Local optima of the Sherrington-Kirkpatrick Hamiltonian

Louigi Addario-Berry*
Department of Mathematics and Statistics, McGill University, Montreal, Canada.

Luc Devroye†
School of Computer Science, McGill University, Montreal, Canada.

Gábor Lugosi‡
ICREA, Department of Economics, Pompeu Fabra University, Barcelona GSE, Barcelona, Spain

Roberto I. Oliveira§
IMPA, Rio de Janeiro, RJ, Brazil

We sudy local optima of the Hamiltonian of the Sherrington-Kirkpatrick model. We compute the exponent of the expected number of local optima and determine the "typical" value of the Hamiltonian.

* louigi.addario@mcgill.ca
† lucdevroye@gmail.com
‡ gabor.lugosi@upf.edu
§ rimfo@impa.br

I. LOCAL OPTIMA OF THE HAMILTONIAN

Let $W = (W_{i,j})_{n\times n}$ be a symmetric matrix with zero diagonal such that the $(W_{i,j})_{1\le i<j\le n}$ are independent standard normal random variables. The *Sherrington-Kirpatrick* model of spin glasses is defined by a *random Hamiltonian*, that is, a random function $H : \{-1,+1\}^n \to \mathbb{R}$. For a configuration $\sigma = (\sigma_i)_{i=1}^n \in \{-1,+1\}^n$, $H(\sigma)$ is defined as follows.

$$H(\sigma) := \sum_{1\le i<j\le n} \sigma_i\, \sigma_j\, W_{ij} \ .$$

We follow the usual convention of calling $\sigma \in \{-1,+1\}^n$ a *spin configuration*, the coordinates of σ *spins*, and the value $H(\sigma)$ the *energy* of configuration σ.

Given $i \in [n]$ (where $[n] = \{1,\ldots,n\}$) and σ as above, we let $\sigma^{(i)}$ denote a new configuration obtained from σ by flipping the i-th spin and leaving other coordinates unchanged. That is, the components of $\sigma^{(i)}$ are defined as

$$\sigma_j^{(i)} := \begin{cases} -\sigma_i, & j = i \ ; \\ \sigma_j, & j \in [n]\backslash\{i\} \ . \end{cases}$$

We say that σ is a *local minimum* or a *local optimum* of H if

$$H(\sigma^{(i)}) \ge H(\sigma) \quad \text{for all } i \in [n] \ .$$

That is, σ is a local minimum if flipping the sign of any individual spin does not decrease the value of the energy.

The global optimum $\min_{\sigma\in\{-1,+1\}^n} H(\sigma)$—called the "ground-state energy"—has been extensively studied. The problem was introduced by Sherrington and Kirkpatrick [11] as a mean-field model for spin glasses. The value of the optimum was determined non-rigorously in the seminal work of Parisi [10], as a consequence of the so-called "Parisi formula". Parisi's formula was proved by Talagrand [12] in a breakthrough paper, see also Panchenko [9] for an overview. It follows from Talagrand's result that

$$n^{-3/2} \min_{\sigma\in\{-1,+1\}^n} H(\sigma) \to -c \quad \text{in probability,}$$

where c is a constant whose value is numerically estimated to be about 0.7632... (Crisanti and Rizzo [4]) and known to be bounded by $\sqrt{2/\pi} \approx 0.797885...$ (Guerra [8]).

In this paper we are interested in locally optimal solutions. An important reason of why local optima are worth considering is because local optima may be computed quickly by simple greedy algorithms, see Etscheid and Röglin [6], Angel, Bubeck, Peres, and Wei [2] and Section I B below. We show that the expected number of local optima grows exponentially and we establish the rate of growth. Also, we examine the conditional distribution of $H(\sigma)n^{-3/2}$ given that σ is locally optimal. We prove that the distribution is concentrated on an interval of width $O(n^{-1/4})$ and determine its location.

A. Results

In order to state the main result of the paper, we need a few definitions.

Let $\Phi(\lambda) = \mathbb{P}\{N \le \lambda\}$ be the distribution function of a standard normal random variable N and introduce $\phi(\lambda) = \log(2\Phi(\lambda))$. For $x \ge \sqrt{2/\pi}$, we let $\mu_*(x)$ denote the Fenchel-Légendre transform

$$\mu_*(x) := \sup_{\lambda\ge 0} \left(\lambda\, x - \frac{\lambda^2}{2} - \phi(\lambda) \right) \ .$$

Lemma 2 below shows that $\mu_* : [\sqrt{2/\pi}, +\infty) \to \mathbb{R}$ is well defined. Lemma 4 shows that the mapping

$$R(x) := \frac{x^2}{4} - \mu_*(x)$$

is strictly concave for $x \ge \sqrt{\frac{2}{\pi}}$ and achieves its global maximum at $x = v^* > \sqrt{\frac{2}{\pi}}$. We let $\alpha^* = R(v^*) > 0$ denote the maximum value of R.

Theorem 1. *For any fixed $n \geq 1$, the probability $\mathbb{P}\{\sigma$ is locally optimal$\}$ is the same for all $\sigma = \sigma(n) \in \{-1,+1\}^n$. Moreover,*

$$\lim_{n\to+\infty} \frac{1}{n} \log \mathbb{P}\{\sigma \text{ is locally optimal}\} = \alpha^* - \log 2 \ .$$

Also, there exist constants $\epsilon_0 > 0$, $L > 0$, and n_0 such that, for $0 < \epsilon < \epsilon_0$ and $n \geq n_0$,

$$\mathbb{P}\left\{ -\frac{v^*}{2} - \epsilon \leq n^{-3/2} H(\sigma) \leq -\frac{v^*}{2} + \epsilon \,\middle|\, \sigma \text{ is locally optimal} \right\} \geq 1 - \exp\left(L\sqrt{n} - \epsilon^2 n\right) .$$

The values of the constants are numerically evaluated to be $\alpha^* \approx 0.199$ and $v^*/2 \approx 0.506$. Tanaka and Bray [13] have found these values analytically using a different derivation involving a complex-valued integral representation for the average number of local minima. Since the global minimum of $H(\sigma)$ is about $-0.763\, n^{3/2}$, the typical value a local optimum $-0.506\, n^{3/2}$ comes fairly close.

Also note that Proposition 2 below implies that α^* is between $1/(2\pi) \approx 0.1591\ldots$ and $2/(3\pi) \approx 0.2122\ldots$.

B. Local minima, greedy algorithms and MaxCut

Our problem is related to finding a local optimum of weighted MaxCut on the complete graph, which was recently studied in Etscheid and Röglin [6] and Angel, Bubeck, Peres, and Wei [2]. Given $S \subset [n]$, we denote the value of the *cut* $(S, [n]\backslash S)$ as

$$\mathrm{Cut}(S, [n]\backslash S) := \sum_{i\in S} \sum_{j\in[n]\backslash S} W_{i,j}.$$

Note that there is a direct correspondence between cuts $(S, [n]\backslash S)$ and spin configurations σ_S given by

$$\sigma_S := \left(2\mathbf{1}_{\{i\in S\}} - 1\right)_{i=1}^n \ ,$$

$$\mathrm{Cut}(S, [n]\backslash S) = \frac{-H(\sigma_S) + \sum_{1\leq i<j\leq n} W_{ij}}{2} \ .$$

In particular, what [2] calls locally optimal cuts correspond exactly to our notion of local minimum and Theorem 1 may be formulated in terms of locally optimal cuts. (Note that $n^{-3/2} \sum_{1\leq i<j\leq n} W_{ij} = O_p(n^{-1/2})$ and therefore this term does not play a significant role in the typical value of a locally optimal cut.)

The papers [6] and [2] study the typical running time simple greedy algorithms take to find locally optimal cuts. Such algorithms start from a given σ and perform a sequence of local "greedy moves"– that is, single spin flips that decrease energy– until no more such moves are available. The main result of [2] is that this process ends at a local minimum after a polynomial number of moves.

Perhaps surprisingly, it is not clear that the *distribution* of the value of this local minimum is similar to the one we study in Theorem 1. In Figure 1 we present numerical evidence that two variants of local greedy search can find spin configurations with energies roughly of order $-0.7n^{3/2}$, which is better than the typical value $-0.506\, n^{3/2}$ coming out of Theorem 1. Eastham et al. [5] found the same energy value $-0.7n^{3/2}$ for random greedy ascent, where the coordinate to be flipped is chosen uniformly from all flips that decrease energy. Those authors explain why this finding contradicts the so-called Edwards hypothesis, whereby "quasiequilibrium steady states" should be describable by a thermodynamic measure over the metastable states (i.e., the local minima). It would be interesting to put these observations on a rigorous footing.

II. THE PROBABILITY OF LOCAL OPTIMALITY

In this section we take the first and crucial step to prove Theorem 1. For any fixed spin configuration $\sigma \in \{-1,+1\}^n$, we establish integral formulae for the probability that σ is locally optimal and for the conditional distribution of its energy.

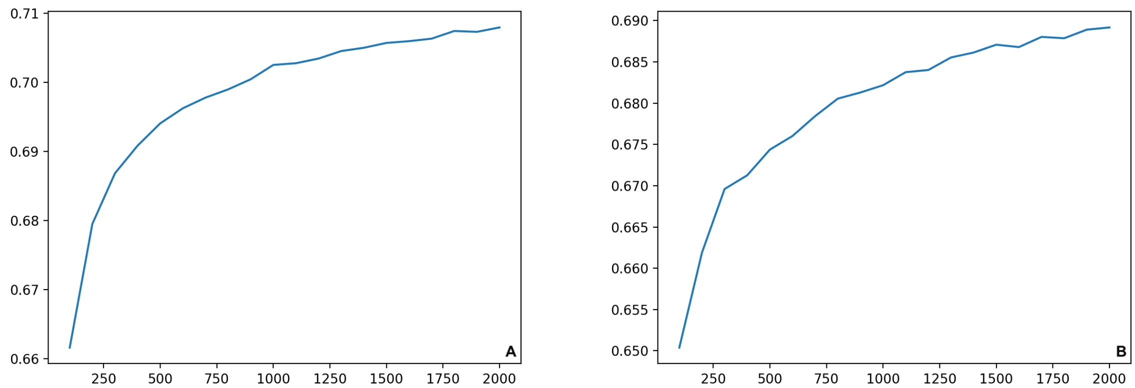

FIG. 1. Energy plots for two variants of the local greedy algorithm for finding a local optimum. The x axis in each plot is the value of the dimension n which takes values $n = 100, 200, 300, \ldots, 2000$. The y axis is the value of $H(\sigma)n^{-3/2}$ when the algorithm halts, averaged over 100 runs of the algorithm. The algorithm used in figure 1.A is the local greedy rule whereby the first coordinate that gives an improvement in energy is flipped at each step. In figure 1.B, the coordinate giving the largest improvement is used at each step.

Proposition 1 (Key integral formulae; proof in §II A)**.** *Let N denote a n-dimensional standard Gaussian random vector. Then for all $\sigma \in \{-1,+1\}^n$ and $\Delta \in \mathbb{R}$,*

$$\mathbb{P}\{\sigma \text{ is locally optimal}\} = 2^{-n}\sqrt{\frac{n-2}{2n-2}}\,\mathbb{E}\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right); \tag{1}$$

$$\mathbb{P}\left\{-\frac{H(\sigma)}{n^{3/2}} \leq \Delta \,\middle|\, \sigma \text{ is locally optimal}\right\} = \frac{\mathbb{E}\left[\mathbf{1}_{\{\|N\|_1 \geq 2\Delta n^{3/2}/\sqrt{n-2}\}}\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right)\right]}{\mathbb{E}\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right)}\ . \tag{2}$$

In Section III, we take a closer look at the integral expression of the probability of local optimality. In fact, we prove that $(1/n)\log\mathbb{P}\{\sigma \text{ is locally optimal}\}$ converges to $\alpha^* - \log 2$ defined in the introduction. Before we do that, we take a brief detour in §II B to give some explicit upper and lower bounds on the probability of local optimality.

Proposition 2 (Proven in §II B)**.** *For all spin configurations $\sigma \in \{-1,1\}^n$,*

$$\frac{1}{2\pi} - \log 2 - O\left(\frac{1}{n}\right) \leq \frac{1}{n}\log\mathbb{P}\{\sigma \text{ is locally optimal}\} \leq \frac{2}{3\pi} - \log 2 + O\left(\frac{1}{n}\right)\ .$$

A. Proof of the integral formulae

We prove here Proposition 1.

Proof of Proposition 1: For $i \in [n]$, define

$$Z_i(\sigma) := \frac{H(\sigma^{(i)}) - H(\sigma)}{2} = -\sum_{j\in[n]\setminus i}\sigma_i\,\sigma_j\,W_{i,j}\ .$$

Note that

$$\sigma \text{ is a local minimum if and only if } Z_i(\sigma) \geq 0 \text{ for all } i \in [n]\ . \tag{3}$$

Moreover,

$$-H(\sigma) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\in[n]\setminus i} -\sigma_i\,\sigma_j\,W_{i,j} = \frac{\sum_{i=1}^n Z_i(\sigma)}{2}\ . \tag{4}$$

Since σ is fixed, we write Z_i instead of $Z_i(\sigma)$ most of the time. A key point in our calculations is that the random vector

$$Z = (Z_1, Z_2, \dots, Z_n)^T$$

is a multivariate normal vector with zero mean and covariance matrix $C = (C_{i,j})_{n\times n}$ such that $C_{i,i} = n-1$ for all $i \in [n]$ and $C_{i,j} = 1$ for all $i \neq j$. In other words,

$$C = (n-2)\mathrm{Id}_n + \mathbf{1}_n \mathbf{1}_n^T \ ,$$

where Id_n is the $n \times n$ identity matrix and $\mathbf{1}_n = (1, 1, \dots, 1)^T$ is the column vector with 1 in each component.

Clearly, the eigenvalues of C are $2n-2$ with multiplicity 1 and $n-2$ with multiplicity $n-1$, and therefore $\det(C) = (2n-2)(n-2)^{n-1}$.

One may use the Sherman-Morrison formula to invert C and obtain

$$C^{-1} = \frac{1}{n-2}\left(\mathrm{Id}_n - \frac{1}{2n-2}\mathbf{1}_n\mathbf{1}_n^T\right) \ .$$

Hence,

$$\begin{aligned}
&\mathbb{P}\{\sigma \text{ is locally optimal}\} \\
&= \frac{1}{(2\pi)^{n/2}\det(C)^{1/2}} \int_{[0,\infty)^n} \exp\left(\frac{-x^T C^{-1} x}{2}\right) dx \\
&= \frac{1}{(2\pi)^{n/2}(2n-2)^{1/2}(n-2)^{(n-1)/2}} \int_{[0,\infty)^n} \exp\left(\frac{-\|x\|_2^2}{2(n-2)} + \frac{\|x\|_1^2}{2(n-2)(2n-2)}\right) dx \\
&= 2^{-n}\frac{1}{(2\pi)^{n/2}(2n-2)^{1/2}(n-2)^{(n-1)/2}} \int_{\mathbb{R}^n} \exp\left(\frac{-\|x\|_2^2}{2(n-2)} + \frac{\|x\|_1^2}{2(n-2)(2n-2)}\right) dx \ .
\end{aligned}$$

The change of variables $u = x/\sqrt{n-2}$ allows us to rewrite this integral as:

$$\mathbb{P}\{\sigma \text{ is locally optimal}\} = 2^{-n}\sqrt{\frac{n-2}{2n-2}}\,\mathbb{E}\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right)$$

where N is a vector of n independent standard normal random variables. This gives the first part of the proposition. To prove the second part, we follow the above calculations with slight modifications.

$$\begin{aligned}
&\mathbb{P}\left\{\sigma \text{ is locally optimal}, n^{-3/2}H(\sigma) \leq -\Delta\right\} \\
&= \mathbb{P}\left\{(\cap_{i=1}^n\{Z_i \geq 0\}) \bigcap \left\{\sum_{i=1}^n Z_i \geq 2\Delta n^{3/2}\right\}\right\} \\
&= \frac{1}{(2\pi)^{n/2}\det(C)^{1/2}} \int_{[0,\infty)^n \cap \{x: \sum_i x_i \geq 2\Delta n^{3/2}\}} \exp\left(\frac{-x^T C^{-1} x}{2}\right) dx \\
&= \frac{1}{(2\pi)^{n/2}(2n-2)^{1/2}(n-2)^{(n-1)/2}} \int_{[0,\infty)^n \cap \{x: \sum_i x_i \geq 2\Delta n^{3/2}\}} \exp\left(\frac{-\|x\|_2^2}{2(n-2)} + \frac{\|x\|_1^2}{2(n-2)(2n-2)}\right) dx \\
&= 2^{-n}\frac{1}{(2\pi)^{n/2}(2n-2)^{1/2}(n-2)^{(n-1)/2}} \int_{\{x: \|x\|_1 \geq 2\Delta n^{3/2}\}} \exp\left(\frac{-\|x\|_2^2}{2(n-2)} + \frac{\|x\|_1^2}{2(n-2)(2n-2)}\right) dx \ .
\end{aligned}$$

Thus, by the same change of variables $u = x/\sqrt{n-2}$ as before, we get:

$$\begin{aligned}
&\mathbb{P}\left\{\sigma \text{ is locally optimal}, n^{-3/2}H(\sigma) \leq -\Delta\right\} \\
&= 2^{-n}\sqrt{\frac{n-2}{2n-2}}\mathbb{E}\left[\mathbf{1}_{\{\|N\|_1 \geq 2\Delta n^{3/2}/\sqrt{n-2}\}} \exp\left(\frac{\|N\|_1^2}{4(n-1)}\right)\right] \ ,
\end{aligned}$$

where N is as above. $\square$

B. Explicit estimate on the probability of local optimality

In what follows, we derive Proposition 2 on upper and lower bounds for $\mathbb{P}\{\sigma \text{ is locally optimal}\}$. We will use the following Lemma.

Lemma 1. *If N is a vector of independent standard normal random variables, then for all $\lambda > 0$,*

$$\lambda \mathbb{E}\|N\|_1^2 \leq \log \mathbb{E} \exp\left(\lambda \|N\|_1^2\right) \leq \lambda \mathbb{E}\|N\|_1^2 \left(1 + \frac{n\lambda}{(1-n\lambda)}\right) .$$

Proof of Lemma 1: The inequality on the left-hand side is obvious from Jensen's inequality. To prove the right-hand side, we use the Gaussian logarithmic Sobolev inequality. In particular, writing $f(x) = \|x\|_1^2$ and $F(\lambda) = \mathbb{E}\exp(\lambda f(N))$, the inequality on page 126 of Boucheron, Lugosi, and Massart [3] asserts that

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq \frac{\lambda^2}{2}\mathbb{E}\left[e^{\lambda f(N)}\|\nabla f(N)\|^2\right] .$$

Since $\|\nabla f(N)\|^2 = 4n\|N\|_1^2$, we obtain the differential inequality

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq 2n\lambda^2 F'(\lambda) .$$

This inequality has the same form as the one at the top of page 191 of [3] with $a = 2n$ and $b = 0$ and Theorem 6.19 implies the result above. □

Proof of Proposition 2: Since

$$\mathbb{E}\|N\|_1^2 = n + n(n-1)\frac{2}{\pi} ,$$

we get

$$\mathbb{P}\{\sigma \text{ is locally optimal}\} \geq 2^{-n}\sqrt{\frac{n-2}{2n-2}}\exp\left(n/(4(n-1)) + \frac{n}{2\pi}\right)$$

and

$$\mathbb{P}\{\sigma \text{ is locally optimal}\} \leq 2^{-n}\sqrt{\frac{n-2}{2n-2}}\exp\left(\left(n/(4(n-1)) + \frac{n}{2\pi}\right)\frac{4n-1}{3n-1}\right) .$$

The Proposition follows from takings logarithms and dividing by n on both sides. □

III. APPROXIMATING THE INTEGRAL

The key fact that emerges from Proposition 1 is that, in order to prove our main result, we need to analyze exponential moments of $\|N\|_1 = \sum_{i=1}^n |N_i|$, a sum of absolute values of i.i.d. standard Gaussians. The next lemma gives a quantitative version of the Large Deviations Principle for such a sum.

Lemma 2 (Quantitative LDP for $\|N\|_1$; proof in §IV B)**.** *For $x \geq \sqrt{2/\pi}$, define $\mu_*(x)$ as in the introduction. Let $N = (N_1, \ldots, N_n)$ be a vector of independent standard normal coordinates. Then*

$$\mathbb{P}\left\{\|N\|_1 \geq nx\right\} = e^{-(\mu_*(x)+r_n(x))\,n} \text{ with } 0 \leq r_n(x) \leq \kappa\left(\frac{x-\sqrt{2/\pi}}{\sqrt{n}} + \frac{1}{n}\right)$$

for some $\kappa > 0$ independent of x and n. Moreover, μ_* is smooth, $\mu_*(\sqrt{2/\pi}) = \mu_*'(\sqrt{2/\pi}) = 0$, and $1 \leq \mu_*''(x) \leq 20$ for all $x \geq \sqrt{2/\pi}$.

Given this Lemma, we will base our proof strategy on a Laplace-type approximation to the expectations in Proposition 1. Modulo some technical assumptions, Varadhan's Lemma (see e.g. [7, page 32]) leads one to expect that, as $n \to \infty$,

$$\frac{1}{n}\log \mathbb{E}\left[\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right)\right] \approx \frac{1}{n}\log \mathbb{E}\left[\exp\left(n\,\frac{(\|N\|_1/n^2)}{4}\right)\right] \to \sup_v\left(\frac{v^2}{4} - \mu_*(v)\right) .$$

In fact, the intuition behind the Varadhan's lemma is that most of the "mass" of the expectation concentrates around $\|N\|_1 \sim v_*\,n$, where v_* achieves the above supremum. This is precisely what we show in the proof of Theorem 1 in §III A. The proofs of some additional Lemmas we use are presented in §III B and §III C.

A. Proof of the main result

Proof of Theorem 1: We recall the formulae from Proposition 1.

$$\mathbb{P}\{\sigma \text{ is locally optimal}\} = 2^{-n}\sqrt{\frac{n-2}{2n-2}}\,\mathbb{E}\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right); \tag{5}$$

$$\mathbb{P}\left\{-\frac{H(\sigma)}{n^{3/2}} \le \Delta \,\middle|\, \sigma \text{ is locally optimal}\right\} = \frac{\mathbb{E}\left[\mathbf{1}_{\{\|N\|_1 \ge 2\Delta n^{3/2}/\sqrt{n-2}\}}\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right)\right]}{\mathbb{E}\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right)}; \tag{6}$$

where N is a vector of n i.i.d. random standard Gaussians. Using Lemma 2, we will be able to prove the following result.

Lemma 3 (Proof in §III B)**.** *Given $c \ge 0$, $x \ge \sqrt{2/\pi}$, define $R_c(x) := cx^2/2 - \mu_*(x)$. Letting $r_n(x)$ be as in Lemma 2, we have that for $b \ge \sqrt{2/\pi}$,*

$$\mathbb{E}\left[\exp\left(\frac{c\|N\|_1^2}{2n}\right)\mathbf{1}_{\{\|N\|_1 \ge bn\}}\right] = \exp\left\{n\left(R_c(b) - r_n(b)\right)\right\} + cn\int_b^{+\infty} x\,\exp\left\{n\left(R_c(x) - r_n(x)\right)\right\}\,dx\ ,$$

whereas for $a \ge \sqrt{2/\pi}$,

$$\mathbb{E}\left[\exp\left(\frac{c\|N\|_1^2}{2n}\right)\mathbf{1}_{\{\|N\|_1 \le an\}}\right] = (I) + (II)\ ,$$

where

$$1 \le (I) \le \exp\left(n\,R_c(\sqrt{2/\pi})\right) \ \ \text{and} \ (II) = cn\int_{\sqrt{2/\pi}}^a x\,\exp(n\left(R_c(x) - r_n(x)\right))\,dx.$$

The upshot is that, in order to understand our expectations, we will need to analyze the behavior of the function $R_c(x)$. Comparison with (5) and (6) shows that the value of interest for us is $c = c_n = n/2(n-1) = 1/2 + o(1)$. We will also consider the value $c = 1/2$ obtained when $n \to +\infty$:

$$R(x) = R_{\frac{1}{2}}(x) := \frac{x^2}{4} - \mu_*(x) \quad \text{for } x \ge \sqrt{2/\pi}\ . \tag{7}$$

The next lemma collects some information about $R(x)$.

Lemma 4 (Proven in §III C)**.** *Let $x \ge \sqrt{2/\pi}$. Define R as in equation (7) and μ_* as in Lemma 2. Then there exists a unique $x = v^* > \sqrt{2/\pi}$ that maximizes $R(x)$ over $x \ge \sqrt{2/\pi}$. Letting $\alpha^* := R(v^*)$ denote the value of the maximum, we have the following strict concavity property: for any $x \ge \sqrt{2/\pi}$,*

$$-10\,(x - v^*)^2 \le R(x) - \alpha^* \le \frac{(x - v^*)^2}{4}\ .$$

From now on, we assume $n \ge 100$ for simplicity, and use the notation L to denote the value of a positive constant independent of n whose value may change from line to line. Finally, we set

$$c = c_n := \frac{n}{2\,(n-1)} = \frac{1}{2} + \frac{1}{2\,(n-1)}\ .$$

The definition of R_c in Lemma 3 may be combined with the concavity property in Lemma 4 to obtain that for all $x \ge \sqrt{\frac{2}{\pi}}$,

$$-10\,(x - v^*)^2 \le R_c(x) - \alpha^* \le -\frac{(x - v^*)^2}{4} + \frac{x^2}{4\,(n-1)}\ . \tag{8}$$

We now apply this to estimate expectations to the left of v^*. That is, we consider, for $a \in \left[\sqrt{\frac{2}{\pi}}, v^*\right]$,

$$\mathbb{E}\left[\exp\left(\frac{c\|N\|_1^2}{2n}\right)\mathbf{1}_{\{\|N\|_1 \le an\}}\right] .$$

In this range $|a - v^*|$ is uniformly bounded, so $x^2 \le L$ and

$$0 \le r_n(x) \le \frac{L}{\sqrt{n}} \quad \text{for all } \sqrt{\frac{2}{\pi}} \le x \le v^* .$$

Combining Lemma 3 with $c \le 1$ and (8), we obtain

$$\begin{aligned}
\frac{\mathbb{E}\left[\exp\left(\frac{c\|N\|_1^2}{2n}\right)\mathbf{1}_{\{\|N\|_1 \le an\}}\right]}{\exp(n\,\alpha^*)} &\le \exp(n\,(R_c(\sqrt{2/\pi}) - \alpha^*)) \\
&\quad + cn \int_{\sqrt{2/\pi}}^{a} x\,\exp\left(n\,(R_c(x) - \alpha^*)\right)\,dx \qquad (9) \\
&\le \exp\left(L - \frac{(v^* - \sqrt{2/\pi})^2}{4}\,n\right) \\
&\quad + n \int_{\sqrt{2/\pi}}^{a} x\,\exp\left(L + n\,\frac{(v^* - x)^2}{4}\right)\,dx \\
&\le L\,(1 + cn)\,\exp\left(L - \frac{(a - v^*)^2\,n}{4}\right) \\
&\le \exp\left(L \log n \, - \frac{(a - v^*)^2\,n}{4}\right) \qquad (10)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\mathbb{E}\left[\exp\left(\frac{c\|N\|_1^2}{2n}\right)\mathbf{1}_{\{\|N\|_1 \le v^* n\}}\right]}{\exp(n\,\alpha^*)} &\ge \exp\left(-L\sqrt{n}\right) \int_{v^* - \frac{1}{\sqrt{n}}}^{v^*} x\,\exp\left(n\,(R_c(x) - \alpha^*)\right)\,dx \\
&\ge \frac{\exp\left(-L\sqrt{n}\, - 10\right)}{\sqrt{n}} \\
&\ge \exp(-L\sqrt{n}) . \qquad (11)
\end{aligned}$$

For bounding the expectation for $b \ge v^*$, we cannot simply use $x^2 \le L$ and $r_n(x) \le L/\sqrt{n}$. However, note that

$$-\frac{1}{4}\,(x - v^*)^2 + \frac{x^2}{4\,(n-1)} \le \begin{cases} -\frac{1}{5}\,(x - v^*)^2 + \frac{L}{\sqrt{n}} & \text{for } x \le (n-1)^{1/4}; \\ -\frac{1}{4}\,(x - v^*)^2 + \frac{2(x - v^*)^2 + 2(v^*)^2}{(n-1)} \le -\frac{1}{5}\,(x - v^*)^2 + \frac{L}{n} & \text{for larger } x. \end{cases}$$

Also, recalling the expression for r_n in Lemma 2,

$$0 \le r_n(x) \le \kappa\left(\frac{x - \sqrt{2/\pi}}{\sqrt{n}} + \frac{1}{n}\right) \le \frac{L}{\sqrt{n}} + \frac{L\,(x - v^*)}{\sqrt{n}} .$$

This allows us to obtain, for $b \le v^* + \epsilon_0$,

$$\frac{\mathbb{E}\left[\exp\left(\frac{c\|N\|_1^2}{2n}\right)\mathbf{1}_{\{\|N\|_1 \ge bn\}}\right]}{\exp(n\,\alpha^*)} \le \exp\left(L\sqrt{n} - \frac{(b - v^*)^2\,n}{4}\right) . \qquad (12)$$

This leads to our main results. Indeed, we may upper bound

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\frac{c\|N\|_1^2}{2n}\right)\right] &= \mathbb{E}\left[\exp\left(\frac{c\|N\|_1^2}{2n}\right)\mathbf{1}_{\{\|N\|_1 \le v^* n\}}\right] \\
&\quad + \mathbb{E}\left[\exp\left(\frac{c\|N\|_1^2}{2n}\right)\mathbf{1}_{\{\|N\|_1 \ge v^* n\}}\right] \\
\text{(use (10) \& (12) w/ } a = b = v^*) &\le \exp\left(n\,\alpha^* + n\,\delta_n\right) ,
\end{aligned}$$

where $|\delta_n| \le L/\sqrt{n}$ for a positive constant L. At the same time, inequality (11) gives:

$$\mathbb{E}\left[\exp\left(\frac{c\|N\|_1^2}{2n}\right)\right] \ge \exp\left(n\,\alpha^* - n\,\delta_n\right)\ ,$$

with δ_n as above. So Proposition 1 implies the first statement in the Theorem.

To obtain conditional concentration of the energy, we use Proposition 1 and obtain:

$$\begin{aligned}\mathbb{P}\left\{-H(\sigma) \le -\frac{v^*}{2} - \epsilon \,\middle|\, \sigma \text{ local optimum}\right\} &\le \frac{\mathbb{E}\left[\mathbf{1}_{\{\|N\|_1 \ge b\,n\}}\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right)\right]}{\mathbb{E}\left[\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right]\right)} \\ &\left(\text{with } b = (v^* + 2\epsilon)\sqrt{\frac{n}{n-1}}\right) \\ &= \exp\left(-L\sqrt{n} - \epsilon^2\, n\right),\end{aligned}$$

and, for ϵ small enough, so that the value of a below is at most $\sqrt{2/\pi}$,

$$\begin{aligned}\mathbb{P}\left\{-H(\sigma) \ge -\frac{v^*}{2} + \epsilon \,\middle|\, \sigma \text{ local optimum}\right\} &\le \frac{\mathbb{E}\left[\mathbf{1}_{\{\|N\|_1 \le a\,n\}}\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right)\right]}{\mathbb{E}\left[\exp\left(\frac{\|N\|_1^2}{4(n-1)}\right)\right]} \\ &\left(\text{with } a = (v^* - 2\epsilon)\sqrt{\frac{n}{n-1}}\right) \\ &= \exp\left(-L\sqrt{n} - \epsilon^2\, n\right)\ .\end{aligned}$$

□

B. Estimates on the integrals

Proof of Lemma 3: Recall the definition of R_c from (7),

$$R_c(x) := \frac{cx^2}{2} - \mu_*(x) \quad \text{for } c \ge 0,\ x \ge \sqrt{2/\pi};$$

and that our goal is to estimate:

$$\mathbb{E}\left[\mathbf{1}_{\{\|N\|_1 \le a\,n\}}\exp\left(\frac{cn\,\|N\|_1^2}{2n}\right)\right] \text{ and } \mathbb{E}\left[\mathbf{1}_{\{\|N\|_1 \ge b\,n\}}\exp\left(\frac{cn\,\|N\|_1^2}{2n}\right)\right]$$

for $a, b \ge \sqrt{2/\pi}$, where N is a n-dimensional vector of independent standard Gaussians. Lemma 2 enters our proof via the fact that:

$$\forall x \ge \sqrt{\frac{2}{\pi}}\ :\ \mathbb{P}\{\|N\|_1 \ge nx\} = \exp(-(\mu_*(x) + r_n(x))\,n) \tag{13}$$

where $r_n(x)$ is as in that Lemma.

Let $\phi_{c,n}(x) = e^{cn\,x^2/2}$. Note that

$$\mathbf{1}_{\{\|N\|_1 \le a\,n\}}\exp\left(\frac{cn\,\|N\|_1^2}{2n}\right) = \phi_{c,n}\left(\frac{\|N\|_1}{n}\right)\mathbf{1}_{\{\frac{\|N\|_1}{n} \le a\}}\ .$$

We may compute the expectation of this expression as follows.

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\{\|N\|_1 \le a\,n\}}\exp\left(\frac{cn\,\|N\|_1^2}{2n}\right)\right] &= 1 + \int_0^a \phi'_{c,n}(x)\,\mathbb{P}\left\{\frac{\|N\|_1}{n} \ge x\right\}dx \\ &= 1 + cn\int_0^a \exp\left(\frac{cn\,x^2}{2}\right)\mathbb{P}\left\{\frac{\|N\|_1}{n} \ge x\right\}dx\ .\end{aligned}$$

We split the above integral in two parts.

$$(I) = 1 + cn \int_0^{\sqrt{2/\pi}} x \exp\left(\frac{cn\,x^2}{2}\right) \mathbb{P}\left\{\frac{\|N\|_1}{n} \geq x\right\} dx$$
$$(II) = cn \int_{\sqrt{2/\pi}}^{a} x \exp\left(\frac{cn\,x^2}{2}\right) \mathbb{P}\left\{\frac{\|N\|_1}{n} \geq x\right\} dx \ .$$

For term (I), we bound the probability in the integral by 1, and obtain:

$$1 \leq (I) \leq 1 + cn \int_0^{\sqrt{2/\pi}} x \exp\left(\frac{cn\,x^2}{2}\right) dx = \exp\left(\frac{cn\,x^2}{2}\right)\Big|_{x=\sqrt{\frac{2}{\pi}}} = \exp\left(n\,R_c(\sqrt{2/\pi})\right)$$

because $\mu_*(\sqrt{2/\pi}) = 0$. Term (II) may be evaluated using (13),

$$(II) = cn \int_{\sqrt{2/\pi}}^{a} x \exp\left(\frac{cn\,x^2}{2} - n\,\mu_*(x) - n\,r_n(x)\right) dx \ ,$$

which has the desired form because

$$\frac{cn\,x^2}{2} - n\,\mu_*(x) = n\,R_c(x) \ .$$

Similarly,

$$\mathbf{1}_{\{\|N\|_1 \geq b\,n\}} \exp\left(\frac{cn\,\|N\|_1^2}{2n}\right) = \phi_{c,n}\left(\frac{\|N\|_1}{n}\right) \mathbf{1}_{\{\frac{\|N\|_1}{n} \geq b\}} \ ,$$

and we finish the proof via the identity

$$\mathbb{E}\left[\mathbf{1}_{\{\|N\|_1 \geq b\,n\}} \exp\left(\frac{cn\,\|N\|_1^2}{2n}\right)\right] = \phi_{c,n}(b)\,\mathbb{P}\left\{\frac{\|N\|_1}{n} \geq b\right\} + \int_b^{+\infty} \phi'_{c,n}(x)\,\mathbb{P}\left\{\frac{\|N\|_1}{n} \geq x\right\} dx$$

and using the bound in (13). □

C. Estimates on the optimization problem

Proof of Lemma 4: Recall that our goal is to understand the function

$$R(x) = \frac{x^2}{4} - \mu_*(x)$$

with μ_* as in Lemma 2. We will strongly use some properties of μ_* obtained in that Lemma:

$$\mu_* \text{ is smooth, } \mu_*'(\sqrt{2/\pi}) = 0 \text{ and } 1 \leq \mu_*''(x) \leq 20 \text{ for } x \geq \sqrt{2/\pi}. \tag{14}$$

This immediately implies:

$$-20 \leq R''(x) = \frac{1}{2} - (\mu_*)''(x) \ \leq -\frac{1}{2}.$$

Note also that derivative of R at $x = \sqrt{2/\pi}$ satisfies

$$R'(\sqrt{2/\pi}) = \frac{1}{2}\sqrt{2/\pi} - \mu_*'(\sqrt{2/\pi}) = \frac{1}{\sqrt{2\pi}} > 0 \ .$$

So R is increasing in an interval to the right of $\sqrt{2/\pi}$. At the same time, it is a strictly concave function whose derivative decreases to $-\infty$ as $x \to +\infty$. It follows that this function achieves its maximum at some $v^* > \sqrt{2/\pi}$, with $R'(v^*) = 0$.

Now consider

$$\alpha^* := R(v^*) = \max_{x \geq \sqrt{2/\pi}} R(x) \ .$$

Since v^* is a critical point, a Taylor expansion shows that, if $x \geq \sqrt{2/\pi}$,

$$R(x) = \alpha^* + \frac{R''(v^* + \theta\,(x - v^*))}{2}\,(x - v^*)^2$$

for some $0 \leq \theta \leq 1$, and the theorem follows because $R'' \in [-20, -1/2]$. $\square$

IV. AUXILIARY RESULTS ON LARGE DEVIATIONS

This section is to prove a series of results on the ℓ^1 norm of a n-dimensional vector of i.i.d. standard Gaussians. These results together imply Lemma 2 – a "finite-sample LDP" for $\|N\|_1 -$, which is proven in §IV B. A estimate on derivatives is left to §IV C.

A. Preliminary estimates

We first find an expression for the Laplace transform of the absolute value of a standard normal random variable:

Lemma 5. *Let N be a standard normal random variable. For all $\lambda > 0$,*

$$\mathbb{E}e^{\lambda|N|} = e^{\lambda^2/2+\phi(\lambda)} \ ,$$

where $\phi(\lambda) = \log(2\Phi(\lambda))$, with $\Phi(\lambda) = \mathbb{P}\{N \leq \lambda\}$.

Proof :

$$\begin{aligned}
\mathbb{E}e^{\lambda|N|} &= \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{\lambda x - x^2/2} dx \\
&= 2e^{\lambda^2/2} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{(x-\lambda)^2/2} dx \\
&= 2e^{\lambda^2/2} \mathbb{P}\{N > -\lambda\} \ .
\end{aligned}$$

$\square$

We need to compute the large deviations rate function for $\sum_{i=1}^n |N_i|$, with N_i independent standard normal random variables. As usual, this is given by the Fenchel-Légendre transform of $\log \mathbb{E}e^{\lambda|N|}$:

$$\mu_*(x) := \sup_{\lambda \geq 0} \left(\lambda x - \log \mathbb{E}e^{\lambda|N|}\right) \ .$$

The next lemma collects technical facts on μ_* and the value $\lambda = \lambda_*$ that achieves the minimum.

Lemma 6. *For each $x \geq \sqrt{2/\pi}$, there exists a unique $\lambda = \lambda_*(x) \geq 0$ such that*

$$\lambda + \phi'(\lambda) = x \ .$$

Defining

$$\mu_*(x) := \lambda_*(x)\, x - \frac{\lambda_*(x)^2}{2} - \phi(\lambda_*(x)) \ ,$$

for $x \geq \sqrt{\frac{2}{\pi}}$, we have that, for each x in this range, $\mu_*(x)$ is the global maximum of

$$\lambda\, x - \frac{\lambda^2}{2} - \phi(\lambda)$$

over $\lambda \in [0, \infty)$ which is uniquely achieved at $\lambda = \lambda_*(x)$. We also have the following properties.

1. Derivative: λ_* *is the derivative of* μ_* *and* $\mu_*(\sqrt{2/\pi}) = \lambda_*(\sqrt{2/\pi}) = 0$.

2. Strict concavity. *For each* $\lambda \geq 0$, $x \geq \sqrt{2/\pi}$,

$$\frac{(\lambda - \lambda_*(x))^2}{40} \leq \mu_*(x) - (\lambda\, x - \log \mathbb{E}\, e^{\lambda|N|}) \leq \frac{(\lambda - \lambda_*(x))^2}{2} \ . \tag{15}$$

3. Derivative bounds for λ_*:

$$1 \leq (\mu_*)''(x) = \lambda_*'(x) \leq 20 \ . \tag{16}$$

Proof : By the previous Lemma,

$$\lambda + \phi'(\lambda) = \frac{d}{d\lambda} \log \mathbb{E} e^{\lambda|N|} \ ,$$

which is a smooth function because $|N|$ has a Gaussian-type tail. Using this "lightness of the tail", one can differentiate under the expectation and obtain

$$\phi'(0) = \frac{d}{d\lambda} \log \mathbb{E}\, e^{\lambda|N|}\Big|_{\lambda=0} = \mathbb{E}\,|N| = \sqrt{\frac{2}{\pi}} \ .$$

Lemma 8 in §IV C implies that

$$-0.95 \leq \phi''(\lambda) \leq 0 \ .$$

Therefore, for all $\lambda \geq 0$,

$$\frac{d}{d\lambda} (\lambda + \phi'(\lambda)) \in [0.05, 1] \ . \tag{17}$$

In particular, $\lambda + \phi'(\lambda)$ is an increasing function that is equal to $\sqrt{2/\pi}$ at $\lambda = 0$ and diverges when $\lambda \nearrow +\infty$. It follows that for all $x \geq \sqrt{2/\pi}$ there exists a unique $\lambda = \lambda_*(x)$ with $\lambda_*(x) + \phi'(\lambda_*(x)) = x$, and $\lambda_*(\sqrt{2/\pi}) = 0$. The implicit function theorem guarantees that λ_* is smooth over $[\sqrt{2/\pi}, +\infty)$ and

$$\lambda_*'(x) = \frac{1}{\frac{d}{d\lambda} (\lambda + \phi'(\lambda)) \,|_{\lambda=\lambda_*(x)}} \in [1, 20] \ . \tag{18}$$

Equation (17) above shows that

$$\lambda\, x - \frac{\lambda^2}{2} - \phi(\lambda) = \lambda\, x - \log \mathbb{E}\, e^{\lambda|N|}$$

is a strictly concave function of λ with second derivative

$$-1 \leq -\frac{d^2}{(d\lambda)^2} \left(\frac{\lambda^2}{2} + \phi(\lambda)\right) \leq -\frac{1}{20} \ .$$

Thus $\lambda_*(x)$, which is a critical point for this function, is the unique global maximum of $\lambda\, x - \log \mathbb{E}\, e^{\lambda|N|}$. The value of the function at that point is precisely $\mu_*(x)$. Note also that:

$$(\mu_*)'(x) = \frac{d}{dx} \left(\lambda_*(x)\, x - \frac{\lambda_*(x)^2}{2} - \phi(\lambda_*(x))\right) = \lambda_*(x)$$

because $x = \lambda_*(x) + \phi'(\lambda_*(x))$.

Let us now prove the estimates in the lemma. The strict concavity property in (15) follows from expanding

$$\lambda\, x - \log \mathbb{E}\, e^{\lambda|N|}$$

around the critical point $\lambda = \lambda_*(x)$ and applying a second-order Taylor expansion:

$$\begin{aligned}\lambda\, x - \log \mathbb{E}\, e^{\lambda|N|} - \mu_*(x) = {} & \frac{d}{d\lambda} (\lambda\, x - \log \mathbb{E}\, e^{\lambda|N|})\Big|_{\lambda=\lambda_*(x)} (\lambda - \lambda_*(x)) \\ & + \frac{1}{2} \frac{d^2}{(d\lambda)^2} \left(\lambda\, x - \log \mathbb{E}\, e^{\lambda|N|}\right)\Big|_{\lambda=\tilde{\lambda}} (\lambda - \lambda_*(x))^2 \\ & \text{with } \tilde{\lambda} = (1-\theta)\lambda_*(x) + \theta\lambda, \text{ for some } \alpha \in [0,1] \ ,\end{aligned}$$

noting that the first derivative is 0 and the second one is between -1 and $-1/20$. Finally, the derivative bound in item 2 is proven in (18). □

B. The quantitative LDP

We now have the tools to prove Lemma 2.

Proof of Lemma 2: The facts about values and derivatives about μ_* are contained in Lemma 6. Our goal, then, is to show that that for any $n \geq 1$,

$$\frac{1}{n}\log \mathbb{P}\left\{\|N\|_1 \geq nx\right\} = -\mu_*(x) - r_n(x) \ ,$$

where μ_* is as in Lemma 6 and

$$0 \leq r_n(x) \leq \kappa\left(\frac{x-\sqrt{2/\pi}}{\sqrt{n}} + \frac{1}{n}\right)$$

for some universal $\kappa > 0$ that is independent of $x \geq \sqrt{2/\pi}$ and $n \geq 1$.

For any $\lambda > 0$, the usual Cramér-Chernoff trick may be combined with Lemma 6 to obtain

$$\frac{1}{n}\log \mathbb{P}\left\{\|N\|_1 \geq nx\right\} \leq \inf_{\lambda\geq 0}\left(\log \mathbb{E}\, e^{\lambda|N|} - \lambda\, x\right) = -\mu_*(x) \ .$$

To give a non-asymptotic lower bound for this probability, we use the following lemma that appears in the fourth edition of the book of Alon and Spencer [1, Theorem A.2.1].

Lemma 7. *Let $u, \lambda, \epsilon > 0$ such that $\lambda > \epsilon$. Let X be a random variable such that the moment generating function $\mathbb{E}e^{cX}$ exists for $c \leq \lambda + \epsilon$. For any $a \in \mathbb{R}$, define $g_a(c) = e^{-ac}\mathbb{E}e^{cX}$. Then*

$$\mathbb{P}\{X \geq a - u\} \geq e^{-\lambda u}\left(g_a(\lambda) - e^{-\epsilon u}\left(g_a(\lambda+\epsilon) + g_a(\lambda-\epsilon)\right)\right) \ .$$

We apply Lemma 7 to the random variable $X = \|N\|_1$ with $\lambda = \lambda_*(a/n)$ and a, u, ε to be chosen below. In the notation of Lemma 7, for each $\lambda \geq 0$,

$$g_a(\lambda) = \exp\left(-n\mu_a(\lambda)\right) \quad \text{where} \quad \mu_a(\lambda) = \left(\lambda\,(a/n) - \log \mathbb{E}\, e^{\lambda\,|N|}\right) \ .$$

Using Lemma 6 to bound this expression, we obtain from (15) that

$$\frac{g_a(\lambda_*(a/n)+\epsilon)}{g_a(\lambda_*(a/n))} \leq e^{n\epsilon^2/2} \quad \text{and} \quad \frac{g_a(\lambda_*(a/n)-\epsilon)}{g_a(\lambda_*(a/n))} \leq e^{n\epsilon^2/2} \ .$$

Moreover, $g_a(\lambda_*(a/n)) = e^{-n\mu_*(a/n)}$. So

$$\mathbb{P}\{\|N\|_1 \geq a-u\} \geq e^{-\lambda_*(a/n)\,u}e^{-n\mu_*(a/n)}\left(1 - 2e^{-\epsilon u + \frac{\varepsilon^2 n}{2}}\right) \ .$$

We now choose $\epsilon = \sqrt{2/n}$ and $u = \epsilon n/2 + 1/\epsilon = \sqrt{2n}$ to obtain

$$\mathbb{P}\{\|N\|_1 \geq a-u\} \geq e^{-\lambda_*(a/n)\,\sqrt{2n}}e^{-n\mu_*(a/n)}\left(1 - \frac{2}{e}\right) \ .$$

Letting $a = nx + \sqrt{2n} = n\,(x+\epsilon)$, we have that

$$\mathbb{P}\{\|N\|_1 \geq nx\} = e^{-\lambda_*(x+\epsilon)\,\sqrt{2n}}e^{-n\,\mu_*(x+\epsilon)}\left(1 - \frac{2}{e}\right) \ .$$

Recall from Lemma 6 that $\lambda_*'(y) \leq 20\,(y - \sqrt{2/\pi})$ and $y - \sqrt{2/\pi} \leq (\mu_*)'(y) \leq 20\,(y - \sqrt{2/\pi})$ for all $y \geq \sqrt{2/\pi}$. Thus,

$$\lambda_*(x+\epsilon)\,\sqrt{2n} \leq 20\,(x+\epsilon-\sqrt{2/\pi})\,\sqrt{2n}$$

and

$$\mu_*(x+\epsilon) \leq \mu_*(x) + 20\,\epsilon\,(x+\epsilon-\sqrt{2/\pi}) \ .$$

Recalling $\epsilon = \sqrt{2/n}$, we may plug theses estimates back in the lower bound for our probability and obtain the theorem.
□

C. One more technical estimate

Lemma 8. *Let* $f(\lambda) = (2\pi)^{-1/2} e^{-\lambda^2/2}$ *be the standard normal density let* $\Phi(\lambda) = \int_{-\infty}^{\lambda} f(x)dx$ *be the corresponding cumulative distribution function. Then for all* $\lambda \geq 0$*,*

$$\frac{f'(\lambda)}{\Phi(\lambda)} - \frac{f(\lambda)^2}{\Phi(\lambda)^2} > -0.95 \ .$$

Proof : Note that $f'(\lambda) = -\lambda f(\lambda)$, so we need only need to prove

$$\sup_{\lambda \geq 0} \frac{f(\lambda)}{\Phi(\lambda)} \left(\lambda + \frac{f(\lambda)}{\Phi(\lambda)} \right) < 0.95 \ .$$

We combine three inequalities, considering three ranges of the value of λ, given by $[0, \lambda_1)$, $[\lambda_1, \lambda_2]$, and (λ_2, ∞), where

$$\lambda_1 = \frac{0.95 - \frac{2}{\pi}}{\sqrt{2/\pi}} \approx 0.3927\ldots \quad \text{and} \quad \lambda_2 = \sqrt{\log \frac{2/\pi}{0.95 - \sqrt{2/(\pi e)}}} \approx 0.5584\ldots \ .$$

First, note that $f(\lambda)/\Phi(\lambda) \leq \sqrt{2/\pi}$ since f/Φ is a decreasing function. Thus,

$$\frac{f(\lambda)}{\Phi(\lambda)} \left(\lambda + \frac{f(\lambda)}{\Phi(\lambda)} \right) \leq \frac{2}{\pi} + \lambda \sqrt{\frac{2}{\pi}} < 0.95 \quad \text{for } \lambda \in [0, \lambda_1).$$

Second, $\lambda e^{-\lambda^2/2} \leq 1/\sqrt{e}$, so

$$\frac{f(\lambda)}{\Phi(\lambda)} \left(\lambda + \frac{f(\lambda)}{\Phi(\lambda)} \right) \leq 2\lambda f(\lambda) + 2f(\lambda)^2 \leq \sqrt{\frac{2}{\pi e}} + \frac{2}{\pi} e^{-\lambda^2} < 0.95 \quad \text{for } \lambda \in (\lambda_2, \infty).$$

Finally, since $\lambda e^{-\lambda^2/2}$ is increasing and $e^{-\lambda^2/2}$ is decreasing on $[\lambda_1, \lambda_2]$, on this interval we have

$$\frac{f(\lambda)}{\Phi(\lambda)} \left(\lambda + \frac{f(\lambda)}{\Phi(\lambda)} \right) \leq 2\lambda_2 f(\lambda_2) + 4f^2(\lambda_1) \approx 0.92685\ldots < 0.95 \ .$$

□

V. ACKNOWLEDGEMENTS

We thank Pascal Maillard for useful discussions and for bringing [5] to our attention. LAB was supported by NSERC Discovery Grant 341845. LD was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. GL was supported by *Ministerio de Economía y Competitividad (MINECO)* of Spain, Grant MTM2015-67304-P and FEDER, EU, "High-dimensional problems in structured probabilistic models - Ayudas Fundación BBVA a Equipos de Investigación Cientifica 2017" and by "Google Focused Award Algorithms and Learning for AI". RIO was supported by a *Bolsa de Produtividade em Pesquisa* from CNPq, Brazil; and by the FAPESP Center for Neuromathematics (grant# 2013/ 07699-0 , FAPESP - S. Paulo Research Foundation). GL and RIO received additional support from CNPq, Brazil via *Ciência sem Fronteiras* grant # 401572/2014-5.

[1] N. Alon and J.H. Spencer. *The Probabilistic Method*. Wiley, New York, 2016. 4th edition.

[2] O. Angel, S. Bubeck, Y. Peres, and F. Wei. Local max-cut in smoothed polynomial time. *arXiv preprint arXiv:1610.04807*, 2016.

[3] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

[4] A. Crisanti and T. Rizzo. Analysis of the ∞-replica symmetry breaking solution of the Sherrington-Kirkpatrick model. *Physical Review E*, 65(4):046137, 2002.

[5] P. R. Eastham, R. A. Blythe, A. J. Bray, and M. A. Moore. Mechanism for the failure of the Edwards hypothesis in the Sherrington-Kirkpatrick spin glass. *Phys. Rev. B*, 74:020406, Jul 2006.

[6] M. Etscheid and H. Röglin. Smoothed analysis of local search for the maximum-cut problem. *ACM Transactions on Algorithms (TALG)*, 13(2):25, 2017.
[7] Frank Den Hollander. *Large Deviations*. Fields Institute Monographs. American Mathematical Society, 2000.
[8] F. Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in Mathematical Physics*, 233(1):1–12, 2003.
[9] D. Panchenko. *The Sherrington-Kirkpatrick model*. Springer Science & Business Media, 2013.
[10] G. Parisi. A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.
[11] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792, 1975.
[12] M. Talagrand. The Parisi formula. *Annals of Mathematics*, 163:221–263, 2006.
[13] F Tanaka and S F Edwards. Analytic theory of the ground state properties of a spin glass. I. Ising spin glass. *Journal of Physics F: Metal Physics*, 10(12):2769, 1980.