mass of the entire partition function, by [DLSS26a, Lemma 7.43].

Putting everything together. Thus by Lemma 3.16, the output of Part 1 of Algorithm 2 is a sample $\widetilde{y}_T \sim \widetilde{p}_T$ with $\mathrm{TV}(\widetilde{p}_T, p_T) = o_n(1)$, in $\mathsf{poly}(n)$ time. By Theorem 7.1, with $1 - o_n(1)$ probability over $y_T \sim p_T$, we obtain a sample $\widehat{\sigma}$ that is $o_n(1)$ in TV distance from $\mu_{\beta A, y_T}$ in polynomial time.

Letting $\widetilde{\mu}$ be the distribution of $\widehat{\sigma}$ for tilt $\widetilde{y}_T$, we have by the chain rule for TV distance that

$$\mathrm{TV}(\widetilde{\mu}, \mu) \leqslant \mathrm{TV}(\widetilde{p}_T, p_T) + \mathop{\mathbb{E}}_{y_T \sim p_T} \mathrm{TV}(\mathcal{L}(\widehat{\sigma}|y_T), \mathcal{L}(\sigma|y_T)) = o_n(1),$$

as needed.

Error in solving (TAP). Robustness to approximating the solution to (TAP) is shown in [DLSS26a, §D]. The same mirror descent analysis works with initialization that is within a ball around the true solution m that is inside the expanded safe set, which is ensured under Brownian excursions at every step being at most $O(1)$. □

4. Uniform TAP Hessian positivity over the stochastic localization process

We will use a "hybrid" argument that combines the local strong convexity of the TAP Hessian $\widehat{Q}(\cdot)$ for $0 < \beta < 1$ in small balls around the AMP iterates $\{\overline{m}_t\}_{t \geqslant 0}$ due to a result of Celentano [Cel24] and the $o(\sqrt{n})$-ℓ_2 bound between the AMP sequence $\{\overline{m}_t\}_{t \geqslant 0}$ and SL process $\{m_t\}_{t \geqslant 0}$ implied by the result of El-Alaoui, Montanari and Sellke [AMS22] with a covering argument. This will immediately imply desideratum (d2a), namely:

Desideratum 2. For every $0 \leqslant T(\beta) < \infty$, there exist $\rho(\beta, T), c_{\mathrm{TAP}}(\beta, T) > 0$ such that

$$\mathop{\mathbb{P}}_{A, B_t} \left\{ \forall t \in [0, T] \,.\, B_{\square}(m_t, \rho\sqrt{n}) \subseteq \mathcal{S}_A(c_{\mathrm{TAP}}) \right\} \geqslant 1 - o_n(1)\,.$$

AMS Hessian, TAP Hessian and AMP iterates. Let $D(m_t) := \mathsf{diag}\left(\frac{1}{1-m_t^2}\right)$. The Hessian used in [AMS22] is given as

$$\overline{Q}(m_t)^{-1} := \beta^2(1 - q^*(t))I_n - \beta A + D(m_t)\,.$$

Recall the Hessian of the SL-tilted planted-SK model used in [DLSS26a, §4] is

$$\nabla^2 \mathcal{F}_{\mathrm{TAP}}(m_t, \cdot) := \beta^2 \left(1 - \frac{\|m_t\|_2^2}{n}\right) I_n - \beta A + D(m_t) - \frac{\beta^2}{n} x_0 x_0^{\mathsf{T}} - \frac{2\beta^2}{n} m_t m_t^{\mathsf{T}}\,.$$

The AMP iterates, run for $k \geqslant k_0(\varepsilon, \beta, t)$ iterations at *every* $t \in [0, T(\beta)]$, start at $\overline{m}_{-1} = 0^n$ and $y_t = t x_0 + g_0$ where $g_0 \sim \mathcal{N}(0, I_n)$, and follow the standard updates

$$\begin{aligned} \overline{m}_j &= \tanh(z_j)\,, \\ z_{j+1} &= \beta \left(A + \frac{\beta}{n} x_0 x_0^{\mathsf{T}}\right) \overline{m}_j + y_t - \beta^2 \left(1 - \frac{\|\overline{m}_j\|_2^2}{n}\right) \overline{m}_{j-1}\,, \end{aligned}$$

run for $j \in [0, k]$ iterations. The result of Celentano [Cel24] considers a comparison with a linear functional of an auxiliary Gaussian field, the definition of which requires considering an approximate isometry $\mathbf{T}$ from the subspace spanned by the AMP magnetization estimates $\{\overline{m}_0, \ldots, \overline{m}_{k-1}\}$ to that of the vectors that come from the "pure" SK component

$$g_j = \beta A \overline{m}_{j-1} - \beta^2 \left(1 - \frac{\|\overline{m}_{j-1}\|_2^2}{n}\right) \overline{m}_{j-2}\,,$$

which remove the contributions from the tilt y_t and the plant $J = \mathbf{1}^n (\mathbf{1}^n)^{\mathsf{T}}$.

4.1. "Hybrid" local strong convexity.

For our use case, the only non-trivial part of this proof is to check that the Schur complement of a 2×2-matrix induced by a specific low-dimensional objective induced by the Hessian $\widehat{Q}(\cdot)$ around test-points in a compact set K corresponding to the AMP iterates is negative. Once the correct reduction to the low-dimensional optimization problem is setup in our "hybrid" model, no change is needed – the proof is exactly that of [Cel24, Proof of Theorem 2] with just one syntactic change.

To briefly recap the proof, we introduce some notation which closely follows the definitions in [DLSS26a, §5] and allows a modular invocation of [Cel24, Proof of Theorem 2]. Let (m^*, q^*) solve the fixed-point equations in [DLSS26a, Theorem 5.1]. Let $G \sim \mathcal{N}(0, \beta^2 q^*)$ and $G_0 \sim \mathcal{N}(0,1)$ be independent, and denote various parts of the functionals in the fixed-point equations as

$$\begin{aligned}
t_\infty &:= \beta^2 q^* + t\,,\\
M &:= \tanh(t_\infty + G + \sqrt{t}G_0) = \tanh(Y^*)\,,\\
b^* &:= \mathbb{E}[M^3] = \mathbb{E}[M^4]^{4}\,,\\
\Delta &:= \mathbb{E}\left[(1-M^2)^2\right] = \mathbb{E}\left[\operatorname{sech}^4(Y^*)\right] = 1 - 2q^* + b^*\,,\\
r &:= 1 - \frac{b^*}{q^*}\,,\\
s &:= 1 - \beta^2\Delta = 1 - \beta^2\,\mathbb{E}[\operatorname{sech}^4(Y^*)]\,.
\end{aligned}$$

In the notational choices above, the dependencies of various constants on t have been suppressed for convenience[5]. The astute reader may observe that the quantity s is a "self-stability" term that shows up in the denominator of various Taylor-expanded terms in the cavity estimates conducted in [DLSS26a, §5.4]. For $0 < \beta < 1$ and $0 < t \leqslant T(\beta) < \infty$, it is straightforward to see that $s > 0$ and $0 < b^* < q^* < 1$. The latter immediately implies that $\Delta, r > 0$.

We now state the result, which shows that after sufficiently many runs of the AMP algorithm initialized with tilt y_t, the Hessian of the TAP free energy for the SL-tilted planted SK model is PSD in a ℓ_2-ball around of radius $\Omega_{\beta,t}(\sqrt{n})$ around the final AMP iterate.

Theorem 4.1 ("Hybrid" local strong convexity [Cel24, Syntactic modification of proof of Theorem 2])**.** *For every $\beta, t > 0$, there exist $c_0(\beta,t), c_1(\beta,t) > 0$, such that, for all $0 < \varepsilon < c_1(\beta,t)$ and $k \geqslant k_0(\varepsilon,\beta,t)$,*

$$\mathop{\mathbb{P}}_{A,g_0}\left\{\inf_{\sigma \in B_\square(\overline{m}_{k-1}, \varepsilon\sqrt{n})} \nabla^2 \mathcal{F}_{\mathrm{TAP}}(\sigma,\cdot) \succeq c_0(\beta,t)\right\} \geqslant 1 - o_n(1)\,.$$

Proof. The proof is exactly analogous to [Cel24, Theorem 2] with the small modification that it tests the quadratic forms of the TAP Hessian considered in our previous work [DLSS26a][6] evaluated at the AMP iterates given above. Using gauge symmetry, we fix $x_0 = \mathbf{1}^n$. Throughout the proof, the AMP law is the one corresponding to the iterates which come from [AMS22, Algorithm 1, (1)–(5)] and the Hessian is $\nabla^2 \mathcal{F}_{\mathrm{TAP}}(\sigma,\cdot)$[7]. For completeness, we briefly recreate the proof of [Cel24, Theorem 2, Pg 11–18] by analyzing the objective $-v^\mathsf{T}\nabla^2\mathcal{F}_{\mathrm{TAP}}(u,\cdot)v$ on the AMP transcript $(\overline{m}_0, \ldots, \overline{m}_{k-1}, g_0, g_1, \ldots, g_k)$ and the limiting joint law $\mathcal{S}_1^{(2)}(\varepsilon) := (V, U, M, G, G_0, \Xi)$ where $\Xi \sim \mathcal{N}(0,1)$ (independently of G and G_0), V and U are the limiting empirical laws of the test vector $v \in B_\square(r,x)$ and the evaluation vector u respectively, and the joint law satisfies the constraints in the definition of the set given in [Cel24, Pg 14].

Reduction to the scalar objective. For every $v \in \mathcal{S}^{n-1}(1)$, expanding the Hessian yields

$$-\frac{1}{n}v^\mathsf{T}\nabla^2\mathcal{F}_{\mathrm{TAP}}(\sigma,\cdot)v = \beta v^\mathsf{T}Av + \frac{\beta^2}{n}\langle \mathbf{1}, v\rangle^2 - \sum_{i=1}^n \frac{v_i^2}{1-u_i^2} - \lambda^2(1-Q(u)) + \frac{2\lambda^2}{n}\langle u, v\rangle^2\,,$$

which is just $f_{\mathrm{FMM}}(v,u)$ as defined on [Cel24, Pg. 11]. We now apply the post-AMP comparison to the AMS transcript $(\overline{m}_0, \ldots, \overline{m}_{k-1})$ and $f_{\mathrm{FMM}}(v,u)$, followed by the empirical-law and fixed-point reductions [Cel24, Corollary 1, Lemma 1 and §4.2–4.3]. These give

$$\limsup_{\epsilon \downarrow 0} \limsup_{k\to\infty} \text{p-}\limsup_{n\to\infty} T_{n,k}(\epsilon) \leqslant \limsup_{\varepsilon\downarrow 0} \sup_{\mu \in \mathcal{S}_1^{(2)}(\varepsilon)} \mathcal{F}_0^{(2)}(\mu)\,, \tag{4.1}$$

[4] This can be proved via a simple extension of the proof of [LSS26b, Lemma 2.1] – see Lemma B.4.

[5] The quantities m^*, q^*, Y^* and b^* all dependent on t.

[6] This is equivalent to the Hessian used by Fan, Mei and Montanari [FMM21].

[7] The original result of Celentano [Cel24, Theorems 1–2] tests AMP iterates designed by Fan et al [FMM21] on the TAP Hessian $\nabla^2\mathcal{F}_{\mathrm{TAP}}(\sigma,\cdot)$ or the AMP iterates designed by El-Alaoui et al [AMS22] on $\overline{Q}^{-1}(\cdot)$. We require testing the AMP iterates of Alaoui et al [AMS22] on the TAP Hessian $\nabla^2\mathcal{F}_{\mathrm{TAP}}(\sigma,\cdot)$ – this leads to the name "hybrid" local convexity.

where $T_{n,k}(\varepsilon) \coloneqq \sup_{u \in B_\square(\overline{m}_{k-1}, \varepsilon\sqrt{n}), \|v\|_2 = 1} -v^\mathsf{T} \widehat{Q}(u) v$ and $\mathcal{F}_0^{(2)}(\mu)$ is as defined in [Cel24, (24)]. Carrying the computations in [Cel24, §4.4] further by rewriting the Lagrange multiplier to combine with the AMS state evolution and FMM TAP Hessian gives

$$\sup_{\mu \in \mathcal{S}_1^{(2)}(\varepsilon)} \mathcal{F}_0^{(2)}(\mu) \leqslant \sup_{\rho, u \in [-1,1]} \min_{(\alpha_\rho, \alpha_u) \in K} \mathsf{L}(\rho, u; \alpha_\rho, \alpha_u, \alpha_v) + C_K \varepsilon \,,$$

for some constant $C_K > 0$. Expanding the Lagrangian as in after [Cel24, (27)] yields

$$\sup_{\mu \in \mathcal{S}_1^{(2)}(\varepsilon)} \mathcal{F}_0^{(2)}(\mu) \leqslant \sup_{\rho, u \in [-1,1]} (\rho, u)^\mathsf{T} S_0(\alpha_v)(\rho, u) + C_K \varepsilon \,,$$

where $S_0(\alpha_v) = A_{11,0}(\alpha_v) - A_{12}(\alpha_v) A_{22}(\alpha_v)^{-1} A_{12}(\alpha_v)^\mathsf{T}$ and these matrices are explicitly derived next[8].

The hybrid Schur-complement calculation. We compute the blocks at $\alpha_v = 0$. We first apply Stein's lemma over G and keep G_0 fixed. Since the variance is $\beta^2 q^*$, and we have $\partial_G M = 1 - M^2$, together with the Nishimori moment identities ($m^* = q^*$ and Lemma B.4), this gives

$$\begin{aligned}
\mathbb{E}[GM(1-M^2)] &= \beta^2 q^* \, \mathbb{E}[(1-M^2)(1-3M^2)] = \beta^2 q^* \, d_0 \,, \\
\mathbb{E}[G(1-M^2)] &= -2\beta^2 q \, \mathbb{E}[M(1-M^2)] = -2\beta^2 q^* (q^* - b^*) \,, \\
\mathbb{E}[G^2(1-M^2)] &= \beta^2 q^* \, \mathbb{E}[1 - M^2 - 2GM(1-M^2)] = \beta^2 q^* (1-q^*) - 2\beta^4 (q^*)^2 d_0 \,,
\end{aligned}$$

where $d_0 = 1 - 4q^* + 3b^*$. Substituting $h_0(M) = 1 - M^2$ yields

$$\begin{aligned}
A_{11,0}(0) &= \begin{pmatrix} 2\beta^2 q^* - 2\beta^4 q^* d_0 & 0 \\ 0 & \beta^2 \end{pmatrix} , \\
A_{12}(0) &= \frac{1}{2} \begin{pmatrix} -1 + \beta^2 d_0 & -2\beta^2 \sqrt{q}^* \, (q^* - b^*) \\ 0 & -1 \end{pmatrix} , \\
A_{22}(0) &= \frac{1}{4} \begin{pmatrix} r & \sqrt{q}^* \, r \\ \sqrt{q}^* \, r & 1 - q^* \end{pmatrix} .
\end{aligned}$$

Only the first block depends explicitly on the Hessian choice, while the other blocks use the AMS scalar law. Using $1 - q^* = \Delta + r q^*$ gives

$$A_{22}(0)^{-1} = 4 \begin{pmatrix} r^{-1} + q/\Delta & -\sqrt{q}^*/\Delta \\ -\sqrt{q}^*/\Delta & 1/\Delta \end{pmatrix} .$$

The $(2,2)$ and $(1,2)$ entries of the Schur complement are therefore

$$\begin{aligned}
S_0(0)_{22} &= \beta^2 - \Delta^{-1} = -\frac{s}{\Delta} \,, \\
S_0(0)_{12} &= \frac{\sqrt{q}^* \, s}{\Delta} \,.
\end{aligned}$$

For the remaining scalar Schur complement, evaluate the quadratic form at $(1, \sqrt{q})^\mathsf{T}$. Substituting the displayed inverse and using $d_0 = \Delta - 2rq^*$ and $s = 1 - \beta^2 \Delta$ yields

$$S_0(0)_{11} + \frac{s q^*}{\Delta} = 3\beta^2 q^* - 2\beta^4 q^* d_0 - \frac{(s + 2\beta^2 r q^*)^2}{r} - q\beta^4 \Delta = -q\beta^2 s - \frac{s^2}{r} \,.$$

Consequently, for $z = (\rho, a)^\mathsf{T}$,

$$z^\mathsf{T} S_0(0) z = -\frac{s}{\Delta} (a - \sqrt{q}^* \, \rho)^2 - \left(q^* \beta s + \frac{s^2}{r} \right) \rho^2 \,. \tag{4.2}$$

Both coefficients are strictly positive, and the two squares vanish simultaneously only at $z = 0$, so $S_0(0) \prec 0$. At this point, a simple continuity argument implies that $S_0(\alpha_v) \prec 0$ for small enough $\alpha_v > 0$, and a choice of $\alpha_v \leqslant \min\{s/2, 1/2\}$ yields a strict upper bound of 0 on the objective function. □

4.2. **Covering the entire SL path via ℓ_2-balls centered at posterior means.** The proof of Theorem 4.1 implies that for every fixed $t \in [0, T(\beta)]$, after taking sufficiently many AMP iterates and outputting $\overline{m}_{k-1}$, the

[8] This is done for completeness, though the calculations in [Cel24, §4.4] also cover our case almost entirely.

TAP Hessian $\nabla^2 \mathcal{F}_{\mathrm{TAP}}(u)$ is PSD at any point $u \in B_\square(\overline{m}_{k-1}, \Omega_{\beta,t}(\sqrt{n}))$. However, this does not yield *uniform* control over the entire SL trajectory $\{m_t\}_{t\in[0,T(\beta)]}$ which is critical to proving the requisite desideratum. We now "boost" the fixed-time, local strong convexity in reasonable-sized ℓ_2-balls around AMP iterates implied by Theorem 4.1 to uniform local convexity along the SL paths by using these balls to cover a significant measure of the SL paths. Doing this requires using the uniform control over the fluctuations of the SL process inside small intervals supplied by [AMS22, Lemma 4.9] along with the ℓ_2-bound

$$\|m_t - \overline{m}_{k-1}\| \leqslant \varepsilon\sqrt{n}\,,$$

for a large-enough AMP iterate corresponding to an initialization at every $t \in [0,T]$ implied by [AMS22, Proposition 4.8].

Theorem 4.2 (Uniform local strong convexity along $\{m_t\}_{0\leqslant t\leqslant T(\beta)}$)**.**
Let $\{m_t\}_{t\geqslant 0}$ evolve according to the HD process, and $S_A(c_{\mathrm{TAP}})$ be as in Definition 2.2. Then,

$$\mathop{\mathbb{P}}_{A,B}\left\{\forall t \in [0,T], \quad B_\square(m_t, \rho\sqrt{n}) \subseteq \mathcal{S}_A(c_{\mathrm{TAP}})\right\} = 1 - o_n(1)\,.$$

Proof. We first show there are small ℓ_2-balls around the true means $\{m_t\}_{t\geqslant 0}$ in which $\widehat{Q}(m_t) \succeq c(\beta,t) I_n$ at every *fixed* t. At $t > 0$, this uses contiguity between the planted and original model for $\beta < 1$ [AMS22, §2] as well as [AMS22, Proposition 4.8] along with a union bound. At $t = 0$, we prove this directly by conditioning on the event that $\|A\|_{\mathsf{op}} \leqslant 2 + \frac{(1-\beta)^2}{8\beta}$ and the radius of the ball $\varepsilon(\beta,0) \leqslant \frac{(1-\beta)^2}{24\beta^2}$. We then use [AMS22, Lemma 4.9] through Doob's inequality to control the fluctuations of m_t in a small time-interval $(t-\delta_t, t+\delta_t) \subseteq [0,T]$ and finally construct a "tube" composing balls for any finite sub-cover of the interval $[0,T]$ conditioned on the fixed-time convexity estimates and the small-time stability estimates.

Fixed-time convexity balls. For every fixed $t > 0$, [AMS22, Proposition 4.8] immediately implies that, in the planted model,

$$\mathop{\mathbb{P}}_{A,B_t}\left\{\left\|m_t - \widehat{m}^{k-1}(A, y_t)\right\|_2 \leqslant \varepsilon'(\beta,t)\right\} \geqslant 1 - o_n(1)\,,$$

for any $\varepsilon'(\beta,t) > 0$ and $\widehat{m}^{k-1}(A, y_t)$ denoting the k-th output of the AMP iteration of AMS initialized with tilt y_t. Invoking Theorem 4.1 we obtain parameters $\varepsilon''(\beta,t)$ and $k'(\beta,t)$ so that

$$\mathop{\mathbb{P}}_{A,g_0}\left\{\forall u \in B_\square(\overline{m}_{k'(\beta,t)-1}, \varepsilon''(\beta,t)\sqrt{n}),\, \nabla^2 \mathcal{F}_{\mathrm{TAP}}(u,\cdot) \succcurlyeq \varepsilon''(\beta,t)\right\} \geqslant 1 - o_n(1)\,.$$

Choose $k(\beta,t) \geqslant \max\{k, k'(\beta,t)\}$ and set $\overline{m}_{k(\beta,t)-1} = \widehat{m}^{k-1}(A, y_t)$, along with a choice of radius $\varepsilon(\beta,t) = \frac{1}{2}\min\{\varepsilon'(\beta,t), \varepsilon''(\beta,t)\}$. Then, a union bound yields

$$\mathop{\mathbb{P}}_{A,B_t}\left\{\forall u \in B_\square(\overline{m}_{k(\beta,t)-1}, \varepsilon(\beta,t)\sqrt{n}),\, \nabla^2 \mathcal{F}_{\mathrm{TAP}}(u,\cdot) \succcurlyeq c(\beta,t) I_n\right\} \geqslant 1 - o_n(1)\,,$$

where we invoke contiguity to transfer back to the original SK law from the planted model.
To obtain the same statement at $t = 0$, use $m_0 = 0$ and choose $\delta_A = \frac{(1-\beta)^2}{8\beta}$ and $\varepsilon(\beta,0) = \frac{(1-\beta)^2}{24\beta^2}$ so that $\mathbb{P}_A\left\{\|A\|_{\mathrm{op}} \leqslant 2 + \frac{(1-\beta)^2}{8\beta}\right\} \geqslant 1 - o_n(1)$. Conditioning on this event, for every $u \in B_\square(0, \varepsilon(\beta,0)\sqrt{n})$, the bounds $D(u) \succcurlyeq I_n$ and $uu^\intercal/n \preccurlyeq \frac{\|u\|_2^2}{n} I_n$ yield

$$\begin{aligned}\nabla^2 \mathcal{F}_{\mathrm{TAP}}(u,\cdot) &= \beta^2\left(1 - \frac{\|u\|_2^2}{n}\right) I_n - \beta A + D(u) - \frac{2\beta^2}{n} uu^\intercal \succcurlyeq \left(1 + \beta^2 - \beta\|A\|_{\mathrm{op}} - \frac{3\beta^2}{n}\|u\|_2^2\right) I_n \\ &\succcurlyeq \left((1-\beta)^2 - \frac{(1-\beta)^2}{8} - \frac{(1-\beta)^2}{8}\right) I_n \succcurlyeq \frac{(1-\beta)^2}{2} I_n\,.\end{aligned}$$

Set $c(\beta,0) = (1-\beta)^2/2$. Thus, for every fixed $t \in [0,T]$, there is a deterministic tuple $(c(\beta,t), \varepsilon(\beta,t))$ for which $\mathbb{P}_{A,B}\left\{\forall u \in B_\square(\overline{m}_{k(\beta,t)-1}, \varepsilon(\beta,t)\sqrt{n}),\, \nabla^2 \mathcal{F}_{\mathrm{TAP}}(u,\cdot) \succcurlyeq c(\beta,t) I_n\right\} \geqslant 1 - o_n(1)$, and we denote these events $\mathcal{E}_{n,t} := \left\{\forall u \in B_\square(\overline{m}_{k(\beta,t)-1}, \varepsilon(\beta,t)\sqrt{n}),\, \nabla^2 \mathcal{F}_{\mathrm{TAP}}(u,\cdot) \succcurlyeq c(\beta,t) I_n\right\}$.

Fluctuations across small time intervals. For each $t \in [0,T]$, choose a deterministic $0 < \delta_t \leqslant \frac{(\varepsilon(\beta,t)(1-\beta))^2}{128}$ and set

$$a_t := \max\{0, t - \delta_t\}\,, \qquad b_t := \min\{T, t + \delta_t\}\,,$$

and $J_t := (t-\delta_t, t+\delta_t) \cap [0,T]$. Note that J_t is an open neighborhood of t relative to $[0,T]$. By the proof of [AMS22, Lemma 4.9] and [AMS22, (4.24)],

$$\mathop{\mathbb{P}}_{A,B_t}\left\{\sup_{r\in[a_t,b_t]} \frac{\|m_r - m_{a_t}\|_2^2}{n} \geqslant q(b_t) - q(a_t) + \frac{\varepsilon(\beta,t)^2}{64}\right\} \leqslant o_n(1)\,,$$
$$q(b_t) - q(a_t) \leqslant \frac{(b_t - a_t)}{(1-\beta)^2} \leqslant \frac{\delta_t}{(1-\beta)^2} \leqslant \frac{\varepsilon(\beta,t)^2}{64}\,.$$

In particular, the event $\mathcal{F}_{n,t} := \left\{\sup_{r\in[a_t,b_t]} \frac{\|m_r - m_{a_t}\|_2^2}{n} \leqslant \frac{\varepsilon(\beta,t)^2}{32}\right\}$ has probability $1 - o_n(1)$. On this event, for every $s \in J_t$,

$$\begin{aligned}\|m_s - m_t\|_2 &\leqslant \|m_s - m_{a_t}\|_2 + \|m_t - m_{a_t}\|_2\\ &\leqslant 2\sqrt{nZ_{n,t}} \leqslant \frac{\varepsilon(\beta,t)}{2}\sqrt{n}\,.\end{aligned}$$

Finite sub-cover and a uniform bound. The open sets $\{J_t : t \in [0,T]\}$ cover $[0,T]$ and compactness implies the existence of a finite subcover $J_{t_1}, \dots, J_{t_m}$. Denote the uniform radius and convexity parameters

$$\varepsilon_M := \frac{1}{2}\min_{j\in[M]} \varepsilon(\beta,t_j) > 0\,,$$
$$c_M := \min_{j\in[M]} c(\beta,t_j) > 0\,.$$

Taking a union bound over the M events $\bigcap_{j=1}^M \mathcal{F}_{n,t_j} \cap \mathcal{E}_{n,t_j}$ gives

$$\mathop{\mathbb{P}}_{A,B}\left\{\bigcap_{j=1}^M \mathcal{F}_{n,t_j} \cap \mathcal{E}_{n,t_j}\right\} \geqslant 1 - o_n(1)\,.$$

Note that each of the j events concern the same disorder and SL path. On this event, fix any $s \in [0,T]$ and choose j with $s \in J_{t_j}$. For every $u \in B_\square(m_s, \varepsilon_M\sqrt{n})$,

$$\begin{aligned}\left\|u - m_{t_j}\right\|_2 &\leqslant \|u - m_s\|_2 + \left\|m_s - m_{t_j}\right\|_2\\ &\leqslant \left(\varepsilon_M + \frac{\varepsilon(\beta,t_j)}{2}\right)\sqrt{n} \leqslant \varepsilon(\beta,t_j)\sqrt{n}\,.\end{aligned}$$

This implies the containment $B_\square(m_s, \varepsilon_M\sqrt{n}) \subseteq B_\square(m_{t_j}, \varepsilon(\beta,t_j)\sqrt{n})$, and $\mathcal{E}_{n,t_j}$ gives $\nabla^2\mathcal{F}_{\mathrm{TAP}}(u,\cdot) \succeq c_M I_n$. Therefore,

$$\mathop{\mathbb{P}}_{A,B}\left\{\forall s \in [0,T],\ \forall u \in B_\square(m_s, \varepsilon_M\sqrt{n}),\ \nabla^2\mathcal{F}_{\mathrm{TAP}}(u,\cdot) \succeq c_M I_n\right\} \geqslant 1 - o_n(1)\,.$$

If $T = 0$, the same conclusion follows from the first part of the proof with $\varepsilon_M := \varepsilon(\beta,0)/2$ and $c_M := c(\beta,0)$.

Löwner order vis-a-vis $D(u)$. Intersect the preceding event $\bigcap_{j=1}^M \mathcal{F}_{n,t_j} \cap \mathcal{E}_{n,t_j}$ with $\{\|A\|_{\mathrm{op}} \leqslant 3\}$, which also has probability $1 - o_n(1)$. Since $\|u\|_2^2/n \leqslant 1$ for any $u \in (-1,1)^n$, this gives the following Löwner order lower bound

$$\nabla^2\mathcal{F}_{\mathrm{TAP}}(u,\cdot) = \beta^2(1 - \|u\|_2^2/n)I_n - \beta A + D(u) - \frac{2\beta^2}{n}uu^\mathsf{T} \succeq D(u) - \beta(2\beta+3)I_n\,.$$

Throughout the tube we also have $\nabla^2\mathcal{F}_{\mathrm{TAP}}(u,\cdot) \succeq c_M I_n$, so combining with the previous inequality gives

$$D(u) \preceq \nabla^2\mathcal{F}_{\mathrm{TAP}}(u,\cdot) + \beta(2\beta+3)I_n \preceq \left(1 + \frac{\beta(2\beta+3)}{c_M}\right)\nabla^2\mathcal{F}_{\mathrm{TAP}}(u,\cdot)\,.$$

Choose

$$c_{\mathrm{TAP}} = \frac{c_M}{c_M + \beta(2\beta+3)} > 0\,,$$

and note the preceding bound immediately yields $\nabla^2\mathcal{F}_{\mathrm{TAP}}(u,\cdot) \succeq c_{\mathrm{TAP}} D(u)$ throughout the tube. Since $D(u)$ is PD, it follows that

$$\mathop{\mathbb{P}}_{A,B_t}\left\{\forall t \in [0,T],\, B_\square(m_t, \rho\sqrt{n}) \subseteq \mathcal{S}_A(c_{\mathrm{TAP}})\right\} \geqslant 1 - o_n(1)\,. \qquad \square$$

4.3. **Löwner order sandwich on the safe set.** The fact that $D(m) \succ 0$ plus the fact that $\widehat{Q}(m_t) \succeq c(\beta,t) I_n$ allows one to conclude the PSD-ness in the primal space (obtained by conjugating by $D(m_t)$).

Lemma 4.3 (Desideratum 2(b) – Löwner sandwich for dual coordinate chart)**.**
Let $m \in \mathcal{S}_A(c_{\mathrm{TAP}})$ and condition on the "master" event $\mathcal{E}_A$. Then, there exist $0 < c_{\mathrm{TAP}}(\beta,T), C_{\mathrm{TAP}}(\beta) < \infty$, such that

$$C_{\mathrm{TAP}}(\beta)^{-1} D^{-1}(m) \preceq \widehat{Q}(m) \preceq c_{\mathrm{TAP}}(\beta,T)^{-1} D^{-1}(m).$$

Proof. Conditioning again on $\|A\|_{\mathsf{op}} \leqslant 3$, and using the fact that $D(m) \succ I_n$ for any $m \in (-1,1)^n$, yields

$$\begin{aligned}
\nabla^2 \mathcal{F}_{\mathrm{TAP}}(m,\cdot) &= \beta^2(1 - \|m\|_2^2/n) I_n - \beta A + D(m) - \frac{2\beta^2}{n} m m^{\mathsf{T}} \\
&\preceq D(m) + \beta \left(\|A\|_{\mathsf{op}} + \beta^2 \right) I_n \\
&\preceq (1 + 3\beta + \beta^2) D(m)\,.
\end{aligned}$$

By Theorem 4.2, $\nabla^2 \mathcal{F}_{\mathrm{TAP}}(m,\cdot) \succeq c_{\mathrm{TAP}} D(m)$ for any $m \in \mathcal{S}_A(c_{\mathrm{TAP}})$, conditioned on the desirable event $\mathcal{E}_A$. Inverting the Hessian and combining the PD-ness with the preceding upper bound yields

$$\frac{1}{1 + 3\beta + \beta^2} D^{-1}(m) \preceq \widehat{Q}(m) \preceq \frac{1}{c_{\mathrm{TAP}}} D^{-1}(m)\,,$$

for any $m \in \mathcal{S}_A(c_{\mathrm{TAP}})$. □

5. Resolvent analysis: Regularization, uniform concentration and interpolation

We now repeat the structure of the argument in [DLSS26a, §B] to develop control over the diagonal sub-algebra of $\widehat{Q}^2(m_t)$ for all points $m_t \in S_A(c)$. While the idea of repeating the same argument restricted to the safe set $S_A(c)$ (Definition 2.2) is natural, it causes technical problems in the chaining argument used for uniform concentration of $E_{\mathcal{D}_n}[\widehat{Q}^2(m_t)]$ around its expected value (§5.2) – this happens because the supremum is taken over a set of points m_t that now depend on A through $S_A(c)$ and any sort of conditioning destroys the structure of the underlying Gaussian randomness. One solution to this is to define a regularization $\phi(\widehat{Q}^{-2}(m_t))$ that agrees with $\widehat{Q}^2(m_t)$ for every $m_t \in S_A(c)$ and is Lipschitz, bounded and continuously twice-differentiable for every $m_t \in \{-1,1\}^n \setminus S_A(c)$ (§5.1). As it turns out, our definition of this regularization ϕ as a function of $\widehat{Q}^{-2}(m_t)$ merges well with a representation for $\widehat{Q}^{-1}(m_t)$ in block-form with each block being an affine function of A — this representation is the key to a comparatively simple free interpolation argument comparing $E_{\mathcal{D}_n}[\varphi(\widehat{Q}^{-2}(m_t))]$ with the ideal free limit (§5.3). After obtaining uniform concentration and the control of the diagonal of the expected squared resolvent, one combines these estimates via a triangle inequality and restricts to the safe set $m \in S_A(c)$ at the very end to obtain a valid spectral cut-off $\gamma_Y(\beta)$ with which to invoke ϕ (§5.4).

5.1. **Bounded extension of matrix inverse.** In [DLSS26a, §B], the authors established a deterministic diagonal approximation to $E_{\mathcal{D}_n}[D(aI_n - \beta A + D)^{-2} D]$, uniformly over $a \in [0,1]$ and $D \in \mathcal{D}_n([1,\infty))$, with high probability over A. To do this, we required uniform Löwner lower bounds on $(aI_n - \beta A + D)$.

In our current setting, with $1/2 < \beta < 1$, it is possible for $(aI_n - \beta A + D)$ to become singular. Therefore, we replace the weighted squared inverse $D(\cdot)^{-2}D$ with a regularized version $\phi(D^{-1}(\cdot)^2 D^{-1})$, where ϕ is a bounded C^2 extension of the reciprocal $(\cdot)^{-1}$ defined, for fixed $\gamma_Y > 0$, as

$$\phi(y) := \left[y + \gamma_Y \left(\frac{(\gamma_Y - y)_+}{\gamma_Y + y} \right)^3 \right]^{-1}, \qquad y > -\gamma_Y. \tag{5.1}$$

This scalar function agrees with y^{-1} when $y \geqslant \gamma_Y$ and has bounded and continuous first and second derivatives on $[0,\infty)$[9]. Following the standard practice in matrix analysis [Bha13], we will lift ϕ to symmetric matrices

[9] This particular regularization ϕ bears technical and conceptual parallel to the regularization used in [JSS25, §2] to "smooth" out the first derivative of the primal Parisi PDE in the event that a PHA iterate's coordinate leaves the solid cube. Both regularizations are of the form $a^{-1} \rightsquigarrow (a + y(a))^{-1}$ where $a = \Phi''(x_\gamma)$ in [JSS25, §2] and $a \in \mathsf{Spec}\left((I_n + D^{-1}(aI_n - \beta A))(I_n + (aI_n - \beta A)D^{-1})\right)$ in §5.1. The regularization of the primal Parisi PDE regularizes the (generalized) TAP free energy at the level of coordinate curvature, whereas the current regularization ϕ regularizes the Hessian of the TAP free energy at the level of spectral invertibility.

Y satisfying $Y \succ -\gamma_Y I_n$ so that $\phi(Y)$ is the matrix with the same eigenvectors as Y but with eigenvalues transformed by ϕ.

Fix $\beta > 0$, $a \in [0,1]$, $A \in \mathrm{Sym}_n(\mathbb{R})$, $W \in \mathcal{D}_n([0,1])$, and $T \in \mathcal{D}_n(\mathbb{R})$. Denote $M := aI_n - \beta A$ and define the test functional over the diagonal sub-algebra

$$\Phi_{a,W,T}(A) := \mathrm{tr}_n[T\,\phi((I_n + WM)(I_n + MW))]\,. \tag{5.2}$$

In [DLSS26a, Lemma B.8], the corresponding function was defined as $\mathrm{tr}_n\big(T\,E_{\mathcal{D}_n}\big[D\,(a - \beta A + D)^{-2} D\big]\big)$, and we will show in Proposition 5.2(e) that these two definitions agree on the safe set defined in Definition 2.2. As in [DLSS26a, Lemma B.8], since T is diagonal, inserting $E_{\mathcal{D}_n}$ around the matrix function leaves (5.2) unchanged.

We continue to use normalized Schatten norms throughout, with $\|\cdot\|_{L^\infty} = \|\cdot\|_{\mathsf{op}}$.

To obtain dimension-independent bounds for the matrix Fréchet derivatives in the required norms, we first prove bounds on derivatives of the function $(\cdot)_+^3$ applied to matrix arguments, using the same Fréchet derivative notation defined in [DLSS26a, §B.2].

Lemma 5.1 (Derivative bounds for the cubed positive part)**.**
The matrix map $Z \mapsto Z_+^3$ is twice continuously Fréchet differentiable on $\mathrm{Sym}_n(\mathbb{R})$. For every symmetric Z with $\|Z\|_{\mathsf{op}} \leqslant 1$, symmetric directions H, K, and $p \in \{2, \infty\}$,

$$\begin{aligned} \big\|d(Z_+^3)[H]\big\|_{L^p} &\leqslant 4\,\|H\|_{L^p}\,, \\ \big\|d^2(Z_+^3)[H,K]\big\|_{L^2} &\leqslant 10\,\|H\|_{L^2}\,\|K\|_{\mathsf{op}}\,. \end{aligned} \tag{5.3}$$

Proof. For $z \neq 0$, the elementary arctangent integral gives

$$\int_0^\infty \frac{z^4}{z^2 + t^2}\,dt = |z|^3 \left[\arctan\left(\frac{t}{|z|}\right)\right]_0^\infty = \frac{\pi}{2}|z|^3.$$

The same identity holds at $z = 0$, since the integrand vanishes for $t > 0$. Thus $z_+^3 = (z^3 + |z|^3)/2$ and this generalizes to the matrix case (consider a basis where Z is diagonal) as

$$Z_+^3 = \frac{Z^3}{2} + \frac{1}{\pi}\int_0^\infty Z^4 (Z^2 + t^2 I_n)^{-1}\,dt. \tag{5.4}$$

We apply a partial fraction decomposition. For $t > 0$,

$$(Z^2 + t^2 I_n)^{-1} = \frac{(Z - itI_n)^{-1} - (Z + itI_n)^{-1}}{2it}\,.$$

Letting $S_\pm = (Z \pm itI_n)^{-1}$, the product and inverse rules and the non-commutative Hölder inequality yield

$$\big\|d[(Z^2 + t^2 I_n)^{-1}][H]\big\|_{L^p} = \left\|\frac{-S_- H S_- + S_+ H S_+}{2it}\right\|_{L^p} \leqslant t^{-3}\,\|H\|_{L^p}\,, \tag{5.5}$$

$$\big\|d^2[(Z^2 + t^2 I_n)^{-1}][H,K]\big\|_{L^2} = \left\|\frac{S_- H S_- K S_- + S_- K S_- H S_- - \dots}{2it}\right\|_{L^2} \leqslant 2t^{-4}\,\|H\|_{L^2}\,\|K\|_{\mathsf{op}}\,. \tag{5.6}$$

For $0 < t \leqslant 1$, the identity $(Z^2 + t^2 I_n)(Z^2 + t^2 I_n)^{-1} = I_n$ gives a cancellation of the integrand of (5.4):

$$\begin{aligned} Z^4 (Z^2 + t^2 I_n)^{-1} &= Z^2 - t^2 I_n + t^4 (Z^2 + t^2 I_n)^{-1}, \\ d[Z^4 (Z^2 + t^2 I_n)^{-1}][H] &= ZH + HZ + t^4 \cdot d[(Z^2 + t^2 I_n)^{-1}][H], \\ d^2[Z^4 (Z^2 + t^2 I_n)^{-1}][H,K] &= KH + HK + t^4 \cdot d^2[(Z^2 + t^2 I_n)^{-1}][H,K]. \end{aligned}$$

For $t \geqslant 1$, instead differentiate the integrand directly with the product rule, using $\big\|(Z^2 + t^2 I_n)^{-1}\big\|_{\mathsf{op}} \leqslant t^{-2}$. Combining both cases therefore gives, when $\|Z\|_{\mathsf{op}} \leqslant 1$,

$$\big\|d[Z^4 (Z^2 + t^2 I_n)^{-1}][H]\big\|_{L^p} \leqslant \begin{cases} 3\,\|H\|_{L^p}\,, & 0 < t \leqslant 1, \\ (4t^{-2} + t^{-3})\,\|H\|_{L^p}\,, & t \geqslant 1, \end{cases}$$

$$\left\|d^2[Z^4(Z^2+t^2I_n)^{-1}][H,K]\right\|_{L^2} \leqslant \begin{cases} 4\,\|H\|_{L^2}\,\|K\|_{\mathsf{op}}\,, & 0<t\leqslant 1,\\ (12t^{-2}+8t^{-3}+2t^{-4})\,\|H\|_{L^2}\,\|K\|_{\mathsf{op}}\,, & t\geqslant 1. \end{cases}$$

The integrand itself has operator norm at most $\min\{1,t^{-2}\}$. The same estimates, with larger numerical constants, hold on $\|Z\|_{\mathsf{op}} < 2$. This gives that the integrand, and its first two derivatives, have locally uniform integrable bounds. This justifies differentiation under the integral and continuity of both derivatives. Integrating these bounds and adding the product-rule bounds for $Z^3/2$ yields

$$\left\|d(Z_+^3)[H]\right\|_{L^p} \leqslant \left[\frac{3}{2}+\frac{1}{\pi}\left(3+4+\frac{1}{2}\right)\right]\|H\|_{L^p} \leqslant 4\,\|H\|_{L^p}\,,$$
$$\left\|d^2(Z_+^3)[H,K]\right\|_{L^2} \leqslant \left[3+\frac{1}{\pi}\left(4+12+4+\frac{2}{3}\right)\right]\|H\|_{L^2}\,\|K\|_{\mathsf{op}} \leqslant 10\,\|H\|_{L^2}\,\|K\|_{\mathsf{op}}\,.$$

By positive homogeneity of degree three, rescaling extends this twice continuous Fréchet differentiability to all of $\mathrm{Sym}_n(\mathbb{R})$. □

Proposition 5.2 (Regularity and block identities for the reciprocal extension)**.**
With ϕ and $\Phi_{a,W,T}$ as above, the following claims hold. All constants below are independent of the matrix dimension.

(a) Cutoff bounds. *The cutoff is well-defined and positive on $(-\gamma_Y,\infty)$. For $y\geqslant 0$, its denominator satisfies $\phi(y)^{-1}\geqslant y$ and $\phi(y)^{-1}\geqslant(\sqrt{6}-2)\gamma_Y\geqslant 4\gamma_Y/9$. Also $\phi(y)=y^{-1}$ for $y\geqslant\gamma_Y$. For every symmetric Z,*

$$\left\|\phi(Z^2)\right\|_{\mathsf{op}} \leqslant \frac{9}{4\gamma_Y}, \qquad \left\|Z\phi(Z^2)^{1/2}\right\|_{\mathsf{op}} \leqslant 1. \tag{5.7}$$

(b) Regularity of the squared-argument cutoff. *The following derivatives are of the whole map $Z\mapsto\phi(Z^2)$. This map is twice continuously Fréchet differentiable on $\mathrm{Sym}_n(\mathbb{R})$ and there is a universal constant $C>0$ such that, for all symmetric $Z,\widetilde{Z},H,K$ and $p\in\{2,\infty\}$,*

$$\left\|d[\phi(Z^2)][H]\right\|_{L^p} \leqslant C\gamma_Y^{-3/2}\,\|H\|_{L^p}\,, \tag{5.8}$$
$$\left\|d^2[\phi(Z^2)][H,K]\right\|_{L^2} \leqslant C\gamma_Y^{-2}\,\|H\|_{L^2}\,\|K\|_{\mathsf{op}}\,, \tag{5.9}$$
$$\left\|\left(d[\phi(Z^2)]-d[\phi(\widetilde{Z}^2)]\right)[H]\right\|_{L^2} \leqslant C\gamma_Y^{-2}\left\|Z-\widetilde{Z}\right\|_{\mathsf{op}}\|H\|_{L^2}\,. \tag{5.10}$$

(c) Block representation and normalization. (5.2) *has the exact representation*

$$\Phi_{a,W,T}(A) = \mathrm{tr}_{2n}\left[\begin{pmatrix}2T & 0\\ 0 & 0\end{pmatrix}\phi\left(\begin{pmatrix}0 & I_n+WM\\ I_n+MW & 0\end{pmatrix}^2\right)\right]. \tag{5.11}$$

For every symmetric direction H,

$$\left\|\begin{pmatrix}2T & 0\\ 0 & 0\end{pmatrix}\right\|_{L^2} = \sqrt{2}\,\|T\|_{L^2}\,, \qquad \left\|\begin{pmatrix}0 & WH\\ HW & 0\end{pmatrix}\right\|_{L^2} = \|WH\|_{L^2} \leqslant \|H\|_{L^2}\,. \tag{5.12}$$

(d) Affine dependence on A. *For every symmetric direction H,*

$$d_A\left[\begin{pmatrix}0 & I_n+WM\\ I_n+MW & 0\end{pmatrix}\right][H] = -\beta\begin{pmatrix}0 & WH\\ HW & 0\end{pmatrix}. \tag{5.13}$$

Its normalized Schatten 2-norm is at most $\beta\,\|H\|_{L^2}$.

(e) Agreement on the safe set. *For $W\succ 0$, set $D=W^{-1}$. If $(I_n+WM)(I_n+MW)\succeq\gamma_Y I_n$, then*

$$\phi((I_n+WM)(I_n+MW)) = ((I_n+WM)(I_n+MW))^{-1} = D(M+D)^{-2}D.$$

Proof. We prove each of the regularity bounds and the block identity and its affine dependence on A in sequence.

Part (a): cutoff bounds. For $y\geqslant 0$, the inequality $x_+^3\geqslant 3x/4-1/4$ gives

$$\phi(y)^{-1} \geqslant y+\gamma_Y\left(\frac{3}{4}\frac{\gamma_Y-y}{\gamma_Y+y}-\frac{1}{4}\right)$$

$$= (\gamma_Y + y) + \frac{3\gamma_Y^2}{2(\gamma_Y + y)} - 2\gamma_Y \geqslant (\sqrt{6} - 2)\gamma_Y \geqslant \frac{4}{9}\gamma_Y .$$

Also $\phi(y)^{-1} \geqslant y$. For $-\gamma_Y < y < 0$, its denominator is greater than $y + \gamma_Y > 0$. The bounds in (5.7) follow by applying the same bound to each eigenvalue by using the commutativity of Z and $\phi(Z^2)$. For $y \geqslant \gamma_Y$, the correction vanishes and $\phi(y) = y^{-1}$.

Part (b): regularity of the squared-argument cutoff. Write

$$\phi(Z^2)^{-1} = Z^2 + \gamma_Y \left(2\gamma_Y(\gamma_Y I_n + Z^2)^{-1} - I_n\right)_+^3 . \tag{5.14}$$

The argument of the positive part is a self-adjoint contraction. Apply (5.5) and (5.6) with $t = \sqrt{\gamma_Y}$ to obtain

$$\begin{aligned} \left\| d[2\gamma_Y(\gamma_Y I_n + Z^2)^{-1} - I_n][H] \right\|_{L^p} &\leqslant 2\gamma_Y^{-1/2} \|H\|_{L^p} , \\ \left\| d^2[2\gamma_Y(\gamma_Y I_n + Z^2)^{-1} - I_n][H,K] \right\|_{L^2} &\leqslant 4\gamma_Y^{-1} \|H\|_{L^2} \|K\|_{\mathsf{op}} . \end{aligned}$$

Apply Lemma 5.1 on the matrix $2\gamma_Y(\gamma_Y I_n + Z^2)^{-1} - I_n$, and then use the chain rule. The correction term in (5.14) consequently satisfies

$$\begin{aligned} \left\| d[\phi(Z^2)^{-1} - Z^2][H] \right\|_{L^p} &\leqslant 8\gamma_Y^{1/2} \|H\|_{L^p} , \\ \left\| d^2[\phi(Z^2)^{-1} - Z^2][H,K] \right\|_{L^2} &\leqslant 56 \|H\|_{L^2} \|K\|_{\mathsf{op}} . \end{aligned}$$

The second constant is $10 \cdot 2^2 + 4 \cdot 4 = 56$. Since $d(Z^2)[H] = ZH + HZ$ and $d^2(Z^2)[H,K] = HK + KH$, (5.7) and Hölder's inequality give

$$\begin{aligned} \left\| \phi(Z^2)^{1/2} d[\phi(Z^2)^{-1}][H] \phi(Z^2)^{1/2} \right\|_{L^p} &\leqslant 21\gamma_Y^{-1/2} \|H\|_{L^p} , \\ \left\| d^2[\phi(Z^2)^{-1}][H,K] \right\|_{L^2} &\leqslant 58 \|H\|_{L^2} \|K\|_{\mathsf{op}} . \end{aligned} \tag{5.15}$$

In the first line, the two terms from $ZH + HZ$ contribute $3\gamma_Y^{-1/2} \|H\|_{L^p}$, and the correction contributes at most $18\gamma_Y^{-1/2} \|H\|_{L^p}$, using Proposition 5.2(a) to bound the factors of $\phi(Z^2)^{1/2}$.

Inverting the denominator. Lemma 5.1, (5.14), and positivity of the denominator imply twice continuous differentiability on all of $\mathrm{Sym}_n(\mathbb{R})$. Applying the inverse rule gives

$$\begin{aligned} d[\phi(Z^2)][H] &= -\phi(Z^2) \cdot d[\phi(Z^2)^{-1}][H] \cdot \phi(Z^2), \\ d^2[\phi(Z^2)][H,K] &= \phi(Z^2) \cdot d[\phi(Z^2)^{-1}][K] \cdot \phi(Z^2) \cdot d[\phi(Z^2)^{-1}][H] \cdot \phi(Z^2) \\ &\quad + \phi(Z^2) \cdot d[\phi(Z^2)^{-1}][H] \cdot \phi(Z^2) \cdot d[\phi(Z^2)^{-1}][K] \cdot \phi(Z^2) \\ &\quad - \phi(Z^2) \cdot d^2[\phi(Z^2)^{-1}][H,K] \cdot \phi(Z^2). \end{aligned}$$

In each product containing a first derivative of the denominator, insert its two adjacent factors $\phi(Z^2)^{1/2}$ and apply (5.15). The remaining outside factors have total operator norm at most $9/(4\gamma_Y)$ by Proposition 5.2(a). In the last term, apply the second line of (5.15) directly. Thus

$$\begin{aligned} \left\| d[\phi(Z^2)][H] \right\|_{L^p} &\leqslant \frac{9}{4} 21 \gamma_Y^{-3/2} \|H\|_{L^p} , \\ \left\| d^2[\phi(Z^2)][H,K] \right\|_{L^2} &\leqslant \left(2 \cdot \frac{9}{4} 21^2 + \frac{81}{16} 58\right) \gamma_Y^{-2} \|H\|_{L^2} \|K\|_{\mathsf{op}} . \end{aligned}$$

This proves (5.8) and (5.9). The square roots here are used only to bound products after differentiation; no square root is differentiated. Finally, integrate the second derivative along the segment from $\widetilde{Z}$ to Z to obtain (5.10).

Part (c): block representation and normalization. The symmetric block matrix satisfies

$$\begin{pmatrix} 0 & I_n + WM \\ I_n + MW & 0 \end{pmatrix}^2 = \begin{pmatrix} (I_n + WM)(I_n + MW) & 0 \\ 0 & (I_n + MW)(I_n + WM) \end{pmatrix}.$$

Applying ϕ and taking the normalized trace against $\left(\begin{smallmatrix} 2T & 0 \\ 0 & 0 \end{smallmatrix}\right)$ proves (5.11). The identities in (5.12) follow by taking the normalized traces of the corresponding products with their transposes; the last inequality uses $\|W\|_{\mathsf{op}} \leqslant 1$.

Part (d): affine dependence on A. Since $M = aI_n - \beta A$, differentiation with a and W fixed gives (5.13). Its norm bound follows from (5.12).

Part (e): agreement on the safe set. The cutoff agrees with the reciprocal on the spectrum of $(I_n + WM)(I_n + MW)$ under the stated assumption. Since $I_n + WM = W(M + D)$, this product is $W(M + D)^2W$. Its invertibility implies that $M + D$ is invertible, and taking the inverse gives $D(M + D)^{-2}D$. □

5.2. **Chaining argument for fluctuations around expected value.** In [DLSS26a, §B], the uniform approximation bound on $E_{\mathcal{D}_n}[D(aI_n - \beta A + D)^{-2}D]$ was divided into two parts, with a free interpolation argument bounding its expected value over random A, then a chaining argument bounding its fluctuations around its expectation. We now tackle the chaining part of the argument.

Since the set of values of D leading to singular or near-singular $(aI_n - \beta A + D)$ depends on the chaining variable A, we cannot simply perform the same argument in a safe subset of its original domain $(-1, 1)^n$. Therefore, we adapt the argument to prove the bound for the regularized extension $\Phi_{a,W,T}(A)$ from §5.1, which is defined over all of $(-1, 1)^n$ and agrees with the original quantity on the safe set.

First, we apply the regularity estimates and block identities from Proposition 5.2 to compare the gradients of the regularized scalar functionals.

Lemma 5.3 (Mixed Lipschitzness for chaining)**.**
Fix $\beta, \gamma_Y, C_{\mathrm{op}} > 0$ and let $C_1 := 1 + \beta C_{\mathrm{op}}$. Let $A \in \mathrm{Sym}_n(\mathbb{R})$ satisfy $\|A\|_{\mathsf{op}} \leqslant C_{\mathrm{op}}$, let $a, \widetilde{a} \in [0, 1]$, and let $W, \widetilde{W} \in \mathcal{D}_n([0, 1])$. Let $T, \widetilde{T} \in \mathcal{D}_n(\mathbb{R})$ satisfy $\|T\|_{L^2}, \|\widetilde{T}\|_{L^2} \leqslant 1$. Set

$$M := aI_n - \beta A, \qquad \widetilde{M} := \widetilde{a}\, I_n - \beta A,$$

and let $\Phi_{a,W,T}$ be the functional defined in (5.2). Let $\nabla_A \Phi_{a,W,T}(A)$ be the gradient on $\mathrm{Sym}_n(\mathbb{R})$ with respect to $\langle H, K\rangle = \mathrm{tr}_n(HK)$. Then there is a universal constant $C > 0$ such that

$$\left\|\nabla_A \Phi_{a,W,T}(A) - \nabla_A \Phi_{\widetilde{a},\widetilde{W},\widetilde{T}}(A)\right\|_{L^2} \leqslant C\beta\Big[\gamma_Y^{-2}|a - \widetilde{a}| + \left(C_1\gamma_Y^{-2} + \gamma_Y^{-3/2}\right)\left\|W - \widetilde{W}\right\|_{\mathsf{op}} + \gamma_Y^{-3/2}\left\|T - \widetilde{T}\right\|_{L^2}\Big]. \tag{5.16}$$

Consequently,

$$\|\nabla_A \Phi_{a,W,T}(A)\|_{L^2} \leqslant C\beta\gamma_Y^{-3/2}\|T\|_{L^2}. \tag{5.17}$$

Proof. We will use the block representation of the test function. We use the fact that the block representation is an affine function of A, whose derivative is easy to compute. Extending this to the full test function is a matter of chain rule, followed by combining standard Fréchet derivative computations, Cauchy–Schwarz and the established regularity bounds.

Differences of the block matrix. With A fixed, let Δ denote the difference between (a, W, T) and $(\widetilde{a}, \widetilde{W}, \widetilde{T})$, consistently across all its uses, so that $\Delta M = (\Delta a)I_n$. Note also that $\|M\|_{\mathsf{op}}, \|\widetilde{M}\|_{\mathsf{op}} \leqslant C_1$. For a symmetric direction H, (5.13) and (5.12) bound the normalized Schatten 2-norm of the A-derivative by $\beta\|H\|_{L^2}$. Then the discrete product rule gives

$$\left\|\Delta\begin{pmatrix} 0 & I_n + WM \\ I_n + MW & 0 \end{pmatrix}\right\|_{\mathsf{op}} = \left\|WM - \widetilde{W}\widetilde{M}\right\|_{\mathsf{op}} \leqslant |\Delta a| + C_1\|\Delta W\|_{\mathsf{op}}, \tag{5.18}$$

$$\left\|\Delta\left(d_A\left[\begin{pmatrix} 0 & I_n + WM \\ I_n + MW & 0 \end{pmatrix}\right][H]\right)\right\|_{L^2} \leqslant \beta\|\Delta W\|_{\mathsf{op}}\|H\|_{L^2}. \tag{5.19}$$

Here the off-diagonal blocks are transposes of one another. In particular, the A-derivative does not change when only a changes.

The scalar functional. Apply Proposition 5.2(b) in dimension $2n$, with its matrix argument Z replaced by

$$Z \leftarrow \begin{pmatrix} 0 & I_n + WM \\ I_n + MW & 0 \end{pmatrix}.$$

By the chain rule,

$$d_A\Phi_{a,W,T}(A)[H] = -\beta \operatorname{tr}_{2n}\left[\begin{pmatrix} 2T & 0 \\ 0 & 0 \end{pmatrix} \cdot d[\phi(Z^2)]\left[\begin{pmatrix} 0 & WH \\ HW & 0 \end{pmatrix}\right]\right].$$

By Cauchy–Schwarz for tr_{2n}, (5.8) in dimension $2n$, (5.12), (5.13), $\|W\|_{\mathsf{op}} \leqslant 1$, and absorbing a factor of $\sqrt{2}$ into C,

$$\begin{aligned} |d_A\Phi_{a,W,T}(A)[H]| &\leqslant C\beta\gamma_Y^{-3/2} \left\|\begin{pmatrix} 2T & 0 \\ 0 & 0 \end{pmatrix}\right\|_{L^2} \left\|\begin{pmatrix} 0 & WH \\ HW & 0 \end{pmatrix}\right\|_{L^2} \\ &= \sqrt{2}\, C\beta\gamma_Y^{-3/2} \|T\|_{L^2} \|WH\|_{L^2} \\ &\leqslant C\beta\gamma_Y^{-3/2} \|T\|_{L^2} \|H\|_{L^2}. \end{aligned}$$

Now keep A and H fixed. In the chain-rule expression for $d_A\Phi_{a,W,T}(A)[H]$, first change T to $\widetilde{T}$, then change the direction supplied to the derivative from the block matrix containing W to that containing $\widetilde{W}$, and finally change the evaluation point of $d[\phi(Z^2)]$ from the block matrix associated with (a, W) to that associated with $(\widetilde{a}, \widetilde{W})$. The discrete product rule and Cauchy–Schwarz for tr_{2n} then give

$$\begin{aligned} &|\Delta(d_A\Phi_{a,W,T}(A)[H])| \\ &\leqslant \sqrt{2}\,\beta \|\Delta T\|_{L^2} \left\|d[\phi(Z^2)]\left[\begin{pmatrix} 0 & WH \\ HW & 0 \end{pmatrix}\right]\right\|_{L^2} \\ &\quad + \sqrt{2}\,\beta \|\widetilde{T}\|_{L^2} \left\|d[\phi(Z^2)]\left[\begin{pmatrix} 0 & (\Delta W)H \\ H(\Delta W) & 0 \end{pmatrix}\right]\right\|_{L^2} \\ &\quad + \sqrt{2}\,\beta \|\widetilde{T}\|_{L^2} \left\|[\Delta(d[\phi(Z^2)])]\left[\begin{pmatrix} 0 & \widetilde{W}H \\ H\widetilde{W} & 0 \end{pmatrix}\right]\right\|_{L^2}. \end{aligned}$$

The $\sqrt{2}$ factors come from $\|\left(\begin{smallmatrix} 2T & 0 \\ 0 & 0 \end{smallmatrix}\right)\|_{L^2} = \sqrt{2}\|T\|_{L^2}$.

For the first and second terms, apply (5.8). For the third term, apply (5.10). The required bounds on the block matrices are

$$\begin{aligned} \left\|\begin{pmatrix} 0 & WH \\ HW & 0 \end{pmatrix}\right\|_{L^2} &\leqslant \|W\|_{\mathsf{op}} \|H\|_{L^2}, \\ \left\|\begin{pmatrix} 0 & (\Delta W)H \\ H(\Delta W) & 0 \end{pmatrix}\right\|_{L^2} &\leqslant \|\Delta W\|_{\mathsf{op}} \|H\|_{L^2}, \\ \left\|\Delta\begin{pmatrix} 0 & I_n + WM \\ I_n + MW & 0 \end{pmatrix}\right\|_{\mathsf{op}} &\leqslant |\Delta a| + C_1 \|\Delta W\|_{\mathsf{op}}, \end{aligned}$$

by (5.12), (5.19), and (5.18). Substituting these estimates into the three terms above and absorbing $\sqrt{2}$ into the universal constant gives

$$\begin{aligned} |\Delta(d_A\Phi_{a,W,T}(A)[H])| \leqslant C\beta\Big[&\gamma_Y^{-3/2} \|W\|_{\mathsf{op}} \|\Delta T\|_{L^2} \\ &+ \gamma_Y^{-3/2} \|\widetilde{T}\|_{L^2} \|\Delta W\|_{\mathsf{op}} \\ &+ \gamma_Y^{-2} \|\widetilde{W}\|_{\mathsf{op}} \|\widetilde{T}\|_{L^2} \left(|\Delta a| + C_1 \|\Delta W\|_{\mathsf{op}}\right)\Big] \|H\|_{L^2}. \end{aligned}$$

Finally, using $\|\widetilde{T}\|_{L^2} \leqslant 1$ and $\|W\|_{\mathsf{op}} \leqslant 1$ and $\|\widetilde{W}\|_{\mathsf{op}} \leqslant 1$, we obtain

$$|\Delta(d_A\Phi_{a,W,T}(A)[H])| \leqslant C\beta\Big[\gamma_Y^{-3/2} \|\Delta T\|_{L^2} + \gamma_Y^{-2}\left(|\Delta a| + C_1 \|\Delta W\|_{\mathsf{op}}\right) + \gamma_Y^{-3/2} \|\Delta W\|_{\mathsf{op}}\Big] \|H\|_{L^2}.$$

Taking the supremum over symmetric H with $\|H\|_{L^2} = 1$ proves (5.16) and then (5.17) follows from the special case when $(\widetilde{a}, \widetilde{W}, \widetilde{T}) = (a, W, 0)$, so that $\nabla_A\Phi_{\widetilde{a},\widetilde{W},\widetilde{T}}(A) = 0$. □

We now follow the proof of [DLSS26a, Theorem B.9], using Lemma 5.3 in place of its mixed-Lipschitz estimate.

Theorem 5.4 (Chaining bound for fluctuations of the reciprocal extension $\Phi_{a,W,T}$)**.** *Fix $\beta, \gamma_Y, C_{\mathrm{op}} > 0$, let $C_1 := 1 + \beta C_{\mathrm{op}}$, and let $A \sim \mathsf{GOE}(n)$. Let $\Pi_{C_{\mathrm{op}}} : \mathrm{Sym}_n(\mathbb{R}) \to \mathrm{Sym}_n(\mathbb{R})$ be the metric projection with respect to $\|\cdot\|_{L^2}$, onto the closed convex set $\{M : \|M\|_{\mathsf{op}} \leqslant C_{\mathrm{op}}\}$, and set $\overline{A} := \Pi_{C_{\mathrm{op}}}(A)$. For the functional in (5.2), define*

$$\mathcal{J} := [0,1] \times \mathcal{D}_n([0,1]) \times \{T \in \mathcal{D}_n(\mathbb{R}) : \|T\|_{L^2} = 1\}, \tag{5.20}$$

$$Z_{a,W,T} := \Phi_{a,W,T}(\overline{A}) - \mathop{\mathbb{E}}_{A}[\Phi_{a,W,T}(\overline{A})]. \tag{5.21}$$

Then there are universal constants $C_{\mathrm{ch}}, C_{\mathrm{tail}} > 0$ such that

$$\mathop{\mathbb{E}}_{A}\left[\sup_{(a,W,T)\in\mathcal{J}} |Z_{a,W,T}|\right] \leqslant \frac{C_{\mathrm{ch}}\beta}{\sqrt{n}}\left((1+C_1)\gamma_Y^{-2} + \gamma_Y^{-3/2}\right), \tag{5.22}$$

and, for every $p \in (0,1)$,

$$\mathop{\mathbb{P}}_{A}\left\{\sup_{(a,W,T)\in\mathcal{J}} |Z_{a,W,T}| \leqslant \frac{C_{\mathrm{ch}}\beta}{\sqrt{n}}\left((1+C_1)\gamma_Y^{-2} + \gamma_Y^{-3/2}\right) + \frac{C_{\mathrm{tail}}\beta\gamma_Y^{-3/2}}{n}\sqrt{\log\frac{2}{p}}\right\} \geqslant 1 - p. \tag{5.23}$$

Proof. The proof will first utilize Lemma 5.3 on the test function induced by the regularized resolvent. After that, the proof is standard by invoking sub-Gaussian concentration via chaining and computing an upper bound on Talagrand's-γ_2 functional using the Dudley entropy integral.

Increment metric. Define the weights

$$\kappa_a := C\gamma_Y^{-2}, \qquad \kappa_W := C\big(C_1\gamma_Y^{-2} + \gamma_Y^{-3/2}\big), \qquad \kappa_T := C\gamma_Y^{-3/2},$$

with a common universal prefactor large enough for (5.16) and (5.17). For $\theta = (a, W, T)$ and $\widetilde{\theta} = (\widetilde{a}, \widetilde{W}, \widetilde{T})$ in $\mathcal{J}$, set

$$\rho(\theta, \widetilde{\theta}) := \kappa_a|a - \widetilde{a}| + \kappa_W \left\|W - \widetilde{W}\right\|_{\mathsf{op}} + \kappa_T \left\|T - \widetilde{T}\right\|_{L^2}. \tag{5.24}$$

By Lemma 5.3,

$$\left\|\nabla_A\Phi_\theta(A) - \nabla_A\Phi_{\widetilde{\theta}}(A)\right\|_{L^2} \leqslant \beta\rho(\theta,\widetilde{\theta}) \qquad (\|A\|_{\mathsf{op}} \leqslant C_{\mathrm{op}}).$$

Integrating along line segments in this convex ball shows that $\Phi_\theta - \Phi_{\widetilde{\theta}}$ is $\beta\rho(\theta,\widetilde{\theta})$-Lipschitz there.

Sub-Gaussian increments. The preceding Lipschitz bound, the 1-Lipschitzness of $\Pi_{C_{\mathrm{op}}}$, and GOE Herbst concentration [DLSS26a, Lemma B.7] give, exactly as in [DLSS26a, proof of Theorem B.9],

$$\mathbb{P}\{|Z_\theta - Z_{\widetilde{\theta}}| > u\} \leqslant 2\exp\left(-\frac{n^2u^2}{4\beta^2\rho(\theta,\widetilde{\theta})^2}\right) = 2\exp\left(-\frac{u^2}{2d(\theta,\widetilde{\theta})^2}\right), \qquad d := \frac{\sqrt{2}\beta}{n}\rho, \tag{5.25}$$

for $u > 0$ and $\theta \neq \widetilde{\theta}$.

Chaining bound. Continuity of $\Phi_\theta(\overline{A})$ in θ and its uniform bound $|\Phi_\theta(\overline{A})| \leqslant 9/(4\gamma_Y)$ from Proposition 5.2(a) imply, by dominated convergence, that the centered process has continuous sample paths on the compact set $\mathcal{J}$ and is therefore separable. Since $Z_{a,W,-T} = -Z_{a,W,T}$, Talagrand's chaining [DLSS26a, Theorem B.6] gives

$$\mathop{\mathbb{E}}_{A}\sup_{\theta\in\mathcal{J}}|Z_\theta| = \mathop{\mathbb{E}}_{A}\sup_{\theta\in\mathcal{J}} Z_\theta \leqslant C\gamma_2(\mathcal{J}, d).$$

Entropy integral. By [DLSS26a, Theorem B.5] and $d = (\sqrt{2}\beta/n)\rho$,

$$\gamma_2(\mathcal{J}, d) \leqslant C\frac{\sqrt{2}\beta}{n}\int_0^{\mathrm{diam}(\mathcal{J},\rho)} \sqrt{\log N(\mathcal{J}, \rho, u)}\, du, \tag{5.26}$$

where N denotes the covering number.

To bound $N(\mathcal{J}, \rho, u)$, note that $C_1 \geqslant 1$ implies $\kappa_a + \kappa_T \leqslant \kappa_W$, and hence

$$\mathrm{diam}(\mathcal{J}, \rho) \leqslant \kappa_a + \kappa_W + 2\kappa_T \leqslant 3\kappa_W.$$

For $0 < u \leqslant \mathrm{diam}(\mathcal{J}, \rho)$, cover each parameter factor at radius $u/(3\kappa_W) \leqslant 1$ in its respective norm. Since ρ is a sum metric and each weight is at most κ_W, the resulting product cover has ρ-radius at most u. The same

small-radius volumetric bounds as in [DLSS26a, proof of Theorem B.9] imply

$$\log N(\mathcal{J}, \rho, u) \leqslant (2n+1)\log\left(\frac{9\kappa_W}{u}\right),$$

for $0 < u \leqslant \operatorname{diam}(\mathcal{J}, \rho)$. Consequently, extending the integral of this nonnegative upper bound to $9\kappa_W$ and using the same substitution $u = 9\kappa_W e^{-t}$ as in [DLSS26a, proof of Theorem B.9], we obtain

$$\begin{aligned}\int_0^{\operatorname{diam}(\mathcal{J},\rho)} \sqrt{\log N(\mathcal{J}, \rho, u)}\, du &\leqslant \sqrt{2n+1} \int_0^{9\kappa_W} \sqrt{\log(9\kappa_W/u)}\, du \\ &= \frac{9\sqrt{\pi}}{2}\kappa_W\sqrt{2n+1}.\end{aligned}$$

Substituting this into (5.26) yields

$$\gamma_2(\mathcal{J}, d) \leqslant C\frac{\beta\kappa_W}{\sqrt{n}} \leqslant \frac{C\beta}{\sqrt{n}}\left(C_1\gamma_Y^{-2} + \gamma_Y^{-3/2}\right),$$

where the last step uses the definition of κ_W. Together with the preceding chaining bound, this proves (5.22).

High-probability bound. By (5.17) and the same line-segment and projection argument, each $A \mapsto \Phi_\theta(\overline{A})$ is globally $\beta\kappa_T$-Lipschitz. Centering preserves this constant, and so does taking the supremum. Thus

$$\mathcal{Z}(A) := \sup_{\theta\in\mathcal{J}} Z_\theta = \sup_{\theta\in\mathcal{J}} |Z_\theta|$$

is $\beta\kappa_T$-Lipschitz. As in [DLSS26a, proof of Theorem B.9], applying GOE Herbst concentration [DLSS26a, Lemma B.7] again to this supremum gives

$$\mathbb{P}\left\{\mathcal{Z}(A) - \mathop{\mathbb{E}}_{A}[\mathcal{Z}(A)] > u\right\} \leqslant 2\exp\left(-\frac{n^2u^2}{4\beta^2\kappa_T^2}\right).$$

Taking $u = 2\beta\kappa_T n^{-1}\sqrt{\log(2/p)}$ and using (5.22) proves (5.23). □

5.3. **Expected value from free probability.** We compare the unconditional expectation of the reciprocal extension with its free counterpart. Using the block representation from Proposition 5.2(c), we compare GOE to GUE and then apply the matrix-valued resolvent estimate of [BBv23, Theorem 2.8] in dimension $2n$, without a growing identity amplification. Smooth envelopes then recover the same reciprocal extension[10].

We retain the spectral-cutoff γ_Y, regularized resolvent ϕ, and diagonal test functional $\Phi_{a,W,T}$ from (5.1) and (5.2). Let S be a semicircular element free from $(M_n(\mathbb{C}), \operatorname{tr}_n)$ in $(\mathcal{M}_n, \tau_n)$, and let E_{M_n} be the trace-preserving conditional expectation onto $M_n(\mathbb{C})$. Fix deterministic $a \in [0,1]$, $\beta > 0$, and $W \in \mathcal{D}_n([0,1])$. The block identities and reciprocal agreement in Proposition 5.2(c), Proposition 5.2(d), and Proposition 5.2(e) are algebraic and hold for Hermitian matrices and S. We retain the normalized Schatten norms from the preliminaries, except for the explicitly unnormalized Frobenius norm in the covariance-parameter calculation.

We introduce the GUE as an intermediate ensemble because its coefficient-defined free model has exactly the M_n-valued law of S, whereas the GOE covariance contains an additional transpose term of order $1/n$. This permits [BBv23, Theorem 2.8] to compare directly with the free limit, while the covariance correction is handled by the GOE–GUE interpolation. This approach is simpler than the previous approach adopted by the authors in [DLSS26a, §B.6] where the interpolation for $\sqrt{t}A \otimes I_k + \sqrt{1-t}B$ to its corresponding free counter-part encounters a covariance structure that cannot be tightly analyzed by the techniques in [BBv23, vH25].

Lemma 5.5 (Comparison of GOE block resolvents with GUE free limit)**.**
Let $A \sim \mathsf{GOE}(n)$ and $A_{\mathrm{GUE}} \sim \mathsf{GUE}(n)$ independently. For $t \in [0,1]$ and $z \in \mathbb{C} \setminus \mathbb{R}$, set

$$A_t = \sqrt{t}\, A + \sqrt{1-t}\, A_{\mathrm{GUE}},$$

and

$$R_t(z) = \left[zI_{2n} - \begin{pmatrix} 0 & I_n + D^{-1}(aI_n - \beta A_t) \\ I_n + (aI_n - \beta A_t)D^{-1} & 0 \end{pmatrix}\right]^{-1}.$$

[10] This replaces the previous approach of the authors in [DLSS26a, §B.6] of taking a real-axis inverse-square limit - see Remark 5.6.

For $b = \Im z \neq 0$, we have

$$\|\mathbb{E}\, R_1(z) - \mathbb{E}\, R_0(z)\|_{\mathsf{op}} \leqslant \frac{2\beta^2}{n|b|^3}, \tag{5.27}$$

$$\left\|\mathbb{E}\, R_0(z) - (\mathrm{id}_2 \otimes E_{M_n})\left[\left(zI_{2n} - \begin{pmatrix} 0 & I_n + D^{-1}(aI_n - \beta S) \\ I_n + (aI_n - \beta S)D^{-1} & 0 \end{pmatrix}\right)^{-1}\right]\right\|_{\mathsf{op}} \leqslant \frac{2\beta^4}{n|b|^5}. \tag{5.28}$$

Proof. We use the characterization of the $M_n(\mathbb{C})$-valued covariances η_1 and η_2 of the $\mathsf{GOE}(n)$ and $\mathsf{GUE}(n)$ ensembles [JLNP25, Remark 1.6] to identify the difference between them, and then invoke [JLNP25, Proposition 1.8] and [BBv23, Theorem 2.8] to get quantitative strong convergence of A_{GUE} to its $(M_n(\mathbb{C}), \eta)$-semicircular limit. After that, one can use the characterization of $R_t(z)$ as an affine function of A to compute its time derivative, and apply the standard operator-norm bound on the resolvent to compute the error between the $\mathsf{GOE}(n)$ and $\mathsf{GUE}(n)$ ensemble.

The free $\mathsf{GUE}(n)$ model. For every $H \in M_n(\mathbb{C})$, the normalized GUE covariance is

$$\eta_2 = \mathbb{E}[A_{\mathrm{GUE}} H A_{\mathrm{GUE}}] = \frac{1}{n}\sum_{i,j=1}^{n} E_{ij} H E_{ji} = \mathrm{tr}_n(H) I_n, \tag{5.29}$$

where E_{ij} are the standard basis elements for $M_n(\mathbb{C})$. The Hermitian coefficient matrices multiplying independent standard real Gaussians are

$$\mathcal{A} := \left\{\frac{E_{ii}}{\sqrt{n}}\right\}_{1\leqslant i\leqslant n} \bigcup \left\{\frac{E_{ij}+E_{ji}}{\sqrt{2n}}\right\}_{1\leqslant i<j\leqslant n} \bigcup \left\{\frac{i(E_{ij}-E_{ji})}{\sqrt{2n}}\right\}_{1\leqslant i<j\leqslant n}.$$

Replacing the standard Gaussians in front of the coefficient matrices by copies of freely independent standard semicirculars gives an M_n-valued semicircular variable with covariance η_2 (5.29) that describes the free limit via [JLNP25, Proposition 1.8]. By the proof of [JLNP25, Proposition 1.8], all cumulants of $M_n(\mathbb{C})$ but the first two vanish under the non-commutative conditional expectation $\mathrm{id} \otimes \tau$. The affine (in A) block expression of $R_t(z)$ preserves this agreement [MS17, Proposition 9.3.13]. The block resolvents consequently agree for sufficiently large $|z|$ by their moment series, and then on both half-planes by analytic continuation. This identifies the coefficient-defined free model in [BBv23, Theorem 2.8] exactly with the free resolvent in (5.28).

For the block affine representation at $t = 0$, the relevant coefficients vis-a-vis the representation in [BBv23, (1.1)] are $A_0 = \begin{pmatrix} 0 & I_n + aD^{-1} \\ I_n + aD^{-1} & 0 \end{pmatrix}$ and $A_{ij} := -\beta \begin{pmatrix} 0 & D^{-1}\mathcal{A}_{ij} \\ \mathcal{A}_{ij}D^{-1} & 0 \end{pmatrix}$. For these matrices, the parameters $\sigma(A_{\mathrm{GUE}})$ and $v(A_{\mathrm{GUE}})$ can be straightforwardly bounded using the fact that $\left\|D^{-1}\right\|_{\mathsf{op}} \leqslant 1$ along with the operator covariance calculation for η_2 in (5.29) with $H = D^{-2}$ and $H = I_n$ to be

$$\begin{aligned}\sigma^2(A_{\mathrm{GUE}}) &= \left\|\sum_{\mathcal{A}_{ij}\in\mathcal{A}} A_{ij}^2\right\|_{\mathsf{op}} = \beta^2 \left\|\sum_{ij}\begin{pmatrix} D^{-1}\mathcal{A}_{ij}^2 D^{-1} & 0 \\ 0 & \mathcal{A}_{ij}D^{-2}\mathcal{A}_{ij}\end{pmatrix}\right\|_{\mathsf{op}} \\ &= \beta^2 \left\|\begin{pmatrix} D^{-2} & 0 \\ 0 & \mathrm{tr}_n[D^{-2}] I_n\end{pmatrix}\right\|_{\mathsf{op}} \leqslant_{\|D^{-1}\|_{\mathsf{op}}\leqslant 1, \mathrm{tr}_n[D^{-1}]\leqslant 1} \beta^2,\end{aligned}$$

and

$$\begin{aligned} v^2(A_{\mathrm{GUE}}) &= \|\mathbf{Cov}(A_{\mathrm{GUE}})\|_{\mathsf{op}} = \sup_{M\in M_n(\mathbb{C}), \|M\|_F^2=1} \sum_{ij} \left|\mathsf{Tr}\left[A_{ij}M\right]\right|^2 \\ &= \beta^2 \sup_{M\in M_n(\mathbb{C}), \|M\|_F^2=1} \sum_{ij} \left|\mathsf{Tr}\left[\begin{pmatrix} 0 & D^{-1}\mathcal{A}_{ij} \\ \mathcal{A}_{ij}D^{-1} & 0\end{pmatrix}\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22}\end{pmatrix}\right]\right|^2 \\ &= \beta^2 \sup_{M\in M_n(\mathbb{C}), \|M\|_F^2=1} \sum_{ij} \left|\mathsf{Tr}\left[D^{-1}\mathcal{A}_{ij}M_{21}\right] + \mathsf{Tr}\left[A_{ij}D^{-1}M_{12}\right]\right|^2 \\ &\leqslant_{\sum_{ij}|\mathsf{Tr}[\mathcal{A}_{ij}H]|^2 = 1/n\|H\|_F^2, \|M_{12}\|_F^2\leqslant 1, \|M_{21}\|_F^2\leqslant 1} \frac{2\beta^2}{n}\left\|D^{-1}\right\|_{\mathsf{op}}^2 \leqslant \frac{2\beta^2}{n}. \end{aligned}$$

Noting that $\widetilde{v}^4(A_{\mathrm{GUE}}) = v^2(A_{\mathrm{GUE}})\sigma^2(A_{\mathrm{GUE}}) \leqslant \frac{2\beta^4}{n}$ and invoking [BBv23, Theorem 2.8] proves (5.28).

The GOE–GUE correction. The underlying entry covariances, without complex conjugation, differ by

$$\mathbb{E}[A_{ij}A_{kl}] - \mathbb{E}[(A_{\mathrm{GUE}})_{ij}(A_{\mathrm{GUE}})_{kl}] = \frac{1}{n}\delta_{ik}\delta_{jl}.$$

Apply the Gaussian covariance interpolation lemma [BBv23, Lemma 4.11] entrywise to every two Gaussians in $R_t(z)$, in real Hermitian coordinates. The variance differences in the directions E_{ii}, $E_{ij}+E_{ji}$, and $i(E_{ij}-E_{ji})$ are $1/n$, $1/(2n)$, and $-1/(2n)$. Combining the latter two directions gives the matrix-unit contraction below. By the affine dependence on A provided by Proposition 5.2(d), the two second-derivative terms from the inverse rule cancel the factor $1/2$ in covariance interpolation. Consequently,

$$\frac{d}{dt}\,\mathbb{E}\,R_t(z) = \frac{\beta^2}{n}\,\mathbb{E}\left[R_t(z)\left\{\sum_{i,j}\begin{pmatrix}0 & D^{-1}E_{ij}\\ E_{ij}D^{-1} & 0\end{pmatrix}R_t(z)\begin{pmatrix}0 & D^{-1}E_{ij}\\ E_{ij}D^{-1} & 0\end{pmatrix}\right\}R_t(z)\right]. \tag{5.30}$$

The integrand is well-defined with bounded derivatives for fixed $b \neq 0$. Using $\sum_{i,j} E_{ij}HE_{ij} = H^{\mathsf{T}}$, the matrix in braces equals

$$\begin{pmatrix}D^{-1}(R_t)_{22}^{\mathsf{T}}D^{-1} & (D^{-1})^2(R_t)_{21}^{\mathsf{T}}\\ (R_t)_{12}^{\mathsf{T}}(D^{-1})^2 & D^{-1}(R_t)_{11}^{\mathsf{T}}D^{-1}\end{pmatrix}.$$

Transposition preserves operator norm, and each block has norm at most $\|R_t\|_{\mathsf{op}}$. A 2×2 block matrix has operator norm at most twice its largest block norm, and since $\|R_t(z)\|_{\mathsf{op}} \leqslant |b|^{-1}$, we have

$$\left\|\frac{d}{dt}\,\mathbb{E}\,R_t(z)\right\|_{\mathsf{op}} \leqslant \frac{2\beta^2}{n}\,\mathbb{E}\,\|R_t(z)\|_{\mathsf{op}}^3 \leqslant \frac{2\beta^2}{n|b|^3}.$$

Integrating $R_z(t)$ over $t \in [0,1]$ gives (5.27). The lower-half-plane version of the GUE–free estimate also follows from the upper-half-plane version by taking adjoints. □

Remark 5.6. The application of [BBv23, Theorem 2.8] and [BBv23, Lemma 5.11] was obstructed in the previous argument of [DLSS26a, §B.6] by the use of $A \otimes I_k$, which has $v(A \otimes I_k) = \sqrt{k}\,v(A)$ by [BBv23, Lemma 8.1]. In the present approach, using the regularizer ϕ applied to the two-block resolvent, one only incurs a factor 2 in v^2. While the previous tensor interpolation argument could've been adapted to this new regularized version of the functional $\Phi_{a,W,T}(A)$, this simplification allows for a much shorter argument.

The lemma below will be invoked to justify comparison between the application of ϕ to the linear block resolvent and the ideal free limit, by using a smooth envelope that will agree under Proposition 5.2(e) and will be shown to have negligible error outside it.

Lemma 5.7 (Smooth test transfer)**.**
Under the hypotheses of Lemma 5.5, every real $f \in L^1(\mathbb{R})$ that is six times weakly differentiable, with all derivatives $f^{(j)} \in L^1(\mathbb{R})$, satisfies

$$\begin{aligned}\Bigg\|\mathop{\mathbb{E}}_{A} f\left(\gamma_Y^{-1/2}\begin{pmatrix}0 & I_n + D^{-1}(aI_n - \beta A)\\ I_n + (aI_n - \beta A)D^{-1} & 0\end{pmatrix}\right)\\ -(\mathrm{id}_2 \otimes E_{M_n})\left[f\left(\gamma_Y^{-1/2}\begin{pmatrix}0 & I_n + D^{-1}(aI_n - \beta S)\\ I_n + (aI_n - \beta S)D^{-1} & 0\end{pmatrix}\right)\right]\Bigg\|_{\mathrm{op}}\\ \leqslant \frac{C}{n}\left[\beta^2\gamma_Y^{-1}\sum_{j=0}^{4}\|f^{(j)}\|_{L^1(\mathbb{R})} + \beta^4\gamma_Y^{-2}\sum_{j=0}^{6}\|f^{(j)}\|_{L^1(\mathbb{R})}\right],\end{aligned} \tag{5.31}$$

where $C > 0$ is a universal constant.

Proof. For a self-adjoint operator K, scaling its resolvent gives

$$(zI - \gamma_Y^{-1/2}K)^{-1} = \sqrt{\gamma_Y}\,(\sqrt{\gamma_Y}zI - K)^{-1}.$$

Therefore, the two resolvent bounds from Lemma 5.5 become $2\beta^2/(n\gamma_Y b^3)$ and $2\beta^4/(n\gamma_Y^2 b^5)$ for the block matrices rescaled by the spectral cutoff γ_Y. Now, choose any $u \in \mathcal{S}_{\mathbb{C}}^{2n-1}(1)$ and consider the (average) u-weighted empirical eigendensities with A and A_{GUE}

$$u^{\mathsf{T}}\left(\mathop{\mathbb{E}}_{A}[R_1(z)] - \mathop{\mathbb{E}}_{\mathrm{GUE}}[R_0(z)]\right)u = g_{\mu_A,u} - g_{\mu_{A_{\mathrm{GUE}},u}} \leqslant_{u^{\mathsf{T}}Cu \leqslant \|C\|_{\mathsf{op}}} \left\|\mathop{\mathbb{E}}_{A}[R_1(z)] - \mathop{\mathbb{E}}_{\mathrm{GUE}}[R_0(z)]\right\|_{\mathsf{op}} \leqslant_{(5.27)} \frac{2\beta^2}{nb^3},$$

and the u-weighted spectral measure with S and A_{GUE}

$$u^{\mathsf{T}}\Big(\mathop{\mathbb{E}}_{\mathrm{GUE}}[R_0(z)] - (\mathrm{id}_2 \otimes E_{M_n})R_S(z)\Big)u = g_{\mu_{A_{\mathrm{GUE},u}}} - g_{\mu_{S,u}} \leqslant \Big\|\mathop{\mathbb{E}}_{\mathrm{GUE}}[R_0(z)] - (\mathrm{id}_2 \otimes E_{M_n})R_S(z)\Big\|_{\mathsf{op}} \leqslant_{(5.28)} \frac{2\beta^4}{nb^5},$$

where $g_{\mu_{A_{\mathrm{GUE},u}}}, g_{\mu_{A_u}}, g_{\mu_{S,u}}$ denote the Cauchy–Stieltjes transforms for the subscripted operators with measure given by the empirical eigendensity reweighed by u[11]. By [BBv23, Lemma 5.11], a Stieltjes-transform difference bounded by Kb^{-m} gives a smooth-test difference at most $C_m K \sum_{j=0}^{m+1} \|f^{(j)}\|_1$. We apply this twice—once to the GOE–GUE pair with $m = 3$, and the second time to the GUE–free pair with $m = 5$. We conclude with a triangle inequality, followed by taking the supremum over unit-vectors u, which proves (5.31). □

Proposition 5.8 (Expectation comparison for the reciprocal extension)**.**
Fix $D \in \mathcal{D}_n(\mathbb{R})$ with $D \geq I_n$ and let $a \in [0,1]$, $\beta > 0$, $\gamma_Y > 0$. Suppose that the spectral gap condition in the free limit

$$(I_n + D^{-1}(aI_n - \beta S))(I_n + (aI_n - \beta S)D^{-1}) \geq 4\gamma_Y I_n \tag{5.32}$$

holds in $\mathcal{M}_n$. Then

$$\begin{aligned}\Big\|\mathop{\mathbb{E}}_{A}\phi\big((I_n + D^{-1}(aI_n - \beta A))(I_n + (aI_n - \beta A)D^{-1})\big) \\ - E_{M_n}\big[D(aI_n - \beta S + D)^{-2}D\big]\Big\|_{\mathrm{op}} \leqslant \frac{C}{n}\left(\beta^2\gamma_Y^{-2} + \beta^4\gamma_Y^{-3}\right).\end{aligned} \tag{5.33}$$

In particular, for every diagonal T,

$$\Big|\mathop{\mathbb{E}}_{A}\Phi_{a,W,T}(A) - \tau_n\big[TD(aI_n - \beta S + D)^{-2}D\big]\Big| \leqslant \frac{C}{n}\left(\beta^2\gamma_Y^{-2} + \beta^4\gamma_Y^{-3}\right)\|T\|_{L^2}, \tag{5.34}$$

where $W := D^{-1}$.

Proof. The main goal of the proof is to "envelope" (or sandwich) the function ϕ inside its range by two functions that satisfy the differentiability requirements of Lemma 5.7. We then show that both the enveloping functions will lead to the same free limit using the spectral-gap condition in (5.32). After that, it is a simple matter of invoking Lemma 5.7 on the sandwiching functions to "lift" them to sandwich the operator-valued function.

Smooth envelopes. To apply Lemma 5.7, which requires derivatives through order six, we sandwich the reciprocal extension $\sqrt{\gamma_Y}x$ by smooth "envelopes". Basic algebra shows that the rescaled function

$$x \longmapsto \gamma_Y\phi\left((\sqrt{\gamma_Y}x)^2\right) = \left(x^2 + \left(\frac{(1-x^2)_+}{1+x^2}\right)^3\right)^{-1}$$

is even and independent of γ_Y. By Proposition 5.2(a), the image of $\gamma_T\phi(\cdot)$ lies in $(0, 9/4]$ and equals x^{-2} for $|x| \geqslant 1$. We set $f_-(x)$ to be a smooth function that is equal to 0 on $[-1,1]$ and to x^{-2} for $|x| \geqslant 2$, and is the lower sandwich for $\gamma_Y\phi(\cdot)$. Similarly, $f_+(x)$ is a smooth function that is equal to $9/4$ on $[-1,1]$ and to x^{-2} for $|x| \geqslant 2$, and is the upper sandwich for $\gamma_Y\phi(\cdot)$. On $1 < |x| < 2$, interpolate smoothly between the respective inner formula and x^{-2} using a weight in $[0,1]$ that is constant near the endpoints[12]. The function $f_-(x)$ is at most x^{-2} there, and the function $f_+(x)$ is at least x^{-2}. These two functions, therefore, sandwich $\gamma_Y\phi(\gamma_Y x^2)$ on $\mathbb{R}$ while being even and smooth. Since $[-2,2]$ is a closed set on which both functions are bounded, and outside that they are equal to x^{-2}, both these functions are integrable. Furthermore, all their derivatives exist, are 0 inside $[-1,1]$, are some bounded smooth function in $(-2,-1] \cup [1,2)$, and are equal to $f^{(k)} = d/dx^k x^{-2} = \prod_{i=1}^k (-i-1)x^{-2-k}$ for $\mathbb{R} \setminus (-2,2)$, and so they are integrable. Consequently, $\sum_{j=0}^6 \|f^{(j)}\|_1 \leqslant C$ for some $C > 0$. Lastly, note that the rescalings $\gamma_Y^{-1} f_\pm(x/\sqrt{\gamma_Y})$ sandwich $\phi(x^2)$ and equal x^{-2} for $|x| \geqslant 2\sqrt{\gamma_Y}$.

[11] To see this, apply the spectral theorem to the respective operators, and take the quadratic form. Then, use the fact that u is unit-norm to construct the u-reweighed empirical eigendensity.

[12] An explicit such choice comes from taking any fixed small $0 < \delta < 1/2$, and a smooth and even function $g : \mathbb{R} \to [0,1]$ that is 0 whenever $|x| \leqslant 1 + \delta$ and 1 whenever $|x| \geqslant 1 + 2\delta$. Then, set $f_-(x) = g(x)x^{-2}$ for $x \neq 0$ (and 0 otherwise) and $f_+(x) = f_-(x) + (1 - g(x))9/4$.

The common free limit. Observe (5.32) implies that $aI_n - \beta S + D$ is invertible, since conjugating by D gives

$$(aI_n - \beta S + D)^2 \succeq 4\gamma_Y D^2 \succeq_{D \succ I_n} 4\gamma_Y I_n \,.$$

Note that (5.32) also implies that $I_n + D^{-1}(aI_n - \beta S)$ is invertible since $D \succeq I_n$, and its two squared singular-value operators are unitarily conjugate[13]. This immediately implies that both diagonal blocks in the square of the free block matrix consequently have spectrum in $[4\gamma_Y, \infty)$ as

$$(aI_n - \beta S + D)D^{-2}(aI_n - \beta S + D) = U(D^{-1}(aI_n - \beta S + D)^2 D^{-1})U^* \succeq U(4\gamma_Y I_n)U^* \succcurlyeq 4\gamma_Y I_n \,,$$

which leads to the conclusion that $\mathsf{Spec}\,(\text{free block}) \notin (-2\sqrt{\gamma_Y}, 2\sqrt{\gamma_Y})$. Both $\gamma_Y^{-1} f_\pm(x/\sqrt{\gamma_Y})$ agree outside the spectrum with x^{-2} (which is $\phi(x^2)$ in this range) - this means their (non-commutative) free expectations are identical.

Operator sandwich. We now invoke Lemma 5.7 on $f_\pm(x/\sqrt{\gamma_Y})$ and multiply by γ_Y^{-1}, which gives

$$\left\| \mathop{\mathbb{E}}_A [f_\pm(\gamma_Y^{-1/2}\mathcal{B}(A))] - \mathrm{id}_2 \otimes E_{M_n}[f_\pm(\gamma^{-1/2}\mathcal{B}(S))] \right\|_{\mathsf{op}} \leqslant \frac{C}{n}\left(\beta^2\gamma_Y^{-1} + \beta^4\gamma_Y^{-2}\right) ,$$

where $\mathcal{B}(\cdot)$ is the block matrix taken as input to f in Lemma 5.7. Multiplying by γ_Y^{-1}, the finite expected value of each rescaled envelope $\gamma_Y^{-1} f(\gamma_Y^{-1/2}x)$ is within $Cn^{-1}(\beta^2\gamma_Y^{-2} + \beta^4\gamma_Y^{-3})$ in operator norm of that common free conditional expectation. Spectral functional calculus at each fixed block matrix, followed by positivity of expectation, places the expected reciprocal extension between these two matrix bounds. It therefore satisfies the same operator-norm estimate. This uses pointwise ordering of scalar functions, not operator monotonicity with respect to different matrix arguments. Taking the upper-left block and applying the block identity and reciprocal agreement from Proposition 5.2(c) and Proposition 5.2(e) at the free endpoint proves (5.33). The definition of $\Phi_{a,W,T}$, trace duality, and $\|T\|_{L^1} \leqslant \|T\|_{L^2}$ then give (5.34), using that E_{M_n} is trace-preserving. □

5.4. **Uniform estimate for the diagonal.** We combine the chaining estimate with the expectation estimate established by the free interpolation between the expected regularized resolvent and the ideal free object, as in [DLSS26a, Corollary B.35], restricting to the safe set in the final step[14]. This allows us to control the deviation between the diagonal of the squared resolvent from its ideal free-probabilistic limit with high probability over A.

Corollary 5.9 (Diagonal of the squared resolvent on the safe set)**.**
Let $A \sim \mathsf{GOE}(n)$ and fix $0 < \beta < 1$, $0 < \gamma_Y \leqslant \frac{(1-\beta)^4}{4(1+2\beta^2)^2}$. For all $n \geqslant n_0(\beta, \gamma_Y)$, uniformly over all

$$\left\{ D \in \mathcal{D}_n([1,\infty)) \mid D^{-1}(aI_n - \beta A + D)^2 D^{-1} \succeq \gamma_Y I_n \right\} , \tag{5.35}$$

the following holds

$$\mathop{\mathbb{P}}_A \left\{ \left\| E_{\mathcal{D}_n}\left[D(\beta^2 \operatorname{tr}_n(D^{-1})I_n - \beta A + D)^{-2} D \right] - \frac{I_n}{1 - \beta^2 \operatorname{tr}_n(D^{-2})} \right\|_{L^2} \leqslant \frac{C\beta}{\gamma_Y^2\sqrt{n}} \right\} \geqslant 1 - 4e^{-n/4}, \tag{5.36}$$

for a universal constant $C > 0$. In particular, on the same event, (5.36) holds uniformly for $D = D(m)$ with $m \in \mathcal{S}_A(c)$ from Definition 2.2, for any fixed $c > 0$ such that

$$\gamma_Y \leqslant \left(1 + \frac{1+3\beta}{c}\right)^{-2} . \tag{5.37}$$

Proof. We first control the deviation of the regularized resolvent. Set $C_{\mathrm{op}} = 3$ and $\overline{A} = \Pi_3(A)$ in Theorem 5.4, so that $C_1 = 1 + 3\beta \leqslant 4$ and invoke Proposition 2.3. Then, invoke (5.23) with $p = 2e^{-n/4}$. Since $\gamma_Y \leqslant 1$ and

[13] This simply means that there exists a unitary U, such that, for any invertible operator Y in the algebra $YY^* = U(Y^*Y)U^*$.
[14] As stated before, the free interpolation in §5.3 is a simplification of the original interpolation argument used in [DLSS26a, §B]. The key observation that allows for this is the representation of the input to the regularized resolvent $\phi(\cdot)$ having a block representation with entries that are affine functions of A (Proposition 5.2(d)). This allows the results of [BBv23, §2] to directly control the error between the ideal free object and (average) regularized resolvent.

$\sqrt{\log(2/p)} = \sqrt{n}/2$,

$$\mathop{\mathbb{P}}_{A}\left\{\sup_{(a,W,T)\in\mathcal{J}} |Z_{a,W,T}| \leqslant \frac{C\beta}{\gamma_Y^2\sqrt{n}}\right\} \geqslant 1 - 2e^{-n/4}.$$

Set a deterministic $W = D \succeq I_n$ and $a = \beta^2 \operatorname{tr}_n(D^{-1}) \in [0,1]$. Lemma A.2 provides a spectral cut-off via the TAP subordination so that (5.32) holds, yielding (5.34) as the free target. Consequently, for every diagonal T,

$$\left|\mathop{\mathbb{E}}_{A}\Phi_{a,W,T}(A) - \frac{\operatorname{tr}_n(T)}{1-\beta^2\operatorname{tr}_n(D^{-2})}\right| \leqslant \frac{C}{n}\left(\beta^2\gamma_Y^{-2} + \beta^4\gamma_Y^{-3}\right)\|T\|_{L^2}.$$

Also, Proposition 5.2(a) and Proposition 2.3 give

$$\left|\mathop{\mathbb{E}}_{A}\Phi_{a,W,T}(\overline{A}) - \mathop{\mathbb{E}}_{A}\Phi_{a,W,T}(A)\right| \leqslant \frac{9}{\gamma_Y}e^{-n/4}\|T\|_{L^2}, \tag{5.38}$$

since the two observables agree when $\|A\|_{\mathsf{op}} \leqslant 3$ and each has absolute value at most $9\|T\|_{L^2}/(4\gamma_Y)$. For fixed β, γ_Y, these two expectation errors are $o(\beta\gamma_Y^{-2}n^{-1/2})$ and so, on the chaining event, the triangle inequality yields

$$\left|\Phi_{a,W,T}(\overline{A}) - \frac{\operatorname{tr}_n(T)}{1-\beta^2\operatorname{tr}_n(D^{-2})}\right| \leqslant \frac{C\beta}{\gamma_Y^2\sqrt{n}},$$

uniformly over valid D and all diagonal T with $\|T\|_{L^2} = 1$, for all $n \geqslant n_0(\beta,\gamma_Y)$. Intersecting the chaining event with $\{\|A\|_{\mathsf{op}} \leqslant 3\}$ gives probability at least $1 - 4e^{-n/4}$. On this intersection $\overline{A} = A$, and Proposition 5.2(e) makes the reciprocal extension equal to $D(aI_n - \beta A + D)^{-2}D$ for every realized D satisfying (5.35). Taking the supremum over T and applying diagonal norm duality (see [JSS25, Proof of Lemma 4.25]) proves (5.36).

Finally, to restrict to the good set, choose $m \in \mathcal{S}_A(c)$ and set $D = D(m)$. Then, $a = \beta^2\operatorname{tr}_n(D^{-1}) = \beta^2(1 - \|m\|_2^2/n)$. Since $\|aI_n - \beta A\|_{\mathsf{op}} \leqslant 1 + 3\beta$ on the same event, (2.1) and Definition 2.2 give

$$aI_n - \beta A + D = \nabla_m^2\mathcal{F}_{\mathrm{TAP}}(m,\cdot) + \frac{2\beta^2}{n}mm^{\mathsf{T}} \succ cD \succeq cI_n,$$

$$\left\|(I_n + D^{-1}(aI_n - \beta A))^{-1}\right\|_{\mathsf{op}} = \left\|I_n - (aI_n - \beta A + D)^{-1}(aI_n - \beta A)\right\|_{\mathsf{op}} \leqslant 1 + \frac{1+3\beta}{c}.$$

Therefore, (5.37) implies the product condition (5.35) uniformly over $m \in \mathcal{S}_A(c)$. □

6. Properties of the algorithmic stochastic process

To prove the desiderata (d3), (d4), and (d5), we need the analog of [DLSS26a, §6], but restricted to the good set $\mathcal{S}_A(c)$ defined in Definition 2.2, as opposed to the uniform bounds over $(-1,1)^n$ that can be obtained when $\beta < 1/2$. Recall the coefficient functions

$$\begin{aligned} D(m) &:= (I_n - \mathsf{diag}(m)^2)^{-1}, & a(m) &:= \beta^2\operatorname{tr}_n(D(m)^{-1}), \\ \overline{Q}(m) &:= (a(m)I_n - \beta A + D(m))^{-1}, & \widehat{Q}(m) &:= \left(a(m)I_n - \beta A + D(m) - \frac{2\beta^2}{n}mm^{\mathsf{T}}\right)^{-1}. \end{aligned} \tag{6.1}$$

Recall also the Itô correction to ASL-TAP

$$f(m) := \left(E_{\mathcal{D}_n}[D(m)\widehat{Q}(m)^2D(m)] - \beta^2\operatorname{tr}_n(\widehat{Q}(m)^2)I_n - \frac{2\beta^2}{n}\widehat{Q}(m)^2\right)m. \tag{6.2}$$

Nearly all of the argument structure in [DLSS26a, §6] is reused, so we deliberately maintain as much of the same document structure as possible: establishing basic resolvent bounds, controlling the rank-one correction and diagonal error, and then deducing regularity of the drift and properties of the trajectories.

The only new difficulty is that the set $\mathcal{S}_A(c)$ that we work in is no longer convex. Thus we can no longer bound Lipschitz constants by integrating bounds on their derivatives along straight line segments. The main workaround is to use the resolvent identity to directly calculate the differences between the endpoints, and Lemma A.1 collects those calculations, replacing [DLSS26a, Lemma B.4]. To obtain the transportation inequality and therefore sub-Gaussian concentration of Lipschitz path functionals as in [DLSS26a, §6.6], we

use the Kirszbraun extension theorem to fill in values for the SDE coefficients outside of $\mathcal{S}_A(c)$. Finally, the main new argument is in Lemma 6.9, where we provide a Lipschitz bound and extension of the JE weight integrand, this being the only place where the resolvent identity doesn't suffice to establish the needed Lipschitz bounds.

As before, vector norms are Euclidean and $\|\cdot\|_{L^2}$ is the normalized Schatten 2-norm, so that $\|\cdot\|_F = \sqrt{n}\,\|\cdot\|_{L^2}$, and all Lipschitz constants for matrix-valued maps in this section use $\|\cdot\|_{L^2}$ unless another norm is specified.

6.1. **High-probability events.** Throughout this section, we work on the event $\{\|A\|_{\mathsf{op}} \leqslant 3\}$, which we recall for GOE A holds with probability at least $1 - 2e^{-n/4}$ by Proposition 2.3.

Fix $0 < \beta < 1$ and $c > 0$, and choose

$$\gamma_Y := \min\left\{\frac{(1-\beta)^4}{4(1+2\beta^2)^2}, \left(1 + \frac{1+3\beta}{c}\right)^{-2}\right\}. \tag{6.3}$$

Let

$$\mathcal{S}_D := \left\{D \in \mathcal{D}_n([1,\infty)) : \beta^2 \operatorname{tr}_n(D^{-1}) I_n - \beta A + D \succeq c I_n\right\}. \tag{6.4}$$

For all sufficiently large n, depending only on β, c, let $\Omega_{\beta,c}$ be the conclusion of Corollary 5.9, so that there is a universal constant C so that

$$\Omega_{\beta,c} := \left\{\forall D \in \mathcal{S}_D \,.\, \left\|E_{\mathcal{D}_n}\left[D(\beta^2 \operatorname{tr}_n(D^{-1}) I_n - \beta A + D)^{-2} D\right] - \frac{I_n}{1 - \beta^2 \operatorname{tr}_n(D^{-2})}\right\|_{L^2} \leqslant \frac{C\beta}{\gamma_Y^2 \sqrt{n}}\right\}. \tag{6.5}$$

Then $\mathbb{P}_A(\Omega_{\beta,c} \cap \{\|A\|_{\mathsf{op}} \leqslant 3\}) \geqslant 1 - 4e^{-n/4}$ by the proof of Corollary 5.9. The choice (6.3) ensures that this diagonal bound also applies to every $m \in (-1,1)^n$ satisfying $a(m) I_n - \beta A + D(m) \succeq c\, I_n$. Constants denoted $C_{\beta,c}$ may increase between occurrences and are independent of n and of A.

6.2. **Basic resolvent and Lipschitzness bounds.** The resolvent operator norm and Lipschitzness bounds exactly echo those of [DLSS26a, §6.2], with the exception of an additional pair of bounds established for $\overline{Q}$ to assist in the Lipschitz extension of the JE weights.

Lemma 6.1 (Basic resolvent bounds on the good set)**.**
Let $D, a, \overline{Q}, \widehat{Q}$ be as in (6.1). Assume $\|A\|_{\mathsf{op}} \leqslant 3$. Then for every $m \in \mathcal{S}_A(c)$, both $\overline{Q}$ and $\widehat{Q}$ are positive definite and satisfy

(a)
$$\left\|\overline{Q}(m)\right\|_{\mathsf{op}}, \ \left\|\widehat{Q}(m)\right\|_{\mathsf{op}} \leqslant c^{-1}.$$

(b)
$$\left\|D(m)\overline{Q}(m)\right\|_{\mathsf{op}}, \ \left\|D(m)\widehat{Q}(m)\right\|_{\mathsf{op}} \leqslant 1 + \frac{3\beta + 3\beta^2}{c}. \tag{6.6}$$

(c)
$$\left\|D(m)\overline{Q}(m)^2\right\|_{\mathsf{op}}, \ \left\|D(m)\widehat{Q}(m)^2\right\|_{\mathsf{op}} \leqslant c^{-1}\big(1 + (3\beta + 3\beta^2)/c\big).$$

(d)
$$\left\|D(m)\overline{Q}(m)^2 D(m)\right\|_{\mathsf{op}}, \ \left\|D(m)\widehat{Q}(m)^2 D(m)\right\|_{\mathsf{op}} \leqslant \big(1 + (3\beta + 3\beta^2)/c\big)^2.$$

Proof. The proof is the same as [DLSS26a, Lemma 6.1], except that the gap now follows from

$$\widehat{Q}(m)^{-1} \succ cD(m) \succeq cI_n, \qquad \overline{Q}(m)^{-1} = \widehat{Q}(m)^{-1} + \frac{2\beta^2}{n} m m^{\mathsf{T}} \succ cD(m).$$

In [DLSS26a, Lemma B.1], use $W = D(m)^{-1}$ and, respectively, $M = a(m)I_n - \beta A$ or $M = a(m)I_n - \beta A - 2\beta^2 mm^{\mathsf{T}}/n$. Both have $\|M\|_{\mathsf{op}} \leqslant 3\beta + 3\beta^2$. The identities $DR = I_n - MR$ and $RD = I_n - RM$ give the weighted bounds. □

Lemma 6.2 (Basic resolvent Lipschitzness on the good set)**.**
In the setting of Lemma 6.1, the following Lipschitz bounds hold for the following functions of m on $\mathcal{S}_A(c)$, with Euclidean distance on the input $m \in \mathcal{S}_A(c)$ and normalized Schatten 2-norm $\|\cdot\|_{L^2}$ on the output:

(a)

$$\left\|\widehat{Q}\right\|_{\mathsf{Lip}},\ \left\|\overline{Q}\right\|_{\mathsf{Lip}} \leqslant \frac{C_{\beta,c}}{\sqrt{n}}. \tag{6.7}$$

(b)

$$\left\|\widehat{Q}^2\right\|_{\mathsf{Lip}} \leqslant \frac{C_{\beta,c}}{\sqrt{n}}.$$

(c)

$$\left\|D\widehat{Q}\right\|_{\mathsf{Lip}} \leqslant \frac{C_{\beta,c}}{\sqrt{n}}.$$

(d)

$$\left\|D\widehat{Q}^2 D\right\|_{\mathsf{Lip}} \leqslant \frac{C_{\beta,c}}{\sqrt{n}}.$$

Proof. Follow [DLSS26a, Lemma 6.2], replacing its derivative bounds by parts (a)–(e) of Lemma A.1 with

$$W \leftarrow I_n - \mathsf{diag}(m)^2, \qquad M \leftarrow a(m)I_n - \beta A - \frac{2\beta^2}{n}mm^{\mathsf{T}},$$

at each of the two values of m required for the Lipschitzness bound, omitting the rank-one term for $\overline{Q}$. The required input bounds are

$$\begin{aligned}\|W(m) - W(w)\|_{L^2} &\leqslant \frac{2}{\sqrt{n}}\|m - w\|_2,\\ \|M(m) - M(w)\|_{L^2} &\leqslant \left(\frac{2\beta^2}{\sqrt{n}} + \frac{4\beta^2}{n}\right)\|m - w\|_2.\end{aligned}$$

□

We skip Lemma 6.3 because the argument has been moved to Lemma 4.3. The next lemma restarts numbering at 6.4 in order to maintain the same lemma numbering as in [DLSS26a, §6].

6.3. **Diagonal error from rank-1 term of TAP Hessian.** The definition of the event $\Omega_{\beta,c}$ from Corollary 5.9 is in terms of $\overline{Q}$, whereas the algorithm uses $\widehat{Q}$. The following analog of [DLSS26a, Lemma 6.4] bounds the difference, substantially simplifying the earlier version of the proof.

Lemma 6.4 (Small diagonal error from rank-1 terms)**.**
In the setting of Lemma 6.1, for all $m \in \mathcal{S}_A(c)$,

$$\left\|E_{\mathcal{D}_n}[D(m)(\widehat{Q}(m)^2 - \overline{Q}(m)^2)D(m)]\right\|_{L^2} \leqslant \frac{C_{\beta,c}}{\sqrt{n}}.$$

Proof. Apply part (e) of Lemma A.1, keeping $W = D(m)^{-1}$ fixed and changing M by $-2\beta^2 mm^{\mathsf{T}}/n$. The $\min\mathrm{Spec}(|M + W^{-1}|) > c$ condition is satisfied at both endpoints and $\|\Delta M\|_{L^2} = \left\|-2\beta^2 mm^{\mathsf{T}}/n\right\|_{L^2} \leqslant 2\beta^2/\sqrt{n}$. Then contractivity of $E_{\mathcal{D}_n}$ concludes the bound. □

6.4. **Control of the diagonal.** As in [DLSS26a, §6.4], define

$$\delta_{\mathsf{diag}}(m) := E_{\mathcal{D}_n}[\widehat{Q}(m)^2]D(m)^2 - (1 + \beta^2\,\mathrm{tr}_n(\widehat{Q}(m)^2))I_n. \tag{6.8}$$

Since D is diagonal, the first term is also $E_{\mathcal{D}_n}[D\widehat{Q}^2 D]$.

Lemma 6.5 (Basic diagonal properties on the good set)**.**
In the setting of Lemma 6.1, for $m, w \in \mathcal{S}_A(c)$,

(a)

$$\|\delta_{\mathsf{diag}}(m)\|_{\mathsf{op}} \leqslant C_{\beta,c}.$$

(b)

$$\|\delta_{\mathsf{diag}}(m) - \delta_{\mathsf{diag}}(w)\|_{L^2} \leqslant \frac{C_{\beta,c}}{\sqrt{n}}\|m - w\|_2. \tag{6.9}$$

Proof. The proof is the same as [DLSS26a, Lemma 6.5], using Lemma 6.1 and Lemma 6.2. Both the diagonal projection and $X \mapsto \mathrm{tr}_n(X) I_n$ are contractions in the relevant norms. □

Lemma 6.6 (Uniform diagonal control)**.**
In the setting of Lemma 6.1, assume also that we are in the event $\Omega_{\beta,c}$. Then for every $m \in \mathcal{S}_A(c)$,

$$\|\delta_{\mathsf{diag}}(m)\|_{L^2} \leqslant \frac{C_{\beta,c}}{\sqrt{n}}. \tag{6.10}$$

The same bound, with every $\widehat{Q}$ replaced by $\overline{Q}$, holds for all $m \in (-1,1)^n$ such that $D(m) \in \mathcal{S}_D$.

Proof. The proof of [DLSS26a, Lemma 6.6] is unchanged after using the definition of $\mathcal{S}_A(c)$ and Lemma 6.4 for the initial diagonal estimate. For $\overline{Q}$, skip the application of Lemma 6.4. □

6.5. **Error between ASL-TAP and PHD is small.** Recall that ASL-TAP and PHD differ by the drift $\widehat{Q}(m)(m - f(m))$, with f given in (6.2). We obtain the analog of [DLSS26a, Lemma 6.7], showing that this drift term is small.

Lemma 6.7 (Closeness of ASL-TAP and PHD)**.**
In the setting of Lemma 6.1, assuming the event $\Omega_{\beta,c}$ defined in (6.5) and with f as in (6.2), the following bounds hold on $m \in \mathcal{S}_A(c)$ in Euclidean norm:

(a)
$$\|f\|_{\mathsf{Lip}} \leqslant C_{\beta,c}. \tag{6.11}$$

(b)
$$\|m - f(m)\|_2 \leqslant C_{\beta,c}. \tag{6.12}$$

(c)
$$\left\|\widehat{Q}(m)(m - f(m))\right\|_2 \leqslant C_{\beta,c}. \tag{6.13}$$

(d)
$$\left\|D(m)\widehat{Q}(m)(m - f(m))\right\|_2 \leqslant C_{\beta,c}.$$

(e)
$$\left\|m \mapsto \widehat{Q}(m)(m - f(m))\right\|_{\mathsf{Lip}} \leqslant C_{\beta,c}. \tag{6.14}$$

Proof. Use the proof of [DLSS26a, Lemma 6.7], with all of the references to intermediate lemmas replaced with their analogs written above, with the same section-internal numbering. The identity expressing f in terms of $\delta_{\mathsf{diag}}(m)$ is

$$f(m) = \left(I_n + \delta_{\mathsf{diag}}(m) - \frac{2\beta^2}{n}\widehat{Q}(m)^2\right) m.$$

In particular, $\|\delta_{\mathsf{diag}}(m)m\|_2 \leqslant \sqrt{n}\,\|\delta_{\mathsf{diag}}(m)\|_{L^2}$, and the same inequality applies to the difference of the diagonal errors acting on any $w \in (-1,1)^n$. All other matrix differences are bounded using $\|\cdot\|_{\mathsf{op}} \leqslant \sqrt{n}\,\|\cdot\|_{L^2}$. The remaining product estimates are unchanged. □

6.6. **Lipschitz extension of the JE weight integrand.** The next two lemmata provide the additional ingredients to adapt the JE-weight argument in [DLSS26a, Corollary 6.14(d5)] to a nonconvex good set. We start by constructing short paths preserving the absolute spectral gap of the resolvent without the rank-one term, so that derivative bounds hold along these paths and can be used to prove Lipschitz bounds.

Lemma 6.8 (Short paths preserving the spectral gap)**.**
Fix $0 < \beta < 1$, $c > 0$, and $A \in \mathrm{Sym}_n(\mathbb{R})$. Let $W_0, W_1 \in \mathcal{D}_n((0,1])$ satisfy

$$W_j^{-1} \in \mathcal{S}_D, \qquad j = 0, 1,$$

where $\mathcal{S}_D$ was defined in (6.4). Then there exists a piecewise C^1 path $W : [0,1] \to \mathcal{D}_n((0,1])$ with $W(0) = W_0$ and $W(1) = W_1$ such that

$$W(t)^{-1} \in \mathcal{S}_D, \qquad 0 \leqslant t \leqslant 1, \tag{6.15}$$

and its normalized Schatten 2-length satisfies

$$\int_0^1 \|W'(t)\|_{L^2}\, dt \leqslant \frac{2}{1-\beta^2}\,\|W_1 - W_0\|_{L^2}\,. \tag{6.16}$$

Proof. Let $\min(W_0, W_1)$ denote their coordinatewise minimum. The path will proceed in a line segment from W_0 to $(1+s)^{-1}\min(W_0,W_1)$ for some scalar s and then in another line segment to W_1. Accordingly, denote the two line segments as

$$s := \frac{\beta^2}{1-\beta^2}\,\|W_1 - W_0\|_{L^2}\,, \qquad W_j(t) := (1-t)W_j + \frac{t}{1+s}\min(W_0,W_1), \quad 0 \leqslant t \leqslant 1, \quad j = 0,1.$$

Both segments end at $(1+s)^{-1}\min(W_0,W_1)$ and satisfy $0 \prec W_j(t) \preceq W_j \preceq I_n$. We now proceed with a few path derivative calculations to show that (6.15) holds throughout the line segments. First,

$$-W_j'(t) = \frac{sW_j + W_j - \min(W_0,W_1)}{1+s} \succeq \frac{s}{1+s}W_j.$$

By diagonality and $0 \prec W_j(t) \preceq W_j \preceq I_n$,

$$\frac{d}{dt}W_j(t)^{-1} = W_j(t)^{-2}(-W_j'(t)) \succeq \frac{s}{1+s}W_j(t)^{-2}W_j \succeq \frac{s}{1+s}I_n.$$

Since $\mathrm{tr}_n(W_j) \leqslant 1$, Cauchy–Schwarz for the normalized trace gives

$$\begin{aligned}\mathrm{tr}_n[-W_j'(t)] &= \frac{s\,\mathrm{tr}_n(W_j) + \mathrm{tr}_n[W_j - \min(W_0,W_1)]}{1+s} \\ &\leqslant \frac{s + \|W_j - \min(W_0,W_1)\|_{L^2}}{1+s} \leqslant \frac{s + \|W_1 - W_0\|_{L^2}}{1+s}.\end{aligned}$$

Combining these estimates yields

$$\frac{d}{dt}\left(W_j(t)^{-1} + \beta^2\,\mathrm{tr}_n(W_j(t))I_n - \beta A\right) \succeq \frac{(1-\beta^2)s - \beta^2\,\|W_1 - W_0\|_{L^2}}{1+s}I_n = 0.$$

Integrating from 0 to t and using the $W_j^{-1} \in \mathcal{S}_D$ assumption gives

$$W_j(t)^{-1} + \beta^2\,\mathrm{tr}_n(W_j(t))I_n - \beta A \succeq W_j^{-1} + \beta^2\,\mathrm{tr}_n(W_j)I_n - \beta A \succeq cI_n,$$

showing (6.15) for both of the two segments. Concatenate $W_0(t)$ with $W_1(t)$ traversed in reverse, and rescale the parameter interval to $[0,1]$ to obtain the required piecewise C^1 path $W(t)$. Since $\|W_j\|_{L^2} \leqslant 1$ and $\|W_j - \min(W_0,W_1)\|_{L^2} \leqslant \|W_1 - W_0\|_{L^2}$, its length satisfies

$$\begin{aligned}\int_0^1 \|W'(t)\|_{L^2}\, dt &= \sum_{j=0}^{1}\int_0^1 \left\|W_j'(t)\right\|_{L^2} dt \\ &\leqslant \frac{2\left(s + \|W_1 - W_0\|_{L^2}\right)}{1+s} \\ &= \frac{2}{(1-\beta^2)(1+s)}\,\|W_1 - W_0\|_{L^2} \leqslant \frac{2}{1-\beta^2}\,\|W_1 - W_0\|_{L^2}\,,\end{aligned}$$

which proves (6.16). □

Lemma 6.9 (Lipschitzness of the JE integrand)**.**
In the setting of Lemma 6.7, let

$$\omega(m) := \frac{1}{2}\left(\mathsf{Tr}[\widehat{Q}(m)] + \|m\|_2^2\right), \qquad (\forall m \in \mathcal{S}_A(c)) \tag{6.17}$$

as in (d5). Then there is an $L_\omega = C_{\beta,c} < \infty$ such that

$$|\omega(m) - \omega(w)| \leqslant L_\omega\,\|m - w\|_2 \qquad (\forall m, w \in \mathcal{S}_A(c)).$$

Proof. Fix A. We will first prove the Lipschitz bound for

$$\overline{\omega}(m) := \frac{1}{2}(\mathsf{Tr}\overline{Q}(m) + \|m\|_2^2) = \frac{n}{2} + \frac{1}{2}(\mathsf{Tr}\overline{Q}(m) - \mathsf{Tr}W), \qquad W = D(m)^{-1},$$

and then add the rank-one term back in at the end. With A held fixed, we use the right-hand expression to regard $\overline{\omega}$ as a function of $W \in \mathcal{D}_n((0,1])$ wherever the following inverse exists:

$$D = W^{-1}, \qquad a = \beta^2 \operatorname{tr}_n(W), \qquad \overline{Q} = \big(W^{-1} + \beta^2 \operatorname{tr}_n(W) I_n - \beta A\big)^{-1}.$$

Differentiating with respect to W in a diagonal direction H gives

$$d_W[\overline{\omega}][H] = d_W\left[\frac{1}{2}(\mathsf{Tr}\overline{Q} - \mathsf{Tr}W)\right][H] = \frac{1}{2}\mathsf{Tr}\Big[\Big(E_{\mathcal{D}_n}[D\overline{Q}^2 D] - (1 + \beta^2 \operatorname{tr}_n(\overline{Q}^2))I_n\Big) H\Big]. \tag{6.18}$$

For every W satisfying (6.15), Lemma 6.6 gives

$$\Big\| E_{\mathcal{D}_n}[D\overline{Q}^2 D] - (1 + \beta^2 \operatorname{tr}_n(\overline{Q}^2))I_n \Big\|_F \leqslant C_{\beta,c}.$$

Cauchy–Schwarz for the unnormalized trace in (6.18) therefore yields, for all W satisfying (6.15) and diagonal H,

$$|d_W[\overline{\omega}][H]| \leqslant C_{\beta,c} \|H\|_F. \tag{6.19}$$

Fix $m, w \in \mathcal{S}_A(c)$ and set $W_0 = D(m)^{-1}$ and $W_1 = D(w)^{-1}$, which implies $W_0^{-1}, W_1^{-1} \in \mathcal{S}_D$. Then let $W(t)$ be the path supplied by Lemma 6.8 between those two endpoints, so that (6.19) applies along this path. Since $\overline{\omega}$ depends only on $D(m)^{-1}$, integrating that bound and using (6.16) gives

$$\begin{aligned} |\overline{\omega}(m) - \overline{\omega}(w)| &\leqslant C_{\beta,c}\sqrt{n} \int_0^1 \|W'(t)\|_{L^2}\, dt \\ &\leqslant \frac{2C_{\beta,c}\sqrt{n}}{1-\beta^2} \|W_1 - W_0\|_{L^2} \leqslant C_{\beta,c} \|m - w\|_2, \end{aligned} \tag{6.20}$$

where the last inequality uses

$$\sqrt{n}\,\|W_1 - W_0\|_{L^2} = \left(\sum_{i=1}^n (m_i^2 - w_i^2)^2\right)^{1/2} \leqslant 2\|m - w\|_2.$$

Finally, each $\widehat{Q} - \overline{Q}$ has rank at most one, so their difference between two endpoints has rank at most two. Therefore

$$\begin{aligned} &\big|\mathsf{Tr}\big[(\widehat{Q}(m) - \overline{Q}(m)) - (\widehat{Q}(w) - \overline{Q}(w))\big]\big| \\ &\quad\leqslant \sqrt{2}\big(\|\widehat{Q}(m) - \widehat{Q}(w)\|_F + \|\overline{Q}(m) - \overline{Q}(w)\|_F\big) \leqslant C_{\beta,c}\|m - w\|_2, \end{aligned}$$

by Lemma 6.2(a), applied to $\widehat{Q}$ and $\overline{Q}$ at $m, w \in \mathcal{S}_A(c)$. Combining this with (6.20) proves the claim. □

6.7. **Implications for the sampling algorithm.**

Corollary 6.10 (Algorithmic desiderata)**.**
Assume we are in the events $\{\|A\|_{\mathsf{op}} \leqslant 3\}$ and $\Omega_{\beta,c}$ as defined in (6.5). Then the following hold with constants independent of n and A.

(d3) *For every $m \in \mathcal{S}_A(c)$, $\|f(m) - m\|_2^2 \leqslant C_{\beta,c}^2$.*

(d4) *(a) There is a constant $L \leqslant C_{\beta,c}$ such that $\|\widehat{Q}(m) - \widehat{Q}(w)\|_F \leqslant L\|m - w\|_2$ for $m, w \in \mathcal{S}_A(c)$.*
(b) There exists a constant L_{drift} such that for all $m, w \in \mathcal{S}_A(c_{\mathrm{TAP}})$, with f defined as in (6.2),

$$\big\|\widehat{Q}(m)(m - f(m)) - \widehat{Q}(w)(w - f(w))\big\|_2 \leqslant L_{\mathrm{drift}}\|m - w\|_2.$$

(d5) *Let*

$$\omega(m) := \frac{1}{2}\big(\mathsf{Tr}[\widehat{Q}(m)] + \|m\|_2^2\big).$$

Then there is an $L_\omega < \infty$ such that

$$|\omega(m) - \omega(w)| \leqslant L_\omega \|m - w\|_2 \qquad (\forall m, w \in \mathcal{S}_A(c)).$$

Proof. Parts (d3)–(d4). As in [DLSS26a, Corollary 6.14], these follow in order from Lemma 6.7(b), Lemma 6.2(a), and Lemma 6.7(e). For (d4), convert normalized Schatten 2-norm to Frobenius norm.

Part (d5). This is Lemma 6.9. □

7. Sampling from the time-T localized distribution

Using a result of Kumar, Sarkar, Tian and Zhu [KSTZ26, Theorem 4 and Corollary 5] we give a thresholding time $T(\beta) > 0$ to inform how long the ASL-TAP-JE procedure should run to obtain a warm-start for the annealing-type sampler in [KSTZ26, Algorithm 2] run on a small wedge around a fixed tilt y_T. This samples from the time-$T(\beta)$ localized distribution in $\mathsf{poly}(n)$-time with $o_n(1)$-TVD error. This approach replaces the polarized–walk based sampler for the time-T localized distribution that was used in [DLSS26a, §7]. The reason for this replacement is that the prior approach uses the needle-decomposition [EKZ22] before an application of the HS transform, which introduces a shift of $2\beta I_n$ to make the interactions PSD, after which a union bound over PSD sub-matrices constrained to have small-enough operator norm forces an upper bound of $\beta < 1/2$ due to the shift. The approach of Kumar, Sarkar, Tian and Zhu bypasses the obstacle in the prior approach of the authors [DLSS26a, §7.2] by reasoning about a Dobrushin-type condition straight away on the wedge around the tilt y_T (which the localized measure will concentrate on with high probability for any $\beta > 0$, see [DLSS26a, §7]).

We invoke [KSTZ26, Corollary 5] with $J \coloneqq A$[15] and $h \coloneqq y_{T(\beta)} = T(\beta)x_0 + \sqrt{T(\beta)}\,B$. We know that with probability $1-e^{-c(\beta)n}$ over the SL process, the localized measure concentrates on a wedge $B(\mathrm{sign}(y_T), \varepsilon(\beta, T)n)$ [DLSS26a, Lemma 7.43]. The following theorem then shows efficient sampling for this wedge.

Theorem 7.1 (Sampling on small wedges via [KSTZ26, Algorithm 2])**.** *Fix $\beta > 0$, $\varepsilon(\beta) = \frac{1}{C\overline{\beta}^2 \log(e\overline{\beta})}$ with $C > 0$ some large enough universal constant and $\overline{\beta} = \max\{1, \beta\}$, and $k \leqslant \varepsilon(\beta)n$. Then, with probability $1 - o_n(1)$, for any $0 < \beta' < \beta$ and $0 < \varepsilon < 1/2$, uniformly over $y \in \mathbb{R}^n$, letting $\sigma_0 = \mathrm{sign}(y) \in \{-1, 1\}^n$, [KSTZ26, Algorithm 2] run for $2\log\left(\frac{16}{\varepsilon^2(\beta')}\right)$ annealing levels, with $\mathsf{poly}(n)\log(1/\varepsilon)$ Markov chain steps for each level, gives a sample that is ε close in TV distance from $\mu_{\beta' A, y}|_{B(\sigma_0, \varepsilon n)}$.*

Proof. Define the diagonal matrix $R_y = \mathsf{diag}(\mathrm{sign}(y))$ and set $x = -R_y\sigma$ for any $\sigma \in B(\mathrm{sign}(y), \varepsilon(\beta')n)$. Clearly, then, σ can be in the wedge around y if, and only if, the number of positive signs in x are no more than $\varepsilon(\beta')n$. This bijection between centering on y and 0^n allows the measure on the wedge to be rewritten as

$$\mu_y(x) \propto \exp\left\{\frac{\beta}{2}\left(x^{\mathsf{T}}(R_y J R_y)x - (R_y y)^{\mathsf{T}}x\right)\right\},$$

where we use the distributional equivalence $R_y J R_y \overset{d}{=} \mathsf{GOE}(n)$. Denote the new parameters $J = \frac{1}{2}R_y A R_y$ for $A \sim \mathsf{GOE}(n)$ and $h = -\frac{2R_y y}{\beta}$. Note that the [KSTZ26, Proof of Theorem 7] applies verbatim with the interaction matrix $R_y J R_y$ due to the distributional equivalence before since the Dobrushin-type bounds based on [KSTZ26, Lemma 8] remain unchanged. Furthermore, we can choose $\varepsilon(\beta) = \frac{1}{C\overline{\beta}^2\log(e\overline{\beta})} \leqslant \frac{1}{C\overline{\beta'}^2\log(e\overline{\beta'})} = \varepsilon(\beta')$ so that the validity of [KSTZ26, Theorem 4] holds when invoking [KSTZ26, Corollary 5], and we can get the mixing to hold simultaneously on the same event for A for every $0 < \beta' < \beta$. □

[15] Note that since we are fixing the tilt y_T, we don't need to invoke contiguity to the planted model. Since we will choose T large enough so that the localized measure concentrates on a desirably small wedge, we will argue about sampling in that wedge *uniformly* across the choice of tilts $\sigma_0 = \mathrm{sign}(y_T)$.

Acknowledgements

JSS thanks David Jekel for introducing him to the coefficient-matrix representation for Gaussian random matrices used in the work of Bandeira, Boedihardjo and van Handel [BBv23], the operator-valued convergence established in the work of Jekel, Lee, Nelson and Pi [JLNP25], and the idea of the block matrix representation used in (5.11) which led to the simplification of the free interpolation in §5.3 compared to the authors' original approach based on [DLSS26a, §B.6].

AI Disclosure: We used generative AI to assist with verification, and choose from different valid choices we developed for Lemma 3.6, (5.1), and short-path constructions in Lemma 6.8 to aid in selecting the most readable options.

Declaration: The authors of this article do not consent to having the contents of this article be used for the training or development of LLMs/AI models.

References

[AJK+22] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 1418–1430, New York, NY, USA, 2022. Association for Computing Machinery. DOI:10.1145/3519935.3520048.

[AKV24] Nima Anari, Frederic Koehler, and Thuy-Duong Vuong. Trickle-Down in Localization Schemes and Applications. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, pages 1094–1105, New York, NY, USA, 2024. Association for Computing Machinery. DOI:10.1145/3618260.3649622.

[AMS22] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the Sherrington-Kirkpatrick Gibbs Measure via Algorithmic Stochastic Localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 323–334, 2022. DOI:10.1109/FOCS54457.2022.00038.

[BAR26] Afonso S Bandeira, Ahmed El Alaoui, and Almut Rödder. Mixing of Glauber Dynamics on High Overlap Gibbs Measures. *arXiv preprint arXiv:2607.06813*, 2026.

[BBv23] Afonso S. Bandeira, March T. Boedihardjo, and Ramon van Handel. Matrix Concentration Inequalities and Free Probability. *Inventiones mathematicae*, 234(1):419–487, 2023. DOI:10.1007/s00222-023-01204-6.

[Bha13] Rajendra Bhatia. *Matrix analysis*. Springer Science & Business Media, 2013.

[Bia98] Philippe Biane. Processes with Free Increments. *Mathematische Zeitschrift*, 227(1):143–174, 1998. DOI:10.1007/PL00004363.

[BKMR25] Afonso S Bandeira, Anastasia Kireeva, Antoine Maillard, and Almut Rödder. Randomstrasse101: Open problems of 2024. *arXiv preprint arXiv:2504.20539*, 2025.

[BLG26] Mathews Boban, Anqi Li, and Shayan Oveis Gharan. Rank-1-perturbed trickledown theorems: Mixing time of Glauber dynamics for the Sherrington-Kirkpatrick model up to $\beta \leqslant 1/2 + \epsilon$. *arXiv preprint arXiv:2609.13138*, 2026.

[CDMF+11] Mireille Capitaine, Catherine Donati-Martin, Delphine Féral, Maxime Février, et al. Free convolution with a semicircular distribution and eigenvalues of spiked deformations of Wigner matrices. *Electron. J. Probab*, 16(64):1750–1792, 2011.

[Cel24] Michael Celentano. Sudakov–Fernique Post-AMP, and a New Proof of the Local Convexity of the TAP Free Energy. *The Annals of Probability*, 52(3), 2024. DOI:10.1214/23-AOP1675.

[Che24] Sinho Chewi. *Log-concave sampling*. 2024. URL:https://chewisinho.github.io/main.pdf.

[CPS23] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. Generalized TAP free energy. *Communications on Pure and Applied Mathematics*, 76(7):1329–1415, 2023.

[dAT78] J. R. L. de Almeida and D. J. Thouless. Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, 11(5):983–990, 1978. DOI:10.1088/0305-4470/11/5/028.

[DGW04] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *The Annals of Probability*, 32(3B):2702 – 2732, 2004. DOI:10.1214/009117904000000531.

[DLSS26a] Ewan Davies, Holden Lee, Juspreet Singh Sandhu, and Jonathan Shi. Potential Hessian Ascent III: Sampling the Sherrington–Kirkpatrick Model at Beta< 1/2. *arXiv preprint arXiv:2605.03718*, 2026.

[DLSS26b] Ewan Davies, Holden Lee, Juspreet Singh Sandhu, and Jonathan Shi. Weak Poincaré Inequalities via Approximate Stochastic Localization: Application to Sampling the Sherrington-Kirkpatrick Model. *arXiv preprint arXiv:2607.08160*, 2026.

[DS01] Kenneth R. Davidson and Stanislaw J. Szarek. Local operator theory, random matrices and Banach spaces. In William B. Johnson and Joram Lindenstrauss, editors, *Handbook of the Geometry of Banach Spaces*, volume 1, pages 317–366. North-Holland, Amsterdam, 2001. DOI:10.1016/S1874-5849(01)80010-3.

[EKZ22] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. A Spectral Condition for Spectral Gap: Fast Mixing in High-Temperature Ising Models. *Probability Theory and Related Fields*, 182(3):1035–1051, 2022. DOI:10.1007/s00440-021-01085-x.

[FMM21] Zhou Fan, Song Mei, and Andrea Montanari. TAP free energy, spin glasses and variational inference. *The Annals of Probability*, 49(1):1–45, 2021. DOI:10.1214/20-AOP1443.

[GSS23] Stephan Gufler, Adrien Schertzer, and Marius A Schmidt. On the concavity of the TAP free energy in the SK model. *Stochastic Processes and their Applications*, 164:160–182, 2023.

[JLNP25] David Jekel, Yoonkyeong Lee, Brent Nelson, and Jennifer Pi. Strong convergence to operator-valued semicirculars. *arXiv preprint arXiv:2506.19940*, 2025.

[JSS] David Jekel, Juspreet Singh Sandhu, and Jonathan Shi. Potential Hessian Ascent II: Duality, Robustness & State Evolution. In progress.

[JSS25] David Jekel, Juspreet Singh Sandhu, and Jonathan Shi. Potential Hessian Ascent: The Sherrington-Kirkpatrick Model. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 5307–5387. Society for Industrial and Applied Mathematics, 2025. DOI:10.1137/1.9781611978322.182.

[KSTZ26] Syamantak Kumar, Purnamrita Sarkar, Kevin Tian, and Yusong Zhu. High-Magnetization Sampling at Low Temperatures: Ising Models and Bayesian Sparse Linear Regression. *arXiv preprint arXiv:2609.08873*, 2026.

[LM19] Marc Lelarge and Léo Miolane. Fundamental Limits of Symmetric Low-Rank Matrix Estimation. *Probability Theory and Related Fields*, 173(3-4):859–929, 2019. DOI:10.1007/s00440-018-0845-x.

[Lop26a] Patrick Lopatto. Full replica symmetry breaking in the Sherrington-Kirkpatrick model. *arXiv e-prints*, pages arXiv–2607, 2026.

[Lop26b] Patrick Lopatto. Replica symmetry up to the de Almeida-Thouless line in the Sherrington-Kirkpatrick model. *arXiv preprint arXiv:2604.11921*, 2026. DOI:10.48550/arXiv.2604.11921.

[LSS26a] Holden Lee, Juspreet Singh Sandhu, and Jonathan Shi. Weak Poincaré Inequalities via Approximate Stochastic Localization II: Local regularity. In progress, 2026.

[LSS26b] Yutong Li, Juspreet Singh Sandhu, and Jonathan Shi. On overlap concentration in the Curie-Weiss Random Field model. *arXiv preprint arXiv:2609.08169*, 2026.

[Mon19] Andrea Montanari. Optimization of the Sherrington-Kirkpatrick Hamiltonian. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1417–1433, Baltimore, MD, USA, 2019. IEEE. DOI:10.1109/FOCS.2019.00087.

[MS17] James A Mingo and Roland Speicher. *Free probability and random matrices*, volume 35. Springer, 2017.

[Øks03] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 6 edition, 2003.

[OR00] James M Ortega and Werner C Rheinboldt. *Iterative solution of nonlinear equations in several variables*. SIAM, 2000.

[Sel25] Mark Sellke. Exponentially Slow Mixing of the Low Temperature SK Model. *arXiv preprint arXiv:2511.22621*, 2025.

[Sub21] Eliran Subag. Following the Ground States of Full-RSB Spherical Spin Glasses. *Communications on Pure and Applied Mathematics*, 74(5):1021–1044, 2021. DOI:10.1002/cpa.21922.

[TAP77] David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of'solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, 1977.

[Üst12] Ali Suleyman Üstünel. Transportation cost inequalities for diffusions under uniform distance. In *Stochastic Analysis and Related Topics: In Honour of Ali Süleyman Üstünel, Paris, June 2010*, pages 203–214. Springer, 2012.

[vH25] Ramon van Handel. The strong convergence phenomenon. *arXiv preprint arXiv:2507.00346*, 2025.

[Wan26] Sihan Wang. Optimal Mixing of Glauber Dynamics for the Sherrington-Kirkpatrick Model at $\beta < 1/2$. *arXiv preprint arXiv:2608.22159*, 2026.

A. Resolvent product regularity and subordination at the TAP value

This section has two goals – the first is a series of bounds on finite differences of matrices made of resolvent products, and the second is a proof of a spectral gap with respect to a particular value for the free convolution of a semicircular operator and a diagonal operator with explicit empirical density.

A.1. Regularity estimates for resolvent products.

The bounds here are finite-difference analogues of those presented in [DLSS26a, Lemma B.4]. They are quantitative estimates for finite differences between various matrices involving products of resolvents, when the minimum of the resolvent's spectrum is bounded to be greater than 0.

Lemma A.1 (Bounds on differences of resolvent products (analog of [DLSS26a, Lemma B.4]))**.** *Let $W_0, W_1 \in \mathbb{R}^{n\times n}$ be invertible symmetric matrices, and let $M_0, M_1 \in \mathbb{C}^{n\times n}$. Suppose that, for $j = 0, 1$,*

$$\min \mathrm{Spec}(|M_j + W_j^{-1}|) \geqslant \gamma, \qquad \|M_j\|_{\mathsf{op}} \leqslant C_1,$$

where $\gamma > 0$ and $C_1 \geqslant 0$. Let $C_2 := (\gamma + C_1)\gamma^{-1}$.

Let $D(W) := W^{-1}$, $X(W, M) := M + D(W)$, and $R(W, M) := X(W, M)^{-1}$ whenever the stated inverses exist.

Let Δ be the difference operator operating on functions of W and M, so that $\Delta f := f(W_1, M_1) - f(W_0, M_0)$. Let $T \in \mathbb{C}^{n\times n}$ be fixed and take all exponents k below to be integers. Then:

(a)
$$\|\Delta R\|_{L^2} \leqslant \gamma^{-2} \|\Delta M\|_{L^2} + C_2^2 \|\Delta W\|_{L^2}.$$

(b) For $k \geqslant 1$,
$$\left\|\Delta(R^k)\right\|_{L^2} \leqslant k\gamma^{-(k+1)} \|\Delta M\|_{L^2} + kC_2^2\gamma^{-(k-1)} \|\Delta W\|_{L^2}.$$

(c)
$$\|\Delta(DR)\|_{L^2} \leqslant C_2\gamma^{-1} \|\Delta M\|_{L^2} + C_1C_2^2 \|\Delta W\|_{L^2}.$$

(d) For $k \geqslant 1$,
$$\left\|\Delta(DR^k)\right\|_{L^2} \leqslant kC_2\gamma^{-k} \|\Delta M\|_{L^2} + (kC_2 - 1)C_2^2\gamma^{-(k-2)} \|\Delta W\|_{L^2}.$$

(e) For $k \geqslant 2$,
$$\left\|\Delta(DR^kD)\right\|_{L^2} \leqslant kC_2^2\gamma^{-(k-1)} \|\Delta M\|_{L^2} + (kC_2 - 2)C_2^3\gamma^{-(k-3)} \|\Delta W\|_{L^2}.$$

Proof. For f a matrix-valued function of W and M, we define the evaluation operators $f_1 := f(W_1, M_1)$ and $f_0 := f(W_0, M_0)$ so that $\Delta f = f_1 - f_0$. Then we have the discrete non-commutative sum, product, power, and inverse rules. Explicitly,

- $\Delta(f + g) = \Delta f + \Delta g$,
- $\Delta(fg) = f_1(\Delta g) + (\Delta f)g_0$,
- $\Delta(f^k) = \sum_{i=1}^{k} f_1^{i-1}(\Delta f)f_0^{k-i}$,
- $\Delta(f^{-1}) = -f_1^{-1}(\Delta f)f_0^{-1}$ (also known as the resolvent identity).

Note by the resolvent identity, $\Delta D = -D_1(\Delta W)D_0$ and $\Delta X = \Delta M - D_1(\Delta W)D_0$. By the argument of [DLSS26a, Lemma B.1], we have $\left\|R^k\right\|_{\mathsf{op}} \leqslant \gamma^{-k}$ and $\left\|DR^{k+1}\right\|_{\mathsf{op}} \leqslant C_2\gamma^{-k}$, which we will use in every following part. We will also repeatedly use the non-commutative Hölder inequality.

Part A.1(a). We apply the just-mentioned operator norm bounds after computing the derivative

$$\Delta R = -R_1(\Delta X)R_0 = -R_1(\Delta M)R_0 + (RD)_1(\Delta W)(DR)_0.$$

Part A.1(b). We use the discrete non-commutative power rule

$$\Delta(R^k) = \sum_{j=1}^{k} R_1^{j-1}(\Delta R)R_0^{k-j},$$

then apply Lemma A.1(a) and the same operator norm bounds from before.

Part A.1(c). By the resolvent identity, $DR = I - MR$. Hence $\Delta(DR) = -M_1(\Delta R) - (\Delta M)R_0$. Then, apply Lemma A.1(a) and $\|M\|_{\mathsf{op}} \leqslant C_1$ and $C_2 = 1 + C_1\gamma^{-1}$.

Part A.1(d). Use the discrete non-commutative product rule to compute

$$\Delta(DR^k) = (DR)_1\Delta(R^{k-1}) + \Delta(DR)R_0^{k-1}.$$

Then apply Lemma A.1(c) and Lemma A.1(b) and, for the $\|\Delta W\|_{L^2}$ term, $C_1C_2^2\gamma^{-(k-1)} + (k-1)C_2^3\gamma^{-(k-2)} = (kC_2 - 1)C_2^2\gamma^{-(k-2)}$.

Part A.1(e). Use the discrete non-commutative product rule to compute

$$\Delta(DR^kD) = (DR)_1\Delta(R^{k-1}D) + \Delta(DR)(R^{k-1}D)_0,$$

then apply Lemma A.1(c) and Lemma A.1(d) and collect terms. □

A.2. **Free analysis at the TAP value of $a = \beta^2 \operatorname{tr}_n[D^{-1}(m)]$.** The following deterministic lemma gives the diagonal control and spectral gap in the free limit used in Corollary 5.9. For the following lemma, D and S are free from each other.

Lemma A.2 (The TAP free limit for $\beta < 1$)**.**
Let $0 < \beta < 1$, $D \geqslant I_n$, and $a = \beta^2 \operatorname{tr}_n(D^{-1})$. Then

$$aI_n - \beta S + D \geqslant (1-\beta)^2 I_n. \tag{A.1}$$

For all

$$0 < \gamma_Y \leqslant \frac{(1-\beta)^4}{4(1+2\beta^2)^2}, \tag{A.2}$$

it is true uniformly in $D \geqslant I_n$ that

$$(I_n + D^{-1}(aI_n - \beta S))(I_n + (aI_n - \beta S)D^{-1}) \geqslant 4\gamma_Y I_n.$$

For every such D and a,

$$E_{M_n}\big[D(aI_n - \beta S + D)^{-2}D\big] = \frac{I_n}{1 - \beta^2 \operatorname{tr}_n(D^{-2})}.$$

Proof. We first show that the range of the subordination function allows one to conclude a gap between $a = \beta^2 \operatorname{tr}_n[D^{-1}]$ and the supremum of the bulk spectrum of the free convolution of βS and $-D$ when $\beta < 1$. We then use the inverse identity and analyticity of the subordination function at $z = a$ to explicitly evaluate F and its derivative to compute the free target.

The free gap. The real inverse-subordination map for $\beta S - D$ [DLSS26a, Lemma B.18] is

$$u \longmapsto u + \beta^2 \operatorname{tr}_n[(D + uI_n)^{-1}].$$

For $u > \beta - 1$, it is defined to the right of $\operatorname{Spec}(-D)$, and its derivative is strictly positive since $\beta^2 \operatorname{tr}_n[(D + uI_n)^{-2}] < 1$. The support characterization in [CDMF+11, Proposition 2.2] identifies this interval with part of the right exterior branch. Letting $u \downarrow \beta - 1$ gives

$$\sup \operatorname{Spec}(\beta S - D) \leqslant \beta - 1 + \beta^2 \operatorname{tr}_n[(D + (\beta-1)I_n)^{-1}].$$

Indeed, the image is an interval connected to $+\infty$ and disjoint from the free support. Subtracting this upper bound from a yields

$$\begin{aligned} a - \sup \operatorname{Spec}(\beta S - D) &\geqslant (1-\beta)\big[1 - \beta^2 \operatorname{tr}_n\big(D^{-1}(D + (\beta-1)I_n)^{-1}\big)\big] \\ &\geqslant (1-\beta)^2, \end{aligned}$$

because $D^{-1}(D + (\beta - 1)I_n)^{-1} \preceq \beta^{-1} I_n$. Faithfulness of the trace identifies the free support with the operator spectrum, proving (A.1).

The product cutoff and the target. The inverse identity and the free gap imply

$$\begin{aligned}\left\|(I_n + D^{-1}(aI_n - \beta S))^{-1}\right\|_{\mathsf{op}} &= \left\|(aI_n - \beta S + D)^{-1} D\right\|_{\mathsf{op}} \\ &\leqslant 1 + \frac{a + 2\beta}{(1-\beta)^2} \leqslant \frac{1 + 2\beta^2}{(1-\beta)^2}.\end{aligned}$$

Thus the free product is at least $(1-\beta)^4(1+2\beta^2)^{-2} I_n$, and (A.2) implies the required buffer.

Let $F = F_{\beta,-D}$ be the subordination function. The inverse-subordination map above has positive derivative at $u = 0$ and takes 0 to a. This implies its (exterior) inverse satisfies

$$F(a) = 0, \qquad F'(a) = \frac{1}{1 - \beta^2 \operatorname{tr}_n(D^{-2})}.$$

Subordination under a non-commutative conditional expectation [Bia98, Theorem 3.1] gives $E_{M_n}[(zI_n - \beta S + D)^{-1}] = (F(z)I_n + D)^{-1}$. Note that $z = a$ is outside the free spectrum by (A.1), and so $\partial_z F(z)$ exists at $z = a$. Evaluating the derivative and conjugating by D proves the asserted free target. □

B. Self-stability for the cavity interpolation of [DLSS26a, §5] at $\beta < 1$

Extension of cavity interpolation theory in [DLSS26a, §5] to $\beta < 1$. To see that desideratum (1) holds for the full range $\beta < 1$ (as opposed to the stated bound of $\beta < 1/2$ in [DLSS26a]) an inspection of the arguments in [DLSS26a, §4, §5, §A] makes it clear only two statements need to be strengthened:

(1) **Nishimori condition up to $\beta < 1$:** The fixed-point equations for (m^*, q^*) stated in [DLSS26a, Theorem 5.1] have unique solutions for $\beta < 1$ which satisfy the Nishimori condition that $m^* = q^*$.
(2) **Spectral self-stability for $\mathbf{M}$ up to $\beta < 1$:** The operator $\mathbf{M}$ in [DLSS26a, §5] should satisfy the spectral condition that $\beta^2 \rho(\mathbf{M}) < 1$ for every $\beta < 1$, as opposed to $\beta < 1/2$ as shown in [DLSS26a, §5.5].

As it turns out, both these conditions are rather elementary to show. The first condition is a direct consequence of [LSS26b, Lemma 2.1]. The second condition follows by avoiding the convenient (but lossy) step of using the Frobenius norm to bound the operator norm in the proof of [DLSS26a, Lemma 5.31] and replacing it with the use of the Nishimori identity ($m^* = q^*$) to simplify the entries of operator $\mathbf{X}$ followed by conjugation with a diagonal matrix to obtain a similar matrix $\mathbf{K}(z)$ which is anti-symmetric and whose largest singular value is bounded by 1. This leads to the bound $\rho(\mathbf{M}) < 1$ from which it straightforwardly follows that $\beta^2 \rho(\mathbf{M}) < 1$ for $\beta < 1$.

Corollary B.1 (Overlap and magnetization concentration for $\beta < 1$)**.**
For $0 < \beta < 1$ and any fixed time $0 \leqslant t \leqslant T(\beta) < \infty$, and every positive integer $k \in \mathbb{Z}_+$,

$$\mathop{\mathbb{E}}_{A,B_t} \left\langle (R_{12} - q^*)^{2k} \right\rangle_t \leqslant \frac{C(k)}{n^k} + C(\beta, k) e^{-c(k,t)n},$$

and

$$\mathop{\mathbb{E}}_{A,B_t} \left\langle (M - m^*)^{2k} \right\rangle_t \leqslant \frac{C'(k)}{n^k} + C'(\beta, k) e^{-c(k,t)n}$$

Proof. This follows straightforwardly after invoking [LSS26b, Lemma 2.1], which implies that there is a unique pair $(m^*, q^*) \in [0,1)^2$ that solve the fixed-point equations of [DLSS26a, Theorem 5.1] for $\beta < 1$. An immediate consequence of this is that the proof of [DLSS26a, Theorem 5.2] implies magnetization concentration for $\beta < 1$ as well. For overlap concentration, note that the entire argument of [DLSS26a, Theorem 5.1] which verifies requirements (1) through (7) in the proof are valid provided that $q^*(\lambda) = q^*$ and $\beta^2 \leqslant \lambda$ even when $\beta < 1$. For the first, we replace the use of [DLSS26a, §5.6] with [LSS26b, Lemma 2.1], and for the latter, observe that $F(\lambda, q)$ [LM19, §2.3] continues to have a unique supremum over q for $\lambda \leqslant 1$. □

Lemma B.2 ($\beta^2\rho(\mathbf{M}) < 1$ for $\beta < 1$, [DLSS26a, Minor generalization of §5.5])**.**
Let Y^* be a random variable that is almost-surely finite. Set $T := \tanh(Y^*) \in [-1,1]$. Define

$$\mu_r := \mathbb{E}[T^r], \qquad r \in \{1,2,3,4\},$$

and

$$\mathbf{M} := \begin{pmatrix} 1 - 4\mu_2 + 3\mu_4 & 2(\mu_1 - \mu_3) \\ \mu_3 - \mu_1 & 1 - \mu_2 \end{pmatrix}.$$

Then for every $\beta < 1$,

$$\rho(\beta^2\mathbf{M}) \;<\; 1.$$

Proof. The proof of [DLSS26a, Lemma 5.31] implies that, with $T = \tanh(Y^*) \in [-1,1]$,

$$\mathbf{M} = \mathbb{E}[\mathbf{X}(T)]\,,$$

where

$$\mathbf{X}(z) := (1-z^2)\begin{pmatrix} 1-3z^2 & 2z \\ -z & 1 \end{pmatrix}.$$

Set $\mathbf{S} := \mathsf{diag}(1,\sqrt{2})$ and define

$$\mathbf{K}(z) := \mathbf{S}\mathbf{X}(z)\mathbf{S}^{-1} = (1-z^2)\begin{pmatrix} 1-3z^2 & \sqrt{2}\,z \\ -\sqrt{2}\,z & 1 \end{pmatrix},$$

which is similar to $\mathbf{X}(z)$. We show that $\|\mathbf{K}(z)\|_{\mathrm{op}} \leqslant 1$ for every $z \in [-1,1]$ by proving that the slack matrix $\mathbf{R}(z) := I - \mathbf{K}(z)^\mathsf{T}\mathbf{K}(z) \succeq 0$. To see this, let $u := z^2 \in [0,1]$ and note that a direct computation gives

$$\mathbf{R}(z)_{22} = u^2(3-2u)\,,$$
$$\det(\mathbf{R}(z)) = u^4(u^2-6u+6)\,.$$

For $0 < u \leqslant 1$, both quantities are positive, since $3-2u \geqslant 1$ and $u^2-6u+6 \geqslant 1$. The Schur-complement criterion, therefore, gives $\mathbf{R}(z) \succeq 0$ (since $\mathbf{K}(0) = I_2$ and $\mathbf{R}(0) = 0$). This yields

$$\mathbf{K}(z)^\mathsf{T}\mathbf{K}(z) \preceq I_2\,.$$

$\mathbf{S}$ is deterministic, and convexity of the operator norm gives

$$\|\mathbf{S}\mathbf{M}\mathbf{S}^{-1}\|_{\mathrm{op}} = \|\,\mathbb{E}[\mathbf{K}(T)]\|_{\mathrm{op}} \leqslant_{\text{Jensen's}} \mathbb{E}\big[\|\mathbf{K}(T)\|_{\mathrm{op}}\big] \leqslant 1\,.$$

Finally, similarity preserves the spectral radius, giving

$$\rho(\beta^2\mathbf{M}) = \beta^2\rho(\mathbf{S}\mathbf{M}\mathbf{S}^{-1}) \leqslant \beta^2\|\mathbf{S}\mathbf{M}\mathbf{S}^{-1}\|_{\mathrm{op}} \leqslant \beta^2 <_{\beta<1} 1\,. \qquad \square$$

We now state a lemma which uses Corollary B.1 and Lemma B.2 to replace invocations of overlap and magnetization concentration, and the validity of the cavity estimates of [DLSS26a, §5], in the proof of [DLSS26a, Theorem 4.1], to work all the way for $\beta < 1$. Consequently, the same covariance estimate as in [DLSS26a, Theorem 4.1] is implied for the range of $0 < \beta < 1$.

Theorem B.3 (Error of covariance estimate for $\beta < 1$)**.**
Assume overlap and magnetization concentration as given in Corollary B.1 and the spectral radius bound for the self-stability operator $\mathbf{M}$ defined in Lemma B.2. Then, for $0 < \beta < 1$,

$$\mathop{\mathbb{E}}_{A,B_t}\big\|\widehat{Q}^{-1}(m_t)Q(m_t) - I_n\big\|_F^2 \leqslant \varepsilon(\beta,t)\,,$$

for every $t \in [0,T(\beta)]$ with $\sup_{t\in[0,T(\beta)]}\varepsilon(\beta,t) < \infty$.

Proof. The proof follows by observing that, in the proof of [DLSS26a, Theorem 4.1], all estimates in [DLSS26a, Lemmata 4.11–4.16] rely only on, after application of Stein's lemma and algebraic identities involving independent replicas, three things:

(1) Magnetization and overlap concentration (cf. [DLSS26a, Theorems 5.1–5.2]).
(2) Enhanced concentration for linear and cubic observables of bulk deviations (cf. [DLSS26a, §5.4]).
(3) Exact cavity estimates for the "heavy" cancellations (cf. [DLSS26a, §5.4]).

The first point is directly valid for $0 < \beta < 1$ due to Corollary B.1. The second and third point rely on the fact that magnetization and overlaps have first moments which fluctuate on the $O(1/n)$ scale [DLSS26a, §5.4] and that various denominators in the second-order Taylor expansions used to compute exact cavity estimates via self-reducibility do not diverge [DLSS26a, §5.4]. A simple inspection of the argument in [DLSS26a, §5] with the strengthened spectral radius bound $\beta^2\rho\mathbf{M} < 1$ and the fact that $\beta^2\,\mathbb{E}_{Y*}[\mathrm{sech}^{2k}(Y^*)] < 1$ for $0 < \beta < 1$ immediately implies the validity of all estimates in [DLSS26a, §5]. □

Nishimori identity for $\mathbb{E}[\tanh^3(Y^*)] = \mathbb{E}[\tanh^4(Y^*)]$. We write a brief identity that shows the Nishimori condition for the third and fourth moments in the presence of a replica–symmetric cavity field Y^* that is induced by the SL process.

Lemma B.4 (Nishimori identity for 3rd and 4th moments)**.**
Fix $t \geqslant 0$ and let $z \sim \mathcal{N}(0,1)$. Denote $Y^* = \beta^2 m^* + t + \sqrt{\beta^2 q^* + t}\,h$ where (m^*, q^*) solve the fixed-point equations in [DLSS26a, Theorem 5.1]. *Then,*

$$\mathop{\mathbb{E}}_{z}\left[\tanh^3(Y^*)\right] = \mathop{\mathbb{E}}_{z}\left[\tanh^4(Y^*)\right].$$

Proof. By [LSS26b, Lemma 2.1], we have $m^* = q^*$ for $0 < \beta < 1$. Following the same notation and strategy as in [LSS26b, Lemma 2.1] and [DLSS26a, §5.6], let $r := t + \beta^2 q^* \geqslant 0$ so that $Y^* \sim \mathcal{N}(r, r)$. If $r = 0$, then $Y^* = 0$ almost surely and both expectations vanish. Therefore, assume that $r > 0$ for the rest of the proof.

Let p_r denote the density of Y^*. As in the proof of [LSS26b, Lemma 2.1] a direct computation gives

$$p_r(-y) = e^{-2y} p_r(y).$$

Since $\tanh(y) = (1 - e^{-2y})/(1 + e^{-2y})$, it follows that

$$p_r(y) - p_r(-y) = \tanh(y)\big(p_r(y) + p_r(-y)\big).$$

Pairing the contributions from y and $-y$, and using the oddness of $\tanh^3$ and the evenness of $\tanh^4$, yields

$$\begin{aligned}\mathbb{E}[\tanh^3(Y^*)] &= \int_0^\infty \tanh^3(y)\big(p_r(y) - p_r(-y)\big)\,dy \\ &= \int_0^\infty \tanh^4(y)\big(p_r(y) + p_r(-y)\big)\,dy \\ &= \mathbb{E}[\tanh^4(Y^*)].\end{aligned}$$

□

C. Algorithmic error analysis

The following is an algorithm and standard guarantee for rejection sampling with unknown normalizing constant.

Algorithm 4 Approximate rejection sampler with unknown normalization

1: **Input:** Oracle for sampling from $\widetilde{Q} \approx Q$, (possibly randomized) function $\widetilde{R} \approx R$ where $\frac{dP}{dQ} \propto R$, cutoff parameters C_1, C_2, failure probability δ'
2: Draw N samples from $\widetilde{Q}$, where $N = \Omega(C_1^2 \log(1/\delta'))$ for appropriate constant.
3: Let R_0 be the pth quantile of the samples, where $p = 1 - \frac{1}{6C_1}$.
4: **repeat**
5: Draw $X \sim \widetilde{Q}$.
6: Draw $U \sim \mathsf{Uniform}([0,1])$.
7: **until** $U \leqslant \frac{\widetilde{R}(X)}{C_3 R_0}$, where $C_3 = 8C_2$.
8: **Output:** X.

Lemma C.1 (Approximate rejection sampling with unknown normalization, [DLSS26a, Lemma 3.8])**.**
Suppose that Algorithm 4 is run with $\widetilde{Q}$, $\widetilde{R}$, and parameters C_1, C_2 satisfying the following.

(1) *(Sampling error)* $\widetilde{Q}$ *is a distribution such that* $\mathrm{TV}(\widetilde{Q}, Q) \leqslant \varepsilon_{\mathrm{sample}}$.
(2) *(Tails of ratio) Let*

$$C(\varepsilon) = \inf_{C} \left\{ C : \mathop{\mathbb{E}}_{Q} \left(\frac{dP}{dQ} \wedge C \right) \geqslant 1 - \varepsilon \right\},$$

and suppose $C_1 \geqslant C(1/2) \vee 1$, $C_2 \geqslant C(\varepsilon_{\mathrm{weight}})$.
(3) *(Ratio error) Let* G_1, G_2 *be events such that*

$$G_1 \subseteq \left\{ \frac{\widetilde{R}}{R} \in [0, e^{\varepsilon_{\mathrm{ratio}}}] \right\}$$

$$G := G_1 \cap G_2 \subseteq \left\{ \frac{\widetilde{R}}{R} \in [e^{-\varepsilon_{\mathrm{ratio}}}, e^{\varepsilon_{\mathrm{ratio}}}] \right\}$$

and $\widetilde{Q}(G_1^c) \leqslant \delta_1$, $P(G_2^c) \leqslant \delta_2$. *(Note these probabilities can involve randomness in the algorithm.)*

Then letting $\widehat{P}$ *be the output distribution, with probability* $\geqslant 1 - \delta'$ *over the* N *initial samples,*

$$\mathrm{TV}(\widehat{P}, P) \leqslant \varepsilon_{\mathrm{weight}} + \delta_2 + 384 C_1 C_2 (\delta_1 + \varepsilon_{\mathrm{ratio}} + \varepsilon_{\mathrm{sample}})$$

and if $\varepsilon_{\mathrm{weight}} + \delta_2 + 96 C_1 C_2 (\delta_1 + \varepsilon_{\mathrm{ratio}} + \varepsilon_{\mathrm{sample}}) \leqslant \frac{1}{2}$, *the acceptance probability is at least* $\frac{1}{192 C_1 C_2}$.

Proof (Proof of Lemma 3.16). Define the good sets on path space (1) the ratio between the ideal and approximate distribution is bounded, (2) the ratio between the scaffolding and approximate distribution is not too small, and (3) the weights are bounded,

$$G_1 = \left\{ \frac{d\mathbf{P}_T^s}{d\widehat{\mathbf{P}}_T}(\gamma_T) \leqslant L \right\} \qquad G_2 = \left\{ \frac{e^{\widehat{w}_T}}{\widehat{Z}_T} \geqslant c_3 \right\} \qquad G_3 = \left\{ \frac{e^{\widehat{w}_T}}{\widehat{Z}_T} \leqslant C_1 \right\}$$

and let their complements be B_1, B_2, B_3, respectively. Now

$$\begin{aligned} &\mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T} \frac{R_2(\widehat{x}_T)}{Z_2} \mathbb{1}_{S_T \cap (B_1 \cup B_2 \cup B_3)} \right] \\ &\qquad \leqslant \mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T} \frac{R_2(\widehat{x}_T)}{Z_2} \mathbb{1}_{S_T \cap B_1} \right] + \mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T} \frac{R_2(\widehat{x}_T)}{Z_2} \mathbb{1}_{S_T \cap G_1 \cap B_2} \right] + \mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T} \frac{R_2(\widehat{x}_T)}{Z_2} \mathbb{1}_{S_T \cap G_1 \cap G_2 \cap B_3} \right]. \end{aligned} \tag{C.1}$$

We bound each term separately. Note that if $\mu \ll \nu$, then

$$\begin{aligned} &\mathrm{KL}(\mu \| \nu) = \mathop{\mathbb{E}}_{\mu}\left(\log \frac{d\mu}{d\nu} \right) \mathbb{1}_{\frac{d\mu}{d\nu} \geqslant 1} + \mathop{\mathbb{E}}_{\mu}\left(\log \frac{d\mu}{d\nu} \right) \mathbb{1}_{\frac{d\mu}{d\nu} < 1} \geqslant \mathop{\mathbb{E}}_{\mu}\left(\log \frac{d\mu}{d\nu} \right) \mathbb{1}_{\frac{d\mu}{d\nu} \geqslant 1} - 1 \geqslant \mu\left(\frac{d\mu}{d\nu} \geqslant L \right) (\log L) - 1 \\ &\implies \mu\left(\frac{d\mu}{d\nu} \geqslant L \right) \leqslant \frac{\mathrm{KL}(\mu \| \nu) + 1}{\log L}. \end{aligned} \tag{C.2}$$

To apply to our setting, we define the stopped process by

$$dx_t^{\mathrm{stop}} = \mathbb{1}_{t < \tau'} f_t \, dt + dB_t$$

and similarly define $\widehat{x}_t^{\mathrm{stop}}$. Let their laws up to time T be $\mathbf{P}_t^{\mathrm{stop}}$ and $\widehat{\mathbf{P}}_t^{\mathrm{stop}}$, respectively. Then

$$\begin{aligned} \mathrm{KL}(\mathbf{P}_T^{\mathrm{stop}} \| \widehat{\mathbf{P}}_T^{\mathrm{stop}}) &= \mathop{\mathbb{E}}_{\mathbf{P}_T^{\mathrm{stop}}} \log \frac{d\mathbf{P}_T^{\mathrm{stop}}}{d\widehat{\mathbf{P}}_T^{\mathrm{stop}}} \\ &= -\mathbb{E} \int_0^{T \wedge \tau'} \langle \widehat{f}_t - f_t, dB_t \rangle + \frac{1}{2} \mathbb{E} \int_0^{T \wedge \tau'} \left\| \widehat{f}_t - f_t \right\|^2 dt \\ &= \frac{1}{2} \int_0^{T \wedge \tau'} \mathbb{E} \left\| \widehat{f}_t - f_t \right\|^2 dt \leqslant \frac{1}{2} \int_0^T \varepsilon(t)^2 \, dt. \end{aligned} \tag{C.3}$$

by assumption 1 and Girsanov's theorem.

For the first term of (C.1), noting that $x_{[0,T]} = x_{[0,T]}^{\mathrm{stop}}$ if $T < \tau'$ so that $\mathbf{P}_T^s = \mathbf{P}_T^{\mathrm{stop}}$ restricted to $\{T < \tau'\}$,

$$\mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T} \frac{R_2(\widehat{x}_T)}{Z_2} \mathbb{1}_{S_T \cap B_1} \right] = \mathbb{E}\left[\frac{d\mathbf{P}_T^s}{d\widehat{\mathbf{P}}_T}(\widehat{x}_{[0,T]}) \mathbb{1}_{S_T \cap B_1}(\widehat{x}_{[0,T]}) \right] = \mathbb{E}\left[\mathbb{1}_{S_T \cap B_1}(x_{[0,T]}) \right]$$

$$\leqslant \mathbb{P}\left[\frac{d\mathbf{P}_T^s}{d\widehat{\mathbf{P}}_T}(x_{[0,T]}) > L\right] \leqslant \mathbb{P}\left[\frac{d\mathbf{P}_T^{\mathrm{stop}}}{d\widehat{\mathbf{P}}_T^{\mathrm{stop}}}(x_{[0,T]}^{\mathrm{stop}}) > L\right] + \mathbf{P}_T[\tau' \leqslant T]$$

$$\leqslant_{(C.2),\ \text{assumption 1}} \frac{\mathrm{KL}(\mathbf{P}_T^{\mathrm{stop}} \| \widehat{\mathbf{P}}_T^{\mathrm{stop}}) + 1}{\log L} + \delta' \leqslant_{(C.3)} \frac{\frac{1}{2}\int_0^T \varepsilon(t)^2 dt + 1}{\log L} + \delta' \leqslant \frac{\varepsilon}{3} + \delta'$$

by choosing $L = \exp\left(\frac{3}{\varepsilon}\left(\frac{1}{2}\int_0^T \varepsilon(t)^2 dt + 1\right)\right)$. Next, note that directly by assumption 3, $\mathbb{P}[S_T \cap B_2] \leqslant \varepsilon_3$. Hence we can bound the second term

$$\mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T}\frac{R_2(\widehat{x}_T)}{Z_2}\mathbb{1}_{S_T\cap G_1\cap B_2}\right] = \mathbb{E}\left[\frac{d\mathbf{P}_T^s}{d\widehat{\mathbf{P}}_T}(\widehat{x}_{[0,T]})\mathbb{1}_{S_T\cap G_1\cap B_2}\right] \leqslant L\,\mathbb{P}[S_T \cap B_2] \leqslant L\varepsilon_3.$$

Finally, note that under $S_T \cap G_1 \cap G_2$, $\frac{R_2(\widehat{x}_T)}{Z_2} = \frac{d\mathbf{P}_T^s}{d\widehat{\mathbf{P}}_T}\Big/\frac{e^{\widehat{w}_T}}{\widehat{Z}_T} \leqslant \frac{L}{c_3}$, so for the third term, by assumption 2,

$$\mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T}\frac{R_2(\widehat{x}_T)}{Z_2}\mathbb{1}_{S_T\cap G_1\cap G_2\cap B_3}\right] \leqslant \frac{L}{c_3}\,\mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T}\mathbb{1}_{S_T\cap B_3}\right] \leqslant \frac{L\varepsilon_1}{c_3}.$$

Putting everything together,

$$\mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T}\frac{R_2(\widehat{x}_T)}{Z_2}\mathbb{1}_{S_T\cap(B_1\cup B_2\cup B_3)}\right] \leqslant \frac{\varepsilon}{3} + \delta' + L\varepsilon_3 + \frac{L\varepsilon_1}{c_3} = \varepsilon + \delta'$$

by choosing $\varepsilon_3 = \frac{\varepsilon}{3L}$ and $\varepsilon_1 = \frac{c_3\varepsilon}{3L}$. Hence

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{S_T}\frac{e^{\widehat{w}_T}}{\widehat{Z}_T}\frac{R_2(\widehat{x}_T)}{Z_2} \wedge \frac{C_1 L}{c_3}\right] &\geqslant \mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T}\frac{R_2(\widehat{x}_T)}{Z_2}\mathbb{1}_{S_T\cap G_1\cap G_2\cap G_3}\right] \\ &= \mathbb{E}\left[\mathbb{1}_{S_T}\frac{d\mathbf{P}_T^s}{d\widehat{\mathbf{P}}_T}(\widehat{x}_T)\right] - \mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T}\frac{R_2(\widehat{x}_T)}{Z_2}\mathbb{1}_{S_T\cap(B_1\cup B_2\cup B_3)}\right] \\ &= \mathbb{E}\left[\mathbb{1}_{S_T}(x_{[0,T]})\right] - \mathbb{E}\left[\frac{e^{\widehat{w}_T}}{\widehat{Z}_T}\frac{R_2(\widehat{x}_T)}{Z_2}\mathbb{1}_{S_T\cap(B_1\cup B_2\cup B_3)}\right] \geqslant 1 - \delta - \delta' - \varepsilon.\end{aligned}$$

The first part now follows from Lemma C.1 applied to path measures $Q = \widehat{\mathbf{P}}_T$ and $P = \mathbf{P}_T$, with $\varepsilon_{\mathrm{weight}} = \varepsilon+\delta+\delta'$, $\delta_1 = \delta_2 = \varepsilon_{\mathrm{ratio}} = \varepsilon_{\mathrm{sample}} = 0$, and then taking the marginal at time T. Note we allow probability ε of failure (in the quantile estimate or rejection sampling taking too many trials).

For the second part, let $\widetilde{Q} = \widetilde{\mathbf{P}}_T$ be the path measure of $(\widetilde{x}_t)_{t\in[0,T]}$. Let $\widetilde{R} = e^{\widetilde{w}_T}\widetilde{R}_2$, and $R = e^{\widehat{w}_T}R_2$. Note that assumption 4 gives assumption 1 of Lemma C.1 with $\varepsilon_{\mathrm{sample}} = \varepsilon_4$. Consider the events

$$G_1' = \left\{\frac{\widetilde{R}_2}{R_2} \in [0, e^{\varepsilon_4/2}]\right\} \cap \left\{|\widetilde{w}_T - \widehat{w}_T| \leqslant \frac{\varepsilon_4}{2}\right\} \subseteq \left\{\frac{\widetilde{R}}{R} \in [0, e^{\varepsilon_4}]\right\}.$$

$$G_2' = \left\{\frac{\widetilde{R}_2}{R_2} \in [e^{-\varepsilon_4/2}, e^{\varepsilon_4/2}]\right\},$$

and note $G_1' \cap G_2' \subseteq \left\{\frac{\widetilde{R}}{R} \in [e^{-\varepsilon_4}, e^{\varepsilon_4}]\right\}$. It remains to check assumption 3 of Lemma C.1. We check

$$\widetilde{\mathbf{P}}_T({G_1'}^c) \leqslant \widetilde{\mathbf{P}}_T^s({G_1'}^c) + \widehat{\delta} \leqslant \mathbb{P}\left(\frac{\widetilde{R}_2(\widetilde{x}_T)}{R_2(\widetilde{x}_T)} \notin [0, e^{\varepsilon_4/2}]\right) + \widetilde{\mathbf{P}}_T^s\left(|\widetilde{w}_T - \widehat{w}_T| > \frac{\varepsilon_4}{2}\right) + \widehat{\delta} \leqslant_{(5),(6)(a)} \frac{\varepsilon_4}{2} + \frac{\varepsilon_4}{2} + \widehat{\delta} = \varepsilon_4 + \widehat{\delta}$$

$$\mathbf{P}_T({G_1'}^c) = \mathbb{P}\left(\frac{\widetilde{R}_2(\widetilde{x}_T)}{R_2(\widetilde{x}_T)} \notin [e^{-\varepsilon_4/2}, e^{\varepsilon_4/2}]\right) \leqslant_{(6)(b)} \varepsilon.$$

Thus the assumption holds with $\varepsilon_{\mathrm{weight}} = \varepsilon + \delta + \delta'$, $\delta_2 = \varepsilon$, $\delta_1 = \varepsilon_{\mathrm{sample}} = \varepsilon_4 + \widehat{\delta}$, and $\varepsilon_{\mathrm{ratio}} = \varepsilon_4$. The final TV error of the sampled distribution $\widehat{P}$ is

$$\begin{aligned}\mathrm{TV}(\widehat{P}, P) &\leqslant \varepsilon_{\mathrm{weight}} + \delta_2 + 384C_1'C_2'(\delta_1 + \varepsilon_{\mathrm{ratio}} + \varepsilon_{\mathrm{sample}}) \\ &\leqslant \varepsilon + \delta + \delta' + \varepsilon + 384{C_1'}^2(\varepsilon_4 + \widehat{\delta} + \varepsilon_4 + \varepsilon_4 + \widehat{\delta}) = 5\varepsilon + \delta + \delta' + 768{C_1'}^2\widehat{\delta}.\end{aligned}$$

Allowing probability ε of failure then gives the result. □