Proceedings of Machine Learning Research vol 313:1–17, 2026 37th International Conference on Algorithmic Learning Theory

Talagrand Meets Talagrand: Upper and Lower Bounds on Expected Soft Maxima of Gaussian Processes with Finite Index Sets

Yifeng Chu YCHU26@ILLINOIS.EDU
University of Illinois

Maxim Raginsky MAXIM@ILLINOIS.EDU
University of Illinois

Editors: Matus Telgarsky and Jonathan Ullman

Abstract

Analysis of extremal behavior of stochastic processes is a key ingredient in a wide variety of applications, including probability, statistical physics, theoretical computer science, and learning theory. In this paper, we consider centered Gaussian processes on finite index sets and investigate expected values of their smoothed, or "soft," maxima. We obtain upper and lower bounds for these expected values using a combination of ideas from statistical physics (the Gibbs variational principle for the equilibrium free energy and replica-symmetric representations of Gibbs averages) and from probability theory (Sudakov minoration). These bounds are parametrized by an inverse temperature $\beta > 0$ and reduce to the usual Gaussian maximal inequalities in the zero-temperature limit $\beta \to \infty$. We provide an illustration of our methods in the context of the Random Energy Model, one of the simplest models of physical systems with random disorder.

Keywords: Gaussian processes; maximal inequalities; information theory; disordered systems

1. Introduction and informal summary of results

The study of extremal behavior of stochastic processes is of fundamental importance in a wide variety of applications, including pure and applied probability, theoretical computer science, mathematical statistics, and statistical physics. For example, the method of generic chaining (Talagrand, 2014) gives sharp upper and lower bounds on the expected supremum of a centered Gaussian process $X = (X_t)_{t\in T}$ in terms of the geometry of the index set T induced by the natural metric $d(s,t) := (\mathbf{E}|X_s - X_t|^2)^{1/2}$. The underlying intuition here is that, for finite T, $\sqrt{\log|T|}$ is a good proxy for the expected supremum of the centered Gaussian process $(X_t)_{t\in T}$, in the sense that

$$a\sqrt{\log|T|} \lesssim \mathbf{E}\Big[\sup_{t\in T} X_t\Big] \lesssim \Delta\sqrt{\log|T|}, \tag{1}$$

where $a = \min\{d(s,t) : s,t \in T, s \neq t\}$ is the minimum separation between the elements of T and $\Delta = \max\{d(s,t) : s,t \in T\}$ is the diameter of T. Here, the upper bound is by the well-known maximal inequality for Gaussian random variables, while the lower bound follows from Sudakov minoration. Another important observation is that the zero-mean Gaussian random variables X_s and X_t are highly correlated when $d(s,t)$ is very small and nearly independent when $d(s,t)$ is very large. The method of generic chaining is, essentially, a multiscale iterative application of these ideas, where at each scale $n = 0,1,2,\ldots$ one approximates the process X by its "subsampled" version at a finite set of indices $T_n \subseteq T$ satisfying the cardinality bound $|T_n| < 2^{2^n}$.

While the expected supremum is a natural measure of the size of X, there are various "smoothed" versions of it that may be of interest in their own right, for example in the context of concentration

© 2026 Y. Chu & M. Raginsky.

and superconcentration phenomena in Gaussian random fields (Chatterjee, 2014). For example, if T is finite, then $\max_{t\in T} X_t$ can be bounded from above and from below by the "β-softmax" for every $\beta > 0$:

$$\max_{t\in T} X_t \le \frac{1}{\beta}\log\sum_{t\in T} e^{\beta X_t} \le \max_{t\in T} X_t + \frac{\log|T|}{\beta}. \tag{2}$$

Another smooth proxy for the maximum is

$$g(X;\beta) := \frac{1}{Z(\beta)}\sum_{t\in T} X_t e^{\beta X_t}, \tag{3}$$

where $Z(\beta) := \sum_{t\in T} e^{\beta X_t}$ is known as the partition function. It satisfies the inequalities

$$\max_{t\in T} X_t - \frac{\log|T|}{\beta} \le g(X;\beta) \le \max_{t\in T} X_t; \tag{4}$$

moreover, as we increase the parameter $\beta > 0$, $g(\beta; X)$ varies from the sample average $\frac{1}{|T|}\sum_{t\in T} X_t$ at $\beta = 0$ to the maximum $\max_{t\in T} X_t$ in the limit as $\beta \to \infty$.

With these considerations in mind, let $X = (X_t)_{t\in T}$ be a centered Gaussian process with a finite index set T. We will assume throughout that $\mathbf{E}|X_s - X_t|^2 > 0$ for all $s \ne t$, i.e., d is a metric and not a pseudometric. For each $\beta > 0$, we introduce the *quenched Gibbs average*

$$g(\beta) := \mathbf{E}\left[\frac{1}{Z(\beta)}\sum_{t\in T} X_t e^{\beta X_t}\right] = \mathbf{E}g(X;\beta) \tag{5}$$

and the *quenched equilibrium free energy*

$$\phi(\beta) := \frac{1}{\beta}\mathbf{E}\log\left(\frac{1}{|T|}\sum_{t\in T} e^{\beta X_t}\right) = \frac{1}{\beta}\mathbf{E}\log Z(\beta) - \frac{\log|T|}{\beta}. \tag{6}$$

The terminology originates in statistical physics of disordered systems (Bovier, 2006), where T is interpreted as the state space of a physical system and $-X_t$ is the random energy of the state $t \in T$. Each realization of X gives rise to a family of random Gibbs measures $(\nu_\beta)_{\beta>0}$ on T:

$$\nu_\beta(A) := \frac{\sum_{t\in A} e^{\beta X_t}}{Z(\beta)}, \qquad A \subseteq T$$

where the parameter β is the inverse temperature. In the infinite-temperature limit $\beta \to 0$, ν_β converges to the uniform distribution on T, which we will denote by ν_0. If we write $\langle f\rangle_\beta$ for the expectation of a function $f : T \to \mathbb{R}$ w.r.t. ν_β, then we can express the quantities (5) and (6) as

$$g(\beta) = \mathbf{E}[\langle X\rangle_\beta] \qquad \text{and} \qquad \phi(\beta) = \mathbf{E}\Big[\langle X\rangle_\beta - \beta^{-1}D(\nu_\beta\|\nu_0)\Big],$$

where $D(\cdot\|\cdot)$ is the relative entropy (or Kullback–Leibler divergence). The term "quenched" refers to the fact that the disorder $(X_t)_{t\in T}$ is frozen (or quenched) when computing the Gibbs averages $\langle\cdot\rangle_\beta$, and the expectation $\mathbf{E}[\cdot]$ w.r.t. the randomness of X is taken at the end.

In light of (2) and (4), both $g(\beta)$ and $\phi(\beta)$ can be viewed as smoothed versions of the expected maximum of X. In particular, both converge to it in the zero-temperature limit $\beta \to \infty$. Since $\sqrt{\log |T|}$ is present on both sides of (1), it is natural to ask which quantity plays the role of $\sqrt{\log |T|}$ for finite values of β. Ideally, it should converge to $\sqrt{\log |T|}$ as $\beta \to \infty$ and to 0 as $\beta \to 0$. In this paper, we show the following:

1. For every $\beta \geq 0$, the quenched Gibbs average $g(\beta)$ can be upper-bounded as

$$g(\beta) \lesssim \sigma \sqrt{\mathbf{E} D(\nu_\beta \| \nu_0)}$$

where σ^2 is the maximum of the variances $\mathrm{Var}[X_t]$. Moreover, a Sudakov-type lower bound holds in the low-temperature regime: There exists a threshold $\beta_* > 0$ that depends on the covariance structure of X, such that

$$g(\beta) \gtrsim a \sqrt{\mathbf{E} D(\nu_\beta \| \nu_0)} \qquad \text{for } \beta \geq \beta_*$$

where $a > 0$ is the minimum separation between the elements of T. When the X_t's are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables, the upper and the lower bounds match up to universal constants and hold for *all* values of β:

$$g(\beta) \asymp \sigma \sqrt{\mathbf{E} D(\nu_\beta \| \nu_0)}.$$

2. The quenched free energy $\phi(\beta)$ can be upper-bounded as

$$\phi(\beta) \lesssim \sigma \sqrt{\mathbf{E} D_{1/2}(\nu_\beta \| \nu_0)},$$

where $D_{1/2}(\cdot \| \cdot)$ is the Rényi divergence of order $1/2$ (see Section 2.1 for the information-theoretic background). When the X_t's are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables, there is a matching lower bound for all values of β:

$$\phi(\beta) \asymp \sigma \sqrt{\mathbf{E} D_{1/2}(\nu_\beta \| \nu_0)}.$$

Since $a > 0$, all the X_t's are almost surely distinct, so the Gibbs measures ν_β converge to a (random) Dirac measure as $\beta \to \infty$. Consequently, the information divergences in the above inequalities converge to $\log |T|$, as expected. Similarly, since ν_β converges to the uniform measure ν_0 as $\beta \to 0$, our inequalities have the right high-temperature behavior as well (wherever applicable).

1.1. Related work

Approximation of the maxima of random processes on finite sets by their smoothed versions is a widely used technique in statistical physics and in probability theory (Bovier, 2006; Adler and Taylor, 2007; Talagrand, 2011; Panchenko, 2013; Chatterjee, 2014). However, in most of these applications, one is interested mainly in either the expected maximum itself (corresponding to the zero-temperature limit) or in various asymptotic quantities arising in the so-called *thermodynamic limit* (when the size of the index set tends to infinity). To the best of our knowledge, the upper and lower bounds presented in this paper for $\beta > 0$ are new.

In the context of Sudakov minoration, we should mention recent work of Liu (2022, 2023), where lower bounds are given for quantities of the form

$$\mathbf{E} \log \int_T e^{\langle t, Y \rangle} \mu(\mathrm{d}t), \tag{7}$$

where μ is a probability measure supported on a compact convex set $T \subset \mathbb{R}^N$, $\langle \cdot, \cdot \rangle$ is the Euclidean inner product, and Y is a random vector in $\mathbb{R}^N$. It is natural to view (7) as a softmax-type relaxation of the expected "Y-width" of T, $\mathbf{E}[\max_{t \in T} \langle t, Y \rangle]$. Liu obtains Sudakov-type lower bounds on (7) by first lower-bounding it by the expected supremum of a related process on a certain subset of T and then using a deep convex-geometric result of Pajor (1984) to lower-bound this expected supremum in terms of the packing numbers of T. The dimension reduction approach to Sudakov minoration due to Mendelson et al. (2019) alsy plays a key role here for obtaining dimension-free lower bounds. In contrast to our results, which are specific to Gaussian processes with finite index sets, the bounds of Liu apply to linear processes of the form $X_t = \langle t, Y \rangle$ with compact convex index sets, and Y is not required to be a Gaussian random vector.

1.2. Organization of the paper

The remainder of the paper is structured as follows. Section 2 introduces some concepts from information theory and from statistical physics that will be used later on. Upper and lower bounds on the quenched Gibbs average $g(\beta)$ are stated and proved in Section 3, followed by the results on the quenched free energy $\phi(\beta)$ in Section 4. Section 5 presents an application of our results to deriving upper and lower bounds on the quenched pressure in the Random Energy Model (REM), one of the basic models of physical systems with random disorder (Bovier, 2006).

2. Preliminaries and auxiliary results

2.1. Some notions from information theory

Let $\mathcal{P}(T)$ denote the set of all probability measures on T. For any two $\mu, \nu \in \mathcal{P}(T)$ with ν strictly positive everywhere, the Rényi divergence of order $\alpha \in (0, \infty)$, $\alpha \neq 1$, is defined as

$$D_\alpha(\mu \| \nu) := \frac{1}{\alpha - 1} \log \sum_{t \in T} \mu^\alpha(t) \nu^{1-\alpha}(t).$$

The limit of $D_\alpha(\mu \| \nu)$ as $\alpha \to 1$ exists and is equal to the Kullback–Leibler divergence $D(\mu \| \nu)$, see, e.g., van Erven and Harremoës (2014). The Shannon entropy of $\mu \in \mathcal{P}(T)$ is

$$H(\mu) := -\sum_{t \in T} \mu(t) \log \mu(t).$$

We will mainly deal with the situation when $\mu = \nu_\beta$ for some $\beta > 0$ and $\nu = \nu_0$, the uniform distribution on T. The expressions for the Rényi divergences $D_\alpha(\nu_\beta \| \nu_0)$ involve the log-partition function $\Lambda(\beta) := \log Z(\beta)$, whose first and second derivatives are given by $\Lambda'(\beta) = \langle X \rangle_\beta$ and $\Lambda''(\beta) = \langle X^2 \rangle_\beta - (\langle X \rangle_\beta)^2$. Since $\Lambda''(\beta) \geq 0$, it follows that $\Lambda(\beta)$ is a convex function of β. One useful consequence of this is the following (see Supplementary Material for the proof):

Lemma 1 *The function $\beta \mapsto \|\nu_\beta\|_2^2 = \sum_{t \in T} \nu_\beta^2(t)$ is nondecreasing.*

The quantity $\|\nu_\beta\|_2^2$, which is sometimes referred to as the *participation ratio* (see, e.g., Section 5.3 of Mézard and Montanari (2009)), can be expressed in terms of the partition function as follows:

$$\|\nu_\beta\|_2^2 = \frac{1}{Z^2(\beta)} \sum_{t \in T} e^{2\beta X_t} = \frac{Z(2\beta)}{Z^2(\beta)}.$$

The negative logarithm of $\|\nu_\beta\|_2^2$ is equal to $H_2(\nu_\beta)$, the Rényi entropy of ν_β of order 2.

With these preliminaries out of the way, we have

$$D(\nu_\beta \| \nu_0) = \log |T| - H(\nu_\beta) = \log |T| + \beta \langle X \rangle_\beta - \Lambda(\beta) \tag{8}$$

and

$$D_\alpha(\nu_\beta \| \nu_0) = \log |T| + \frac{1}{\alpha - 1} \big(\Lambda(\alpha\beta) - \alpha \Lambda(\beta) \big). \tag{9}$$

In particular, specializing (9) to $\alpha = 1/2$ gives

$$D_{1/2}(\nu_\beta \| \nu_0) = \log |T| + \log \|\nu_{\beta/2}\|_2^2. \tag{10}$$

2.2. A replica-symmetric representation of the quenched Gibbs average

The following lemma relates the quenched Gibbs average $g(\beta)$ to the geometry of (T, d) and to the Gibbs measure at inverse temperature β (see Supplementary Materials for the proof):

Lemma 2 *The quenched Gibbs average can be expressed as*

$$g(\beta) = \frac{\beta}{2} \sum_{s,t \in T} d^2(s,t) \mathbf{E}[\nu_\beta(s) \nu_\beta(t)]. \tag{11}$$

In particular, if the X_t's are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables, then

$$g(\beta) = \beta \sigma^2 (1 - r(\beta)), \tag{12}$$

where $r(\beta) := \mathbf{E}\|\nu_\beta\|_2^2$.

Results of this type are common in the literature on statistical physics of disordered systems (see, e.g., Lemma 1.3.11 in Talagrand (2011) or Lemma 1.1 in Panchenko (2013) in the context of the Sherrington–Kirkpatrick model), where they are referred to as *replica-symmetric representations*: Let (τ^1, τ^2) be a random element of $T \times T$ sampled from the product measure $\nu_\beta \otimes \nu_\beta$; τ^1 and τ^2 are independent replicas of the system at inverse temperature β. Then we can write (11) as

$$g(\beta) = \mathbf{E}\langle X \rangle_\beta = \frac{\beta}{2} \mathbf{E} \langle d^2(\tau^1, \tau^2) \rangle_\beta,$$

where $\langle \cdot \rangle_\beta$ is the expectation w.r.t. $\nu_\beta \otimes \nu_\beta$ and where $\mathbf{E}[\cdot]$ is the expectation w.r.t. X.

3. Upper and lower bounds on the quenched Gibbs average

Our analysis of the quenched Gibbs average $g(\beta)$ relies crucially on the fact that it can be expressed as an expectation of X first w.r.t. the Gibbs measure ν_β and then w.r.t. the randomness of X. To obtain an upper bound on $g(\beta)$, we make use of the Gibbs variational principle (Dupuis and Ellis, 1997) which, in the context of statistical physics, captures the Legendre–Fenchel duality between the relative entropy and the free energy. For the lower bound, we make fundamental use of the replica-symmetric representation of $g(\beta)$ in Lemma 2 and of Sudakov minoration.

3.1. Upper bound via the Gibbs variational principle

The Gibbs variational principle states that, for any function $f : T \to \mathbb{R}$,

$$-\log\langle \nu_0, e^{-f}\rangle = \inf_{\mu\in\mathcal{P}(T)} \left\{\langle \mu, f\rangle + D(\mu\|\nu_0)\right\},$$

where $\langle \mu, f\rangle$ denotes the expected value of f w.r.t. μ.

Theorem 3 *The quenched Gibbs average is bounded from above as*

$$g(\beta) \le \sqrt{2\sigma^2 \mathbf{E} D(\nu_\beta\|\nu_0)}$$

where σ^2 is the maximum of the variances $\mathrm{Var}[X_t]$.

Proof Applying the Gibbs variational principle to the (random) function $f(t) := -\lambda X_t$ for some $\lambda > 0$ (to be chosen later) gives

$$-\log\left(\frac{1}{|T|}\sum_{t\in T} e^{\lambda X_t}\right) \le -\lambda\langle X\rangle_\beta + D(\nu_\beta\|\nu_0).$$

Rearranging and taking expectations w.r.t. X, we obtain $g(\beta) \le \frac{1}{\lambda}\mathbf{E}[D(\nu_\beta\|\nu_0)] + \phi(\lambda)$, where $\phi(\lambda)$ can be estimated using Jensen's inequality and properties of Gaussian random variables:

$$\phi(\lambda) = \frac{1}{\lambda}\mathbf{E}\log\left(\frac{1}{|T|}\sum_{t\in T} e^{\lambda X_t}\right) \le \frac{1}{\lambda}\log\left(\frac{1}{|T|}\sum_{t\in T}\mathbf{E}e^{\lambda X_t}\right) \le \frac{\lambda\sigma^2}{2}.$$

Combining the above and optimizing over λ yields

$$g(\beta) \le \inf_{\lambda>0}\left\{\frac{1}{\lambda}\mathbf{E}D(\nu_\beta\|\nu_0) + \frac{\lambda\sigma^2}{2}\right\} = \sqrt{2\sigma^2\mathbf{E}D(\nu_\beta\|\nu_0)},$$

and the proof is complete. ■

Using the expression (8) for $D(\nu_\beta\|\nu_0)$, we immediately obtain the following:

Corollary 4 $g(\beta) \le \sqrt{2\sigma^2\big(\log|T| - \mathbf{E}H(\nu_\beta)\big)}$.

Since the X_t's are distinct a.s., the quenched entropy $\mathbf{E}H(\nu_\beta)$ vanishes in the zero-temperature limit $\beta \to \infty$, and we recover the well-known Gaussian maximal inequality with the sharp constant:

$$\mathbf{E}\Big[\max_{t\in T} X_t\Big] \le \sqrt{2\sigma^2\log|T|}. \tag{13}$$

3.2. A low-temperature lower bound via Sudakov minoration

Since the X_t's are a-separated, i.e., $d(s,t) \geq a > 0$ for all distinct $s,t \in T$, we can lower-bound the expected maximum of X via Sudakov minoration (see, e.g., Lemma 2.4.2 in Talagrand (2014)):

$$\mathbf{E}\Big[\max_{t\in T} X_t\Big] \geq ca\sqrt{\log|T|}, \tag{14}$$

where $0 < c < 1$ is a universal constant (Lemma 5.5.6 in Marcus and Rosen (2006) gives $c = \frac{1}{17}$). Our next result is a low-temperature relaxation of Sudakov minoration:

Theorem 5 *There exists an inverse temperature threshold $0 < \beta_* < \infty$ that depends on the covariance structure of $(X_t)_{t\in T}$, such that*

$$g(\beta) \geq ca\sqrt{\mathbf{E}D(\nu_\beta\|\nu_0)} \qquad \textit{for all } \beta \geq \beta_*$$

where $c > 0$ is the same constant as in (14).

Proof Let $\Delta := \max\{d(s,t) : s,t \in T\}$ denote the diameter of (T,d). Then Lemma 2 allows us to bound $g(\beta)$ from above and from below as follows:

$$\frac{\beta a^2}{2}(1 - r(\beta)) \leq g(\beta) \leq \frac{\beta\Delta^2}{2}(1 - r(\beta)),$$

where $r(\beta) = \mathbf{E}\|\nu_\beta\|_2^2$. The function $r(\beta)$ is nondecreasing by Lemma 1 and bounded between $r(0) = \frac{1}{|T|}$ and 1. Therefore, there exists some value $\beta_* > 0$, such that $1 - r(\beta) \leq \frac{c^2a^2}{2\Delta^2}$ for all $\beta \geq \beta_*$. This implies that

$$0 \leq \frac{g(\beta)}{\beta} \leq \frac{c^2a^2}{4}, \qquad \beta \geq \beta_*. \tag{15}$$

Differentiating the function

$$\tilde{g}(\beta) := g(\beta) - ca\sqrt{\mathbf{E}D(\nu_\beta\|\nu_0)} = g(\beta) - ca\sqrt{\beta(g(\beta) - \phi(\beta))}$$

w.r.t. β gives

$$\begin{aligned}
\tilde{g}'(\beta) &= g'(\beta) - \frac{ca}{2}\frac{\frac{\mathrm{d}}{\mathrm{d}\beta}[\beta(g(\beta) - \phi(\beta))]}{\sqrt{\beta(g(\beta) - \phi(\beta))}} \\
&= g'(\beta)\left(1 - \frac{ca}{2}\frac{\beta}{\sqrt{\beta(g(\beta) - \phi(\beta))}}\right) \\
&\leq g'(\beta)\left(1 - \frac{ca}{2}\sqrt{\frac{\beta}{g(\beta)}}\right) \\
&\leq 0,
\end{aligned}$$

where we have used (15) and the fact that $g'(\beta) \geq 0$, which follows from the convexity of $\beta \mapsto \beta\phi(\beta)$. Thus, for any $\beta \geq \beta_*$,

$$\tilde{g}(\beta) \geq \lim_{\beta\to\infty} \tilde{g}(\beta) = \mathbf{E}\Big[\max_{t\in T} X_t\Big] - ca\sqrt{\log|T|} \geq 0,$$

where the inequality is by Sudakov minoration. ■

3.3. A lower bound for i.i.d. processes

When the X_t's are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables, the separation condition holds with $a^2 = 2\sigma^2$. In this setting, we can extend the range of our Sudakov-type lower bound on $g(\beta)$ to *all* values of β, but with a slightly worse constant:

Theorem 6 *When X_t's are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables, we have*

$$g(\beta) \geq c\sigma\sqrt{\mathbf{E}D(\nu_\beta \| \nu_0)} \qquad \textit{for all } \beta \geq 0.$$

Remark 7 *As will become evident from the proof, the lower bound with the worse constant $c/2$ holds only in the high-temperature regime $\beta \leq \beta_*$, where the threshold β_* is determined in exactly the same manner as in Theorem 5. In the low-temperature regime $\beta > \beta_*$, the constant is still c.*

Proof In the i.i.d. case, $a^2 = \Delta^2 = 2\sigma^2$. Since $r(\beta)$ is nondecreasing, the same choice of β_* as in the proof of Theorem 5 ensures that

$$1 - r(\beta) \geq \frac{c^2}{2}, \qquad \beta \leq \beta_*.$$

This is equivalent to the inequality $1 - r(\beta) \geq \frac{c}{\sqrt{2}}\sqrt{1 - r(\beta)}$ for $\beta \leq \beta_*$. For any such β we have

$$g(\beta) = \beta\sigma^2(1 - r(\beta)) \geq c\sigma\sqrt{\frac{1}{2}\beta^2\sigma^2(1 - r(\beta))} = c\sigma\sqrt{\frac{1}{2}\beta g(\beta)} \geq c\sigma\sqrt{\frac{1}{2}\mathbf{E}D(\nu_\beta \| \nu_0)},$$

where we have used Lemma 2 and the fact that $\mathbf{E}D(\nu_\beta \| \nu_0) = \beta(g(\beta) - \phi(\beta)) \leq \beta g(\beta)$. This establishes the lower bound in the high-temperature regime. The analysis of the low-temperature regime $\beta \geq \beta_*$ is exactly the same as in the proof of Theorem 5. ■

4. Upper and lower bounds on the quenched free energy

The analysis of the quenched free energy $\phi(\beta)$ is complicated by the fact that, unlike the quenched Gibbs average $g(\beta)$, it cannot be expressed as an expected value of X w.r.t. a random measure on T. The appropriate strategy here is to capitalize on the easily verified fact that $g(\beta)$ is equal to the derivative of $\beta\phi(\beta)$ and use the fundamental theorem of calculus in conjunction with the already available upper and lower bounds on $g(\beta)$.

4.1. Upper bound

The following upper bound is a counterpart of Theorem 3 for $g(\beta)$:

Theorem 8 *The quenched free energy is bounded from above by*

$$\phi(\beta) \leq \sqrt{2\sigma^2\mathbf{E}D_{1/2}(\nu_\beta \| \nu_0)},$$

where $D_{1/2}(\cdot\|\cdot)$ is the Rényi divergence of order $1/2$.

Proof It is convenient to work with the function $\psi(\beta) := \beta\phi(\beta) = \Lambda(\beta) - \log|T|$, since $g(\beta) = \psi'(\beta)$. Define the shorthand $D(\beta) := \mathbf{E}D(\nu_\beta\|\nu_0)$. A straightforward computation gives

$$D(\beta) = \beta\psi'(\beta) - \psi(\beta). \tag{16}$$

Then

$$\psi(\beta) = \int_0^\beta g(\lambda)\,\mathrm{d}\lambda \le \int_0^\beta \sqrt{2\sigma^2 D(\lambda)}\,\mathrm{d}\lambda \le \beta\sqrt{2\sigma^2}\cdot\sqrt{\frac{1}{\beta}\int_0^\beta D(\lambda)\,\mathrm{d}\lambda},$$

where the first step is by the fundamental theorem of calculus, the second uses Theorem 3, and the third is by Jensen's inequality. Using (16), integration by parts, and convexity of ψ, we obtain

$$\frac{1}{\beta}\int_0^\beta D(\lambda)\,\mathrm{d}\lambda = \frac{1}{\beta}\int_0^\beta (\lambda\psi'(\lambda) - \psi(\lambda))\,\mathrm{d}\lambda = \psi(\beta) - \frac{2}{\beta}\int_0^\beta \psi(\lambda)\,\mathrm{d}\lambda \le \psi(\beta) - 2\psi\left(\frac{\beta}{2}\right).$$

Moreover, by the definition of ψ,

$$\begin{aligned}\psi(\beta) - 2\psi\left(\frac{\beta}{2}\right) &= \log|T| + \mathbf{E}\log\Big(\sum_{t\in T} e^{\beta X_t}\Big) - 2\,\mathbf{E}\log\Big(\sum_{t\in T} e^{\beta X_t/2}\Big)\\ &= \log|T| + \mathbf{E}\log Z(\beta) - 2\,\mathbf{E}\log\sum_{t\in T}(Z(\beta)\nu_\beta(t))^{1/2}\\ &= \log|T| - 2\,\mathbf{E}\log\sum_{t\in T}(\nu_\beta(t))^{1/2}.\end{aligned}$$

Recognizing the latter expression as the expectation of the Rényi divergence $D_{1/2}(\nu_\beta\|\nu_0)$ in (10) and using this in our bound for $\psi(\beta)$, we obtain the desired estimate. ■

4.2. A lower bound for i.i.d. processes

When the X_t's are $\mathcal{N}(0,\sigma^2)$ random variables, the inequality in Theorem 8 can be reversed:

Theorem 9 *When X_t's are i.i.d. $\mathcal{N}(0,\sigma^2)$ random variables, we have*

$$\phi(\beta) \ge \frac{c\sigma}{2}\sqrt{\mathbf{E}D_{1/2}(\nu_\beta\|\nu_0)} \qquad \textit{for all } \beta \ge 0$$

where c is the constant in the Sudakov minoration bound (14).

Proof We start by applying the fundamental theorem of calculus, but then invoke Lemma 2:

$$\psi(\beta) = \int_0^\beta g(\lambda)\,\mathrm{d}\lambda = \sigma^2\int^\beta \lambda(1 - r(\lambda))\,\mathrm{d}\lambda = \sigma^2\Big(\frac{\beta^2}{2} - \int_0^\beta \lambda r(\lambda)\,\mathrm{d}\lambda\Big).$$

Integrating by parts and using the monotonicity of $r(\lambda)$, we obtain

$$\int_0^\beta \lambda r(\lambda)\,\mathrm{d}\lambda = \frac{\beta^2}{2}r(\beta) - \int_0^\beta \frac{\lambda^2}{2}r'(\lambda)\,\mathrm{d}\lambda \le \frac{\beta^2}{2}r(\beta).$$

Substituting this into our expression for $\psi(\beta)$ and using Theorem 6 gives

$$\psi(\beta) \geq \frac{\beta^2\sigma^2}{2}(1 - r(\beta)) = \frac{\beta g(\beta)}{2} \geq \frac{c\beta\sigma}{2}\sqrt{\mathbf{E}D(\nu_\beta\|\nu_0)},$$

where c is the constant in (14). Since the Rényi divergence $D_\alpha(\cdot\|\cdot)$ is an increasing function of α (van Erven and Harremoës, 2014), we obtain the claimed lower bound. ■

The appearance of the Rényi divergence in upper and lower bounds on the quenched free energy is rather natural from the viewpoint of statistical physics, where Rényi entropy is used to quantify the sensitivity of equilibrium free energy to changes in temperature (Baez, 2022).

4.3. A Sudakov-type lower bound without independence

When the X_t's are no longer independent, it is still possible obtain a Sudakov-type lower bound on a quantity closely related to the quenched free energy $\phi(\beta)$. To state it, we first need to introduce additional notation. For a centered Gaussian process $X = (X_t)_{t\in T}$ and for a subset $A \subseteq T$, let $\Phi_\beta(X; A)$ denote the β-softmax of $(X_t)_{t\in A}$:

$$\Phi_\beta(X; A) := \frac{1}{\beta}\log\Big(\sum_{t\in A} e^{\beta X_t}\Big).$$

It is easy to see that $\Phi_\beta(X;\cdot)$ is monotone w.r.t. set inclusion: if $A \subseteq A'$, then $\Phi_\beta(X; A) \leq \Phi_\beta(X; A')$. We then have the following inequality, which can be viewed as a "softmax" relaxation of a result due to Talagrand (1992):

Theorem 10 *Let S be a 4σ-packing of (T, d), i.e., $d(s, t) \geq 4\sigma$ for all $s, t \in S$ with $s \neq t$. Then*

$$\mathbf{E}\Phi_\beta\Big(X; \bigcup_{s\in S} B(s,\sigma)\Big) \geq \sigma\mathbf{E}\Phi_{\beta\sigma}(G; S) + \frac{1}{|S|}\sum_{s\in S}\mathbf{E}\Phi_\beta(X; B(s,\sigma)), \tag{17}$$

where $G = (G_t)_{t\in T}$ consists of independent $\mathcal{N}(0, 1)$ random variables and where $B(s,\sigma) := \{t \in T : d(s,t) \leq \sigma\}$ is the closed ball of radius σ centered at s.

The proof, which is somewhat technical, can be found in the Supplementary Materials.

Let us now examine some consequences of Theorem 10. First, since $\Phi_\beta(X;\cdot)$ is monotone w.r.t. set inclusion, it follows that the right-hand side of (17) is automatically a lower bound for $\mathbf{E}\Phi_\beta(X; T)$. Second, we recover Talagrand's minoration result (for finite T) in the zero-temperature limit. Indeed, using (2) and Sudakov minoration in (17) gives

$$\begin{aligned}\mathbf{E}\Big[\max_{t\in T} X_t\Big] &\geq \sigma\mathbf{E}\Big[\max_{s\in S} G_s\Big] + \min_{s\in S}\mathbf{E}\Big[\max_{t\in B(s,\sigma)} X_t\Big] - \frac{\log|T|}{\beta} \\ &\gtrsim \sigma\sqrt{\log|S|} + \min_{s\in S}\mathbf{E}\Big[\max_{t\in B(s,\sigma)} X_t\Big] - \frac{\log|T|}{\beta}.\end{aligned}$$

Taking $\beta \to \infty$, we obtain Talagrand's result. Finally, since $\phi(\beta) = \mathbf{E}\Phi_\beta(X; T) - \frac{\log|T|}{\beta}$, Theorem 10 can be used to lower-bound the quenched free energy without assuming independence.

5. Application: the quenched pressure in the random energy model

We now illustrate the use of our results in the context of the Random Energy Model (REM), introduced by Derrida (1981). The REM is the simplest model of a physical system with random disorder; yet, despite this apparent simplicity, its behavior exhibits many nontrivial features, including a phase transition (see Chapter 5 of Mézard and Montanari (2009) or Chapter 9 of Bovier (2006) for more details on the REM).

The state space of the REM is the binary cube $T = \{-1,+1\}^N$. Each of the points $t = (t_1,\dots,t_N) \in \{-1,+1\}^N$ corresponds to a configuration of N spins, each of which can point up $(+)$ or down $(-)$. The energies of the configurations $t \in T$ are i.i.d. $\mathcal{N}(0, N/2)$ random variables X_t, where the independence translates into the absence of interactions between the N spins. One of the quantities of interest is the *quenched pressure*

$$P_N(\beta) := \frac{1}{N}\mathbf{E}\log Z_N(\beta),$$

where

$$Z_N(\beta) = \sum_{t\in\{-1,+1\}^N} e^{\beta X_t}$$

is the (random) partition function. The REM exhibits a phase transition in the infinite system limit $N \to \infty$: For $\beta_c = 2\sqrt{\log 2}$,

$$\lim_{N\to\infty} P_N(\beta) = \begin{cases} \log 2 + \frac{\beta^2}{4}, & \beta < \beta_c \\ \beta\sqrt{\log 2}, & \beta \geq \beta_c \end{cases}$$

(see Giardinà and Starr (2007) for a self-contained proof of this using variational methods). The phase transition manifests itself in the different behavior of $\lim_{N\to\infty} P_N(\beta)$ as a function of β: quadratic for $\beta \leq \beta_c$ and linear for $\beta \geq \beta_c$. We now show that Theorems 3 and 5 can be used to give an elementary derivation of upper and lower bounds on the quenched pressure for each finite N. While the constants in these bounds are not sharp, the bounds still pinpoint the different scaling of $P_N(\beta)$ with β (quadratic vs. linear) in high- and low-temperature regimes.

Theorem 11 *For each N and each $\beta_0 \geq 0$, define the functions*

$$\underline{\mathcal{Q}}_N(\beta;\beta_0) := \begin{cases} \log 2 + \frac{c^2\beta^2}{8}, & \beta \leq \beta_0 \\ \log 2 + \frac{c^2\beta_0^2}{8} + c(\beta-\beta_0)\sqrt{\frac{1}{2N}\mathbf{E}D(\nu_{N,\beta_0}\|\nu_{N,0})}, & \beta \geq \beta_0 \end{cases}$$

and

$$\overline{\mathcal{Q}}_N(\beta;\beta_0) := \begin{cases} \log 2 + \frac{\beta^2}{4}, & \beta \leq \beta_0 \\ \log 2 + \frac{\beta_0^2}{4} + (\beta-\beta_0)\sqrt{\frac{1}{N}\mathbf{E}D(\nu_{N,\beta}\|\nu_{N,0})}, & \beta \geq \beta_0 \end{cases}$$

where c is the constant in the Sudakov minoration bound (14) and where $\nu_{N,\beta}(t) = e^{\beta X_t}/Z_N(\beta)$ is the REM Gibbs measure at inverse temperature β. Then there exists an inverse temperature threshold $\beta_* = \beta_{N,*} > 0$, such that

$$\underline{\mathcal{Q}}_N(\beta;\beta_*) \leq P_N(\beta) \leq \inf_{\beta_0\geq 0} \overline{\mathcal{Q}}_N(\beta;\beta_0).$$

Remark 12 *Notice that the quadratic (high-temperature) and the linear (low-temperature) pieces of $\underline{\mathcal{Q}}_N(\beta;\beta_0)$ agree when $\beta = \beta_0$. The same applies to $\overline{\mathcal{Q}}_N(\beta;\beta_0)$, although its low-temperature portion contains a nonlinear factor $\sqrt{\mathbf{E}D(\nu_{N,\beta}\|\nu_{N,0})} = \sqrt{N\log 2 - \mathbf{E}H(\nu_{N,\beta})}$, which can be further upper-bounded by $\sqrt{N\log 2}$. In particular, $\overline{\mathcal{Q}}_N(\beta;\beta_c)$ is equal to $\log 2 + \frac{\beta^2}{4}$ for $\beta \le \beta_c$ and bounded from above by $\beta\beta_c = \beta\sqrt{\log 2}$ for $\beta \ge \beta_c$. This gives the correct upper bound for every finite N (Giardinà and Starr, 2007). See also the work of Olivieri and Picco (1984) for finite-N corrections to the asymptotic value of the free energy in the REM.*

Proof Let $g_N(\beta) := \mathbf{E}[\langle X, \nu_{N,\beta}\rangle]$. Then

$$P_N(\beta) = P_N(\beta_0) + \frac{1}{N}\int_{\beta_0}^{\beta} g_N(\lambda)\,\mathrm{d}\lambda, \qquad \text{for all } \beta \ge \beta_0 \ge 0. \tag{18}$$

Using Jensen's inequality and the properties of Gaussian random variables, we obtain

$$P_N(\beta) \le \log 2 + \frac{\beta^2}{4}. \tag{19}$$

On the other hand, from (18) and Theorem 3 it follows that

$$\begin{aligned} P_N(\beta) &\le P_N(\beta_0) + \int_{\beta_0}^{\beta} \sqrt{\frac{1}{N}\mathbf{E}D(\nu_{N,\lambda}\|\nu_{N,0})}\,\mathrm{d}\lambda \\ &\le \log 2 + \frac{\beta_0^2}{4} + (\beta-\beta_0)\sqrt{\frac{1}{N}\mathbf{E}D(\nu_{N,\beta}\|\nu_{N,0})}, \end{aligned}$$

where the second step follows from (19) and from the fact that $\lambda \mapsto \mathbf{E}D(\nu_{N,\lambda}\|\nu_{N,0})$ is nondecreasing. This gives $P_N(\beta) \le \overline{\mathcal{Q}}_N(\beta;\beta_0)$, which we can tighten further by minimizing w.r.t. β_0.

To obtain the lower bound, let $\beta_* > 0$ be such that $r_N(\beta) = \mathbf{E}\|\nu_{N,\beta}\|_2^2 \le 1 - \frac{c^2}{2}$ for all $\beta \le \beta_*$. Then, for $\beta \le \beta_*$, using (18) with $\beta_0 = 0$ and Lemma 2 we have

$$\begin{aligned} P_N(\beta) &= \log 2 + \frac{1}{2}\int_0^{\beta} \lambda(1 - r_N(\lambda))\,\mathrm{d}\lambda \\ &= \log 2 + \frac{\beta^2}{4} - \frac{1}{2}\int_0^{\beta} \lambda r_N(\lambda)\,\mathrm{d}\lambda \\ &\ge \log 2 + \frac{c^2\beta^2}{8}. \end{aligned}$$

For $\beta > \beta_*$, we can use (18) with $\beta_0 = \beta_*$ and the fact that $\beta \mapsto g_N(\beta)$ is increasing to write

$$\begin{aligned} P_N(\beta) &\ge P_N(\beta_*) + \frac{1}{N}(\beta-\beta_*)g_N(\beta_*) \\ &\ge \log 2 + \frac{c^2\beta_*^2}{8} + \frac{1}{N}(\beta-\beta_*)g_N(\beta_*). \end{aligned}$$

Moreover, observing that β_* is exactly the inverse temperature threshold Theorem 5 gives for $\nu_{N,\beta}$, we can bound $g_N(\beta_*)$ from below by

$$g_N(\beta_*) \ge c\sqrt{\frac{N}{2}\mathbf{E}D(\nu_{N,\beta_*}\|\nu_{N,0})}.$$

Substituting this into our lower bound for $P_N(\beta)$ for $\beta \ge \beta_*$ and then assembling the two pieces for $\beta \le \beta_*$ and for $\beta \ge \beta_*$, we see that $P_N(\beta) \ge \underline{\mathcal{Q}}_N(\beta;\beta_*)$. ■

Acknowledgments

This work was supported by the NSF under awards CCF-2348624 ("Towards a control framework for neural generative modeling") and CCF-2106358 ("Analysis and Geometry of Neural Dynamical Systems").

References

Robert J. Adler and Jonathan E. Taylor. *Random Fields and Geometry*. Springer, 2007.

John C. Baez. Rényi entropy and free energy. *Entropy*, 24(5):706, May 2022.

Antoine Bovier. *Statistical Mechanics of Disordered Systems: A Mathematical Introduction*. Cambridge University Press, 2006.

Sourav Chatterjee. *Superconcentration and Related Topics*. Springer, 2014.

Bernard Derrida. Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, 24(5):2613–2626, 1981.

Paul Dupuis and Richard S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley, 1997.

Cristian Giardinà and Shannon Starr. Variational bounds for the generalized random energy model. *Journal of Statistical Physics*, 127(1):1–20, 2007.

Jingbo Liu. Minoration via mixed volumes and Cover's problem for general channels. *Probability Theory and Related Fields*, 183(1):315–357, 2022.

Jingbo Liu. From soft-minoration to information-constrained optimal transport and spiked tensor models. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 666–671, 2023.

Michael B. Marcus and Jay Rosen. *Markov Processes, Gaussian Processes, and Local Times*. Cambridge University Press, 2006.

Shahar Mendelson, Emanuel Milman, and Grigoris Paouris. Generalized dual Sudakov minoration via dimension reduction – a program. *Studia Mathematica*, 244(159-202), 2019.

Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.

Enzo Olivieri and Pierre Picco. On the existence of thermodynamics for the random energy model. *Communications in Mathematical Physics*, 96(1):125–144, 1984.

Alain Pajor. *Sous-espaces ℓ_1^n des espaces de Banach*. PhD thesis, Université Pierre et Marie Curie, 1984.

Dmitry Panchenko. *The Sherrington–Kirkpatrick Model*. Springer, 2013.

Michel Talagrand. A simple proof of the majorizing measure theorem. *Geometric & Functional Analysis GAFA*, 2(1):118–125, 1992.

Michel Talagrand. *Mean Field Models for Spin Glasses, Volume I: Basic Examples*. Springer, 2011.

Michel Talagrand. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. Springer, 2014.

Tim van Erven and Peter Harremoës. Rényi divergence and Kullback–Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

Appendix A. Omitted proofs

A.1. Proof of Lemma 1

Using the definition of $\nu_\beta(t)$, we calculate

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\beta}\|\nu_\beta\|_2^2 &= 2\sum_{t\in T}\nu_\beta(t)\frac{\mathrm{d}}{\mathrm{d}\beta}\nu_\beta(t) \\
&= 2\sum_{t\in T}\nu_\beta^2(t)\left(X_t - \Lambda'(\beta)\right) \\
&= 2\|\nu_\beta\|_2^2\left(\frac{1}{\|\nu_\beta\|_2^2}\sum_{t\in T}X_t\nu_\beta^2(t) - \Lambda'(\beta)\right) \\
&= 2\|\nu_\beta\|_2^2\left(\Lambda'(2\beta) - \Lambda'(\beta)\right) \\
&\geq 0,
\end{aligned}$$

where the inequality follows from the fact that $\Lambda'(\beta)$ is nondecreasing, as $\Lambda(\beta)$ is convex.

A.2. Proof of Lemma 2

Fix a realization $x = (x_t)_{t\in T}$ of X. For each $s \in T$, we can view $\nu_\beta(s)$ as a function of x:

$$\nu_\beta(s) = f_s(x) = \frac{e^{\beta x_s}}{\sum_{t\in T} e^{\beta x_t}},$$

with

$$\begin{aligned}
\frac{\partial f_s}{\partial x_t}(x) &= \beta\left(\frac{e^{\beta x_t}\mathbf{1}\{s=t\}}{\sum_{t'\in T} e^{\beta x_{t'}}} - \frac{e^{\beta x_s}e^{\beta x_t}}{\left(\sum_{t'\in T} e^{\beta x_{t'}}\right)^2}\right) \\
&= \beta\big(\mathbf{1}\{s=t\}\nu_\beta(s) - \nu_\beta(s)\nu_\beta(t)\big).
\end{aligned}$$

Gaussian integration by parts (Adler and Taylor, 2007, Lemma 2.2.4) then gives

$$\begin{aligned}
\mathbf{E}[X_s\nu_\beta(s)] &= \sum_{t\in T}\mathbf{E}[X_sX_t]\mathbf{E}\Big[\frac{\partial f_s}{\partial x_t}(X)\Big] \\
&= \beta\sum_{t\in T}\mathbf{E}[X_sX_t]\mathbf{E}\big[\mathbf{1}\{s=t\}\nu_\beta(s) - \nu_\beta(s)\nu_\beta(t)\big].
\end{aligned}$$

Summing over all $s \in T$, we obtain

$$\begin{aligned} g(\beta) &= \sum_{s \in T} \mathbf{E}[X_s \nu_\beta(s)] \\ &= \beta \sum_{s,t \in T} \mathbf{E}[X_s X_t] \mathbf{E}\big[\mathbf{1}\{s=t\}\nu_\beta(s) - \nu_\beta(s)\nu_\beta(t)\big] \\ &= \beta \sum_{s,t \in T} \Big(\mathbf{E}[X_s^2] - \mathbf{E}[X_s X_t]\Big) \mathbf{E}[\nu_\beta(s)\nu_\beta(t)] \\ &= \frac{\beta}{2} \sum_{s,t \in T} \mathbf{E}[(X_s - X_t)^2] \mathbf{E}[\nu_\beta(s)\nu_\beta(t)] \\ &= \frac{\beta}{2} \sum_{s,t \in T} d^2(s,t) \mathbf{E}[\nu_\beta(s)\nu_\beta(t)], \end{aligned}$$

where we have made use of the relation

$$\sum_{s,t \in T} \mathbf{E}[X_s^2 - X_t^2] \mathbf{E}[\nu_\beta(s)\nu_\beta(t)] = 0$$

(which follows by symmetry) and of the definition of $d^2(s,t)$.

When the X_t's are i.i.d. $\mathcal{N}(0,\sigma^2)$ random variables, we have $d^2(s,t) = 2\sigma^2 \mathbf{1}\{s \neq t\}$. Substituting this into (11) and using the fact that ν_β is a probability distribution results in (12). Moreover, as can be seen from the proof of Lemma 1, $\frac{\mathrm{d}}{\mathrm{d}\beta}\|\nu_\beta\|_2^2$ is an L^2 random variable, so $r'(\beta) = \frac{\mathrm{d}}{\mathrm{d}\beta}\mathbf{E}\|\nu_\beta\|_2^2 \geq 0$, where the interchange of derivative and expectation is justified by the dominated convergence theorem.

A.3. Proof of Theorem 10

We adapt the proof strategy from Marcus and Rosen (2006, Lemma 5.5.7). Since S is a 4σ-packing subset of T, the balls $B(s,\sigma)$ centered on the points $s \in S$ are disjoint. Let $X^{(s)} = (X_t^{(s)})_{t \in T}$ be independent copies of $(X_t)_{t \in T}$, one for each $s \in S$. Also, let $G = (G_t)_{t \in T}$ be a collection of $\mathcal{N}(0,1)$ random variables independent of everything else. Each $t \in T' := \cup_{s \in S} B(s,\sigma)$ belongs to exactly one of the balls $B(s,\sigma)$; for such t, define $Z_t := X_t^{(s)} - X_s^{(s)} + \sigma G_s$. If t, t' are in the same ball $B(s,\sigma)$, then $Z_t - Z_{t'}$ is equal in law to $X_t - X_{t'}$, so $\mathbf{E}|Z_t - Z_{t'}|^2 = \mathbf{E}|X_t - X_{t'}|^2$. If t and t' belong to disjoint balls, then

$$\mathbf{E}|Z_t - Z_{t'}|^2 = \mathbf{E}|X_t^{(s)} - X_s^{(s)}|^2 + \mathbf{E}|X_{t'}^{(s')} - X_{s'}^{(s')}|^2 + \sigma^2 \mathbf{E}|G_s - G_{s'}|^2$$

for some $s, s' \in S$ with $s \neq s'$. Then $\mathbf{E}|Z_t - Z_{t'}|^2 \leq 4\sigma^2$, but, since t, t' are in disjoint balls, we also have $\mathbf{E}|X_t - X_{t'}|^2 \geq 4\sigma^2$. Consequently, the centered Gaussian process $Z = (Z_t)_{t \in T'}$ satisfies

$$\mathbf{E}|Z_t - Z_{t'}|^2 \leq \mathbf{E}|X_t - X_{t'}|^2, \qquad \forall t, t' \in T'.$$

Then a Gaussian interpolation argument due to Chatterjee (see, e.g., the proof of Theorem 2.2.5 in Adler and Taylor (2007)) shows that $\mathbf{E}\Phi_\beta(X;T') \geq \mathbf{E}\Phi_\beta(Z;T')$. We now proceed to lower-bound

$\mathbf{E}\Phi_\beta(Z;T')$. Define the random Gibbs measure $\tilde{\nu}(s) := \frac{e^{\beta\sigma G_s}}{\sum_{s'\in S} e^{\beta\sigma G_{s'}}}$ on the set S. By construction of Z, we write

$$\begin{aligned}
\Phi_\beta(Z;T') &= \frac{1}{\beta}\log\Big(\sum_{t\in T'} e^{\beta Z_t}\Big) \\
&= \frac{1}{\beta}\log\Big(\sum_{s\in S} e^{\beta\sigma G_s}\sum_{t\in B(s,\sigma)} e^{\beta(X_t^{(s)}-X_s^{(s)})}\Big) \\
&= \sigma\Phi_{\beta\sigma}(G;S) + \frac{1}{\beta}\log\left(\sum_{s\in S}\tilde{\nu}(s)\sum_{t\in B(s,\sigma)} e^{\beta(X_t^{(s)}-X_s^{(s)})}\right) \\
&\geq \sigma\Phi_{\beta\sigma}(G;S) + \sum_{s\in S}\tilde{\nu}(s)\Phi_\beta(X^{(s)};B(s,\sigma)) - \sum_{s\in S}\tilde{\nu}(s)X_s^{(s)},
\end{aligned}$$

where the last step is by Jensen's inequality. Since G is independent of everything else, using the law of iterated expectation gives

$$\begin{aligned}
\mathbf{E}\Phi_\beta(Z;T') &= \mathbf{E}[\mathbf{E}[\Phi_\beta(Z;T')|G]] \\
&\geq \sigma\mathbf{E}\Phi_{\beta\sigma}(G;S) + \sum_{s\in S}\mathbf{E}[\tilde{\nu}(s)]\mathbf{E}[\Phi_\beta(X^{(s)};B(s,\sigma))] \\
&= \sigma\mathbf{E}\Phi_{\beta\sigma}(G;S) + \frac{1}{|S|}\sum_{s\in S}\mathbf{E}\Phi_\beta(X;B(s,\sigma)),
\end{aligned}$$

where $\mathbf{E}[\tilde{\nu}(s)] = \frac{1}{|S|}$ follows by symmetry and where we have used the fact that the processes $X^{(s)}$ and X have the same probability law.