The Annals of Probability
2015, Vol. 43, No. 6, 3468–3493
DOI: 10.1214/14-AOP963

ON MULTIPLE PEAKS AND MODERATE DEVIATIONS FOR THE SUPREMUM OF A GAUSSIAN FIELD

BY JIAN DING[1], RONEN ELDAN AND ALEX ZHAI

University of Chicago, Microsoft Research and Stanford University

We prove two theorems concerning extreme values of general Gaussian fields. Our first theorem concerns the phenomenon of *multiple peaks*. Consider a centered Gaussian field whose sites have variance at most 1, and let ρ be the standard deviation of its supremum. A theorem of Chatterjee states that when such a Gaussian field is *superconcentrated* (i.e., $\rho \ll 1$), it typically attains values near its maximum on multiple almost-orthogonal sites and is said to exhibit *multiple peaks*. We improve his theorem in two respects: (i) the number of peaks attained by our bound is of the order $\exp(c/\rho^2)$ (as opposed to Chatterjee's polynomial bound in $1/\rho$) and (ii) our bound does not assume that the correlations are nonnegative. We also prove a similar result based on superconcentration of the free energy. As primary applications, we infer that for the S–K spin glass model on the n-hypercube and directed polymers on $\mathbb{Z}_n^2$, there are polynomially (in n) many almost-orthogonal sites that achieve values near their respective maxima.

Our second theorem gives an upper bound on moderate deviation for the supremum of a general Gaussian field. While the Gaussian isoperimetric inequality implies a sub-Gaussian concentration bound for the supremum, we show that the exponent in that bound can be improved under the assumption that the expectation of the supremum is of the same order as that of the independent case.

1. Introduction. A *Gaussian field* (or *Gaussian process*) is a collection $\mathbf{X} = \{X_\alpha, \alpha \in I\}$ of random variables such that every finite subset of this collection is distributed according to a multivariate normal law. The topic of this paper revolves around the behavior of extremal and near-extremal values of Gaussian fields.

Extremal values of Gaussian fields have been intensively studied by a variety of communities spanning probability, statistical physics and computer science. A cornerstone of the theory is the Gaussian concentration inequality of Sudakov–Cirel'son [18] and Borell [1], stating that for a (not necessarily centered) Gaussian field $\{X_i : 1 \le i \le N\}$ with $\sigma^2 = \sup_{1\le i\le N} \operatorname{Var} X_i$, we have

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\sup_{1\le i\le N} X_i - \mathbb{E}\Big(\sup_{1\le i\le N} X_i\Big)\Big| \ge z\Big) \\
&\qquad \le \frac{2}{\sqrt{2\pi}\sigma}\int_z^\infty \mathrm{e}^{-y^2/(2\sigma^2)}\,dy \qquad \text{for all } z \ge 0
\end{aligned} \tag{1}$$

Received January 2014.
[1]Supported in part by NSF Grant DMS-13-13596.
MSC2010 subject classifications. 60G15, 60F10, 60K35.
Key words and phrases. Gaussian processes, large deviations, statistical mechanics type models.

(see, e.g., [11], Theorem 7.1, equation (7.4)). An immediate consequence of (1) is that $\mathrm{Var}(\sup_{1\le i\le N} X_i) \le \sigma^2$. Despite being an extremely general and powerful inequality, it was observed by probabilists and statistical physicists that the bound (1) is far from sharp in most canonical examples of Gaussian fields, such as the KPZ universality class [10] and the class of log-correlated Gaussian fields (see, e.g., [12] and references therein). By being far from sharp, we mean, for example, that $\mathrm{Var}(\sup_{1\le i\le N} X_i) \ll \sigma^2$ or that equation (1) holds with a constant smaller than $\frac{1}{2\sigma^2}$ in the exponent. The former property is sometimes referred to as *superconcentration* and the latter fits under the umbrella of large deviation estimates. In this paper, we study the structure of Gaussian fields concerning the following two questions related to (1):

(a) When (1) is not sharp, what extra information can be deduced about the Gaussian field?
(b) Are there some simple and explicit conditions that guarantee an improvement upon (1)?

The rigorous study of question (a) in its full generality was pioneered in [3], where a connection between the so-called *superconcentration*, *chaos* and *multiple peaks* (or *multiple valleys*) phenomena for centered Gaussian fields was established. Multiple peaks is the following phenomenon observed by physicists in many natural settings of Gaussian fields (motivated by the study of energy landscapes of spin glasses): typically there exist many almost-orthogonal sites whose values are very close to the global maximum. This phenomenon was first rigorously established in [3] under the assumption of the aforementioned *superconcentration* property and the assumption that the correlations of the field are nonnegative. The phenomenon of *chaos* refers to an instability of the location of the maximizer with respect to small perturbations of the Gaussian field and was shown to be equivalent to superconcentration in some sense.

Our first goal in this paper is to further explore the connection between superconcentration and multiple peaks. We obtain a quantitative improvement of the number of such peaks (thus attaining an optimal bound in a certain sense) and we also remove the assumption that the correlations are nonnegative.

In order to state our result properly, we need a rigorous definition of the multiple peaks property. We shall use the same definition as introduced in [3]: consider a sequence of *centered* Gaussian fields $\mathbf{X}_N = \{X_{N,i} : 1 \le i \le N\}$. Denote by $\sigma_N^2 = \sup_{1\le i\le N} \mathrm{Var}\, X_{N,i}$ and write $[N] = \{1, \dots, N\}$. Write $R_N(i, j) = \mathrm{Cov}(X_{N,i}, X_{N,j})$ for all $i, j \in [N]$. In addition, define $M(\mathbf{X}_N) = \sup_{1\le i\le N} X_{N,i}$, $m(\mathbf{X}_N) = \mathbb{E}M(\mathbf{X}_N)$ and $\rho_N^2 = \mathrm{Var}(M(\mathbf{X}_N))$.[1]

DEFINITION 1.1. A sequence of Gaussian fields $\mathbf{X}_N$ exhibits multiple peaks if there exist $\ell_N \to \infty$, $\varepsilon_N = o(\sigma_N^2)$, $\delta_N = o(m(\mathbf{X}_N))$ and $\gamma_N \to 0$ such that with

[1]For convenience, we have introduced the variable ρ, which corresponds to $\sqrt{v}$ in [3].

probability at least $1-\gamma_N$, there is a set $A_N \subseteq [N]$ of cardinality at least ℓ_N satisfying

(M.1) $|R_N(i,j)| \le \varepsilon_N$ for all $i \neq j \in A_N$.
(M.2) $X_{N,i} \ge m(\mathbf{X}_N) - \delta_N$ for all $i \in A_N$.

Although the preceding definition applies to any sequence of Gaussian fields, we usually consider those where $\min_{1\le i\le N} \operatorname{Var} X_{N,i}$ is within a constant factor of σ_N^2. We have the following theorem.

THEOREM 1.2. *Fix any positive sequences $\delta_N \le m(\mathbf{X}_N)$, $\varepsilon_N \le \sigma_N^2$ and $\zeta_N \le 1$. Then for all $N \in \mathbb{N}$, with probability at least $1 - \frac{C_1\rho_N^2}{\delta_N^2} - \zeta_N$ there exists $A_N \subseteq [N]$ of cardinality at least $\exp(\frac{C_2\varepsilon_N^2\delta_N\zeta_N}{m(\mathbf{X}_N)\sigma_N^2\rho_N^2})$ such that* (M.1) *and* (M.2) *hold. Here, C_1, C_2 are positive universal constants.*

Quantitatively, [3], Theorem 3.7, guarantees existence of such a set A_N with cardinality at least $(\frac{\delta_N\varepsilon_N}{m(\mathbf{X}_N)\rho_N^2})^{1/3}$ as opposed to the exponential bound in Theorem 1.2. In addition, our result does not require the nonnegative correlation assumption, thereby solving Open Problem 5 in [3].

Another quantity that has received a significant amount of attention in the statistical physics community is the *free energy* of the field at an inverse-temperature $\beta > 0$, defined as

$$F_{N,\beta} = \frac{1}{\beta}\log\left(\sum_{i=1}^{N} \mathrm{e}^{\beta X_{N,i}}\right). \tag{2}$$

Evidently, as $\beta \to \infty$, this quantity approaches $M(\mathbf{X}_N)$. In view of this, it may be natural to look into the property that the quantity $F_{N,\beta}$ is concentrated around its mean for finite values of β. This phenomenon is referred to as the *superconcentration of free energy* of the process at inverse temperature β. In some cases, the free-energy for certain values of β seems to be a more tractable quantity than the supremum, and it may be easier to establish concentration bounds for the free energy than for the supremum of the field, as witnessed in [4] regarding the Sherrington–Kirkpatrick (or S–K) model for spin glasses (see definition below). The result [4] in which Chatterjee deduced multiple-peaks from superconcentration of the free energy, can be seen as an adaptation of the result in [3]. In this paper, we also give an adaptation of Theorem 1.2 to the free energy. We use the notation $\rho_N^2(\beta) = \operatorname{Var}(F_{N,\beta})$.

THEOREM 1.3. *For each $N \in \mathbb{N}$ and $\beta \ge 0$, suppose that $\tilde\rho_N(\beta)$ is an upper bound on $\rho_N(\beta)$. For any positive sequences $\delta_N \le m(\mathbf{X}_N)$, $\varepsilon_N \le \rho_N^2$, $\zeta_N \le 1$ and*

$$\beta_N \ge C_1 \max\left(\frac{\log N}{\delta_N}, \frac{1}{\tilde\rho_N(\beta_N)}, \frac{\delta_N\varepsilon_N^2}{m(\mathbf{X}_N)(\tilde\rho_N(\beta_N))^3\sigma_N^2}\right), \tag{3}$$

with probability at least $1 - \frac{C_2\sigma_N^2}{\delta_N^2} - \zeta_N$ *there exists* $A_N \subset [N]$ *with cardinality at least*

$$\exp\left(\frac{C_3\varepsilon_N^2\delta_N\zeta_N}{m(\mathbf{X}_N)(\tilde{\rho}_N(\beta_N))^2\sigma_N^2}\right)$$

such that (M.1) *and* (M.2) *holds. Here*, C_1, C_2, C_3 *are positive universal constants.*

REMARK 1. We have stated the preceding theorem in terms of the upper bound $\tilde{\rho}_N(\beta)$ rather than $\rho_N(\beta)$ so that in order to verify the assumption (3) one does not need an *a priori* lower bound on $\rho_N(\beta_N)$.

Let us now briefly discuss some applications of Theorems 1.2 and 1.3. Our first application is for directed polymers. Let $\mathbb{Z}_n^2$ denote the graph whose vertices are $\{0, 1, \dots, n\}^2$ and where two vertices are connected by an edge if they differ by 1 in exactly one coordinate. Let $\mathcal{P}_n$ be the collection of all the $N = \binom{2n}{n}$ monotone paths on $\mathbb{Z}_n^2$ joining the left bottom corner $(0, 0)$ and the right top conner (n, n). Associate i.i.d. standard Gaussian variables Z_e to each edge $e \in \mathbb{Z}_n^2$. The directed polymer is defined to be a Gaussian field $\{X_{N,P} : P \in \mathcal{P}_n\}$ where $X_{N,P} = \sum_{e \in P} Z_e$. For this model, [3], Theorem 8.1, provided an upper bound of $O(n/\log n)$ on ρ_N^2. Combined with Theorem 1.2, it gives the following corollary.

COROLLARY 1.4. *There exist absolute constants* $C_1, C_2 > 0$ *such that the following statement holds for directed polymers* [*recall that* $N = \binom{2n}{n}$]. *For any* $0 < \delta_N \le n$, $0 < \varepsilon_N \le n$, $0 < \zeta_N < 1$, *with probability at least* $1 - \frac{C_1 n}{\delta_N^2 \log n} - \zeta_N$ *there exists* $A_N \subseteq [N]$ *of cardinality at least* $n^{C_2\varepsilon_N^2\delta_N\zeta_N/n^3}$ *satisfying* (M.1) *and* (M.2).

We next discuss an application for the S–K model. For a hypercube $H_n = \{-1, 1\}^n$ (write $N = |H_n| = 2^n$), the S–K model introduced in [17] can be viewed for our purposes as a Gaussian field $\{X_{N,s} : s \in H_n\}$ with $X_{N,s} = \frac{1}{\sqrt{2n}} \sum_{i,j \in [n]} s_i s_j Z_{i,j}$ where $Z_{i,j}$'s are i.i.d. standard Gaussian variables. It is easy to see that the variance for each individual $X_{N,s}$ is precisely $n/2$ and the expected supremum is of order n. The asymptotics of the supremum as well as the free energy, however, had been a major challenge until it was established in a celebrated paper of Talagrand [20] which verifies the well-known prediction of the Parisi formula [16]. As for concentration, [4], Theorem 1.5, established an upper bound of $O(\beta n/\log n)$ on $\rho_N^2(\beta)$. Combining the variance bound and Theorem 1.3, we obtain the following [where we set β_N to be of order n/δ_N and $(\tilde{\rho}(\beta_N))^2$ to be of order $n\beta_N/\log n$].

COROLLARY 1.5. *There exist absolute constants* $C_1, C_2 > 0$ *such that the following statement holds for the S–K model* (*recall* $N = 2^n$). *For any positive* $0 <$

$\delta_N \leq n$, $0 < \varepsilon_N \leq n$, $0 < \zeta_N < 1$, *with probability at least* $1 - \frac{C_1 n}{\delta_N^2 \log n} - \zeta_N$ *there exists* $A_N \subseteq [N]$ *of cardinality at least* $n^{C_2 \varepsilon_N^2 \delta_N^2 \zeta_N / n^4}$ *satisfying* (M.1) *and* (M.2).

In particular, for both models we obtain that for fixed $\delta, \varepsilon, \zeta > 0$ with probability at least $1 - \zeta$ there exist $n^{c_{\delta,\varepsilon,\zeta}}$ (for a constant $c_{\delta,\varepsilon,\zeta} > 0$ depending only on δ, ε and ζ) sites such that the Gaussian values on these sites are within additive δn of the expected supremum and the pairwise covariances are at most εn. This improves the corresponding polynomial in $\log n$ sites obtained in [3, 4]. While polynomially many large and almost-orthogonal sites may still be far from satisfactory from the point of view of statistical physics, we remark that a stretched exponentially many large and almost-orthogonal sites can be deduced from our results provided a verification of the prediction that the variances for the supremums (or the free energy at low temperatures) in both directed polymers and the S–K model are polynomial in n with powers strictly less than 1 [2, 10, 15].

We conclude the discussion on multiple peaks by remarking that our results are optimal in the sense that one can construct Gaussian fields consisting of N centered variables of variance 1 whose supremum has variance of order σ_N^2, such that the typical number almost-orthogonal sites whose value is close to the supremum is of the same order as the bound in Theorem 1.2 up to the constant appearing in the exponent. Indeed, for a fixed value of N and of $\rho > 0$, define

$$K = \lfloor \mathrm{e}^{1/\rho^2} \rfloor.$$

Now, let $\mathbf{X}_N$ be a the Gaussian process constructed by taking K independent standard Gaussian variables, and duplicating N/K identical copies of each of them to obtain N variables. It is easy to check that this construction satisfies $\rho_N \sim \rho$. Moreover, it is easily verified that for any $\varepsilon_N \leq 1/2$, and $\delta_N \leq m(\mathbf{X}_N)$, the set of almost-orthogonal peaks [i.e., the cardinality of A_N satisfying (M.1) and (M.2)] will be of order at most $e^{(c\delta_N)/(m(\mathbf{X}_N)\rho^2)}$ with probability at least $1/2$ (for some absolute constant $c > 0$), which shows that the dependence on ρ_N and δ_N is tight in the sense described above. We remark, however, that there exist Gaussian fields which have significantly more almost extremal and almost-orthogonal sites than what is proved in Theorem 1.2. For instance, it was shown in [5] that any sequence of extremal Gaussian fields exhibit multiple peaks with exponentially many peaks (see [5], Theorem 1.6, for details).

Next, we turn to discuss question (b). There are a number of directions for possible improvement upon (1). For instance, it was recently proved in [6, 8, 13, 14] that the unique minimizer that achieves equality in the isoperimetric inequality [from which (1) is deduced] is the half space and any set that genuinely differs from a half space (in some geometric sense) has a strictly larger Gaussian surface area, and consequently will satisfy a stronger version of (1). In this paper, we approach question (b) from a related but slightly different perspective, elaborated below.

One important direction of research concerned with Gaussian processes is finding sharp estimates for the expectation of the supremum. Using the generic chaining technique and building upon the entropy bound in [7], a celebrated result (known as the *majorizing measure* theorem) was developed by Fernique and Talagrand in [9, 19], which provides an estimate of the expected supremum up to a universal multiplicative constant factor. One of the two major ingredients employed in the proof of the majorizing measure theorem is (1). In view of this, it seems plausible that improving (1) based on information on the expected supremum may shed light toward sharpening the lost constant factor in the majorizing measure theorem, and in particular could hopefully help in determining whether a sequence of Gaussian fields is extremal in the sense that its expected supremum is nearly as large as possible with respect to N.

In this paper, we prove that the exponent in the moderate deviation bound in (1) can be improved under the assumption that the expected supremum is of the same order as that of the independent case, namely of order $\sqrt{\log N}$. While this may seem like a rather strong assumption, we would like to draw the reader's attention to the fact that it is actually satisfied by many important examples of Gaussian processes (including the directed polymer and the S–K model discussed above). The theorem reads the following.

THEOREM 1.6. *Let $\{X_i\}_{i=1}^N$ be a centered Gaussian process with* $\mathrm{Var}[X_i] \le 1$ *for all $1 \le i \le N$ and suppose that $\mathbb{E}\sup_{1\le i\le N} X_i \ge \alpha\sqrt{\log N}$ for a fixed $\alpha > 0$. Then there exists an absolute constant $C > 0$ and a constant $c(\alpha) > 0$ depending only on α such that for all $0 < \beta \le \alpha/100$ and all $N \in \mathbb{N}$ one has*

$$\mathbb{P}\Big(\Big|\sup_{1\le i\le N} X_i - \mathbb{E}\sup_{1\le i\le N} X_i\Big| \ge \beta\sqrt{\log N}\Big) \le CN^{-\beta^2/(2-c(\alpha))}.$$

We remark that our current method does not provide a sharp $c(\alpha)$, and thus we did not attempt to optimize its value. The main point of Theorem 1.6 is to suggest a new direction of research by demonstrating the possibility to improve (1) under the assumption of large expected supremum. We believe that it is of significant interest to obtain a sharp estimate on $c(\alpha)$, thus we suggest the following question.

QUESTION 1.7. Under the assumptions of Theorem 1.6, is it true that for all β with $\mathbb{E}\sup_{1\le i\le N} X_i + \beta\sqrt{\log N} \le \sqrt{2\log N}$, we have

$$\mathbb{P}\Big(\sup_{1\le i\le N} X_i \ge \mathbb{E}\sup_{1\le i\le N} X_i + \beta\sqrt{\log N}\Big) \le N^{-(\beta^2+o_N(1))/(2-\alpha^2)}?$$

Note that the exponent in Question 1.7 is achieved by the Gaussian field $X_i = Z + Z_i$ where Z and Z_i's are independent Gaussian variables such that $\mathrm{Var}\, Z = 1 - \alpha^2/2$ and $\mathrm{Var}\, Z_i = \alpha^2/2$ for all $1 \le i \le N$. In spirit, Theorem 1.6 suggests that large expected supremum implies a good concentration property for the supremum.

It turns out that the converse also holds in some sense. That is, a good concentration for the supremum implies that the expected supremum has to be large.

THEOREM 1.8. *There exists an absolute constant $c > 0$ such that for any centered Gaussian field $\{X_i : 1 \le i \le N\}$ with $\operatorname{Var} X_i = 1$ for all $1 \le i \le N$, we have*

$$\Bigl(\operatorname{Var}\Bigl[\sup_{1\le i\le N} X_i\Bigr]\Bigr)^{1/2}\Bigl(1+\mathbb{E}\Bigl[\sup_{1\le i\le N} X_i\Bigr]\Bigr) \ge c.$$

In general, the expected supremum can be bounded from above by (c.f., [3], Lemma 2.1)

$$\mathbb{E}\sup_{1\le i\le N} X_i \le \sqrt{2\log N}\cdot \sup_{1\le i\le N}\sqrt{\operatorname{Var} X_i}. \tag{4}$$

Combined with Theorem 1.8, it yields the following corollary.

COROLLARY 1.9. *Under the assumption of Theorem 1.8, we have that for an absolute $c > 0$*

$$\operatorname{Var}\Bigl[\sup_{1\le i\le N} X_i\Bigr] \ge \frac{c}{\log N}.$$

The structure of the rest of this paper is as follows: in Section 2, we establish some preliminary lemmas and prove Theorem 1.8. In Section 3, we prove Theorem 1.2 and Theorem 1.3. Finally, in Section 4, we prove Theorem 1.6.

2. Level sets of Gaussian fields. In this section, we prove several lemmas related to level sets of Gaussian fields, which we will use in later sections. We also derive Theorem 1.8 along the way.

Let us now establish some notation that will be used throughout this section and the next. Consider a Gaussian field $\mathbf{X} = \{X_i : i \in S\}$. By rescaling, we can assume without loss of generality that $\operatorname{Var} X_i \le 1$ for all $i \in S$. To lighten notation, this normalization will be assumed in the rest of the paper.

Recall that we defined $M(\mathbf{X}) = \sup_{i\in S} X_i$, $m(\mathbf{X}) = \mathbb{E}M(\mathbf{X})$, and $\rho(\mathbf{X}) = \sqrt{\operatorname{Var} M(\mathbf{X})}$. For any real number t, we also define the (random) superlevel set

$$U_t(\mathbf{X}) = \bigl\{i \in S : X_i \ge tm(\mathbf{X})\bigr\}.$$

Since the Gaussian field under consideration is often denoted by $\mathbf{X}$, for brevity we will simply write m, ρ, and U_t. We extend this shorthand also to Gaussian fields like $\mathbf{X}' = \{X_i' : i \in S\}$, so that $U_t' = U_t(\mathbf{X}')$, $U_t'' = U_t(\mathbf{X}'')$, etc.

For a set $V \subset S$, define $\mathbf{X}_V = \{X_i : i \in V\}$, so that X_V is the process $\mathbf{X}$ restricted to V. Extending our notation for the supremum of $\mathbf{X}$, we write $M(\mathbf{X}_V) = \sup_{i\in V} X_i$, with $M(\mathbf{X}_V) = 0$ if $V = \varnothing$. The results of this section are inspired by the following theorem of Chatterjee, Dembo and Ding [5].

THEOREM 2.1 ([5], Theorem 1.1). *For any* $\alpha \in (0,1)$, *let* $\mathbf{X}$ *be a centered Gaussian process, and let* $\mathbf{X}'$ *be an independent copy of* $\mathbf{X}$. *Then*

$$\frac{\mathbb{E}(M(\mathbf{X}'_{U_\alpha}))}{m} \to \sqrt{1-\alpha^2}$$

in probability as $m \to \infty$.

Theorem 2.1 is not directly applicable for our purposes, because it lacks quantitative bounds. However, the results we now prove can be seen as quantitative variants of Theorem 2.1 and use some of the same ideas.

LEMMA 2.2. *Let* $\mathbf{X}$ *be a centered Gaussian process with* $\operatorname{Var} X_i \le 1$. *For any real number* $t \in (0,1)$, *define* $U_t = \{i | X_i \ge tm\}$. *Then for any* $\lambda > 0$,

$$\mathbb{P}\left(M(\mathbf{X}'_{U_t}) \ge \sqrt{1-t^2} \cdot m + \frac{\lambda}{\sqrt{1-t^2}}\right) \le \frac{\rho^2}{\lambda^2},$$

where $M(\mathbf{X}'_{U_t}) = \sup_{i \in U_t} X'_i$ *and* $\mathbf{X}'$ *is an independent copy of* $\mathbf{X}$.

PROOF. Let $\mathbf{X}'' = t\mathbf{X} + \sqrt{1-t^2}\mathbf{X}'$, so that $\mathbf{X}''$ is equal in distribution to $\mathbf{X}$. Then

$$M(\mathbf{X}'') \ge \sup_{i \in U_t} \mathbf{X}''_i \ge t^2 m + \sqrt{1-t^2} M(\mathbf{X}'_{U_t}),$$

$$M(\mathbf{X}'_{U_t}) \le \frac{M(\mathbf{X}'') - t^2 m}{\sqrt{1-t^2}}.$$

By Chebyshev's inequality,

$$\mathbb{P}(M(\mathbf{X}'') \ge m + \lambda) \le \frac{\rho^2}{\lambda^2},$$

which gives the desired inequality. □

We can now give a proof of Theorem 1.8 using Lemma 2.2.

PROOF OF THEOREM 1.8. In the present notation, our goal is to prove that if $\mathbf{X} = \{X_i, i \in S\}$ is any centered Gaussian process with $\operatorname{Var} X_i = 1$, then we have $\rho \cdot (1+m) \ge c$ for some universal constant $c > 0$.

Let us first consider the relatively uninteresting case when $m \le \frac{1}{2}$. We have

$$\mathbb{P}(M(\mathbf{X}) \ge 1) \ge \mathbb{P}(X_1 \ge 1) \ge \tfrac{1}{9},$$

and so

$$\rho^2 \ge \mathbb{P}(M(\mathbf{X}) \ge 1) \cdot (1-m)^2 \ge \tfrac{1}{36},$$

which yields the desired bound with $c = \frac{1}{6}$.

We now assume that $m > \frac{1}{2}$. Let $t = \sqrt{1 - \frac{1}{4m^2}}$. Note that

$$\sqrt{1-t^2} \cdot m + \frac{(2m)^{-1}}{\sqrt{1-t^2}} = 1,$$

so by Lemma 2.2,

$$\mathbb{P}\big(M\big(\mathbf{X}'_{U_t}\big) \geq 1\big) \leq 4m^2\rho^2, \tag{5}$$

where $\mathbf{X}'$ is an independent copy of $\mathbf{X}$.

On the other hand, note that $t \leq 1 - \frac{1}{8m^2}$. Thus,

$$\mathbb{P}(U_t \neq \varnothing) = \mathbb{P}\big(M(\mathbf{X}) \geq tm\big) \geq \mathbb{P}\left(M(\mathbf{X}) \geq m - \frac{1}{8m}\right) \geq 1 - 64m^2\rho^2,$$

where the last step uses Chebyshev's inequality. Note that when U_t is nonempty, $M(\mathbf{X}'_{U_t})$ stochastically dominates a standard Gaussian. Consequently,

$$\mathbb{P}\big(M\big(\mathbf{X}'_{U_t}\big) \geq 1 | U_t \neq \varnothing\big) \geq \tfrac{1}{8},$$

and so

$$\mathbb{P}\big(M\big(\mathbf{X}'_{U_t}\big) \geq 1\big) = \mathbb{P}\big(M\big(\mathbf{X}'_{U_t}\big) \geq 1 | U_t \neq \varnothing\big) \cdot \mathbb{P}(U_t \neq \varnothing) \geq \tfrac{1}{8}\big(1 - 64m^2\rho^2\big). \tag{6}$$

Combining equations (5) and (6), we obtain

$$4m^2\rho^2 \geq \tfrac{1}{8}\big(1 - 64m^2\rho^2\big),$$

from which it follows that $m\rho \geq \frac{1}{\sqrt{96}}$, as desired. □

The next two lemmas will be used to prove the existence of multiple peaks.

LEMMA 2.3. *Let* $\mathbf{X}$ *be a centered Gaussian process with* $\operatorname{Var} X_i \leq 1$. *Consider any* $t \in (1/2, 1)$. *We may write*

$$\mathbf{X} = t\mathbf{X}' + \sqrt{1-t^2}\mathbf{X}'',$$

where $\mathbf{X}'$ *and* $\mathbf{X}''$ *are independent copies of* $\mathbf{X}$. *Then, there exist universal constants* $c, C > 0$ *such that for any* $\lambda \in (1-t, t)$,

$$\mathbb{P}\big(U_{1-c\lambda} \not\subseteq U'_{t-\lambda}\big) \leq \frac{C\rho^2}{m^2\lambda^2}.$$

PROOF. Define $\mathbf{X}''' = -\sqrt{1-t^2}\mathbf{X}' + t\mathbf{X}''$, so that $\mathbf{X}'''$ and $\mathbf{X}$ are independent, and $\mathbf{X}'''$ has the same law as $\mathbf{X}$. Then, by a simple algebraic manipulation, we may write

$$\mathbf{X}' = t\mathbf{X} + (1-t^2)\mathbf{X}' - t\sqrt{1-t^2}\mathbf{X}'' = t\mathbf{X} - \sqrt{1-t^2}\mathbf{X}'''. \tag{7}$$

For convenience, define $s = 1 - \lambda/18$ and note that the condition $\lambda \geq 1 - t$ implies

$$\sqrt{1-t^2} \leq \sqrt{2(1-t)} \leq \sqrt{2\lambda} \leq 6\sqrt{1-s^2}.$$

Define the event

$$E = \left\{ M(\mathbf{X}'''_{U_s}) < \sqrt{1-s^2} \cdot m + \frac{\lambda m}{24\sqrt{1-s^2}} \right\},$$

which by Lemma 2.2 satisfies $\mathbb{P}(E^C) \leq \frac{576\rho^2}{\lambda^2 m^2}$. If E holds, then (7) implies

$$\begin{aligned}
\inf_{i \in U_s} X'_i &\geq t \inf_{i \in U_s} X_i - \sqrt{1-t^2} \cdot M(\mathbf{X}'''_{U_s}) \\
&\geq tsm - \sqrt{1-t^2} \cdot \left(\sqrt{1-s^2} + \frac{\lambda}{24\sqrt{1-s^2}} \right) m \\
&\geq tm - \frac{1}{18}\lambda m - 6(1-s^2)m - \frac{1}{4}\lambda m \geq tm - \left(\frac{1}{18} + \frac{2}{3} + \frac{1}{4} \right) \lambda m \\
&\geq (t - \lambda)m.
\end{aligned}$$

Thus,

$$\mathbb{P}(U_{1-\lambda/18} \subseteq U'_{t-\lambda}) = \mathbb{P}\Big(\inf_{i \in U_{1-\lambda/18}} X'_i \geq (t-\lambda)m \Big) \geq \mathbb{P}(E) \geq 1 - \frac{576\rho^2}{m^2\lambda^2},$$

which proves the lemma upon taking complements. $\square$

COROLLARY 2.4. *Let* $\mathbf{X}$ *be a centered Gaussian process with* $\operatorname{Var} X_i \leq 1$. *For a given* $\delta > 0$, *let* $\alpha = 1 - \frac{\delta}{4}$. *Write*

$$\mathbf{X} = \alpha \mathbf{X}' + \sqrt{1-\alpha^2}\mathbf{X}'',$$

where $\mathbf{X}'$ *and* $\mathbf{X}''$ *have the same distribution as* $\mathbf{X}$ *and are independent of each other. Then*

$$\mathbb{P}(\mathbf{X}'_{i(\mathbf{X})} \geq (1-\delta)m) \geq 1 - \frac{C\rho^2}{m^2\delta^2}$$

for a universal constant $C > 0$, *where* $\rho^2 = \operatorname{Var}[M(\mathbf{X})]$.

PROOF. We use the notation of Lemma 2.3, with $t = \alpha$ and $\lambda = \frac{3}{4}\delta$. Then

$$\begin{aligned}
&\mathbb{P}(\mathbf{X}'_{i(\mathbf{X})} < (1-\delta)m) \\
&\quad = \mathbb{P}(i(\mathbf{X}) \notin U'_{1-\delta}) \\
&\quad \leq \mathbb{P}\left(M(\mathbf{X}) \leq \left(1 - \frac{3}{4}c\delta\right)m \right) + \mathbb{P}\left(\sup_{i \in S, i \notin U'_{1-\delta}} \mathbf{X}_i \geq \left(1 - \frac{3}{4}c\delta\right)m \right) \\
&\quad \leq \frac{16\rho^2}{9c^2m^2\delta^2} + \frac{16C\rho^2}{9m^2\delta^2},
\end{aligned}$$

by Chebyshev's inequality and Lemma 2.3. $\square$

3. Superconcentration implies multiple peaks. This section is devoted to the proofs of Theorems 1.2 and 1.3. We use the notation introduced in Section 2 and additionally define for a Gaussian field $\mathbf{X} = \{X_i : i \in S\}$ the quantities $R(i, j) = \mathrm{Cov}(X_i, X_j)$ for $i, j \in S$.

3.1. *Proof of Theorem* 1.2. Theorem 1.2 can be directly deduced by applying the following result to each Gaussian field $\mathbf{X}_N$ in the sequence.

THEOREM 3.1. *There exist absolute positive constants C_1, C_2 such that the following holds. For any $0 < \varepsilon, \delta, \zeta \le 1$ with probability at least $1 - \frac{C_1 \rho^2}{m^2 \delta^2} - \zeta$ there exists $A \subset S$ with cardinality at least $\exp(\frac{C_2 \varepsilon^2 \delta \zeta}{\rho^2})$ such that $X_i \ge (1-\delta) m(\mathbf{X})$ for each $i \in A$, and $|R(i, j)| < \varepsilon$ for each distinct $i, j \in A$.*

Proving Theorem 3.1 amounts to showing that $U_{1-\delta}$ has a large almost-orthogonal subset (i.e., a subset where the pairwise correlations are at most ε). A preliminary and seemingly innocent question is whether we are able to find a large subset of S of almost-orthogonal variables. It turns out that the concentration property for the supremum of the Gaussian field on S guarantees the existence of a large almost-orthogonal subset of S, as shown in Lemma 3.2 below.

We can actually leverage this deterministic result by fixing the random set $U_{1-\delta}$ and considering an independent copy $\mathbf{X}'$ of $\mathbf{X}$. If the supremum of $\mathbf{X}'$ over $U_{1-\delta}$ exhibits a good concentration property, then $U_{1-\delta}$ has a large almost-orthogonal subset. Of course, some work is required to show that $M(\mathbf{X}'_{U_{1-\delta}})$ has the required concentration property with high probability. A key ingredient is Corollary 2.4.

We begin with the deterministic claim that any Gaussian process which exhibits superconcentration has a large subset of almost-orthogonal variables.

LEMMA 3.2. *Let $\mathbf{X} = \{X_i : i \in S\}$ be a (not necessarily centered) Gaussian process such that $\mathrm{Var}(X_i) \le 1$ for all $i \in S$. For a given $\varepsilon > 0$, if $[r, s]$ is an interval of length at most $\frac{\varepsilon}{8}$ such that*

$$\mathbb{P}\big(M(\mathbf{X}) \notin [r, s]\big) < \tfrac{1}{4}, \tag{8}$$

then there exists $A \subset S$ such that

$$|A| \ge \mathrm{e}^{\varepsilon^2/(32(r-s)^2)}, \tag{9}$$

and for every distinct $i, j \in A$, $|R(i, j)| \le \varepsilon$.

PROOF. It is sufficient, and for technical reasons convenient, to consider the case where, ranging over all distinct $i, j \in S$, the values $|R(i, j)|$ are all distinct [except $R(i, j) = R(j, i)$]. Indeed, correlations satisfying this property are dense among all possible correlations, and the claim of the lemma is a closed property.

Let $A \subset S$ be a maximal set (with respect to inclusion) satisfying $|R(i, j)| \le \varepsilon$ for all $i, j \in A$. We will show that such a set must satisfy (9). For each $i \in S$, let $b(i)$ denote the element of A which maximizes $|R(i, b(i))|$. Note that by the maximality of A, we necessarily have $|R(i, b(i))| \ge \varepsilon$ for all $i \in S$.

Letting $N = |S|$, we now consider the probability space underlying $\mathbf{X}$ as a standard N-dimensional Gaussian Γ with density γ. Let $\{v_i\}_{i\in S}$ be vectors with norm at most 1 such that $X_i = \langle \Gamma, v_i\rangle + \mu_i$, so that $\mu_i = \mathbb{E}(X_i)$ and $\langle v_i, v_j\rangle = R(i, j)$. Define for $x \in \mathbb{R}^N$,

$$m(x) = \sup_{i\in S} \langle x, v_i\rangle + \mu_i$$

so that $m(\Gamma) \sim M(\mathbf{X})$. In addition, we define

$$i(x) = \arg\max_{i\in S} \bigl(\langle x, v_i\rangle + \mu_i\bigr).$$

For a positive constant $c > 0$ to be specified later, define a piecewise linear mapping $f_c : \mathbb{R}^N \to \mathbb{R}^N$ as follows: for a point $x \in \mathbb{R}^N$, let (a, η) be the element of $A \times \{-1, 1\}$ which maximizes $m(x + c\eta v_a)$, and define $f_c(x) = x + c\eta v_a$.

Our next goal is to show that the function f_c is injective. To do this, we fix $x \in \mathbb{R}^N$, and let (a, η) be as above. For notational convenience, write $y = f_c(x)$. Then, by the definition of f_c,

$$\langle x + c\eta v_a, v_{i(y)}\rangle + \mu_{i(y)} = \langle y, v_{i(y)}\rangle + \mu_{i(y)} = m(y) \ge \langle x \pm c v_{b(i(y))}, v_{i(y)}\rangle + \mu_{i(y)},$$

where the plus or minus indicates that the inequality holds for either choice of sign. It follows that

$$\bigl|\langle v_a, v_{i(y)}\rangle\bigr| \ge \bigl|\langle v_{b(i(y))}, v_{i(y)}\rangle\bigr|.$$

On the other hand, by the definition of $b(i(y))$ we have

$$\bigl|\langle v_{b(i(y))}, v_{i(y)}\rangle\bigr| \ge \bigl|\langle v_a, v_{i(y)}\rangle\bigr|,$$

so in fact $|\langle v_{b(i(y))}, v_{i(y)}\rangle| = |\langle v_a, v_{i(y)}\rangle|$.

Recalling the assumption that the $|R(i, j)| = |\langle v_i, v_j\rangle|$ are distinct, this implies that $b(i(y)) = a$. Therefore,

$$x = y - c v_{b(i(y))} \operatorname{sign}\bigl(\langle v_{b(i(y))}, y\rangle\bigr),$$

where $\operatorname{sgn}(x) = x/|x|$ if $x \neq 0$ and $\operatorname{sgn}(0) = 0$. This recovers x from $y = f_c(x)$, showing that f_c is injective.

Next, we fix $c = \frac{s-r}{\varepsilon}$ and consider the region

$$\mathcal{R} = \Bigl\{x \in \mathbb{R}^N : m(x) \ge r, \sup_{a\in A} \langle x, v_a\rangle \le \sqrt{2\log|A|} + 3\Bigr\}.$$

We claim that

$$x \in \mathcal{R} \quad \Longrightarrow \quad m\bigl(f_c(x)\bigr) \ge s. \tag{10}$$

Indeed, we have for all $x \in \mathbb{R}^N$,

$$m\big(f_c(x)\big) \geq m(x \pm cv_{b(i(x))}) \geq \langle x \pm cv_{b(i(x))}, v_{i(x)}\rangle + \mu_{i(x)}$$
$$\geq \langle x, v_{i(x)}\rangle + c\varepsilon + \mu_{i(x)} = m(x) + c\varepsilon.$$

In addition, under the assumption $x \in \mathcal{R}$, we have

$$m\big(f_c(x)\big) \geq m(x) + c\varepsilon \geq r + c\varepsilon = s,$$

thereby proving (10). Now, equation (10) implies that

$$\mathbb{P}\big(M(\mathbf{X}) \geq s\big) \geq \gamma\big(f_c(\mathcal{R})\big).$$

Therefore, we conclude from the assumption (8) that necessarily

$$\gamma\big(f_c(\mathcal{R})\big) \leq \tfrac{1}{4}. \tag{11}$$

In the following, we will suppose for the sake of contradiction that equation (9) is not satisfied and conclude that $\gamma(f_c(\mathcal{R})) > \frac{1}{4}$, thus concluding the lemma.

By (4), we have $\mathbb{E}(\sup_{i \in A}\langle \Gamma, v_i\rangle) \leq \sqrt{2\log|A|}$. Furthermore, a simple application of (1) show that the above maximum is relatively concentrated around its expectation in the sense that

$$\mathbb{P}\Big(\sup_{i \in A}\langle \Gamma, v_i\rangle \geq \sqrt{2\log|A|} + 3\Big) \leq \tfrac{1}{8}.$$

We also have by hypothesis

$$\mathbb{P}\big(M(\mathbf{X}) \leq r\big) \leq \tfrac{1}{4}.$$

Thus, $\gamma(\mathcal{R}) \geq \frac{5}{8}$. Note that for any $x, v \in \mathbb{R}^N$ with $|v| \leq 1$ and $\langle x, v\rangle \leq R$, we have

$$\mathrm{e}^{-\|x+cv\|^2/2} \geq \mathrm{e}^{-cR-c^2/2}\mathrm{e}^{-\|x\|^2/2}.$$

Thus,

$$\gamma\big(f_c(\mathcal{R})\big) \geq \mathrm{e}^{-c(\sqrt{2\log|A|}+3)-c^2/2}\gamma(\mathcal{R}).$$

Recall the hypothesis that $s - r \leq \frac{\varepsilon}{8}$, so $c \leq \frac{1}{8}$. Also, note that $c\sqrt{2\log|A|} \leq \frac{1}{4}$. Hence,

$$\gamma\big(f_c(\mathcal{R})\big) \geq \mathrm{e}^{-c\sqrt{2\log|A|}-1/2}\gamma(\mathcal{R}) \geq \mathrm{e}^{-3/4} \cdot \tfrac{5}{8} > \tfrac{1}{4}$$

which contradicts (11), and the lemma is proven. □

We are now ready to prove the connection between superconcentration and multiple peaks.

PROOF OF THEOREM 3.1. We begin by fixing some $\zeta > 0$, writing $\alpha = 1 - \frac{\delta}{4}$, and considering the decomposition $\mathbf{X} = \alpha\mathbf{X}' + \sqrt{1-\alpha^2}\mathbf{X}''$, where $\mathbf{X}'$ and $\mathbf{X}''$ are independent copies of $\mathbf{X}$. Recall the definition

$$U'_{1-\delta} = \big\{i \in S : X'_i \geq (1-\delta)\mathbb{E}\big(M(\mathbf{X}')\big)\big\}.$$

To lighten notation, we will write U' for $U'_{1-\delta}$ in the rest of the proof.

Since $\mathbf{X}'$ has the same distribution as $\mathbf{X}$, it is enough to show that with probability at least $1 - 4C\rho^2/(m^2\delta^2) - \zeta$ there exists a subset $A \subset U'$ such that

$$|A| \geq \mathrm{e}^{(\zeta\varepsilon^2(1-\alpha^2))/(32\rho^2)} \geq \mathrm{e}^{(\zeta\varepsilon^2\delta)/(128\rho^2)} \tag{12}$$

and

$$i, j \in A, i \neq j \quad \Longrightarrow \quad \big|R(i,j)\big| \leq \varepsilon, \tag{13}$$

where $C > 0$ is a universal constant. To this end, we consider

$$\mathbf{Y}(\mathbf{X}') = \frac{\alpha \mathbf{X}'_{U'}}{\sqrt{1-\alpha^2}} + \mathbf{X}''_{U'}.$$

It is convenient to separate the two sources of randomness in $\mathbf{Y}(\mathbf{X}')$. In what follows, we will condition on the realization of $\mathbf{X}'$ (therefore also U'), and consider $\mathbf{Y}(\mathbf{X}')$ as a noncentered Gaussian process indexed over set U', where the randomness comes from the process $\mathbf{X}''$ and the mean vector is given by $\frac{\alpha \mathbf{X}'_{U'}}{\sqrt{1-\alpha^2}}$. Define

$$g(\mathbf{X}') = \mathbb{P}\biggl(M\bigl(\mathbf{Y}(\mathbf{X}')\bigr) \in \biggl[\frac{m - \rho\zeta^{-1/2}}{\sqrt{1-\alpha^2}}, \frac{m + \rho\zeta^{-1/2}}{\sqrt{1-\alpha^2}}\biggr] \bigg| \mathbf{X}'\biggr)$$

and let E be the event that $\{g(\mathbf{X}') \geq 3/4\}$. Note that E is measurable in the ρ-field generated by $\mathbf{X}'$. Clearly, whenever the event E holds, the Gaussian process $(\mathbf{Y}(\mathbf{X}'))$ (conditioned on $\mathbf{X}'$ in the aforementioned manner) satisfies the assumption (8) with

$$s - r = \frac{2\rho}{\sqrt{\zeta}\sqrt{1-\alpha^2}}.$$

Therefore, by applying Lemma 3.2 we learn that whenever E holds, there exists a subset $A \subset U'$ satisfying (12) and (13). It thus remains to show that

$$\mathbb{P}(E) > 1 - 4C\rho^2/\bigl(m^2\delta^2\bigr) - \zeta. \tag{14}$$

To this end, we define another event

$$F = \bigl\{X'_{i(\mathbf{X})} \geq (1-\delta)m\bigr\}.$$

By Corollary 2.4, we have

$$\mathbb{P}(F) > 1 - \frac{C\rho^2}{m^2\delta^2}, \tag{15}$$

where $C > 0$ is a universal constant. Now, observe that whenever the event F holds, one has

$$\sqrt{1-\alpha^2}\, M\bigl(\mathbf{Y}(X')\bigr) = M(\mathbf{X}).$$

It follows from the definition of $g(\cdot)$ that

$$\mathbb{E}g(\mathbf{X}') \geq \mathbb{P}\biggl(M(\mathbf{X}) \in \biggl[m - \frac{\rho}{\sqrt{\zeta}}, m + \frac{\rho}{\sqrt{\zeta}}\biggr]\biggr) - \mathbb{P}(F^C) \geq 1 - C\rho^2/(m^2\delta^2) - \zeta,$$

where in the second passage, we used (15) and Chebyshev's inequality. An application of Markov's inequality with the last equation establishes (14), and the proof is complete. □

3.2. *Proof of Theorem* 1.3. For $\beta > 0$, define $F_\beta = (\frac{1}{\beta}\log\sum_{i\in S} \mathrm{e}^{\beta X_i})$, and write $\rho_\beta^2 = \operatorname{Var} F_\beta$. Theorem 1.3 can be deduced by applying the following result to each Gaussian field in the sequence.

THEOREM 3.3. *There exist constants $C_1, C_2, C_3 > 0$ such that the following holds. Suppose that $\rho_\beta \leq \tilde{\rho}_\beta$ for all $\beta \geq 0$. For any $0 < \delta, \varepsilon, \zeta < 1$ and all $\beta \geq C_1 \max(\frac{\log N}{\delta m}, \frac{1}{\tilde{\rho}_\beta}, \frac{\delta\varepsilon^2}{\tilde{\rho}_\beta^3})$, with probability at least $1 - \frac{C_2}{m^2\delta^2} - \zeta$ there exists $A \subset S$ with cardinality at least $\exp(\frac{C_3\varepsilon^2\delta\zeta}{\tilde{\rho}_\beta^2})$ such that $X_i \geq (1-\delta)m(\mathbf{X})$ for each $i \in A$, and $|R(i,j)| < \varepsilon$ for each distinct $i, j \in A$.*

The proof is similar to that of Theorem 1.2, but a few changes are needed. In what follows, we will omit details which are repeated from the proof of Theorem 1.2 and highlight the differences. The reader is advised to become familiar with the previous subsection before reading this one.

First, we need an analogous version of Lemma 3.2 in which the assumption of a concentrated supremum is replaced by concentration of the free energy.

LEMMA 3.4. *Let $\mathbf{X} = \{X_i : i \in S\}$ be a (not necessarily centered) Gaussian process such that $\operatorname{Var}(X_i) \leq 1$ for all $i \in S$. For a given $\varepsilon > 0$, suppose that $[r, s]$ is an interval of length at most $\frac{\varepsilon}{8}$ such that*

$$\mathbb{P}\bigl(F_\beta(\mathbf{X}) \notin [r,s]\bigr) < \tfrac{1}{4}. \tag{16}$$

Furthermore, suppose that

$$\beta \geq \frac{\varepsilon^2}{128(s-r)^3}. \tag{17}$$

Then there exists $A \subset S$ such that

$$|A| \geq \mathrm{e}^{\varepsilon^2/(128(r-s)^2)}, \tag{18}$$

and for every distinct $i, j \in A$, $|R(i,j)| \leq \varepsilon$.

PROOF. As in the proof of Lemma 3.2, it suffices to handle the case where the $|R(i,j)|$ are all distinct.

Define $A \subset S$ and $b : S \to A$ as in the proof of Lemma 3.2. Our goal is to show that A satisfies (18). Suppose for the sake of contradiction that it does not.

As before, letting $N = |S|$, consider the probability space underlying $\mathbf{X}$ as a standard N-dimensional Gaussian Γ with density γ. Let $\{v_i\}_{i\in S}$ be vectors with norm at most 1 such that $X_i = \langle \Gamma, v_i\rangle + \mu_i$, so that $\mu_i = \mathbb{E}(X_i)$ and $\langle v_i, v_j\rangle = R(i,j)$. Define for $x \in \mathbb{R}^N$,

$$m_\beta(x) = \frac{1}{\beta}\log\sum_{i\in S} \mathrm{e}^{\beta(\langle x, v_i\rangle + \mu_i)}$$

so that $m_\beta(\Gamma) \sim F_\beta(\mathbf{X})$. Let us also define for $a \in A$ and $\chi \in \{-1,1\}$ the quantities

$$g_\beta^\chi(x,a) = \sum_{\substack{j\in S, b(j)=a\\ \operatorname{sgn}(v_a, v_j)=\chi}} \mathrm{e}^{\beta(\langle x, v_j\rangle + \mu_j)},$$

$$\hat{m}_\beta(x) = \frac{1}{\beta}\sup_{a\in A, \chi\in\{-1,1\}} g_\beta^\chi(x,a).$$

Evidently, we have

$$m_\beta(x) \ge \hat{m}_\beta(x) \ge m_\beta(x) - \beta^{-1}\log(2|A|).$$

For a positive constant $c > 0$ to be specified later, define a piecewise linear mapping $f_c : \mathbb{R}^N \to \mathbb{R}^N$ as follows: for a point $x \in \mathbb{R}^N$, let (a,η) be the element of $A \times \{-1,1\}$ which maximizes $\hat{m}_\beta(x + c\eta v_a)$, and define $f_c(x) = x + c\eta v_a$. We next verify that f_c is injective outside of a set of probability zero.

Write $y = f_c(x)$, and let

$$\bigl(\hat{i}(y), \chi(y)\bigr) = \operatorname*{arg\,max}_{a\in A, \chi\in\{-1,1\}} g_\beta^\chi(y,a).$$

By definition of (a,η), we get that

$$g_\beta^{\chi(y)}\bigl(x + c\eta v_a, \hat{i}(y)\bigr) = g_\beta^{\chi(y)}\bigl(y, \hat{i}(y)\bigr) \ge g_\beta^{\chi(y)}\bigl(x \pm c v_{\hat{i}(y)}, \hat{i}(y)\bigr).$$

On the other hand, by definition of $b(i(y))$, we see that

$$\max\bigl(g_\beta^{\chi(y)}\bigl(x + c v_{\hat{i}(y)}, \hat{i}(y)\bigr), g_\beta^{\chi(y)}\bigl(x - c v_{\hat{i}(y)}, \hat{i}(y)\bigr)\bigr) \ge g_\beta^{\chi(y)}\bigl(x + c\eta v_a, \hat{i}(y)\bigr).$$

Altogether, we deduce that

$$\begin{aligned} &\max\bigl(g_\beta^{\chi(y)}\bigl(x + c v_{\hat{i}(y)}, \hat{i}(y)\bigr), g_\beta^{\chi(y)}\bigl(x - c v_{\hat{i}(y)}, \hat{i}(y)\bigr)\bigr) \\ &\quad = g_\beta^{\chi(y)}\bigl(x + c\eta v_a, \hat{i}(y)\bigr). \end{aligned} \tag{19}$$

Using the assumption that the $|R(i,j)| = |\langle v_i, v_j\rangle|$ are distinct, we see that almost surely with respect to $x \sim \gamma$ we have that the values $\{g_\beta^{\pm 1}(x + \chi c v_i, \hat{i}(x +$

$\chi c v_i)) : \chi \in \{-1, 1\}, i \in A\}$ are all distinct. Combined with (19) [note that the left-hand side of (19) is a function of y], it follows that for almost surely every given y we could reconstruct (a, η) (and thus x), thereby completing the verification of the injectivity of f_c.

We now take $c = \frac{\beta^{-1}\log(2|A|)+s-r}{\varepsilon}$ and consider the region

$$\mathcal{R} = \Big\{x \in \mathbb{R}^N : m_\beta(x) \geq r, \sup_{a\in A}\langle x, v_a\rangle \leq \sqrt{2\log|A|} + 3\Big\}.$$

We have for all $x \in \mathbb{R}^N$,

$$\begin{aligned} m_\beta\big(f_c(x)\big) &\geq \hat{m}_\beta\big(f_c(x)\big) \geq \hat{m}_\beta(x) + c\varepsilon \\ &= \hat{m}_\beta(x) + \beta^{-1}\log\big(2|A|\big) + (s-r) \geq m_\beta(x) + (s-r). \end{aligned}$$

Thus, we have

$$x \in \mathcal{R} \quad \Longrightarrow \quad m\big(f_c(x)\big) \geq s.$$

The rest of the proof proceeds in exactly the same manner as the proof of Lemma 3.2. The only difference is that we have chosen a different value of c. However, by the hypothesis that (18) is not satisfied, we have

$$\log|A| < \frac{\varepsilon^2}{128(s-r)^2}.$$

The lower bound condition on β then implies

$$c = \frac{\beta^{-1}\log|A| + s - r}{\varepsilon} \leq \frac{2(s-r)}{\varepsilon}. \tag{20}$$

We thus have

$$c\sqrt{2\log|A|} \leq \tfrac{1}{4}.$$

Also, combining (20) with the condition that $s - r \leq \frac{\varepsilon}{16}$, we obtain

$$c \leq \tfrac{1}{8}.$$

These are the only properties of c needed in the proof of Lemma 3.2, so the same argument works. □

To prove Theorem 3.3, we use the same setup as the proof of Theorem 3.1. Let $\alpha = 1 - \frac{\delta}{4}$, and make the decomposition $\mathbf{X} = \alpha\mathbf{X}' + \sqrt{1-\alpha^2}\mathbf{X}''$. Let $U' = U'_{1-\delta} = \{i \in S : \mathbf{X}'_i \geq (1-\delta)m\}$. We are not assuming any superconcentration of $M(\mathbf{X})$, but we will implicitly use the bound $\rho^2(M(\mathbf{X})) \leq 1$.

Recall that in the proof of Theorem 3.1, we used Corollary 2.4 to show that the maximum comes from the indices in U' with high probability. The next lemma establishes a similar statement for the free energy; although all indices contribute to the free energy, we show that with high probability, most of the contribution comes from indices in U'.

LEMMA 3.5. *There exists a universal constant C such that whenever $\beta > \frac{C \log N}{\delta m}$, we have*

$$F_\beta(\mathbf{X}) \ge F_\beta(\mathbf{X}_{U'}) \ge F_\beta(\mathbf{X}) - \frac{1}{\beta},$$

with probability at least $1 - \frac{C}{\delta^2 m^2}$.

PROOF. Clearly, $F_\beta(\mathbf{X}) \ge F_\beta(\mathbf{X}_{U'})$ holds deterministically. Thus, we focus our attention on the second inequality. According to Lemma 2.3, we have $M(\mathbf{X}_{S \setminus U'}) \le (1 - c'\delta)m$ with probability at least $1 - \frac{C'}{\delta^2 m^2}$ for universal constants $c', C' > 0$. Furthermore, Chebyshev's inequality tells us that $M(\mathbf{X}) \ge (1 - c'\delta/2)m$ with probability at least $1 - \frac{1}{c'^2 \delta^2 m^2}$.

Let $\delta' = c'\delta/2$. Then, excluding events of probability at most $\frac{C}{\delta^2 m^2}$, where C is a universal constant, we may assume that $M(\mathbf{X}) \ge (1 - \delta')m$ and $M(\mathbf{X}_{S \setminus U'}) \le (1 - 2\delta')m$. In that case,

$$\begin{aligned}
F_\beta(\mathbf{X}) &= \frac{1}{\beta} \log \sum_{i \in S} e^{\beta X_i} = \frac{1}{\beta} \log\biggl(\sum_{i \in U'} e^{\beta X_i} + \sum_{i \in S \setminus U'} e^{\beta X_i} \biggr) \\
&\le \frac{1}{\beta} \log\biggl(\sum_{i \in U'} e^{\beta X_i} + N e^{\beta(1 - 2\delta')m} \biggr) \\
&\le \frac{1}{\beta} \log\biggl(\sum_{i \in U'} e^{\beta X_i} + N e^{-\beta\delta' m} e^{\beta(1 - \delta')m} \biggr) \\
&\le \frac{1}{\beta} \log\biggl(\sum_{i \in U'} e^{\beta X_i} + N e^{-\beta\delta' m} \cdot e^{\beta M(\mathbf{X})} \biggr) \\
&\le \frac{1}{\beta} \log\biggl(\sum_{i \in U'} e^{\beta X_i} + N e^{-\beta\delta' m} \sum_{i \in U'} e^{\beta X_i} \biggr) \\
&= F_\beta(\mathbf{X}_{U'}) + \frac{1}{\beta} \log\bigl(1 + N e^{-\beta\delta' m}\bigr).
\end{aligned}$$

If C is taken to be sufficiently large, the assumption $\beta > \frac{C \log N}{\delta m}$ implies the second term in the last expression is bounded by $\frac{1}{\beta}$. This proves the lemma. $\square$

PROOF OF THEOREM 3.3. Recall the strategy of proving Theorem 3.1: we first show that concentration of $F_\beta(\mathbf{X})$ implies with high probability a concentration of $F_\beta(\mathbf{X}''_{U'} + \mu)$ for some μ, and then we apply Lemma 3.4 to show the existence of many almost-orthogonal indices in U', which proves the theorem.

Let $f_\beta = \mathbb{E}(F_\beta(\mathbf{X}))$. By Chebyshev's inequality,

$$\mathbb{P}\bigl(F_\beta(\mathbf{X}) \in [f_\beta - \tilde{\rho}_\beta \zeta^{-1/2}, f_\beta - \tilde{\rho}_\beta \zeta^{-1/2}]\bigr) \ge 1 - \frac{\rho_\beta^2 \zeta}{\tilde{\rho}_\beta^2} \ge 1 - \zeta. \tag{21}$$

Define the process

$$\mathbf{Y}(\mathbf{X}') = \frac{1}{\sqrt{1-\alpha^2}} \cdot \mathbf{X}_{U'} = \frac{\alpha \mathbf{X}'_{U'}}{\sqrt{1-\alpha^2}} + \mathbf{X}''_{U'},$$

(where, as above, $\alpha = 1 - \delta/4$) and let $\beta' = \sqrt{1-\alpha^2}\beta$. Recall the hypotheses that $\beta \geq \frac{C_1 \log N}{\delta m}$ and $\beta \geq \frac{C_1}{\tilde{\rho}_\beta}$, and take $C_1 \geq \max(1, C)$, where C is the constant of Lemma 3.5. Then Lemma 3.5 tells us that

$$\left|\sqrt{1-\alpha^2} \cdot F_{\beta'}(\mathbf{Y}) - F_\beta(\mathbf{X})\right| = \left|F_\beta(\mathbf{X}_{U'}) - F_\beta(\mathbf{X})\right| \leq \frac{1}{\beta} \leq \tilde{\rho}_\beta \zeta^{-1/2}$$

with probability at least $1 - \frac{C}{\delta^2 m^2}$. Thus, combining with (21), $F_{\beta'}(\mathbf{Y})$ lies in an interval $[r, s]$ of size

$$\frac{4\tilde{\rho}_\beta}{\sqrt{\zeta}\sqrt{1-\alpha^2}}$$

with probability at least $1 - \zeta - \frac{C}{\delta^2 m^2}$. As in the proof of Theorem 3.1, define

$$g(\mathbf{X}') = \mathbb{P}\big(F_{\beta'}\big(\mathbf{Y}(\mathbf{X}')\big) \in [r, s] | \mathbf{X}'\big),$$

so according to the above we have

$$\mathbb{E}\big[g(\mathbf{X}')\big] \geq 1 - \zeta - \frac{C}{\delta^2 m^2}. \tag{22}$$

Recall that $\beta' = \sqrt{1-\alpha^2}\beta$. By our assumption on β, we get that

$$\begin{aligned} \beta' &\geq \frac{C_1\sqrt{1-\alpha^2} \cdot \delta\varepsilon^2}{\tilde{\rho}_\beta^3} \geq \frac{2C_1(\sqrt{1-\alpha^2})^3\varepsilon^2}{\tilde{\rho}_\beta^3} \\ &\geq \frac{128C_1(\sqrt{1-\alpha^2})^3\zeta^{3/2}\varepsilon^2}{64\tilde{\rho}_\beta^3} = \frac{128C_1\varepsilon^2}{(s-r)^3}. \end{aligned}$$

Thus, on the event $g(\mathbf{X}') > \frac{3}{4}$ and taking C_1 sufficiently large, the hypotheses of Lemma 3.4 are fulfilled with $s - r = \frac{4\tilde{\rho}_\beta}{\sqrt{\zeta}\sqrt{1-\alpha^2}}$ and inverse temperature β'. It follows that U' contains at least

$$\exp\left(\frac{\varepsilon^2}{128(s-r)^2}\right) = \exp\left(\frac{\varepsilon^2\zeta(1-\alpha^2)}{2048\tilde{\rho}_\beta^2}\right) \geq \exp\left(\frac{C_3\varepsilon^2\zeta\delta}{\tilde{\rho}_\beta^2}\right)$$

indices whose pairwise covariances do not exceed ε in magnitude, as required. Finally equation (22), combined with Markov's inequality, teaches us that $g(\mathbf{X}') > \frac{3}{4}$ occurs with probability at least $1 - 4\zeta - \frac{4C}{\delta^2 m^2}$, which completes the proof after a renormalization of constants. $\square$

4. A moderate deviation bound based on expectation. This section is devoted to the proof of Theorem 1.6. The proof is based on stochastic calculus, and we need some preliminary notation. For a continuous martingale M_t adapted to a filtration $\mathcal{F}_t$, we denote by $[M]_t$ the quadratic variation of M_t between time 0 and t. By dM_t we denote the Itô differential of M_t, which we understand as a predictable process ρ_t such that M_t satisfies the stochastic differential equation $dM_t = \rho_t dW_t$ where W_t is a standard Wiener process.

Fix a Gaussian field $\mathbf{X} = \{X_i, 1 \le i \le N\}$ such that $\operatorname{Var}[X_i] \le 1$ for all $1 \le i \le N$. Now, take $(B_t)_{t\ge 0}$ to be a be a standard Brownian motion in $\mathbb{R}^N$ with a corresponding filtration $\mathcal{F}_t$. Clearly, there exist vectors $\{v_i : 1 \le i \le N\}$ of Euclidean norms at most 1 such that we can represent the Gaussian field $\mathbf{X}$ by

$$X_i = \langle v_i, B_1\rangle \qquad \text{for every } 1 \le i \le N.$$

Define $f : \mathbb{R}^N \mapsto \mathbb{R}$ by $f(x) = \sup_{1\le i\le N}\langle v_i, x\rangle$ so that

$$f(B_1) \sim \sup_{1\le i\le N} X_i.$$

Our goal is to derive a moderate deviation bound for $f(B_1)$. A central component of the proof will be the Doob martingale

$$S_t = \mathbb{E}\big[f(B_1)|\mathcal{F}_t\big],$$

generated by the random variable $f(B_1)$ and filtration $\mathcal{F}_t$. Thanks to the Dambis/Dubins–Schwartz theorem, we can then view $(S_t)_{0\le t\le 1}$ as a time change of (one-dimensional) Brownian motion stopped at some random time $\tau = [S]_1$ (which corresponds to $t = 1$). The main idea is that, due to the Gaussian concentration of the maximum for a Brownian motion stopped before time T, it will suffice to prove that with overwhelming probability τ is strictly less than 1. To this end, we will try to calculate $d[S]_t$ by means of Itô calculus, in what follows.

For $v \in \mathbb{R}^N$ and $\sigma > 0$, define

$$\gamma_{v,\sigma}(x) = \frac{1}{\sigma^N(2\pi)^{N/2}}\exp\left(-\frac{1}{2\sigma^2}|x-v|^2\right) \qquad \text{for } x \in \mathbb{R}^N.$$

An elementary property of the Brownian motion is that the distribution of B_1 conditioned on $\mathcal{F}_t$ has density $\gamma_{B_t,\sqrt{1-t}}(x)$. Therefore, we have that

$$S_t = \int_{\mathbb{R}^N} f(x)\gamma_{B_t,\sqrt{1-t}}(x)\,dx.$$

For convenience of notation, we write

$$F_t(x) = \gamma_{B_t,\sqrt{1-t}}(x).$$

A direct calculation carried out in [8], Lemma 7, gives that

$$dF_t(x) = (1-t)^{-1}F_t(x)\langle x - B_t, dB_t\rangle.$$

As a result of the above equation, we can calculate

$$dS_t = d\int_{\mathbb{R}^N} f(x)F_t(x)\,dx = (1-t)^{-1}\left\langle \int_{\mathbb{R}^N} f(x)(x-B_t)F_t(x), dB_t\right\rangle,$$

and, therefore, we obtain

$$d[S]_t = (1-t)^{-2}\left|\int_{\mathbb{R}^N}(x-B_t)f(x)F_t(x)\,dx\right|^2 dt,$$

where we recall that $[S]_t$ denotes the quadratic variation for process (S_t). Substituting $y = \frac{x-B_t}{\sqrt{1-t}}$ in the last equation, we get that

$$d[S]_t = (1-t)^{-1}\left|\int_{\mathbb{R}^N} yf(\sqrt{1-t}\,y+B_t)\,d\gamma(y)\right|^2 dt. \tag{23}$$

For convenience, we denote

$$g_t(x) = \frac{f(\sqrt{1-t}\,x+B_t)-f(B_t)}{\sqrt{1-t}}.$$

Plugging this definition into (23), and using the fact that $\int_{\mathbb{R}^N} x\,d\gamma(x)=0$ gives

$$d[S]_t = V_t\,dt,$$

where V_t is defined as

$$V_t = \left|\int_{\mathbb{R}^N} xg_t(x)\,d\gamma(x)\right|^2. \tag{24}$$

We wish to show that V_t is strictly less than 1 for a strictly positive time interval. To this end, let $\varepsilon,\delta>0$ be two small numbers to be fixed, and define two events

$$\begin{aligned} &E_1 = \{V_t \le 1-\varepsilon, \forall 0\le t\le\delta\} \quad \text{and} \\ &E_2 = \left\{f(B_t)\le \frac{\alpha}{2}\sqrt{\log N}, \forall 0\le t\le \delta\right\} \end{aligned} \tag{25}$$

(recall that by our assumption we have $\mathbb{E}f(B_1)\ge\alpha\sqrt{\log N}$). In order to bound $\mathbb{P}(E_1)$, we will need the next lemma whose point is that if $|V_t|$ is at some point close to 1, then $\mathbb{E}[f(B_1)-f(B_t)|\mathcal{F}_t]$ cannot be too large.

LEMMA 4.1. *Let $\{\mu_i\}_{i=1}^N$ be such that $\mu_i\le 0$ for all $1\le i\le N$. Define $\tilde f:\mathbb{R}^N\mapsto\mathbb{R}$ by (recall that $|v_i|\le 1$)*

$$\tilde f(x) = \sup_{1\le i\le N}(x\cdot v_i+\mu_i) \qquad \text{for all } x\in\mathbb{R}^N. \tag{26}$$

Also define $\varepsilon = 1-\sup_{\theta\in S^{N-1}}\int_{\mathbb{R}^N}\langle x,\theta\rangle \tilde f(x)\,d\gamma(x)$. Then we have

$$\int_{\mathbb{R}^N}\tilde f(x)\,d\gamma(x)\le 10(1+\sqrt{\varepsilon\log N}) \qquad \text{for all } N\in\mathbb{N}$$

in particular, one has

$$\varepsilon\ge 0. \tag{27}$$

PROOF. Pick $\theta \in S^{N-1}$ such that

$$1 - \int_{\mathbb{R}^N} \langle x, \theta \rangle \tilde{f}(x)\, d\gamma(x) = \varepsilon. \tag{28}$$

For each $x \in \mathbb{R}^N$, consider the unique representation $x = y + z\theta$ where $z = \langle x, \theta \rangle$ and $y \in \theta^\perp$ (i.e., $\langle y, \theta \rangle = 0$). Denote by γ^1 and γ^{N-1} standard Gaussian measures in dimension 1 and $N-1$, respectively, we can view γ^1 as a measure on span$\{\theta\}$ and γ^{N-1} a measure on $\theta^\perp$. It is clear that if $x \sim \gamma$, we have $(z, y) \sim \gamma^1 \otimes \gamma^{N-1}$. Therefore, we get that

$$\int_{\mathbb{R}^N} \langle x, \theta \rangle \tilde{f}(x)\, d\gamma(x) = \int_{\theta^\perp} \int_{\mathbb{R}} z \tilde{f}(y + z\theta)\, d\gamma^1(z)\, d\gamma^{N-1}(y).$$

Applying integration by parts to $\int_{\mathbb{R}} z \tilde{f}(y + z\theta)\, d\gamma^1(z)$, we obtain that

$$\begin{aligned} \int_{\mathbb{R}^N} \langle x, \theta \rangle \tilde{f}(x)\, d\gamma(x) &= \int_{\theta^\perp} \int_{\mathbb{R}} \left(\frac{\partial}{\partial z} f(y + z\theta) \right) d\gamma^1(z)\, d\gamma^{n-1}(y) \\ &= \int_{\mathbb{R}^N} \langle \nabla \tilde{f}(x), \theta \rangle\, d\gamma(x). \end{aligned} \tag{29}$$

For $x \in \mathbb{R}^N$, write

$$i^*(x) = \arg\sup_{1 \le i \le N} (x \cdot v_i + \mu_i)$$

(note that the maximizer is unique with probability 1 when we sample $x \sim \gamma$ and thus $i^*(x)$ is well-defined almost surely. Here, we use the legitimate assumption that the vectors $\{v_i\}$ are distinct). By definition of $\tilde{f}$, we see that $\nabla \tilde{f}(x) = v_{i^*(x)}$. Combined with (28) and (29), it follows that

$$\mathbb{E}\langle v_{i^*(\Gamma)}, \theta \rangle = 1 - \varepsilon,$$

where Γ is a standard Gaussian random vector in $\mathbb{R}^N$. Recall that $|v_i| \le 1$ for all $1 \le i \le N$. In view of the last equation, this fact gives $\varepsilon \ge 0$. As a consequence of Markov's inequality, this fact also teaches us that

$$\mathbb{P}\big(\langle v_{i^*(\Gamma)}, \theta \rangle \ge 1 - 10\varepsilon\big) \ge 9/10. \tag{30}$$

Let $I \subset [N]$ be the set of indices i such that $\langle v_i, \theta \rangle \ge 1 - 10\varepsilon$. For every $i \in I$, write $v_i = u_i \theta + \tilde{v}_i$ where $u_i = \langle v_i, \theta \rangle$ and $\tilde{v}_i \in \theta^\perp$. By our assumption on I, we have $|\tilde{v}_i| \le \sqrt{20\varepsilon}$ for all $i \in I$. Therefore, we have

$$\mathbb{E} \sup_{i \in I} \langle \Gamma, v_i \rangle \le \mathbb{E}\big|\langle \Gamma, \theta \rangle\big| + \mathbb{E} \sup_{i \in I} \langle \tilde{v}_i, \Gamma \rangle \le 1 + \sqrt{40\varepsilon \log |I|},$$

where the last inequality follows from (4). Combined with (1), it then follows that (note that $|I| \le N$)

$$\mathbb{P}\Big(\sup_{i \in I} \langle \Gamma, v_i \rangle \ge \sqrt{40\varepsilon \log N} + 10 \Big) \le 1/5.$$

Combined with (30), using a union bound we get that

$$\mathbb{P}\Big(\sup_{i\in[N]} \langle \Gamma, v_i\rangle \geq \sqrt{40\varepsilon \log N} + 10\Big) \leq 1/2.$$

Together with another application of (1), it completes the proof of the lemma. $\square$

The next lemma applies the above in order to show that with high probability, either V_t remains bounded from 1 for a finite inteval of time, or $f(B_t)$ becomes rather large within a short time.

LEMMA 4.2. *Let E_1, E_2 be defined as in (25). For $\varepsilon \leq \alpha^2 \cdot 10^{-4}$ and an absolute constant $C > 0$, we have*

$$\mathbb{P}(E_2 \setminus E_1) \leq CN^{-\alpha^2/32}.$$

PROOF. Suppose that $E_2 \setminus E_1$ holds, and denote by

$$T = \min\{t \geq 0 : V_t \geq 1 - \varepsilon\}$$

to the first time in which $V_t \geq 1 - \varepsilon$. By definition of E_1^C, we have $T \leq \delta$. Using the decomposition that $B_1 = B_t + (B_1 - B_t)$ where $\frac{B_1 - B_t}{\sqrt{1-t}}$ has density function γ and is independent of B_t, we get that

$$\int_{\mathbb{R}^N} g_t(x)\, d\gamma(x) = \frac{\mathbb{E}[f(B_1) - f(B_t)|\mathcal{F}_t]}{\sqrt{1-t}}. \tag{31}$$

Consequently, we have

$$\begin{aligned}\mathbb{E}\big[f(B_1)|\mathcal{F}_T\big] &= f(B_T) + \mathbb{E}\big[f(B_1) - f(B_T)|\mathcal{F}_T\big] \\ &= f(B_T) + \sqrt{1-t}\int_{\mathbb{R}^N} g_T(x)\, d\gamma(x).\end{aligned}$$

Recalling (24), we see that

$$\left|\int_{\mathbb{R}^N} x g_T(x)\, d\gamma(x)\right|^2 \geq 1 - \varepsilon.$$

Therefore, there exists $\theta \in S^{N-1}$ such that

$$\int_{\mathbb{R}^N} g_T(x)\langle x, \theta\rangle\, d\gamma(x) \geq \sqrt{1-\varepsilon} \geq 1 - \varepsilon.$$

We claim that Lemma 4.1 can be applied with the function $g_T(x)$ (conditioning on the filtration $\mathcal{F}_T$) used in place of the function $\tilde{f}$. Indeed, since $f(cx) = cf(x)$ for all $x \in \mathbb{R}^N$ and $c > 0$, we get that

$$\begin{aligned}g_T(x) &= f(x + B_T/\sqrt{1-T}) - f(B_T/\sqrt{1-T}) \\ &= \sup_{1\leq i\leq N} \langle x + B_T/\sqrt{1-T}, v_i\rangle - \sup_{1\leq i\leq N} \langle B_T/\sqrt{1-T}, v_i\rangle \\ &= \sup_{1\leq i\leq N}\left(x \cdot v_i + \frac{B_T}{\sqrt{1-T}} \cdot v_i - \sup_{1\leq i\leq N} \frac{B_T}{\sqrt{1-T}} \cdot v_i\right).\end{aligned}$$

This implies that it admits the form (26). Applying Lemma 4.1, and using the assumption $\varepsilon \le \alpha^2 \cdot 10^{-4}$, we get

$$\int_{\mathbb{R}^N} g_T(x)\,d\gamma(x) \le 10(\sqrt{\varepsilon \log N} + 1) \le \frac{\alpha}{10}\sqrt{\log N} + 10.$$

This implies that on the event $E = \{T \le \delta\} \cap \{f(B_t) \le \alpha\sqrt{\log N}/2, \forall 0 \le t \le T\}$ (note that $E \supseteq E_2 \setminus E_1$), we have

$$\mathbb{E}[f(B_1)|\mathcal{F}_T] \le \frac{3\alpha}{4}\sqrt{\log N} + 10.$$

Applying (1) (the noncentered version) to

$$f(B_1) = \sup_{1\le i\le N}\{\langle v_i, B_1 - B_T\rangle + \langle v_i, B_T\rangle\},$$

(where we treat B_T as deterministic numbers as we conditioned on $\mathcal{F}_T$) we obtain that

$$\mathbb{P}\biggl(f(B_1) \le \frac{3\alpha}{4}\sqrt{\log N} + 20\bigg|E\biggr) \ge 1/2. \tag{32}$$

Recall the definition of α, according to which

$$\mathbb{E}[f(B_1)] \ge \alpha\sqrt{\log N}.$$

Another application of (1) gives that

$$\mathbb{P}\biggl(f(B_1) \le \frac{3\alpha}{4}\sqrt{\log N} + 20\biggr) \le CN^{-\alpha^2/32},$$

where $C > 0$ is an absolute constant. Combined with (32), we see that

$$\mathbb{P}(E_2 \setminus E_1) \le \mathbb{P}(E) \le 2CN^{-\alpha^2/32}. \qquad \square$$

PROOF OF THEOREM 1.6. We first bound $\mathbb{P}(E_2)$ from below, and we will employ the idea from reflection principle of Brownian motion. Defining

$$T' = \min\{t : f(B_t) \ge \alpha\sqrt{\log N}/2\},$$

we see that $E_2 = \{T' > \delta\}$. Let us denote by $i^*_{T'}$ the maximizer of $f(B_{T'})$. That is, $f(B_{T'}) = \langle v_{i^*_{T'}}, B_{T'}\rangle$. Then we have, on the event $T' \le \delta$,

$$f(B_\delta) \ge f(B_{T'}) + \langle v_{i^*_{T'}}, B_\delta - B_{T'}\rangle.$$

Observe that whenever the event $T' \le \delta$ holds, then $(B_\delta - B_{T'})$ has a origin-symmetric distribution conditioned on $\mathcal{F}_{T'}$. We infer that

$$\mathbb{P}(f(B_\delta) \ge \alpha\sqrt{\log N}/2) \ge \mathbb{P}(T' \le \delta)/2.$$

Combined with an application of (1), it follows that

$$\mathbb{P}(E_2^c) = \mathbb{P}(T' \le \delta) \le 4N^{-\alpha^2},$$

where we choose $\delta = 1/100$. Choosing $\varepsilon = 10^{-4}\alpha^2$, it follows from an application of Lemma 4.2 that

$$\mathbb{P}(E_1^c) \le C' N^{-\alpha^2/32} \tag{33}$$

for an absolute constant $C' > 0$.

In order to complete the proof, note that $S_t - \mathbb{E}S_1$ is a mean-zero continuous-time martingale, so according to the Dambis/Dubins–Schwartz theorem, there exists standard a Brownian motion $\{W_t\}_{t\ge 0}$ such that

$$W_{[S]_t} = S_t \qquad \forall 0 \le t \le 1.$$

An elementary fact about the one-dimensional Brownian motion is that

$$\mathbb{P}\Big(\sup_{0\le t\le \tau} |W_t| \ge s\Big) \le 4\mathrm{e}^{-s^2/(2\tau)} \qquad \forall \tau, S \ge 0. \tag{34}$$

As a consequence of equation (27) we know that $V_t \le 1$ for all $0 \le t \le 1$. Therefore, on E_1 we have $[S]_1 \le 1 - \varepsilon\delta \le 1 - 10^{-6}\alpha^2$, and combined with the last inequality,

$$\begin{aligned}
&\mathbb{P}\big(\{|S_1 - ES_1| \ge \beta\sqrt{\log N}\} \cap E_1\big) \\
&\quad \le \mathbb{P}\Big(\sup_{0\le t\le 1-10^{-6}\alpha^2} |W_t| \ge \beta\sqrt{\log N}\Big) \le 4N^{-\beta^2/2(1-10^{-6}\alpha^2)}.
\end{aligned}$$

Combining with (33) and using a union bound finally gives

$$\mathbb{P}\big(|S_1 - ES_1| \ge \beta\sqrt{\log N}\big) \le 4N^{-\beta^2/2(1-10^{-6}\alpha^2)} + 2CN^{-\alpha^2/32}.$$

This completes the proof of Theorem 1.6. □

Acknowledgements. We thank Antonio Auffinger, Wei-Kuo Chen, Galyna Livshyts, Michel Ledoux, Elchanan Mossel and Ofer Zeitouni for helpful discussions. We also thank Ivan Nourdin for pointing out an error in an earlier version. This work was initiated when Jian Ding and Alex Zhai were visiting theory group of Microsoft Research at Redmond. We thank MSR for the hospitality.

REFERENCES

[1] BORELL, C. (1975). The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* **30** 207–216. MR0399402

[2] CASTELLANA, M. and ZARINELLI, E. (2011). Role of tracy-widom distribution in finite-size fluctuations of the critical temperature of the Sherrington–Kirkpatrick spin glass. *Phys. Rev. B* **84** 144417.

[3] CHATTERJEE, S. (2008). Chaos, concentration, and multiple valleys. Preprint. Available at arXiv:0810.4221.

[4] CHATTERJEE, S. (2009). Disorder chaos and multiple valleys in spin glasses. Preprint. Available at arXiv:0907.3381v4.

[5] CHATTERJEE, S., DEMBO, A. and DING, J. (2013). On level sets of Gaussian fields. Preprint. Available at arXiv:1310.5175.

[6] CIANCHI, A., FUSCO, N., MAGGI, F. and PRATELLI, A. (2011). On the isoperimetric deficit in Gauss space. *Amer. J. Math.* **133** 131–186. MR2752937

[7] DUDLEY, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.* **1** 290–330. MR0220340

[8] ELDAN, R. (2013). A two-sided estimate for the Gaussian noise stability deficit. Preprint. Available at arXiv:math/0510424.

[9] FERNIQUE, X. (1971). Régularité de processus gaussiens. *Invent. Math.* **12** 304–320. MR0286166

[10] KARDAR, M., PARISI, G. and ZHANG, Y.-C. (1986). Dynamic scaling of growing interfaces. *Phys. Rev. Lett.* **56** 889–892.

[11] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347

[12] MADAULE, T. (2013). Maximum of a log-correlated Gaussian field. Preprint. Available at arXiv:1307.1365.

[13] MOSSEL, E. and NEEMAN, J. (2015). Robust dimension free isoperimetry in Gaussian space. *Ann. Probab.* **43** 971–991. MR3342656

[14] MOSSEL, E. and NEEMAN, J. (2015). Robust optimality of Gaussian noise stability. *J. Eur. Math. Soc.* (*JEMS*) **17** 433–482. MR3317748

[15] PALASSINI, M. (2008). Ground-state energy fluctuations in the Sherrington–Kirkpatrick model. *J. Stat. Mech. Theory Exp.* **2008** P10005.

[16] PARISI, G. (1980). A sequence of approximated solutions to the S–K model for spin glasses. *J. Phys. A: Math. Gen.* **13** 115–121.

[17] SHERRINGTON, D. and KIRKPATRICK, S. (1975). Solvable model of a spin-glass. *Phys. Rev. Lett.* **35** 1792–1796.

[18] SUDAKOV, V. N. and CIREL'SON, B. S. (1974). Extremal properties of half-spaces for spherically invariant measures. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **41** 14–24, 165. Problems in the theory of probability distributions, II. MR0365680

[19] TALAGRAND, M. (1987). Regularity of Gaussian processes. *Acta Math.* **159** 99–149. MR0906527

[20] TALAGRAND, M. (2006). The Parisi formula. *Ann. of Math.* (2) **163** 221–263. MR2195134

J. DING
DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 S. UNIVERSITY AVENUE
ECKHART 104B
CHICAGO, ILLINOIS 60637
USA

R. ELDAN
WEIZMANN INSTITUTE OF SCIENCE
DEPARTMENT OF MATHEMATICS
POB 26
REHOVOT 76100
ISRAEL
E-MAIL: roneneldan@gmail.com

A. ZHAI
STANFORD UNIVERSITY
1075 SPACE PARK WAY SPC 350
MOUNTAIN VIEW, CALIFORNIA 94043
USA