

The TAP free energy for high-dimensional linear regression

Jiaze Qiu* Subhabrata Sen†

February 25, 2022

Abstract

We derive a variational representation for the log-normalizing constant of the posterior distribution in Bayesian linear regression with a uniform spherical prior and an i.i.d. Gaussian design. We work under the “proportional” asymptotic regime, where the number of observations and the number of features grow at a proportional rate. This rigorously establishes the Thouless-Anderson-Palmer (TAP) approximation arising from spin glass theory, and verifies a conjecture of [22] in the special case of the spherical prior.

1 Introduction

The analysis of high-dimensional probability distributions is a central challenge in modern Statistics and Machine Learning. This is particularly true in the context of Bayesian Statistics, where scientists carry out inference based on the posterior distribution. In modern applications, the posterior distribution is typically high-dimensional, and analytically intractable. Variational Inference (VI) has emerged as an attractive option to approximate these intractable distributions, facilitating fast, parallel computations in state-of-the-art applications [10, 32]. In this approach, the distribution of interest is approximated (in KL divergence) by distributions from a pre-specified, more tractable collection. The simplest version of VI is the Naive Mean-field approximation (NMF), where the distribution of interest is approximated by a product distribution. This strategy has roots in classical approximations for spins-systems in statistical physics [23].

Despite the rapidly growing popularity of variational approximations in Statistics and Machine Learning, the corresponding theoretical guarantees for these approximations are quite limited. In this paper, we study Bayesian linear regression, a standard workhorse of modern Statistics, via the lens of (advanced) Mean-field approximations. Given data $\{(y_i, x_i) : 1 \leq i \leq n\}$, $y_i \in \mathbb{R}$, $x_i \in \mathbb{R}^p$, the scientist posits a linear model $y_i = x_i^\top \beta + \varepsilon_i$, where $\beta \in \mathbb{R}^p$ and $\varepsilon_i \sim N(0, \Delta)$ are iid. Given a prior distribution π_0 , setting $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ and $X^\top = [x_1, \dots, x_n] \in \mathbb{R}^{p \times n}$, the scientist constructs the posterior

$$\frac{d\mathbb{P}}{d\pi_0}(\beta|y, X) = \frac{1}{\mathcal{Z}_p} e^{-\frac{1}{2\Delta} \|y - X\beta\|^2}, \quad (1)$$

where \mathcal{Z}_p is the normalizing constant of the posterior distribution. Borrowing terminology from statistical physics, we will refer to \mathcal{Z}_p as the partition function of the model.

Using the classical Gibbs variational principle [32], one has

$$\log \mathcal{Z}_p = \sup_{Q \ll \pi_0} \left(\mathbb{E}_Q \left[-\frac{1}{2\Delta} \|y - X\beta\|_2^2 \right] - D_{KL}(Q \| \pi_0) \right),$$

with equality if Q is the posterior distribution. Restricting the supremum above to product measures, one obtains the NMF approximation to the log-partition function. Note that the NMF approximation is always

*Department of Statistics, Harvard University, Cambridge, MA 02138, U.S.A.

†Department of Statistics, Harvard University, Cambridge, MA 02138, U.S.A.

a valid lower bound:

$$\log \mathcal{Z}_p \geq \sup_{Q=\prod_i Q_i} \left(\mathbb{E}_Q \left[-\frac{1}{2\Delta} \|y - X\beta\|_2^2 \right] - D_{KL}(Q\|\pi_0) \right).$$

We say that the NMF approximation is *correct to leading order* if

$$\log \mathcal{Z}_p - \sup_{Q=\prod_i Q_i} \left(\mathbb{E}_Q \left[-\frac{1}{2\Delta} \|y - X\beta\|_2^2 \right] - D_{KL}(Q\|\pi_0) \right) = o(p). \quad (2)$$

In recent joint work with Sumit Mukherjee [25], the second author leveraged recent advances in the theory of non-linear large deviations to derive sufficient conditions on X for the accuracy of NMF approximation when $\pi_0 = \prod^{\otimes p} \pi$ is a product distribution. In this special case, it is easy to see that the supremum in (2) is obtained by $Q^* = \prod_i Q_i^*$, with $\frac{dQ_i^*}{d\pi}(\beta_i) \propto \exp\left(-\frac{(\beta_i - R_i)^2}{2\Sigma_i^2}\right)$. The approximation (2) can thus be equivalently parametrized in terms of $a_i = \mathbb{E}_{Q_i^*}[\beta_i]$ and $c_i = \text{Var}_{Q_i^*}[\beta_i]$, so that

$$\log \mathcal{Z}_p = \sup_{a,c} \left(-\frac{1}{2\Delta} \|y - Xa\|_2^2 - D_{KL}(Q^*\|\pi_0) \right) + o(p) \quad (3)$$

if the NMF approximation is accurate to leading order.

In the special case where x_i 's are iid standard gaussian, [25] establishes the accuracy of the NMF approximation as long as $p = o(n)$. This result is expected to be tight—using insights from spin glasses, physicists predict that Bayes linear regression is *not* approximately NMF if the number of samples n and the number of features p grow proportionally [22]. In fact, using the Thouless-Anderson-Palmer (TAP) formulation, one obtains the following prediction for the log-partition function.

Conjecture 1.1 ([22]). *Assume $\pi_0 = \prod^{\otimes p} \pi$, where π is a probability distribution on $[-1, 1]$. Then*

$$\log \mathcal{Z}_p = \max_{\|a\|_\infty \leq 1} \left[-\frac{\|y - Xa\|_2^2}{2\Delta} - \frac{n}{2} \log \left(1 + \frac{\bar{c}}{\Delta} \right) + \sum_{i=1}^p \log \mathbb{E}_\pi \left[\exp\left(-\frac{(\beta_i - R_i)^2}{2\Sigma_*^2}\right) \right] \right] + o(p), \quad (4)$$

where the parameters R_i , c_i and Σ_* are specified as follows: define $(dQ_i/d\pi)(\beta) \propto \exp(-(\beta - R_i)^2/(2\Sigma_*^2))$. R_i is chosen such that $\mathbb{E}_{Q_i}[\beta] = a_i$. Set $c_i = \text{Var}_{Q_i}(\beta)$. Finally, the parameter Σ_* depends on Δ and α , and is specified implicitly as the smallest solution to a fixed point equation.

The interval $[-1, 1]$ is arbitrary, the conjecture generalizes directly to any prior π supported on a bounded set. The main difference between (3) and (4) is the second term in the RHS of (4). This term is referred to as the ‘‘Onsager Correction’’ term in statistical physics. The representation (4) is crucial for a number of reasons:

- (i) It suggests that the widely popular NMF approximation is actually incorrect in certain regimes, and thus practitioners should exercise caution before using the NMF approximation.
- (ii) In settings where the TAP formalism is conjectured to be correct, it is further conjectured that the variational problem (4) has a unique global optimizer. Further, this optimizer is expected to be an asymptotically optimal estimator (in a Bayesian sense).
- (iii) This suggests the following approach—optimize the RHS of (4) directly using an out of the box algorithm such as gradient descent, and then use the resulting estimator for inference.

Exploring the validity of the TAP framework in this problem thus has vital practical implications. From a theoretical perspective, the posterior distribution (1) can be viewed as a ‘‘planted’’ spin glass model. Due to special symmetries present in these models (e.g. the Nishimori identities), posterior distributions naturally provide examples of ‘‘replica symmetric’’ models in spin glass theory. They provide ideal test beds to study

the broader applicability of ideas developed to study mean-field spin glass models such as the Sherrington-Kirkpatrick (SK) and mixed p-spin models. Techniques developed originally in the context of mean-field spin glasses have already been utilized to characterize the limiting mutual information between the data and the underlying signal in the Bayesian linear regression model [6]. In this paper, we adopt a complementary approach, and initiate a study of Bayes linear regression model through the lens of the TAP approximation.

1.1 Main Result

Our main result in this paper is the analogue of Conjecture 1.1 under a uniform spherical prior on β . Specifically, we consider Bayesian linear regression with a uniform prior on $S^{p-1}(\sqrt{p}) := \{\beta \in \mathbb{R}^p : \|\beta\| = \sqrt{p}\}$. Crucially, we assume throughout that we are in the “well-specified” setting, i.e. the data $\{(y_i, x_i) : 1 \leq i \leq n\}$ is also generated from the model $y_i = x_i^\top \beta_0 + \varepsilon_i$, where $\beta_0 \sim \pi_0$, and π_0 henceforth denotes the uniform distribution on $S^{p-1}(\sqrt{p})$.

We assume that $x_i \sim N_p(0, I/n)$ are iid. We will work under the “proportional asymptotics” setting, i.e., the number of samples n and the number of features p grow proportionally, with $n/p \rightarrow \alpha \in (0, \infty)$. For convenience, we will track the asymptotics in terms of p , and will think of n as a function of p . Note that in this setting, the *partition function* is

$$\mathcal{Z}_p = \int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{2\Delta} \|y - X\beta\|^2} d\pi_0(\beta). \quad (5)$$

The partition function \mathcal{Z}_p is a function of both y and X , so we use the notation $\mathcal{Z}_p(y, X)$ when we want to highlight its dependence on y and X . We refer to

$$F_p(y, X) = \frac{1}{p} \ln \mathcal{Z}_p(y, X)$$

as the *free energy* and

$$\mathcal{H}(\beta) = \frac{1}{2\Delta} \|y - X\beta\|^2 \quad (6)$$

as the *Hamiltonian*, borrowing terminology from statistical physics. Armed with these notions, we can state our main result.

Theorem 1. *Fix $\alpha \in (0, \infty)$. There exists $\Delta_0 > 0$ such that for all $\Delta > \Delta_0$, as $p \rightarrow \infty$,*

$$\left| F_p - \sup_{a \in \mathbb{R}^p, \|a\| \leq \sqrt{p}} f_{TAP}(a) \right| \xrightarrow{P} 0,$$

where

$$f_{TAP}(a) := -\frac{1}{2\Delta p} \|y - Xa\|^2 - \frac{\alpha}{2} \ln \left(1 + \frac{1 - \|a\|^2/p}{\Delta\alpha} \right) + \frac{1}{2} \ln \left(1 - \frac{\|a\|^2}{p} \right). \quad (7)$$

Remark 1. Note from (5) that $1/(2\Delta)$ is similar to an inverse temperature parameter in this model. Thus our TAP representation is valid “at high temperature”. The high-temperature requirement comes up when we lower bound the free energy in terms of the TAP formula. We emphasize that the matching upper bound is valid in general, and does not require any additional conditions. Establishing a valid TAP representation at any $\Delta > 0$ remains an open problem in this setting.

Remark 2. A close look at our proof indicates that the TAP representation is valid for any $\Delta > 0$ if α is sufficiently large. In this case, the Signal-to-Noise Ratio is high, and the model is again effectively at high-temperature. We state all our results at fixed α , with $\Delta > 0$ sufficiently large, but note that all results transfer immediately to the “fixed Δ large α ” regime.

1.2 Background and Challenges

We describe here the broader context of our result. We also take this opportunity to highlight the main differences from the existing literature, and describe the key challenges in our setting.

The TAP formalism was originally introduced for the Sherrington-Kirkpatrick model by Thouless-Anderson-Palmer in [29], predating Parisi’s groundbreaking “replica-symmetry breaking” solution. The TAP approach yields a variational representation for the free energy, as well as a fixed point equation for the global magnetizations, referred to as the TAP equations. The TAP equations were established rigorously using a number of different approaches—Stein’s method [12], cavity method [28], dynamical methods [1] etc. At high-temperature, the Gibbs measure is in one “pure-state”, and the magnetizations satisfy the TAP equations. The behavior at low temperatures is significantly more complicated—the Gibbs measure decomposes into exponentially many “pure states”, and a version of the TAP equations are expected to hold within each pure state. A rigorous version of this conjecture was verified in [2] using the approximate ultrametric decomposition from [21].

The progress on the TAP formula for the free energy is more recent. [13] verified the TAP representation by evoking the Parisi formula. More recently, direct proofs of the TAP representation have been derived in the literature. In this regard, [9] utilizes Random Matrix Theory to establish the variational formula for the spherical SK model, while subsequent developments establish these results for mixed p-spin models [14, 15]. In this regard, the notion of “multi-samplability” plays a crucial role. We will also crucially exploit a related idea in our analysis. We emphasize this connection in our discussion of the proof strategy below. We note that in sharp contrast with the model studied in this paper, the models studied in the prior literature are “null” models, i.e. do not have a latent signal. The planted setting is arguably more natural in the context of Statistics and Machine Learning. To the best of our knowledge, this is the very first proof of a TAP formula for the free energy for a planted spin glass model. We note that the random hamiltonians in the study of the Sherrington-Kirkpatrick and mixed p-spin models are centered Gaussian processes—this directly facilitates the use of tools tailored for Gaussian variables e.g. interpolation and comparison techniques [14, 15]. This is no longer true in our setting, and thus necessitates new ideas. In this setting, the hamiltonian roughly looks like a two-spin model, with the sample covariance matrix as the couplings among the spins. We directly utilize this structure in our proofs, exploiting spectral properties of this matrix. We expect our proof ideas to be applicable for more general random matrix ensembles, e.g., the orthogonally invariant matrix ensembles [7, 26].

The behavior of the TAP free energy has been examined in the setting of finite rank spiked matrix models in [18]. [18] establishes that the Naive Mean Field approximation exhibits an undesirable instability property below the Information Theoretic threshold, thus rendering variational methods based on this approximation untrustworthy in practice. In turn, it suggests using the TAP framework to carry out variational inference in this setting. In follow up work [11, 16], the accuracy of TAP-approximation based variational inference has been established in the context of the \mathbb{Z}_2 synchronization problem, a special case of the general setting examined in [18].

1.3 Proof strategy

Our proof has two separate arguments—first, we lower bound the log-partition function $\log \mathcal{Z}_p$ in terms of the TAP free energy $f_{\text{TAP}}(a)$, as defined in (7). To this end, fix $a \in \mathbb{R}^p$, $\|a\| \leq \sqrt{p}$. We re-center the hamiltonian $\mathcal{H}(\beta)$ around the vector a ; this yields the term $-\|y - Xa\|^2/(2\Delta p)$ in $f_{\text{TAP}}(a)$. To obtain a valid lower bound, we restrict the integral in \mathcal{Z}_p to the intersection of $S^{p-1}(\sqrt{p})$ and a subspace that is orthogonal to a and the resulting external magnetization $X^\top(y - Xa)$ from the previous re-centering step. The volume of this “band” is $[\ln(1 - \|a\|^2/p)]/2$ —this contributes the third term in $f_{\text{TAP}}(a)$. Finally, the model restricted to the “band” is a spherical two-spin model with hamiltonian $H(\beta) = -\|X(\beta - a)\|^2/(2\Delta)$. The free energy of this model is calculated using the second moment method, and contributes $-(\alpha/2) \ln[1 + (1 - \|a\|^2/p)/(\Delta\alpha)]$ in $f_{\text{TAP}}(a)$. This yields a natural interpretation of the Onsager correction term in this problem. We note that this step explicitly uses the “high-temperature” condition stated in the Theorem statement. This shows that $(\ln \mathcal{Z}_p)/p \geq f_{\text{TAP}}(a)$ *uniformly* over all $a \in \mathbb{R}^p$ such that $\|a\|/\sqrt{p}$ is bounded away from one. We argue

separately that the end-point can be neglected. Our argument is similar to the one of [9] in the context of the spherical two-spin Sherrington-Kirkpatrick model.

Our main contribution is the corresponding upper bound on the free energy in terms of the TAP free energy functional. To this end, we crucially leverage “overlap concentration” properties under the well-specified posterior distributions [5]. Roughly, this implies that given a vanishing amount of additional side information, with high-probability (over the law of the disorder (y, X)), we can sample many iid “replicas” $\beta^1, \dots, \beta^{m_p}$ from the posterior distribution such that all the pairwise overlaps are concentrated around a fixed value. Such overlap values are also termed as “multi-samplable” in [27] in the study of the TAP formalism for mixed p-spin spherical spin glasses. This overlap concentration property allows us to derive an upper bound to the log-partition function $\ln \mathcal{Z}_p$ in terms of a restricted partition function for a “replicated” system, where all the pairwise overlaps are restricted to approximately the same value. We now set a as the sample average of these replicas, and re-center the hamiltonian around this vector. The concentration of all pairwise overlaps around a fixed value implies that the $\beta^i - a$ ’s are approximately orthogonal, and this allows us to cancel the external magnetic field $X^\top(y - Xa)$ created from the re-centering step. Finally, due to the approximate orthogonality of the $\beta^i - a$ vectors, the integral in the replicated system approximately de-couples, and provides m_p copies of the Onsager correction and volume contributions. In particular, this establishes that $(\ln \mathcal{Z}_p)/p \leq f_{\text{TAP}}(a) + o(1)$ for some (random) $a \in \mathbb{R}^p$, and thus yields the desired upper bound.

We note that the overlap concentration property plays a very crucial role in our proof. This concentration property has been established recently for well-specified Bayesian posteriors (see e.g. [8]). It should be possible to combine the ideas in this paper with overlap concentration to establish similar TAP formulae in other Bayesian inference problems. We leave this for future work.

1.4 Directions for future research

We collect here some open questions for future research.

- (i) An immediate open question is to extend our TAP representation to any $\Delta > 0$. We note that the high-temperature condition is explicitly used to calculate the free energy on the band—this is the only step in our proof which uses this condition.
- (ii) The main question for future research concerns a proof of Conjecture 1.1. We believe that the ideas introduced in this paper will be crucial in this effort. The spherical prior crucially allows us to use ideas from random matrix theory; specifically, after the re-centering step described above, one has to compute the free energy of a two spin model on the sphere. Further, one requires this estimate to be “uniform” over the choice of the center. Exploiting the spherical symmetry, we use asymptotics of spherical integrals [19] to derive these uniform estimates. This symmetry is crucially absent for other natural priors, e.g., product priors. Thus our arguments do not generalize directly to these settings.
- (iii) In Statistics and Machine Learning, one is often interested in Variational Inference based on the TAP free energy. Specifically, given Theorem 1, a natural recipe is to start at some appropriate initialization, and subsequently optimize f_{TAP} directly by some out-of-the-box optimization algorithm such as gradient descent. Of course, f_{TAP} is non-convex in general, and thus the algorithm will usually reach a local optimum, rather than the global maximum. This yields a natural estimator, and one naturally wishes to study the statistical performance of this estimator. It would be of interest to characterize this performance. Such results have been achieved recently in the context of the \mathbb{Z}_2 synchronization problem [11, 17]. We expect these insights for \mathbb{Z}_2 synchronization to also be useful in the context of Bayesian linear regression.

Notations: We use the usual Bachmann-Landau notation $O(\cdot)$, $o(\cdot)$, $\Theta(\cdot)$ for sequences. For a sequence of random variables $\{X_p : p \geq 1\}$, we say that $X_p = o(1)$ if $X_p \xrightarrow{P} 0$ as $p \rightarrow \infty$.

We use $\mathbb{P}[\cdot]$ and $\mathbb{E}[\cdot]$ to denote the probability and expectation under the joint distribution of the data (y, X) . For other measures, we add a suitable subscript to the probability and expectation operators to emphasize the distribution under consideration. In our computations, we will often restrict the calculations to a “good” set. We collect this set here. Let $\kappa = 100$. Define a set

$$\Omega := \left\{ (X, \epsilon, \beta_0) : \sigma_{\max}(X) < 1 + \alpha^{-1/2} + \kappa, \|\epsilon\| < \kappa\sqrt{p\Delta} \right\}, \quad (8)$$

where $\sigma_{\max}(X)$ denotes the largest singular value of X . By Lemma 15, one has

$$\lim_{p \rightarrow \infty} \mathbb{P}((X, \epsilon, \beta_0) \in \Omega) = 0.$$

Note that with a slight abuse of notations, we will use $(X, y) \in \Omega$ and $(X, \epsilon, \beta_0) \in \Omega$ exchangeably. Throughout this paper, unless otherwise specified, a sequence of events is said to occur “with high probability” if it has probability $1 - o(1)$ under the joint distribution (X, y) .

Throughout, we use $C, c, C_1, c_1, C_2, c_2 \dots$ to denote positive constants independent of n, p , possibly depending on the parameters Δ and α . Further, these constants can change from line to line. For any square symmetric matrix A , $\|A\|_2$ and $\|A\|_F$ denote the matrix operator norm and the Frobenius norm respectively. Finally, $\langle \cdot \rangle$ will denote the expectation operator under the posterior distribution (1).

Acknowledgments: SS thanks Sumit Mukherjee for his encouragement during the completion of this manuscript. SS was partially supported by a Harvard Dean’s Competitive Fund Fellowship.

Outline: The rest of the article is structured as follows: We derive a lower bound on the log-partition function via the TAP free energy in Section 2, while Section 3 establishes a matching upper bound. Section 4 combines these two bounds, and establishes Theorem 1. We crucially utilize the notion of overlap concentration in our upper bound proof. The proof of overlap concentration is a direct adaptation of existing arguments in the literature. For the sake of completeness, we include a proof in our setting in the Appendix.

2 Lower bound on the free energy

To get a lower bound of the free energy $F_p = \frac{1}{p} \ln \mathcal{Z}_p$, we restrict the integral in (5) to specific subsets of $S^{p-1}(\sqrt{p})$, which yields a natural lower bound. Following this strategy, we will first state Theorem 3 in Section 2.1; this specifies the subsets that we look at. We then prove Theorem 2 assuming Theorem 3. The proof of Theorem 3 itself will be deferred to Section 2.2. Finally, in Section 2.3, we provide proofs of some intermediate lemmas that are used to prove Theorem 3.

Theorem 2 (Lower bound). *Fix $\alpha \in (0, \infty)$. There exists $\Delta_0 > 0$ such that for all $\Delta > \Delta_0$, any $\eta > 0$, $c \in (0, 1)$,*

$$\mathbb{P}\left(\frac{1}{p} \ln \mathcal{Z}_p \leq \sup_{a \in \mathbb{R}^p, \|a\| \leq (1-c)\sqrt{p}} f_{TAP}(a) - \eta\right) \rightarrow 0.$$

2.1 Proof of Theorem 2 using Theorem 3

Theorem 3. *Let $\{v_1, v_2\}$ be an orthogonal basis of length \sqrt{p} of the 2-dim linear subspace $\text{Span}(a, X^\top(y - Xa)) \subset \mathbb{R}^p$. For small positive ϵ and $a \in \mathbb{R}^p$, define*

$$B(a, \epsilon) = \{\beta : v_i^\top(\beta - a) \in (-p\epsilon, p\epsilon), \forall i \in \{1, 2\}\}. \quad (9)$$

There exists $\Delta_0 > 0$ such that for all $\Delta > \Delta_0$, for any $c \in [0, 1]$, there exists $C > 0$ such that as $p \rightarrow \infty$,

$$\mathbb{P}\left(\sup_{\|a\| \leq (1-c)\sqrt{p}} \left| \frac{1}{p} \ln \left(\int_{B(a, \epsilon)} e^{-\frac{1}{2\Delta}(\beta - a)^\top X^\top X(\beta - a)} d\pi_0(\beta) \right) - \left[-\frac{\alpha}{2} \ln \left(1 + \frac{(1 - \|a\|^2/p)}{\Delta\alpha} \right) + \frac{1}{p} \ln \text{Vol}(a, \epsilon) \right] \right| \leq C\epsilon \right) \rightarrow 1,$$

where $\text{Vol}(a, \epsilon) := \text{Vol}(\text{B}(a, \epsilon)) = \mathbb{E}_{\beta \sim \pi_0} [1_{\beta \in \text{B}(a, \epsilon)}]$.

Proof of Theorem 2: Note that by the definition of Ω , i.e. (8), for any $(X, y) \in \Omega$,

$$\|X^\top(y - Xa)\|^2 \leq \sigma_{\max}^2(X) \|X(\beta_0 - a) + \varepsilon\|^2 \leq \sigma_{\max}^2(X) [4p\sigma_{\max}^2(X) + \|\varepsilon\|^2 + 4\sqrt{p}\sigma_{\max}(X)\|\varepsilon\|] = O(p). \quad (10)$$

From now on, we only consider $(X, y) \in \Omega$. Since the integrand in (5) is always strictly positive and $\text{B}(a, \epsilon)$ is a subset of $S^{p-1}(\sqrt{p})$, naturally we have the following lower bound

$$\mathcal{Z}_p = \int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{2\Delta} \|y - X\beta\|^2} d\pi_0(\beta) \geq \int_{\text{B}(a, \epsilon)} e^{-\frac{1}{2\Delta} \|y - X\beta\|^2} d\pi_0(\beta).$$

Upon recentering the exponent around a , noting that $X^\top(y - Xa) \in \text{Span}(v_1, v_2)$, where $\text{Span}(v_1, v_2)$ denotes the linear subspace of \mathbb{R}^p generated by v_1 and v_2 , by the definition of $\text{B}(a, \epsilon)$, together with (10), we have

$$\begin{aligned} \|y - X\beta\|^2 &= \|y - Xa\|^2 + \|X(\beta - a)\|^2 - 2(X^\top(y - Xa))^\top(\beta - a) \\ &= \|y - Xa\|^2 + \|X(\beta - a)\|^2 + 2\epsilon O(p) \end{aligned}$$

for any $\beta \in \text{B}(a, \epsilon)$. Thus we have

$$\mathcal{Z}_p \geq e^{-\frac{1}{2\Delta} \|y - Xa\|^2 + O(p\epsilon)} \int_{\text{B}(a, \epsilon)} e^{-\frac{1}{2\Delta} \|X(\beta - a)\|^2} d\pi_0(\beta).$$

Equivalently, there is a positive constant C' , such that

$$\ln \mathcal{Z}_p \geq -\frac{1}{2\Delta} \|y - Xa\|^2 - C'\epsilon + \ln \left(\int_{\text{B}(a, \epsilon)} e^{-\frac{1}{2\Delta} \|X(\beta - a)\|^2} d\pi_0(\beta) \right).$$

Note that for any 2-dimensional linear subspace $\mathcal{L} \subset \mathbb{R}^p$, length of the projection of β on to \mathcal{L} divided by \sqrt{p} has density

$$f_{\text{Proj}}(x) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p-2}{2})} (1 - x^2)^{\frac{p-4}{2}}$$

(see for instance of [9, (2.8)]). Without loss of generality, we only consider the non-degenerate case, namely when a and $X^\top(y - Xa)$ are not linearly dependent. Thus when ϵ is small enough,

$$\frac{1}{p} \ln \text{Vol}(a, \epsilon) \geq \frac{1}{p} \ln \left[\sqrt{2\epsilon} \cdot f_{\beta^\perp} \left(\frac{\|a\|}{\sqrt{p}} - \frac{\epsilon}{\sqrt{2}} \right) \cdot \min \left\{ \frac{\epsilon}{\sqrt{2\pi}\|a\|}, 1 \right\} \right] = \frac{1}{2} \ln \left(1 - \frac{\|a\|^2}{p} \right) + o(1).$$

Note that the $o(1)$ term is uniform over different choices of a such that $\|a\| \leq (1 - c)\sqrt{p}$. By Theorem 3, denoting the event on the LHS of (10) as Ω_p ,

$$\begin{aligned} &\mathbb{P} \left(\frac{1}{p} \ln \mathcal{Z}_p \geq \sup_{a \in \mathbb{R}^p, \|a\| \leq (1-c)\sqrt{p}} f_{\text{TAP}}(a) - 2(C + C')\epsilon + o(1) \right) \\ &\geq \mathbb{P} \left(\left\{ \frac{1}{p} \ln \mathcal{Z}_p \geq \sup_{a \in \mathbb{R}^p, \|a\| \leq (1-c)\sqrt{p}} f_{\text{TAP}}(a) - 2(C + C')\epsilon + o(1) \right\} \cap \Omega \right) \\ &\geq \mathbb{P} \left(\left\{ \frac{1}{p} \ln \mathcal{Z}_p \geq -\frac{1}{2\Delta} \|y - Xa\|^2 - C'\epsilon + \ln \left(\int_{\text{B}(a, \epsilon)} e^{-\frac{1}{2\Delta} \|X(\beta - a)\|^2} d\pi_0(\beta) \right) \right\} \right. \\ &\quad \left. \forall a \in \mathbb{R}^p : \|a\| \leq (1 - c)\sqrt{p} \right\} \cap \Omega \cap \Omega_p \Big) \\ &= \mathbb{P}(\Omega \cap \Omega_p) \rightarrow 1, \end{aligned}$$

where C is the one in statement of Theorem 3. In other words, since ϵ can be chosen to be arbitrarily small, we have for any $\eta > 0$,

$$\lim_{p \rightarrow \infty} \mathbb{P} \left(\frac{1}{p} \ln \mathcal{Z}_p \leq \sup_{a \in \mathbb{R}^p, \|a\| \leq (1-c)\sqrt{p}} f_{\text{TAP}}(a) - \eta \right) = 0.$$

□

2.2 Proof of Theorem 3

Building towards the proof of Theorem 3, we first introduce the following lemma, which gives an explicit formula for the high-dimensional limit of the free energy in a system without external field. We establish this using a second moment method argument (see Section 2.3 for its proof). Note that we use the notation $\beta^\top X^\top X \beta$ instead of $\|X\beta\|^2$ in this subsection since we would like to emphasise the fact that our proofs depend strongly on the properties of the sample covariance matrix $X^\top X$.

Lemma 1 (High-dimensional limit of free energy: a model without external field). *Fix $\alpha \in (0, \infty)$. There exists $\Delta_0 > 0$ such that for all $\Delta > \Delta_0$, as $p \rightarrow \infty$*

$$\frac{1}{p} \ln \left(\int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{2\Delta} \beta^\top X^\top X \beta} d\pi_0(\beta) \right) \xrightarrow{P} -\frac{\alpha}{2} \ln \left(1 + \frac{1}{\Delta\alpha} \right).$$

Note the left hand side is Lipschitz in $T = \frac{1}{\Delta}$ with high probability. Thus for any $\Delta_0 > 0$, we can divide the interval $[0, \frac{1}{\Delta_0}]$ into a sufficiently fine grid. Lemma 1 holds simultaneously for each grid-point; combining this with the Lipschitz continuity of the free energy in $\frac{1}{\Delta}$, we immediately obtain the following uniform convergence statement.

Lemma 2. *Fix $\alpha \in (0, \infty)$. There exists $\Delta_0 > 0$ such that as $p \rightarrow \infty$*

$$\sup_{\Delta \geq \Delta_0} \left| \frac{1}{p} \ln \left(\int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{2\Delta} \beta^\top X^\top X \beta} d\pi_0(\beta) \right) - \left(-\frac{\alpha}{2} \ln \left(1 + \frac{1}{\Delta\alpha} \right) \right) \right| \xrightarrow{P} 0.$$

As the definition of $B(a, \epsilon)$ in (9) suggests, we want to consider integral over the intersection of the sphere and a $p - 2$ dimensional linear space, which can be formalized into the following uniform convergence statement.

Lemma 3 (High-dimensional limit of free energy without external field: on a linear subspace). *Fix $\alpha \in (0, \infty)$. There exists $\Delta_0 > 0$ such that for all $\Delta > \Delta_0$, as $p \rightarrow \infty$*

$$\sup_{\Delta \geq \Delta_0, u, v \in \mathbb{R}^p} \left| \frac{1}{p} \ln \left(\int_{S^{p-1}(\sqrt{p}) \cap \text{Span}(u, v)^\perp} e^{-\frac{1}{2\Delta} \beta^\top X^\top X \beta} d\pi^{\text{Span}(u, v)^\perp}(\beta) \right) - \left(-\frac{\alpha}{2} \ln \left(1 + \frac{1}{\Delta\alpha} \right) \right) \right| \xrightarrow{P} 0,$$

where $\text{Span}(u, v)$ denotes the linear subspace of \mathbb{R}^p generated by u and v . Note that $\pi^{\text{Span}(u, v)^\perp}$ refers to the uniform measure on $S^{p-1}(\sqrt{p}) \cap \text{Span}(u, v)^\perp$.

Now we proceed to prove Theorem 3, assuming the convergence results mentioned above. The proofs of these results are included in Section 2.3.

Proof of Theorem 3: Recall the definition of Ω in (8). From now on, we only consider $(X, y) \in \Omega$ for the rest of this proof. Fix any $a \in \mathbb{R}^p$ with $\|a\| \leq (1-c)\sqrt{p}$. Let $\hat{\beta} = \beta - a$. Define

$$H(\hat{\beta}) := \frac{1}{2\Delta} \hat{\beta}^\top X^\top X \hat{\beta} = \frac{1}{2\Delta} (\beta - a)^\top X^\top X (\beta - a). \quad (11)$$

Note that H is different from the Hamiltonian \mathcal{H} defined in (6). By (10), $\|X^\top(y - Xa)\| = O(\sqrt{p})$. By the definition of $B(a, \epsilon)$ in (9), for any $\beta \in B(a, \epsilon)$, one has

$$|a^\top \hat{\beta}| = |a^\top(\beta - a)| < c' p \epsilon$$

and

$$|(X^\top(y - Xa))^\top \hat{\beta}| = |(y - Xa)^\top X(\beta - a)| < c' p \epsilon,$$

for some positive constant c' . Therefore

$$\|\hat{\beta}\|^2 = \|\beta\|^2 - \|a\|^2 - 2(\beta - a)^\top a = p - \|a\|^2 + O(p\epsilon)$$

Let $\hat{\beta}^\perp$ be the projection of $\hat{\beta}$ onto the hyper-plane $\text{Span}(v_1, v_2)^\perp$. For the sake of notation clarity, let $\beta^\perp = \sqrt{p} \hat{\beta}^\perp / \|\hat{\beta}^\perp\|$ and $\gamma = \|\hat{\beta}^\perp\| / \sqrt{p}$. Note that, again by the definition of $B(a, \epsilon)$

$$\|\hat{\beta} - \gamma \beta^\perp\| = \|\hat{\beta} - \hat{\beta}^\perp\| < c' \sqrt{p} \epsilon,$$

and

$$\gamma = \gamma_0 + O(\epsilon),$$

where $\gamma_0 := \sqrt{p - \|a\|^2} / \sqrt{p}$. Therefore,

$$H(\hat{\beta}) = H(\gamma \beta^\perp) + O(p\epsilon) = H(\gamma_0 \beta^\perp) + O(p\epsilon).$$

which is due to the fact that $\|\nabla H(\beta)\| = \|-X^\top X \beta / \Delta\| \leq (\sqrt{p} / \Delta) \|X^\top X\|_2 = O(\sqrt{p})$ for $(X, y) \in \Omega$. Now we come back to the targeted integral

$$\int_{B(a, \epsilon)} e^{-\frac{1}{2\Delta}(\beta - a)^\top X^\top X(\beta - a)} d\pi_0(\beta) = \mathbb{E}_{\pi_0}[\mathbf{1}_{B(a, \epsilon)} e^{-H(\gamma_0 \beta^\perp) + O(p\epsilon)}].$$

Since β^\perp is uniformly distributed on the unit sphere intersected with $\text{Span}(a, X^\top(y - Xa))^\perp$ under π_0 and γ is independent of $\mathbf{1}_{B(a, \epsilon)}$, one has

$$\begin{aligned} & \mathbb{E}_{\pi_0}[\mathbf{1}_{B(a, \epsilon)} e^{-H(\gamma_0 \beta^\perp) + O(p\epsilon)}] \\ &= e^{O(p\epsilon)} \mathbb{E}_{\pi_0}[\mathbf{1}_{B(a, \epsilon)} \int_{S^{p-1}(\sqrt{p - \|a\|^2}) \cap \text{Span}(v_1, v_2)^\perp} e^{-\frac{1}{2\Delta} \beta^\top X^\top X \beta} d\pi^{\text{Span}(v_1, v_2)^\perp}(\beta)] \\ &= e^{O(p\epsilon)} \text{Vol}(a, \epsilon) \int_{S^{p-1}(\sqrt{p}) \cap \text{Span}(v_1, v_2)^\perp} e^{-\frac{1}{2\Delta} \beta^\top X^\top X \beta} d\pi^{\text{Span}(v_1, v_2)^\perp}(\beta) \\ &= e^{O(p\epsilon)} \text{Vol}(a, \epsilon) \mathbb{E}^{\text{Span}(v_1, v_2)^\perp} \left[\exp\left(-\frac{1}{2\tilde{\Delta}} \beta^\top X^\top X \beta\right) \right], \end{aligned}$$

where

$$\tilde{\Delta} = \frac{p}{p - \|a\|^2} \Delta.$$

Note that for any a such that $\|a\| \leq (1 - c)\sqrt{p}$, we have $\tilde{\Delta} > \Delta/2$. By Lemma 3 and the fact that $\mathbb{P}(\Omega) \rightarrow 1$ as $p \rightarrow \infty$, we finally get with high probability

$$\frac{1}{p} \ln \int_{B(a, \epsilon)} e^{-\frac{1}{2\tilde{\Delta}}(\beta - a)^\top X^\top X(\beta - a)} d\pi_0(\beta) = O(\epsilon) + \frac{1}{p} \ln \text{Vol}(a, \epsilon) - \frac{\alpha}{2} \ln\left(1 + \frac{(1 - \|a\|^2/p)}{\Delta\alpha}\right) + o(1),$$

in which the $o(1)$ terms are uniform with respect to the choice of $\tilde{\Delta}$, equivalently, with respect to the choice of a . \square

2.3 Concentration of free energy in the model without external field

In this subsection, we will prove Lemma 1 and Lemma 3.

2.3.1 Proof of Lemma 1

Recall that $H(\beta) = H(X, \beta) = (1/2\Delta)\|X\beta\|^2$, which was first defined in (11). We define the partition function $Z_p(\Delta, X)$ as

$$Z_p(\Delta, X) := \int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{2\Delta}\beta^\top X^\top X\beta} d\pi_0(\beta).$$

which is different from \mathcal{Z}_p defined in (5). Let

$$\Phi_p(\Delta, X) := \frac{1}{p} \ln \int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{2\Delta}\beta^\top X^\top X\beta} d\pi_0(\beta) = \frac{1}{p} \log \mathbb{E}_{\pi_0}[\exp(H(\beta))] = \frac{1}{p} \ln Z_p(\Delta, X)$$

be the *quenched* free energy of the model without external field and

$$\phi_p(\Delta) := \frac{1}{p} \ln \mathbb{E} \int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{2\Delta}\beta^\top X^\top X\beta} d\pi_0(\beta) = \frac{1}{p} \ln \mathbb{E}[Z_p(\Delta, X)]$$

be the *annealed* free energy.

To prove Lemma 1, we need the following concentration result about $\Phi_p(\Delta, X)$.

Lemma 4. *For any $\Delta > 0$, $\alpha \in (0, \infty)$ and $\delta > 0$, there exists a positive constant C and positive integer p_0 (depending only on Δ, α, δ) such that for any $p > p_0$*

$$\mathbb{P}(|\Phi_p(\Delta, X) - \mathbb{E}\Phi_p(\Delta, X)| > \delta) \leq e^{-Cp}.$$

Proof: Let $\sigma_{\max}(X)$ be the largest singular value of X . Define

$$\mathcal{G} := \left\{ X : \sigma_{\max}(X) < \frac{1}{\sqrt{n}}(\sqrt{p} + \sqrt{n} + t) \right\}.$$

First, we prove that $\Phi_p(\Delta, X)$ is Lipschitz as a function of X on \mathcal{G} , with Lipschitz constant $K_p = 2(1 + \alpha^{-1/2} + t/\sqrt{n})/\sqrt{p}$. By triangle inequality, for any $X_1, X_2 \in \mathcal{G}$,

$$\begin{aligned} |H(X_1, \beta) - H(X_2, \beta)| &\leq |\beta^\top (X_1 - X_2)^\top X_2 \beta| + |\beta^\top X_1^\top (X_1 - X_2) \beta| \\ &\leq \|\beta\|^2 [\|(X_1 - X_2)^\top X_2\|_2 + \|X_1^\top (X_1 - X_2)\|_2] \\ &\leq \sqrt{p} [(\sigma_{\max}(X_1) + \sigma_{\max}(X_2)) \|\sqrt{p}X_1 - \sqrt{p}X_2\|_2] \\ &\leq 2\alpha^{-1/2}(\sqrt{p} + \sqrt{n} + t) \|\sqrt{p}X_1 - \sqrt{p}X_2\|_F, \end{aligned}$$

which implies

$$\begin{aligned} |\Phi_p(\Delta, X_1) - \Phi_p(\Delta, X_2)| &= \frac{1}{p} \left| \ln \frac{\int_{S^{p-1}(\sqrt{p})} e^{-H(X_1, \beta)} d\pi_0(\beta)}{\int_{S^{p-1}(\sqrt{p})} e^{-H(X_2, \beta)} d\pi_0(\beta)} \right| \\ &\leq \frac{1}{p} \left[2\alpha^{-1/2}(\sqrt{p} + \sqrt{n} + t) \|\sqrt{p}X_1 - \sqrt{p}X_2\|_F \right] \\ &= \frac{2\alpha^{-1/2}(1 + \sqrt{\alpha} + t/\sqrt{p})}{\sqrt{p}} \|\sqrt{p}X_1 - \sqrt{p}X_2\|_F. \end{aligned}$$

On the other hand, by Lemma 15, $\mathbb{P}(\mathcal{G}^c) \leq 2e^{-t^2}$. Note that $\Phi(\Delta, 0) = 0$ and

$$\mathbb{E} [\Phi_p^2(\Delta, X)] \leq \mathbb{E} \left[\frac{1}{2\Delta} \sigma_{\max}^2(X) \right] < \infty.$$

Therefore, by Lemma 14, there exists a positive constant C such that for any $r > 6(C + \sqrt{npK_p})\sqrt{\mathbb{P}(\mathcal{G}^c)}$ we have

$$\mathbb{P}(|\Phi_p(\Delta, X) - \mathbb{E}\Phi_p(\Delta, X)| > r) \leq 2e^{-\frac{r^2}{16K_p^2}} + \mathbb{P}(\mathcal{G}^c).$$

Now let $\delta > 0$ be any fixed constant. Let $t = \sqrt{p}$. Then there exists $p_0 \in \mathbb{N}$ such that for any $p > p_0$, $\delta > 6(C + \sqrt{npK_p})\sqrt{\mathbb{P}(\mathcal{G}^c)}$ and

$$\mathbb{P}(|\Phi_p(\Delta, X) - \mathbb{E}\Phi_p(\Delta, X)| > \delta) \leq 2e^{-C_1\delta^2p} + 2e^{-p} \leq e^{-Cp},$$

which concludes the proof of lemma 4. \square

Concerning large- p behaviour of $\phi_p(\Delta)$, we have the following characterization.

Lemma 5. Fix $\alpha \in (0, \infty)$ and $\Delta \in (0, \infty)$. One has

$$\lim_{p \rightarrow \infty} \phi_p(\Delta, X) = -\frac{\alpha}{2} \ln\left(1 + \frac{1}{\Delta\alpha}\right).$$

Proof: We compute the LHS explicitly, namely

$$\begin{aligned} \phi_p(\Delta, X) &= \frac{1}{p} \ln \mathbb{E}_X \mathbb{E}_\beta e^{-\frac{1}{2\Delta} \beta^\top X^\top X \beta} \\ &= \frac{1}{p} \ln \mathbb{E}_\beta \mathbb{E}_X e^{-\frac{1}{2\Delta} \text{tr}(X^\top X \beta \beta^\top)} \\ &= \frac{1}{p} \ln \mathbb{E}_\beta \mathbb{E}_X e^{-\frac{1}{2\Delta} \|\beta\|^2 (X^\top X)_{11}} \\ &= \frac{1}{p} \ln \mathbb{E}_X e^{-\frac{p}{2\Delta} (X^\top X)_{11}} \\ &= \frac{1}{p} \ln \mathbb{E}_{V \sim \chi^2(n)} e^{-\frac{p}{2\Delta n} V}. \end{aligned}$$

Note that moment generating function of χ^2 exists for any $t < \frac{1}{2}$, for any $\Delta > 0$ we have

$$\phi_p(\Delta, X) = \frac{1}{p} \ln\left(\left(1 + \frac{p}{\Delta n}\right)^{-n/2}\right) \rightarrow -\frac{\alpha}{2} \ln\left(1 + \frac{1}{\Delta\alpha}\right).$$

\square

A point worth noting is that there is no non-trivial restriction on the value of Δ . Now we proceed to use second moment method to show $\Phi(\Delta, X)$ also converges to the same quantity. Let σ and β be independent samples from the prior distribution π_0 . Then

$$\begin{aligned} \mathbb{E}(Z_p(\Delta, X)^2) &= \mathbb{E}_X \mathbb{E}_\sigma e^{-\frac{1}{2\Delta} H(\sigma)} \mathbb{E}_\beta e^{-\frac{1}{2\Delta} H(\beta)} \\ &= \mathbb{E}_X \mathbb{E}_{\beta, \sigma} e^{-\frac{1}{2\Delta} (H(\sigma) + H(\beta))} \\ &= \mathbb{E}_X \mathbb{E}_{\beta, \sigma} \exp\left(-\frac{1}{2\Delta} \text{tr}\{X^\top X (\sigma\sigma^\top + \beta\beta^\top)\}\right) \\ &= \mathbb{E}_X \mathbb{E}_{\beta, \sigma} \exp\left(-\frac{1}{2\Delta} \{(p + \beta^\top \sigma)(X^\top X)_{11} + (p - \beta^\top \sigma)(X^\top X)_{22}\}\right), \end{aligned}$$

where the last line follows from $\text{rank}(\sigma\sigma^\top + \beta\beta^\top) = 2$ and in fact its non-zero eigenvalues are $\lambda_1 = p + \beta^\top \sigma$ and $\lambda_2 = p - \beta^\top \sigma$. For $i = 1, 2$, let $V_i = (X^\top X)_{ii}$. Note that V_1 and V_2 are iid with $V_i \sim \frac{1}{n} \chi^2(n)$. Therefore,

$$\begin{aligned} \mathbb{E}(Z_p(\Delta, X)^2) &= \mathbb{E}_V \mathbb{E}_{\beta, \sigma} \exp\left(-\frac{1}{2\Delta} \{(p + \beta^\top \sigma)V_1 + (p - \beta^\top \sigma)V_2\}\right) \\ &= \mathbb{E}_{\beta, \sigma} \mathbb{E}_V \exp\left(-\frac{1}{2\Delta} (\lambda_1 V_1 + \lambda_2 V_2)\right). \end{aligned}$$

Again, using MGF of χ^2 distribution, we further have

$$\begin{aligned}
\mathbb{E}(Z_p(\Delta, X)^2) &= \mathbb{E}_{\beta, \sigma} \mathbb{E}_V \exp\left(-\frac{1}{2\Delta}(\lambda_1 V_1 + \lambda_2 V_2)\right) \\
&= \mathbb{E}_{\beta, \sigma} \left\{ \left(1 + \frac{\lambda_1}{\Delta n}\right)^{-n/2} \left(1 + \frac{\lambda_2}{\Delta n}\right)^{-n/2} \right\} \\
&= \mathbb{E}_{\beta, \sigma} \left\{ \left[\left(1 + \frac{p}{\Delta n}\right)^2 - \left(\frac{\beta^\top \sigma}{\Delta n}\right)^2 \right]^{-n/2} \right\}.
\end{aligned} \tag{12}$$

Let

$$\gamma_0 = \frac{4\mathbb{E}(Z_p(\Delta, X)^2)}{(\mathbb{E}(Z_p(\Delta, X)))^2}.$$

In order to apply the second moment method, we first establish the following characterization of γ_0 .

Lemma 6. *Fix $\alpha \in (0, \infty)$. There exists $\Delta_0 > 0$ such that for all $\Delta > \Delta_0$,*

$$\lim_{p \rightarrow \infty} \frac{1}{p} \ln \gamma_0 = 0.$$

Proof: Due to symmetry, $\beta^\top \sigma \stackrel{d}{=} \sum_{i=1}^p \beta_i$, equivalently, we can assume $\sigma = (1, 1, \dots, 1)$ without loss of generality. By Lemma 16, letting $Q = \beta^\top \sigma / p \in (-1, 1)$, for $t \in (0, 1)$,

$$\mathbb{P}_{\beta, \sigma}(|Q| > t) < e^{\pi - pt^2/4}.$$

Armed with this deviation bound, we derive an upper bound on $\mathbb{E}_X(Z_p(\Delta, X)^2)$. Setting $r = p/(\Delta n)$, one has

$$\begin{aligned}
\mathbb{E}(Z_p(\Delta, X)^2) &= \mathbb{E}_{\beta, \sigma} \left\{ \left[\left(1 + \frac{p}{\Delta n}\right)^2 - \left(\frac{\beta^\top \sigma}{\Delta n}\right)^2 \right]^{-n/2} \right\} \\
&= \mathbb{E}_Q \left\{ \left[(1+r)^2 - (Qr)^2 \right]^{-n/2} \right\} := \mathbb{E}_Q \left[e^{pf(Q)} \right],
\end{aligned} \tag{13}$$

where $f(t) := -(\alpha/2) \ln \left[(1+r)^2 - (tr)^2 \right]$. Let $g(t) = f(t) - t^2/4$. Note that $g'(0) = 0$ and

$$g''(t) = \frac{\alpha r^2 \left[(1+r)^2 + t^2 r^2 \right]}{\left[(1+r)^2 - t^2 r^2 \right]^2} - \frac{1}{2} < g''(1) =: -C_Q, \tag{14}$$

for any $t \in (0, 1)$. Note that $r \rightarrow 0$ for any fixed α as $\Delta \rightarrow \infty$, and thus $C_Q > 0$ for $\Delta > 0$ sufficiently large. We have $g(t) \leq -C_Q t^2/2 + g(0) = -C_Q t^2/2 - \alpha \ln(1+r)$. By (12), integration by parts gives

$$\begin{aligned}
\mathbb{E}(Z_p(\Delta, X)^2) &= - \int_0^1 e^{pf(t)} d\mathbb{P}_{\beta, \sigma}(|Q| > t) \\
&\leq - \left[e^{pf(t)} \mathbb{P}_{\beta, \sigma}(|Q| > t) \right]_{t=0}^{t=1} - \int_0^1 \left[e^{pf(t)} \right]' e^{\pi - t^2/4} dt \\
&\leq -e^{pf(0)} + e^\pi \int_0^1 \frac{p\alpha r^2 t}{(1+r)^2 - t^2 r^2} e^{pg(t)} dt \\
&\leq -e^{pf(0)} + e^\pi \frac{p\alpha r^2}{(1+r)^2 - r^2} \int_0^1 e^{pg(t)} dt \\
&\leq e^{\pi - p\alpha \ln(1+r)} \left[1 + Cp \int_0^1 e^{-pC_Q t^2/2} dt \right] \\
&\leq e^{\pi - p\alpha \ln(1+r)} (C\sqrt{n} + 1).
\end{aligned}$$

Thus

$$\frac{1}{p} \ln \gamma_0 \leq \frac{1}{p} \ln \left\{ \frac{4e^\pi (1+r)^{-n} (C\sqrt{n}+1)}{(1+r)^{-n}} \right\} \rightarrow 0, \quad (15)$$

as $p \rightarrow \infty$. \square

For the readers' convenience, we re-state [24, Lemma 4.1.1] here.

Lemma 7.

$$\mathbb{P} \left(|\Phi_p(\Delta, X) - \phi_p(\Delta)| < \frac{1}{p} \ln \gamma_0 \right) \geq \frac{1}{\gamma_0}.$$

Now we are finally ready to prove Lemma 1.

Proof of Lemma 1: Putting Lemma 4, Lemma 7, and Lemma 6 together, we have for any $\delta \in (0, 1)$, there exists a large $p_0 \in \mathbb{N}$ such that for any $p > p_0$

$$\mathbb{P} (|\Phi_p(\Delta, X) - \mathbb{E}\Phi_p(\Delta, X)| \leq \delta) + \mathbb{P} (|\Phi_p(\Delta, X) - \phi_p(\Delta)| < q_p) \geq 1 - e^{Cp} + e^{q_p} > 1,$$

where $q_p := \frac{1}{p} \ln \gamma_0 \rightarrow 0$ as $p \rightarrow \infty$. It implies

$$|\mathbb{E}\Phi_p(\Delta, X) - \phi_p(\Delta)| \leq \delta,$$

which further implies

$$\limsup_{p \rightarrow \infty} |\mathbb{E}\Phi_p(\Delta, X) - \phi_p(\Delta)| \leq \delta.$$

Since it holds for any $\delta \in (0, 1)$, proof of Lemma 1 is therefore complete. \square

Now we establish Lemma 3 based on Lemma 2.

Proof of Lemma 3: For any $u, v \in \mathbb{R}^p$, let w_1, w_2, \dots, w_p be an orthonormal basis such that $\text{Span}(u, v) = \text{Span}(w_{p-1}, w_p)$. Let $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{p-2}\}$ be the eigenvalues of Q , where Q is the top left $(p-2)$ by $(p-2)$ minor of $X^\top X$ under this basis. Namely, let $W = (w_1, w_2, \dots, w_p) \in \mathbb{R}^{p \times p}$, $X = W\tilde{X}$ and $\beta = W\tilde{\beta}$, where $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_p)$, with $\tilde{\beta}_p = \tilde{\beta}_{p-1} = 0$ for $\beta \in \text{Span}(u, v)^\perp$. Equivalently, \tilde{X} and $\tilde{\beta}$ are nothing but X and β under this new basis and $Q = (\tilde{X}^\top \tilde{X})_{1:(p-2), 1:(p-2)}$.

$$\begin{aligned} \mathbb{E}_\beta^{\text{Span}(u, v)^\perp} [\exp(-H(\beta))] &= \mathbb{E}_\beta^{\text{Span}(u, v)^\perp} \left[\exp \left(-\frac{1}{2\Delta} \beta^\top X^\top X \beta \right) \right] \\ &= \mathbb{E}_\beta^{\text{Span}(u, v)^\perp} \left[\exp \left(-\frac{1}{2\Delta} \tilde{\beta}^\top \tilde{X}^\top \tilde{X} \tilde{\beta} \right) \right] \\ &= \mathbb{E}_\beta^{\text{Span}(u, v)^\perp} \left[\exp \left(-\frac{1}{2\Delta} \tilde{\beta}_{1:(p-2)}^\top Q \tilde{\beta}_{1:(p-2)} \right) \right] \\ &= \mathbb{E}_{\tilde{\beta}_{1:(p-2)}}^{p-2} \left[\exp \left(-\frac{1}{2\Delta} \sum_{i=1}^{p-2} \lambda_i \tilde{\beta}_i^2 \right) \right] \\ &= \mathbb{E}_\beta^{p-2} \left[\exp \left(-\frac{1}{2\Delta} \sum_{i=1}^{p-2} \lambda_i \beta_i^2 \right) \right], \end{aligned} \quad (16)$$

where \mathbb{E}^{p-2} denotes expectation under $\text{Unif}(S^{(p-3)}(\sqrt{p}))$. Note that the second last line is due to the distributional invariance of $\tilde{\beta}$ under rotations, which itself is inherited from that of $\beta \sim \text{Unif}(S^{p-1}(\sqrt{p}))$ and the last line above is nothing but a change of notation. Similarly, let S be the top left $(p-2)$ by $(p-2)$ minor of $X^\top X$ when written in the standard basis $\{e_1, e_2, \dots, e_p\}$. let $s_1 \geq s_2 \geq \dots \geq s_{p-2}$ be the corresponding eigenvalues of S . For $\beta \in \mathbb{R}^{p-2}$, by Lemma 1 we have

$$\mathbb{E}_\beta^{p-2} e^{-\frac{1}{2\Delta} \beta^\top S \beta} = e^{p(-\frac{\pi}{2} \ln(1+\frac{1}{\Delta}) + o(1))}, \quad (17)$$

in which the $o(1)$ term is uniform over any $\Delta \geq \Delta_0$. At the same time

$$\mathbb{E}_\beta^{p-2} e^{-\frac{1}{2\Delta} \beta^\top S \beta} = \mathbb{E}_\beta^{p-2} e^{-\frac{1}{2\Delta} \sum_{i=1}^{p-2} s_i \beta_i^2}. \quad (18)$$

By the eigenvalue interlacing inequality [20], we have almost surely

$$\sup_{i=1,2,\dots,p} |\lambda_i - s_i| \rightarrow 0.$$

Therefore almost surely

$$\sup_{\beta \in S^{p-2}(\sqrt{p})} \frac{1}{p} \left| \sum_{i=1}^{p-2} \lambda_i \beta_i^2 - \sum_{i=1}^{p-2} s_i \beta_i^2 \right| \leq \frac{\|\beta\|^2}{p} \sup_{i=1,2,\dots,p} |\lambda_i - s_i| \rightarrow 0. \quad (19)$$

Combining (16,17,18,19), we obtain

$$\mathbb{E}^{\text{Span}(u,v)^\perp} e^{-\frac{1}{2\Delta} \beta^\top X^\top X \beta} = e^{p(-\frac{\alpha}{2} \ln(1+\frac{1}{\Delta}) + o(1))}. \quad (20)$$

Note the $o(1)$ term is uniform with respect to not only $\Delta \geq \Delta_0$ but also the choice of u and v , since the convergence in (20) is independent of u and v , which concludes the proof. \square

3 A matching upper bound on the free energy

We derive an upper bound to the TAP free energy in terms of the log-partition function in this section. We collect this bound in Theorem 4 below. To establish this result, we first introduce a model perturbation and discuss overlap concentration under the perturbed model in Section 3.1. These results are subsequently used to establish Theorem 4 in Section 3.2.

Theorem 4 (Upper bound). *Fix $\Delta > 0$ and $\alpha \in (0, \infty)$. For any $\eta > 0$ as $p \rightarrow \infty$,*

$$\mathbb{P} \left(\frac{1}{p} \ln \mathcal{Z}_p \geq \sup_{a \in \mathbb{R}^p, \|a\| \leq \sqrt{p}} f_{TAP}(a) + \eta \right) \rightarrow 0.$$

3.1 Perturbed model and overlap concentration

Assume that apart from X and y , we observe the following additional side information

$$y^{\text{Pert}} = \sqrt{\lambda_0 \epsilon_p} \beta_0 + Z,$$

with $Z \sim \mathcal{N}(0, I_p)$, $\lambda_0 \in [1/2, 1]$, $\epsilon_p \rightarrow 0$, and $p\epsilon_p \rightarrow +\infty$. We refer to (y^{Pert}, y, X) as the perturbed model. Having observed (y^{Pert}, y, X) , one constructs the posterior distribution

$$\frac{d\mathbb{P}}{d\pi_0}(\beta | y^{\text{Pert}}, y, X) = \frac{1}{\mathcal{Z}_p^{\text{Pert}}} e^{-\mathcal{H}_p^{\text{Pert}}(\beta)}, \quad (21)$$

where $\mathcal{H}_p^{\text{Pert}} := \mathcal{H}_p(\beta) + \mathcal{H}_p^{\text{Gauss}}(\beta)$ is the *Hamiltonian* of the perturbed model, and $\mathcal{H}_p^{\text{Gauss}}(\beta) := -\lambda_0 \epsilon_p \beta_0^\top \beta - \sqrt{\lambda_0 \epsilon_p} Z^\top \beta + \lambda_0 \epsilon_p \|\beta\|^2 / 2$ is the Hamiltonian of the Gaussian side channel. The normalizing constant of this posterior distribution is

$$\mathcal{Z}_p^{\text{Pert}} := \int_{S^{p-1}(\sqrt{p})} e^{-\mathcal{H}_p^{\text{Pert}}(\beta)} d\pi_0(\beta). \quad (22)$$

Finally, the free energy of the perturbed model is

$$F_p^{\text{Pert}}(\lambda_0) := \frac{1}{p} \ln \mathcal{Z}_p^{\text{Pert}}(y^{\text{Pert}}, y, X). \quad (23)$$

In the subsequent discussion, we abuse notation slightly, and use $\langle \cdot \rangle$ to also denote the expectation under the perturbed posterior (21). The notion of overlap will play a crucial role in our subsequent discussion.

Definition 1. Let β^1, β^2 be two iid samples from the posterior distribution (21). We refer to $R_{12} := (\beta^1)^\top \beta^2 / p$ as the overlap between two replicas.

Theorem 5. For any $\Delta > 0$ and $\alpha \in (0, \infty)$, as $p \rightarrow \infty$,

$$\frac{1}{p} |\mathbb{E} \ln \mathcal{Z}_p^{\text{Pert}} - \mathbb{E} \ln \mathcal{Z}_p| \leq O(\epsilon_p).$$

Furthermore, assuming $\lambda_0 \sim \text{Unif}(\frac{1}{2}, 1)$, there exist $C_1, C_2 > 0$ such that

$$\mathbb{E}_{\lambda_0} \mathbb{E} \langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle \leq \frac{C_1}{\epsilon_p} \left(\frac{v_p}{p \epsilon_p} + \frac{1}{p} \right)^{1/3}, \quad (24)$$

where $v_p := p \sup_{\lambda_0 \in [\frac{1}{2}, 1]} \{ \mathbb{E} (F_p^{\text{Pert}}(\lambda_0) - \mathbb{E} F_p^{\text{Pert}}(\lambda_0))^2 \}$ and $v_p < C_2$ for any $p \geq 1$.

In our application of Theorem 5 we will choose $\epsilon_p \rightarrow 0$ slow enough such that the RHS in (24) converges to zero as $p \rightarrow \infty$. Our next result characterizes the limiting behavior of $\mathbb{E}[\langle R_{1,2} \rangle]$.

Lemma 8. For any $\Delta, \alpha > 0$, as $p \rightarrow \infty$, there exists $C > 0$ depending on α and Δ such that

$$1 - C \geq \limsup_{p \rightarrow \infty} \mathbb{E}[\langle R_{1,2} \rangle] \geq \liminf_{p \rightarrow \infty} \mathbb{E}[\langle R_{1,2} \rangle] \geq 0.$$

Note that $\mathbb{E}[\langle R_{1,2} \rangle] \in [-1, 1]$ —thus any subsequence has a further subsequence such that $\mathbb{E}[\langle R_{1,2} \rangle]$ converges to a limit along the subsequence. Further, Lemma 8 implies that this subsequential limit lies in $[0, 1 - C]$. Armed with these results, we relate the log-partition function to that of a “replicated” system, where the overlaps have been restricted to have a specific value.

Lemma 9. For $m \geq 1$ and $c > 0$, define

$$\mathcal{S}_p(m, c) = \left\{ (\beta^1, \beta^2, \dots, \beta^m) : \left| \frac{1}{p} (\beta^i)^\top \beta^j - \mathbb{E} \langle R_{1,2} \rangle \right| \leq c, \forall i \neq j \right\}. \quad (25)$$

There exists $c_p \rightarrow 0$ and $m_p \rightarrow \infty$, such that

$$\frac{1}{p m_p} \ln \left(\int e^{-\sum_i \mathcal{H}(\beta^i)} 1_{\{(\beta^1, \beta^2, \dots, \beta^{m_p}) \in \mathcal{S}\}} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) = \frac{1}{p} \ln \mathcal{Z}_p + o(1).$$

In the equation above and henceforth, we suppress the dependence of \mathcal{S} on p, c_p and m_p for the sake of notational simplicity.

Proof: Using Theorem 5, we can choose a sequence $\{\lambda_0^p : p \geq 1\}$ such that for any $p \geq 1$ we have

$$\mathbb{E} \langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle \leq a_p,$$

where $a_p \rightarrow 0$ as $p \rightarrow \infty$. For any positive sequence $\{b_p : p \geq 1\}$, Markov inequality yields

$$\mathbb{P} \left(\langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle > b_p \right) \leq \frac{a_p}{b_p}.$$

We choose $\{b_p : p \geq 1\}$ such that $b_p \rightarrow 0, a_p/b_p \rightarrow 0$. Let $m_p \geq 1$ (to be chosen later) and let $\beta^1, \beta^2, \dots, \beta^{m_p}$ be iid samples from the posterior distribution (21). Define $R_{i,j} = (\beta^i)^\top \beta^j / p$ for $i, j \in \{1, 2, \dots, m_p\}$ to be the overlaps among the samples. By symmetry, we have $\langle R_{1,2} \rangle = \langle R_{i,j} \rangle$ and $\mathbb{E} \langle R_{1,2} \rangle = \mathbb{E} \langle R_{i,j} \rangle$ for any $i \neq j$. Define the event $\mathcal{C}_p := \{(y^{\text{Pert}}, y, X) : \langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle > b_p\}$. Note that $\mathbb{P}(\mathcal{C}_p) \leq a_p/b_p \rightarrow 0$. On \mathcal{C}_p^c we have

$$\langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle \leq b_p.$$

Using Chebychev inequality, we have, on \mathcal{C}_p^c

$$\langle \mathbf{1}_{\{|R_{1,2} - \mathbb{E}\langle R_{1,2} \rangle| > c_p\}} \rangle \leq \frac{b_p}{c_p^2}.$$

By union bound, on \mathcal{C}_p^c

$$\langle \mathbf{1}_{S^c} \rangle \leq \sum_{i \neq j} \langle \mathbf{1}_{\{|R_{i,j} - \mathbb{E}\langle R_{i,j} \rangle| > c_p\}} \rangle = \sum_{i \neq j} \langle \mathbf{1}_{\{|R_{1,2} - \mathbb{E}\langle R_{1,2} \rangle| > c_p\}} \rangle \leq \frac{m_p^2 b_p}{c_p^2}.$$

We choose $m_p \rightarrow \infty$ and $c_p \rightarrow 0$ such that the RHS above will converge to 0 as $p \rightarrow \infty$. This implies, on \mathcal{C}_p^c ,

$$1 \geq \langle \mathbf{1}_S \rangle = \frac{1}{(\mathcal{Z}_p^{\text{Pert}})^{m_p}} \int e^{-\sum_i \mathcal{H}^{\text{Pert}}(\beta^i)} \mathbf{1}_{\{(\beta^1, \beta^2, \dots, \beta^{m_p}) \in \mathcal{S}\}} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \geq 1 - \frac{m_p^2 b_p}{c_p^2} = 1 - o(1). \quad (26)$$

Since $\lim_{p \rightarrow \infty} \mathbb{P}_Z(\|Z\| < 2\sqrt{p}) = 1$, we have with high probability with respect to Z

$$\begin{aligned} |\mathcal{H}(\beta) - \mathcal{H}^{\text{Pert}}(\beta)| &= |\mathcal{H}^{\text{Gauss}}(\beta)| \\ &\leq |\lambda_0 \epsilon_p \beta_0^\top \beta| + \left| \sqrt{\lambda_0 \epsilon_p} Z^\top \beta \right| + \left| \frac{1}{2} \lambda_0 \epsilon_p \|\beta\|^2 \right| \\ &\leq O(\sqrt{\epsilon_p}) \lambda_0 p. \end{aligned}$$

This implies that with high probability with respect to Z ,

$$|\ln \mathcal{Z}_p^{\text{Pert}} - \ln \mathcal{Z}_p| \leq O(\sqrt{\epsilon_p}) \lambda_0 p, \quad (27)$$

and

$$\left| \frac{1}{pm_p} \ln \left(\frac{\int e^{-\sum_i \mathcal{H}^{\text{Pert}}(\beta^i)} \mathbf{1}_{\{(\beta^1, \dots, \beta^{m_p}) \in \mathcal{S}\}} \prod_{i=1}^{m_p} d\pi_0(\beta^i)}{\int e^{-\sum_i \mathcal{H}(\beta^i)} \mathbf{1}_{\{(\beta^1, \dots, \beta^{m_p}) \in \mathcal{S}\}} \prod_{i=1}^{m_p} d\pi_0(\beta^i)} \right) \right| \leq O(\sqrt{\epsilon_p}). \quad (28)$$

Putting (26, 27, 28) together, with high probability we have

$$\frac{1}{pm_p} \ln \left(\int e^{-\sum_i \mathcal{H}(\beta^i)} \mathbf{1}_{\{(\beta^1, \beta^2, \dots, \beta^{m_p}) \in \mathcal{S}\}} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) = \frac{1}{p} \ln \mathcal{Z}_p + O(\sqrt{\epsilon_p}).$$

Recalling $\epsilon_p \rightarrow 0$ as $p \rightarrow \infty$, the claim follows. \square

3.2 Proof of Theorem 4

We prove Theorem 4 in this section. To this end, recall that $\mathcal{H}(\beta) = -\frac{1}{2\Delta} \|y - X\beta\|^2$. Let $\beta^1, \dots, \beta^{m_p}$ be iid samples from the posterior distribution, and set $a = \sum_{i=1}^{m_p} \beta^i / m_p$ to be the sample average. This implies

$$\begin{aligned} \frac{1}{m_p} \sum_{i=1}^{m_p} \|y - X\beta^i\|^2 &= \|y - Xa\|^2 + \frac{1}{m_p} \sum_{i=1}^{m_p} \|X(\beta^i - a)\|^2 - \frac{2}{m_p} (X^\top (y - Xa))^\top \sum_{i=1}^{m_p} (\beta^i - a) \\ &= \|y - Xa\|^2 + \frac{1}{m_p} \sum_{i=1}^{m_p} \|X(\beta^i - a)\|^2, \end{aligned}$$

since $\sum_{i=1}^{m_p} (\beta^i - a) = 0$. Thus with high probability,

$$\frac{1}{p} \ln \mathcal{Z}_p = \frac{1}{pm_p} \ln \left(\int e^{-\frac{m_p}{2\Delta} \|y - Xa\|^2 - \frac{1}{2\Delta} \sum_{i=1}^{m_p} \|X(\beta^i - a)\|^2} \mathbf{1}_{\{(\beta^1, \beta^2, \dots, \beta^{m_p}) \in \mathcal{S}\}} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) + o(1)$$

$$\begin{aligned}
&\leq \frac{1}{pm_p} \ln \left(\int e^{-\frac{m_p}{2\Delta} \|y-Xa\|^2 - \frac{1}{2\Delta} \sum_{i=2}^{m_p} \|X(\beta^i - a)\|^2} 1_{\{(\beta^1, \beta^2, \dots, \beta^{m_p}) \in \mathcal{S}\}} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) + o(1) \\
&:= R_p + o(1).
\end{aligned} \tag{29}$$

Our next result, Lemma 10 will be critical for the proof of Theorem 4. We defer its proof to the next subsection.

Lemma 10. *Fix $\Delta > 0$, $\alpha \in (0, \infty)$ and any $\eta > 0$. For any subsequence $\{p_k : k \geq 1\}$ such that $\mathbb{E}[\langle R_{1,2} \rangle]$ converges to a positive constant, there exist $c' > 0$ and $c > 0$, $c + c' < 1$, such that as $k \rightarrow \infty$,*

$$\mathbb{P} \left(R_{p_k} \leq \sup_{a \in \mathbb{R}^{p_k}, c' \sqrt{p_k} \leq \|a\| \leq (1-c) \sqrt{p_k}} f_{\text{TAP}}(a) + \eta \right) \rightarrow 1.$$

Before proceeding further, we prove Theorem 4 assuming Lemma 10.

Proof of Theorem 4: Assume, if possible, that there exists a subsequence p_k and $\epsilon, \eta > 0$ such that for all $k \geq 1$,

$$\mathbb{P} \left(\frac{1}{p_k} \ln \mathcal{Z}_{p_k} \geq \sup_{a \in \mathbb{R}^{p_k}, \|a\| \leq \sqrt{p_k}} f_{\text{TAP}}(a) + \eta \right) \geq \epsilon.$$

We can extract a further subsequence such that $\mathbb{E}[\langle R_{1,2} \rangle]$ converges along the subsequence. For notational simplicity, we refer to this subsequence as $\{p_k : k \geq 1\}$ as well.

Consider first the case $\lim \mathbb{E}[\langle R_{1,2} \rangle] > 0$. Combining (29) and Lemma 10, we obtain that along this subsequence,

$$\mathbb{P} \left(\frac{1}{p_k} \ln \mathcal{Z}_{p_k} \geq \sup_{a \in \mathbb{R}^{p_k}, \|a\| \leq \sqrt{p_k}} f_{\text{TAP}}(a) + \eta \right) \rightarrow 0.$$

This is a contradiction, and completes the proof in this case.

We now turn to the case where $\lim \mathbb{E}[\langle R_{1,2} \rangle] = 0$. In this case,

$$\sup_{a \in \mathbb{R}^{p_k}, \|a\| \leq \sqrt{p_k}} f_{\text{TAP}}(a) \geq f_{\text{TAP}}(0) = -\frac{1}{2\Delta p_k} \|y\|^2 - \frac{\alpha}{2} \ln \left(1 + \frac{1}{\Delta \alpha} \right). \tag{30}$$

On the other hand, setting $a = \sum_i \beta^i / m_p$, we note that for $(\beta^1, \dots, \beta^{m_p}) \in \mathcal{S}$, $\|a\|^2 / p \leq 1 / (m_p^2) + c_p$, where we use that $\lim \mathbb{E}[\langle R_{1,2} \rangle] = 0$. Further, using (29), we have

$$\begin{aligned}
\frac{1}{p_k} \ln \mathcal{Z}_{p_k} &\leq \frac{1}{p_k m_{p_k}} \ln \left(\int e^{-\frac{m_{p_k}}{2\Delta} \|y-Xa\|^2 - \frac{1}{2\Delta} \sum_{i=2}^{m_{p_k}} \|X(\beta^i - a)\|^2} 1_{\{(\beta^1, \beta^2, \dots, \beta^{m_{p_k}}) \in \mathcal{S}\}} \prod_{i=1}^{m_{p_k}} d\pi_0(\beta^i) \right) + o(1) \\
&\leq \frac{1}{p_k m_{p_k}} \ln \left(\int e^{-\frac{m_{p_k}}{2\Delta} \|y\|^2 - \frac{1}{2\Delta} \sum_{i=2}^{m_{p_k}} \|X\beta^i\|^2} 1_{\{(\beta^1, \beta^2, \dots, \beta^{m_{p_k}}) \in \mathcal{S}\}} \prod_{i=1}^{m_{p_k}} d\pi_0(\beta^i) \right) + o(1),
\end{aligned}$$

where we have used the fact that $\|a\| = o(\sqrt{p})$. In turn, this implies

$$\begin{aligned}
\frac{1}{p_k} \ln \mathcal{Z}_{p_k} &\leq -\frac{\|y\|^2}{2\Delta p} + \frac{m_{p_k} - 1}{p_k m_{p_k}} \ln \left(\int \exp\left(-\frac{1}{2\Delta} \|X\beta\|^2\right) d\pi_0(\beta) \right) + o(1) \\
&\leq -\frac{\|y\|^2}{2\Delta p_k} - \frac{\alpha}{2} \ln \left(1 + \frac{1}{\Delta \alpha} \right) + o(1),
\end{aligned}$$

where the last step uses Lemma 1. Combining this bound with (30) we immediately obtain that along this subsequence,

$$\mathbb{P}\left(\frac{1}{p_k} \ln \mathcal{Z}_{p_k} \geq \sup_{a \in \mathbb{R}^{p_k}, \|a\| \leq \sqrt{p_k}} f_{\text{TAP}}(a) + \eta\right) \rightarrow 0$$

which is a contradiction. This completes the proof. \square

It remains to prove Lemma 10. To this end, we will assume without loss of generality that as $p \rightarrow \infty$

$$\mathbb{E}[\langle R_{1,2} \rangle] \rightarrow C_0 > 0. \quad (31)$$

It will be useful to express the uniform distribution on $S^{p-1}(\sqrt{p})$ in terms of the gaussian measure. Specifically, let $g^1, g^2, \dots, g^{m_p} \sim \mathcal{N}(0, I_p)$ be iid Gaussian vectors and set $\beta^i = g^i \sqrt{p} / \|g^i\|_2$. The rotational symmetry of the gaussian measure ensures that the β^i vectors are iid uniform on $S^{p-1}(\sqrt{p})$. Set $b = \sum_{i=1}^{m_p} g^i / m_p$ to be sample mean of the g^i vectors, and let $\tilde{b} := b \sqrt{p} / \|b\|$ be the scaled version of b with radius \sqrt{p} . We introduce a counterpart of \mathcal{S} for the g^i -vectors:

$$\tilde{\mathcal{S}}_\epsilon := \left\{ (g^1, g^2, \dots, g^{m_p}) : \left(\sqrt{p} \frac{g^1}{\|g^1\|}, \sqrt{p} \frac{g^2}{\|g^2\|}, \dots, \sqrt{p} \frac{g^{m_p}}{\|g^{m_p}\|} \right) \in \mathcal{S}, \|g^i\| \in (\sqrt{p}(1-\epsilon), \sqrt{p}(1+\epsilon)), \forall i \right\}. \quad (32)$$

We first collect some geometric properties of the configurations in \mathcal{S} and $\tilde{\mathcal{S}}_\epsilon$. The following lemma shows that, $\beta^i - a$ and a are approximately orthogonal on the set \mathcal{S} . Its proof follows directly from the definition of \mathcal{S} and Lemma 8.

Lemma 11. *There exists $C > 0$ depending only on Δ and α such that for any $(\beta^1, \beta^2, \dots, \beta^{m_p}) \in \mathcal{S}$, we have for any $i \in \{1, 2, \dots, m_p\}$,*

$$|(\beta^i - a)^\top \tilde{a}| \leq C p c_p,$$

where $a = \sum_{i=1}^{m_p} \beta^i / m_p$ and $\tilde{a} := a \sqrt{p} / \|a\|$ is a normalized version of a with radius \sqrt{p} .

We also establish an analogous property for the configurations in $\tilde{\mathcal{S}}_\epsilon$ for small $\epsilon > 0$.

Lemma 12. *There exist $C_1, C_2, C_3 > 0$ such that if $\epsilon \in (0, C_1)$ and $g = (g^1, g^2, \dots, g^{m_p}) \in \tilde{\mathcal{S}}_\epsilon$, then*

$$|(g^i - b)^\top \tilde{b}| \leq (C_2 c_p + C_3 \epsilon) p.$$

Proof: Note that, for any $g \in \tilde{\mathcal{S}}_\epsilon$

$$\|b - a\| = \frac{1}{m_p} \left\| \sum_{i=1}^{m_p} (1 - \sqrt{p} / \|g^i\|) g^i \right\| \leq \frac{1}{m_p} \sum_{i=1}^{m_p} \|(1 - \sqrt{p} / \|g^i\|) g^i\| = O(\epsilon \sqrt{p}). \quad (33)$$

Similarly we also have $\|g^i - \beta^i\| = O(\epsilon \sqrt{p})$ on $\tilde{\mathcal{S}}_\epsilon$. By Lemma 11 and triangle inequality,

$$\begin{aligned} |(g^i - b)^\top \tilde{b}| &\leq |(g^i - b)^\top a| + |(g^i - b)^\top (b - a)| \\ &\leq |(g^i - \beta^i)^\top a| + |(\beta^i - a)^\top a| + |(a - b)^\top a| + \|g^i - b\| \|b - a\| \\ &\leq \|a\| (\|g^i - \beta^i\| + \|a - b\| + \|\beta^i - a\|) + O(\epsilon p) \\ &\leq O(\epsilon p) + 2p c_p \\ &= (2c_p + O(\epsilon)) p. \end{aligned}$$

Moreover, note that Lemma 8 together with (31) imply there is some positive constant C such that if $(\beta^1, \dots, \beta^{m_p}) \in \mathcal{S}$ then $(1 - C) \sqrt{p} \geq \|a\| \geq C \sqrt{p}$. Thus, for small enough ϵ , there always exists positive constant C' such that $(1 - C') \geq \sqrt{p} \|b\| \geq C' \sqrt{p}$, which finishes the proof. \square

Let $w \sim \mathcal{N}(0, I_p/m_p)$, independent of $\{g^i : 1 \leq i \leq m_p\}$. Let $h^i = g^i + w$ and $q^i = h^i - b$. Observe that $b, q^2, q^3, \dots, q^{m_p}$ are independent. Moreover, conditioning on $b, h_2, h_3, \dots, h_{m_p}$ are conditionally independent. The following lemma shows that adding a small perturbation w has a negligible effect.

Lemma 13. *There exists $C > 0$ such that with high probability (with respect to X), as $p \rightarrow \infty$*

$$\mathbb{P}_{g,w} \left(\left| \sum_{i=2}^{m_p} (\|X(g^i - b)\|^2 - \|X(g^i - b + w)\|^2) \right| \leq C m_p^{1/2} p \right) \rightarrow 1.$$

Proof: We start with

$$\begin{aligned} \left| \sum_{i=2}^{m_p} (\|X(g^i - b)\|^2 - \|X(g^i - b + w)\|^2) \right| &\leq (m_p - 1) \|Xw\|^2 + 2 \sum_{i=2}^{m_p} |w^\top X^\top X (g^i - b)| \\ &\leq (m_p - 1) \|X^\top X\|_2 \|w\|^2 + 2 \sum_{i=2}^{m_p} \|X^\top X\|_2 \|w\| \|g^i - b\|. \end{aligned}$$

Since $\|w\|^2 = O(p/m_p)$, $\|g^i - b\| = O(\sqrt{p})$, and $\|X^\top X\|_2 = O(1)$ with high probability, we have as $p \rightarrow \infty$

$$\mathbb{P}_{g,w} ((m_p - 1) \|X^\top X\|_2^2 \|w\|^2 \leq Cp) \rightarrow 1,$$

and

$$\mathbb{P}_{g,w} \left(\left| 2 \sum_{i=2}^{m_p} \|X^\top X\|_2 \|w\| \|g^i - b\| \right| \leq C m_p^{1/2} p \right) \rightarrow 1.$$

The lemma follows upon combining these ingredients. \square

3.3 Proof of Lemma 10

Proof of Lemma 10: Throughout this proof, we assume without loss of generality that $\mathbb{E}[\langle R_{1,2} \rangle]$ converges to a positive constant. Fix any $\epsilon \in (0, C_1)$, where C_1 is defined in the statement of Lemma 12. Recall the sets \mathcal{S} and $\tilde{\mathcal{S}}_\epsilon$ from (25) and (32) respectively, and R_p from (29). Since $\{\sqrt{p}g^i/\|g^i\| : 1 \leq i \leq m_p\}$ and $\{\|g^i\| : 1 \leq i \leq m_p\}$ are independent, we have

$$\begin{aligned} R_p &= \frac{1}{pm_p} \ln \left(\mathbb{E}_{g^1, g^2, \dots, g^{m_p}} \left[e^{-G(g)} \mathbf{1}_{\{(g^1/\|g^1\|, g^2/\|g^2\|, \dots, g^{m_p}/\|g^{m_p}\|) \in \mathcal{S}\}} \right] \right) \\ &= \frac{1}{pm_p} \ln \left\{ \mathbb{E}_g \left[e^{-G(g)} \mathbf{1}_{\{(g^1/\|g^1\|, g^2/\|g^2\|, \dots, g^{m_p}/\|g^{m_p}\|) \in \mathcal{S}\}} \mathbf{1}_{\{\|g^i\| \in (\sqrt{p}(1-\epsilon), \sqrt{p}(1+\epsilon)), \forall i\}} \right] \right. \\ &\quad \left. \cdot \mathbb{P}^{-1}(\|g^i\| \in (\sqrt{p}(1-\epsilon), \sqrt{p}(1+\epsilon)), \forall i) \right\} \\ &= \frac{1}{pm_p} \ln \left\{ \mathbb{E}_g \left[e^{-G(g)} \mathbf{1}_{\{g \in \tilde{\mathcal{S}}_\epsilon\}} \right] \right\} + O(\epsilon^2), \end{aligned} \tag{34}$$

where

$$G(g) = G(g^1, g^2, \dots, g^{m_p}) := \frac{m_p}{2\Delta} \left\| y - X \left(\frac{1}{m_p} \sum_{i=1}^{m_p} \frac{g^i \sqrt{p}}{\|g^i\|} \right) \right\|^2 - \frac{1}{2\Delta} \sum_{i=2}^{m_p} \left\| X \left(\frac{g^i \sqrt{p}}{\|g^i\|} - \frac{1}{m_p} \sum_{i=1}^{m_p} \frac{g^i \sqrt{p}}{\|g^i\|} \right) \right\|^2.$$

Define

$$\tilde{G}(g) = \tilde{G}(g^1, g^2, \dots, g^{m_p}) := \frac{m_p}{2\Delta} \|y - Xb\|^2 - \frac{1}{2\Delta} \sum_{i=2}^{m_p} \|X(g^i - b)\|^2.$$

Note that, for any $g \in \tilde{\mathcal{S}}_\epsilon$, by (33), we have $\|b - a\| = O(\epsilon\sqrt{p})$ and $\|g^i - \beta^i\| = O(\epsilon\sqrt{p})$ for any i . Since $\mathbb{P}(\Omega) \rightarrow 1$ as $p \rightarrow \infty$, with high probability

$$\begin{aligned}
|G(g) - \tilde{G}(g)| &\leq \frac{m_p}{2\Delta} \left| \|y - Xb\|^2 - \|y - Xa\|^2 \right| + \frac{1}{2\Delta} \sum_{i=2}^{m_p} \left| \|X(g^i - b)\|^2 - \|X(\beta^i - a)\|^2 \right| \\
&\leq \frac{m_p}{2\Delta} |y^\top X(b - a)| + \frac{m_p}{2\Delta} \left| \|Xb\|^2 - \|Xa\|^2 \right| + \frac{1}{2\Delta} \sum_{i=2}^{m_p} \left\{ |(g^i - b)^\top X^\top X(g^i - b - \beta^i + a)| \right. \\
&\quad \left. + |(\beta^i - a)^\top X^\top X(g^i - b - \beta^i + a)| \right\} \\
&\stackrel{(i)}{\leq} O(\sqrt{pm_p}) \|b - a\| + \frac{1}{2\Delta} \sum_{i=2}^{m_p} \left\{ \|X^\top X\|_2 [\|g^i - b\|(\|g^i - \beta^i\| + \|b - a\|) \right. \\
&\quad \left. + \|\beta^i - a\|(\|g^i - \beta^i\| + \|b - a\|)] \right\} \\
&= O(\sqrt{pm_p}) \|b - a\| + O(\sqrt{pm_p}) \|g^i - \beta^i\| \\
&= O(m_p p \epsilon),
\end{aligned}$$

where (i) follows from the fact that $\|y\| = O(1)$ and $\|X\|_2 = O(1)$ on Ω . This allows us to replace $G(g)$ in (34) by $\tilde{G}(g)$ with an additional cost $O(\epsilon)$. Thus with high probability,

$$R_p = \frac{1}{pm_p} \ln \left(\mathbb{E}_g \left[e^{-\tilde{G}(g)} 1_{\{g \in \tilde{\mathcal{S}}_\epsilon\}} \right] \right) + O(\epsilon). \quad (35)$$

Let $\delta_p := c_p/C_2 \rightarrow 0$, where C_2 is a positive constant defined in the statement of Lemma 12. Recalling the RHS of the last equation, by Lemma 12, the definition of $\tilde{\mathcal{S}}_\epsilon$ and (33), we know that there exist $c', c > 0$ such that $\{g = (g^1, g^2, \dots, g^{m_p}) : |\langle g^i - b, \tilde{b} \rangle| \leq (C_2 c_p + C_3 \epsilon)p, \forall i\} \cap \{\|g^i\| \in ((1 - \epsilon)\sqrt{p}, (1 + \epsilon)\sqrt{p}), \forall i\} \cap \{\|b\|/\sqrt{p} : (c', 1 - c)\}$ contains $\tilde{\mathcal{S}}_\epsilon$, thus

$$\begin{aligned}
&\mathbb{E}_g \left[e^{-\tilde{G}(g)} 1_{\{g \in \tilde{\mathcal{S}}_\epsilon\}} \right] \\
&\leq \mathbb{E}_b \left[1_{\{\frac{\|b\|}{\sqrt{p}} \in (c', 1 - c)\}} \mathbb{E}_{g^2, g^3, \dots, g^{m_p} | b} \left(e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} 1_{\{|\langle g^i - b, \tilde{b} \rangle| \leq (C\delta_p + O(\epsilon))p, \|g^i\| \in ((1 - \epsilon)\sqrt{p}, (1 + \epsilon)\sqrt{p})\}} \right) \right] \\
&= \mathbb{E}_b \left[1_{\{\frac{\|b\|}{\sqrt{p}} \in (c', 1 - c)\}} \mathbb{E}_{g^2, g^3, \dots, g^{m_p}, w | b} \left(e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} 1_{\{|\langle g^i - b, \tilde{b} \rangle| \leq (C\delta_p + O(\epsilon))p, \|g^i\| \in ((1 - \epsilon)\sqrt{p}, (1 + \epsilon)\sqrt{p})\}} 1_{\{\|w\| \leq 2\sqrt{\frac{p}{m_p}}\}} \right) \right] \\
&\quad \cdot \mathbb{P} \left(\|w\| \leq 2\sqrt{\frac{p}{m_p}} \right)^{-1}, \quad (36)
\end{aligned}$$

where the last equation follows from the independence of w and g . Recall that $h^i = g^i + w$ and $q^i = h^i - b$.

We thus have

$$\begin{aligned}
& \mathbb{E}_b \left[\mathbf{1}_{\{\frac{\|b\|}{\sqrt{p}} \in (c', 1-c)\}} \mathbb{E}_{g^2, g^3, \dots, g^{m_p}, w|b} \left(e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} \mathbf{1}_{\{|\langle g^i - b, \tilde{b} \rangle| \leq (C\delta_p + O(\epsilon))p, \|g^i\| \in ((1-\epsilon)\sqrt{p}, (1+\epsilon)\sqrt{p})\}} \mathbf{1}_{\{\|w\| \leq 2\sqrt{\frac{p}{m_p}}\}} \right) \right] \\
& \leq \mathbb{E}_b \left[\mathbf{1}_{\{\frac{\|b\|}{\sqrt{p}} \in (c', 1-c)\}} \mathbb{E}_{h^2, h^3, \dots, h^{m_p}|b} \left(e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} \mathbf{1}_{\{|\langle h^i - b, \tilde{b} \rangle| \leq (C\delta_p + O(\epsilon) + \frac{2}{\sqrt{m_p}})p, \|h^i\| \in ((1-\epsilon - \frac{2}{\sqrt{m_p}})\sqrt{p}, (1+\epsilon + \frac{2}{\sqrt{m_p}})\sqrt{p})\}} \right) \right] \\
& = \mathbb{E}_b \left[\mathbf{1}_{\{\frac{\|b\|}{\sqrt{p}} \in (c', 1-c)\}} e^{-\frac{m_p}{2\Delta} \|y - Xb\|^2} \times \right. \\
& \quad \left. \mathbb{E}_{h^2, h^3, \dots, h^{m_p}|b} \left(e^{-\frac{1}{2\Delta} \sum_{i=2}^{m_p} \|X(h^i - b)\|^2} \prod_{i=2}^{m_p} \mathbf{1}_{\{|\langle h^i - b, \tilde{b} \rangle| \leq (C\delta_p + O(\epsilon) + \frac{2}{\sqrt{m_p}})p, \|h^i\| - \sqrt{p} \leq (\epsilon + \frac{2}{\sqrt{m_p}})\sqrt{p}\}} \right) \right] \\
& \stackrel{(i)}{=} \mathbb{E}_b \left\{ \mathbf{1}_{\{\frac{\|b\|}{\sqrt{p}} \in (c', 1-c)\}} e^{-\frac{m_p}{2\Delta} \|y - Xb\|^2} \times \right. \\
& \quad \left. \left[\mathbb{E}_{h^2|b} \left(e^{-\frac{1}{2\Delta} \|X(h^2 - b)\|^2} \mathbf{1}_{\{|\langle h^2 - b, \tilde{b} \rangle| \leq (C\delta_p + O(\epsilon) + \frac{2}{\sqrt{m_p}})p, \|h^2\| - \sqrt{p} \leq (\epsilon + \frac{2}{\sqrt{m_p}})\sqrt{p}\}} \right) \right]^{m_p-1} \right\} \\
& \leq \mathbb{E}_b \left\{ \mathbf{1}_{\{\frac{\|b\|}{\sqrt{p}} \in (c', 1-c)\}} e^{-\frac{m_p}{2\Delta} \|y - Xb\|^2} \times \right. \\
& \quad \left. \left[\mathbb{E}_{q^2} \left(e^{-\frac{1}{2\Delta} \|Xq^2\|^2} \mathbf{1}_{\{|\langle q^2, \tilde{b} \rangle| \leq (C\delta_p + O(\epsilon) + \frac{2}{\sqrt{m_p}})p, \|q^2\|^2 - (p - \|b\|^2) \leq (C\delta_p + O(\epsilon) + \frac{2}{\sqrt{m_p}})p\}} \right) \right]^{m_p-1} \right\}. \tag{37}
\end{aligned}$$

where (i) follows from the fact that h^i 's are identically distributed and independent conditioned on b . Note that $\epsilon > 0$ is a fixed positive constant. Since $\delta_p \rightarrow 0$ and $m_p \rightarrow \infty$ as $p \rightarrow \infty$, for p large enough we can rewrite the RHS above as

$$\begin{aligned}
& \mathbb{E}_b \left\{ \mathbf{1}_{\{\|b\|/\sqrt{p} \in (c', 1-c)\}} e^{-\frac{m_p}{2\Delta} \|y - Xb\|^2} \left[\mathbb{E}_{q^2} \left(e^{-\frac{1}{2\Delta} \|Xq^2\|^2} \mathbf{1}_{\{|\langle q^2, \tilde{b} \rangle| \leq O(\epsilon)p, \|q^2\|^2 \in (p - \|b\|^2 \pm O(\epsilon)p)\}} \right) \right]^{m_p-1} \right\} \\
& \leq \sup_{b: \|b\|/\sqrt{p} \in (c', 1-c)} \left\{ e^{-\frac{m_p}{2\Delta} \|y - Xb\|^2} \left[\mathbb{E}_{q^2} \left(e^{-\frac{1}{2\Delta} \|Xq^2\|^2} \mathbf{1}_{\{|\langle q^2, \tilde{b} \rangle| \leq O(\epsilon)p, \|q^2\|^2 \in (p - \|b\|^2 \pm O(\epsilon)p)\}} \right) \right]^{m_p-1} \right\}, \tag{38}
\end{aligned}$$

where the last line is obtained by Laplace Method. Note that $\|q^2\|$ and $q^2/\|q^2\|$ are independent. Defining $\chi = \|q^2\|$, we further simplify the quantity inside the square bracket above

$$\begin{aligned}
& \mathbb{E}_{q^2} \left(e^{-\frac{1}{2\Delta} \|Xq^2\|^2} \mathbf{1}_{\{|\langle q^2, \tilde{b} \rangle| \leq O(\epsilon)p, \|q^2\|^2 \in (p - \|b\|^2 \pm O(\epsilon)p)\}} \right) \\
& = \mathbb{E}_{\beta \sim \text{Unif}(S^{p-1}(\sqrt{p})), \chi \sim \chi(p)} \left(e^{-\frac{\chi^2/p}{2\Delta} \|X\beta\|^2} \mathbf{1}_{\{|\frac{\chi}{\sqrt{p}} \langle \beta, \tilde{b} \rangle| \leq O(\epsilon)p\}} \mathbf{1}_{\{\chi^2 \in (p - \|b\|^2 \pm O(\epsilon)p)\}} \right) \\
& \leq \mathbb{E}_{\beta \sim \text{Unif}(S^{p-1}(\sqrt{p})), \chi \sim \chi(p)} \left(e^{-\frac{\chi^2/p}{2\Delta} \|X\beta\|^2} \mathbf{1}_{\{|\langle \beta, \tilde{b} \rangle| \leq O(\epsilon)p\}} \mathbf{1}_{\{\chi^2 \in (p - \|b\|^2 \pm O(\epsilon)p)\}} \right) \\
& \leq \mathbb{E}_{\chi \sim \chi(p)} \left[\mathbb{E}_{\beta \sim \text{Unif}(S^{p-1}(\sqrt{p}))} \left(e^{-\frac{\chi^2/p}{2\Delta} \|X\beta\|^2} \mathbf{1}_{\{|\langle \beta, \tilde{b} \rangle| \leq O(\epsilon)p\}} \right) \mathbf{1}_{\{\chi^2 \in (p - \|b\|^2 \pm O(\epsilon)p)\}} \right] := \kappa_p(b). \tag{39}
\end{aligned}$$

Finally, combining (36, 37, 38, 39), we have

$$\begin{aligned}
& \frac{1}{pm_p} \ln \left(\mathbb{E}_g \left[e^{-\tilde{G}(g)} \mathbf{1}_{\{g \in \tilde{\mathcal{S}}_\epsilon\}} \right] \right) \\
& \leq \frac{1}{pm_p} \ln \left(\sup_{b \in \mathbb{R}^p, c' \sqrt{p} \leq \|b\| \leq (1-c)\sqrt{p}} \left\{ e^{-\frac{m_p}{2\Delta} \|y - Xb\|^2} \kappa_p^{m_p-1}(b) \right\} \right) - \frac{1}{pm_p} \ln \mathbb{P} \left(\|w\| \leq 2\sqrt{\frac{p}{m_p}} \right)^{-1} \\
& = \sup_{b \in \mathbb{R}^p, c' \sqrt{p} \leq \|b\| \leq (1-c)\sqrt{p}} \left\{ \frac{-1}{2p\Delta} \|y - Xb\|^2 + \frac{m_p - 1}{pm_p} \ln \kappa_p(b) \right\} + o(1) \\
& \leq \sup_{b \in \mathbb{R}^p, c' \sqrt{p} \leq \|b\| \leq (1-c)\sqrt{p}} \left\{ \frac{-1}{2p\Delta} \|y - Xb\|^2 + \frac{1}{p} \ln \left[\mathbb{E}_\beta \left(e^{-\frac{(p - \|b\|^2 - C\epsilon p)/p}{2\Delta} \|X\beta\|^2} \mathbf{1}_{\{|\langle \beta, \tilde{b} \rangle| \leq C\epsilon p\}} \right) \right] \right. \\
& \quad \left. + \frac{1}{p} \ln \mathbb{P}(\chi^2 \in (p - \|b\|^2 \pm C\epsilon p)) \right\} + o(1)
\end{aligned}$$

By Lemma 3 and noting that as $p \rightarrow \infty$,

$$\left| \frac{1}{p} \ln \mathbb{P}(\chi^2 \in (p - \|b\|^2 \pm C\epsilon p)) - \frac{1}{2} \ln \left(1 - \frac{\|b\|^2}{p} \right) \right| \leq O(\epsilon) + o(1)$$

uniformly in $b \in \{b \in \mathbb{R}^p : c' \sqrt{p} \leq \|b\| \leq (1-c)\sqrt{p}\}$, we get with high probability

$$\frac{1}{pm_p} \ln \left(\mathbb{E}_g \left[e^{-\tilde{G}(g)} \mathbf{1}_{\{g \in \tilde{\mathcal{S}}_\epsilon\}} \right] \right) \leq \sup_{b \in \mathbb{R}^p, c' \sqrt{p} \leq \|b\| \leq (1-c)\sqrt{p}} \{f_{\text{TAP}}(b) + O(\epsilon)\}.$$

Lemma 10 is then established by (35). □

3.4 Restricted free energy and proof of Lemma 8

We turn to the proof of Lemma 8.

Proof of Lemma 8: First, note that using Lemma 9, there exists $c_p \rightarrow 0$ and $m_p \rightarrow \infty$ such that for any constant $C > 0$, with high probability, for large enough p ,

$$\langle \mathbf{1}_{(\beta^1, \dots, \beta^{m_p}) \in \mathcal{S}} \rangle \geq \exp(-Cpm_p) > 0.$$

In particular, this implies that \mathcal{S} is nonempty with high-probability. For any $(\beta^1, \dots, \beta^{m_p}) \in \mathcal{S}$,

$$0 \leq \left\| \frac{1}{m_p} \sum_i \beta^i \right\|^2 = \frac{p}{m_p^2} + \frac{1}{m_p^2} \sum_{i \neq j} (\beta^i)^\top \beta^j \leq \frac{p}{m_p^2} + p(\mathbb{E}[\langle R_{1,2} \rangle] + c_p)$$

which implies $\mathbb{E}[\langle R_{1,2} \rangle] > -c_p - 1/(m_p^2)$. This establishes the lower bound.

Next, we turn to the upper bound. Without loss of generality, we assume

$$\beta_0 = \sqrt{p}e_1.$$

Define the *restricted free energy*

$$f_p(\delta) = \lim_{\epsilon \rightarrow 0^+} \frac{1}{p} \ln \left\{ \frac{1}{2\epsilon} \int_{|\beta_1 - \sqrt{p}\delta| < \epsilon} e^{-\frac{1}{2\Delta} \|y - X\beta\|^2 + \mathcal{H}_p^{\text{Gauss}}(\beta)} d\pi_0(\beta) \right\}.$$

By the Nishimori identity (see e.g. [8, (2.25)])

$$\mathbb{E} \langle R_{1,2} \rangle = \frac{1}{p} \mathbb{E} \langle \beta^\top \beta_0 \rangle, \tag{40}$$

so it suffices to show there exists $C > 0$ such that

$$(1 - C) > \frac{1}{p} \mathbb{E} \langle \beta^\top \beta_0 \rangle = \mathbb{E} \left\langle \frac{\beta_1}{\sqrt{p}} \right\rangle, \quad (41)$$

The restricted free energy, as defined in (40), can be simplified to

$$f_p(\delta) = \frac{1}{p} \ln \left\{ \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p-1}{2})} (1 - \delta^2)^{\frac{p-3}{2}} \int_{\beta_1 = \sqrt{p}\delta} e^{-\frac{1}{2\Delta} \|y - X\beta\|^2 + \mathcal{H}_p^{\text{Gauss}}(\beta)} d\pi_{\beta_1 = \sqrt{p}\delta}(\beta) \right\},$$

where $\pi_{\beta_1 = \sqrt{p}\delta}$ refers to the uniform distribution on the $p - 2$ dimensional sphere $\{\beta \in S^{p-1}(\sqrt{p}) : \beta_1 = \sqrt{p}\delta\}$. Denote $X = [X_1, X_2, \dots, X_p]$. Define $X_{(-1)} := [X_2, \dots, X_p]$, $h = \sqrt{p}X_1 + z - \beta_1 X_1$, and $\beta_{(-1)} := (\beta_2, \beta_3, \dots, \beta_p)$. Then

$$y - X\beta = \sqrt{p}X_1 + z - \beta_1 X_1 - X_{(-1)}\beta_{(-1)} = h - X_{(-1)}\beta_{(-1)}.$$

Setting $u := \beta_{(-1)}/\sqrt{p(1 - \delta^2)} \in S^{p-2}(1)$, we have

$$\begin{aligned} \|y - X\beta\|^2 &= \|h\|^2 + \beta_{(-1)}^\top X_{(-1)}^\top X_{(-1)}\beta_{(-1)} - 2h^\top X_{(-1)}\beta_{(-1)} \\ &= \|h\|^2 + p(1 - \delta^2)u^\top X_{(-1)}^\top X_{(-1)}u - 2\sqrt{p(1 - \delta^2)}h^\top X_{(-1)}u. \end{aligned}$$

With high probability with respect to Z , $\sup_\beta |\mathcal{H}_p^{\text{Gauss}}(\beta)| = O(\epsilon_p)\lambda_0 p = o(p)$. Denoting the uniform measure on $S^{p-2}(1)$ as μ_{p-2} , we have,

$$\begin{aligned} f_p(\delta) &= \frac{1}{p} \left\{ \frac{1}{2} \ln(1 - \delta^2) + \ln \left(\int_{S^{p-2}(1)} e^{-\frac{1}{2\Delta} (\|h\|^2 + p(1 - \delta^2)u^\top X_{(-1)}^\top X_{(-1)}u - 2\sqrt{p(1 - \delta^2)}h^\top X_{(-1)}u)} d\mu_{p-2}(u) \right) \right\} + o(1) \\ &= \left\{ \frac{1}{2} \ln(1 - \delta^2) - \frac{\|h\|^2}{2p\Delta} + \frac{1}{p} \ln \int_{S^{p-2}(1)} e^{-\frac{1}{2\Delta} (p(1 - \delta^2)u^\top X_{(-1)}^\top X_{(-1)}u - 2\sqrt{p(1 - \delta^2)}h^\top X_{(-1)}u)} d\mu_{p-2}(u) \right\} + o(1). \end{aligned} \quad (42)$$

By Law of Large Numbers, almost surely

$$\frac{1}{p} \|h\|^2 \rightarrow \alpha\Delta + (1 - \delta)^2. \quad (43)$$

As $\log(1 - \delta^2) \rightarrow -\infty$ as $\delta \rightarrow 1$, for any $C_1 > 0$, there exists $\delta < 1$ such that with high probability, $\sup_{\delta' \in [\delta, 1]} f_p(\delta') < -C_1$. On the other hand, for any fixed $\delta < 1$, there exists $C' > 0$ (depending on α , Δ and δ) such that with high probability, $\sup_{\delta' \in [-1, \delta]} f_p(\delta') > -C'$. We now proceed as follows: fix any $\delta_0 < 1$, and assume $\sup_{\delta' \in [-1, \delta_0]} f_p(\delta') > -C_{\delta_0}$. On the other hand, let $\delta_1 < 1$ be such that with high probability, $\sup_{\delta' \in [\delta_1, 1]} f_p(\delta') < -2C_{\delta_0}$. Thus with high-probability,

$$\langle 1(\beta_1 > \delta_1 \sqrt{p}) \rangle = \frac{\int_{\delta_1}^1 \exp(p f_p(\delta')) d\delta'}{\int_{-1}^1 \exp(p f_p(\delta')) d\delta'} \leq \frac{\int_{\delta_1}^1 \exp(p f_p(\delta')) d\delta'}{\int_{-1}^{\delta_0} \exp(p f_p(\delta')) d\delta'} \leq \frac{(1 - \delta_1) \exp(-2C_{\delta_0} p)}{(\delta_0 + 1) \exp(-pC_{\delta_0})} \leq \frac{1 - \delta_1}{1 + \delta_0} \exp(-pC_{\delta_0}).$$

Thus with probability $1 - o(1)$,

$$\langle 1(\beta_1 < \delta_1 \sqrt{p}) \rangle > 1 - \frac{1 - \delta_1}{1 + \delta_0} \exp(-pC_{\delta_0}). \quad (44)$$

This implies, with probability at least $1 - o(1)$,

$$\langle \beta_1 \rangle \leq \delta_1 \sqrt{p} + \sqrt{p} \cdot \frac{1 - \delta_1}{1 + \delta_0} \exp(-pC_{\delta_0}).$$

Finally, this implies

$$\mathbb{E}[\langle \beta_1 \rangle] \leq \delta_1 \sqrt{p} + \sqrt{p} \cdot \frac{1 - \delta_1}{1 + \delta_0} \exp(-pC_{\delta_0}) + \sqrt{p}o(1) = (\delta_1 + o(1))\sqrt{p}.$$

The required upper bound follows upon setting $C := (1 - \delta_1)/2$. □

4 Proof of Theorem 1

In this section we prove Theorem 1 from Theorem 2 and Theorem 4 by arguing the small constant c within Theorem 2 has negligible impact.

Proof of Theorem 1: First, the upper bound, namely

$$\mathbb{P} \left(\frac{1}{p} \ln \mathcal{Z}_p \geq \sup_{a \in \mathbb{R}^p, \|a\| \leq \sqrt{p}} f_{\text{TAP}}(a) + \eta \right) \rightarrow 0,$$

follows immediately from Theorem 4. For the matching lower bound, for any $a \in \mathbb{R}^p$, $\|a\| \leq \sqrt{p}$, define

$$a^* := \arg \min_{x \in \mathbb{R}^p, \|x\| \leq (1-c)\sqrt{p}} \|a - x\|,$$

then

$$\begin{aligned} & \sup_{a \in \mathbb{R}^p, \|a\| \leq \sqrt{p}} f_{\text{TAP}}(a) - \sup_{a \in \mathbb{R}^p, \|a\| \leq (1-c)\sqrt{p}} f_{\text{TAP}}(a) \\ & \leq \max \left\{ 0, \sup_{a \in \mathbb{R}^p, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} (f_{\text{TAP}}(a) - f_{\text{TAP}}(a^*)) \right\}, \end{aligned} \quad (45)$$

the second term of which can be further controlled by

$$\begin{aligned} & \sup_{a \in \mathbb{R}^p, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} (f_{\text{TAP}}(a) - f_{\text{TAP}}(a^*)) \\ & \leq \sup_{a \in \mathbb{R}^p, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} \left\{ \frac{1}{2\Delta p} \left| \|y - Xa\|^2 - \|y - Xa^*\|^2 \right| \right. \\ & \quad \left. - \frac{\alpha}{2} \left[\ln \left(1 + \frac{1 - \|a\|^2/p}{\Delta\alpha} \right) - \ln \left(1 + \frac{1 - \|a^*\|^2/p}{\Delta\alpha} \right) \right] + \frac{1}{2} \left[\ln \left(1 - \frac{\|a\|^2}{p} \right) - \ln \left(1 - \frac{\|a^*\|^2}{p} \right) \right] \right\} \\ & \leq \sup_{a \in \mathbb{R}^p, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} \left\{ \frac{1}{2\Delta p} \left[|2y^\top X(a - a^*)| + |a^\top X^\top Xa - (a^*)^\top X^\top Xa^*| \right] \right\} \\ & \quad + \sup_{a \in \mathbb{R}^p, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} \left\{ \frac{\alpha}{2} \left[\ln \left(1 + \frac{1 - \|a^*\|^2/p}{\Delta\alpha} \right) - \ln \left(1 + \frac{1 - \|a\|^2/p}{\Delta\alpha} \right) \right] \right\} \\ & \quad + \sup_{a \in \mathbb{R}^p, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} \left\{ \frac{1}{2} \left[\ln \left(1 - \frac{\|a\|^2}{p} \right) - \ln \left(1 - \frac{\|a^*\|^2}{p} \right) \right] \right\} \\ & \leq \sup_{a \in \mathbb{R}^p, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} \left\{ \frac{1}{2\Delta p} \left[2c\sqrt{p}\|y^\top X\| + 2p(2c - c^2)\|X^\top X\|_2 \right] + \tau(c) \right\}, \end{aligned}$$

where

$$\tau(c) := \frac{\alpha}{2} \ln \left(1 + \frac{2c - c^2}{\Delta\alpha} \right),$$

as a deterministic function, goes to 0 as $c \rightarrow 0$. Since with high probability, $\|X^\top X\|_2 = O(1)$ and $\|y^\top X\| = O(\sqrt{p})$, we have

$$\sup_{a \in \mathbb{R}^p, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} (f_{\text{TAP}}(a) - f_{\text{TAP}}(a^*)) \rightarrow 0$$

as $c \rightarrow 0$, with high probability. Thus, there exists positive constant c^* such that if $c \leq c^*$, then with high probability,

$$\sup_{a \in \mathbb{R}^p, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} (f_{\text{TAP}}(a) - f_{\text{TAP}}(a^*)) \leq \frac{\eta}{3}. \quad (46)$$

Finally, choosing $c \leq c^*$, Theorem 2 together with (45) and (46) give us

$$\mathbb{P} \left(\frac{1}{p} \ln \mathcal{Z}_p \leq \sup_{a \in \mathbb{R}^p, \|a\| \leq \sqrt{p}} f_{\text{TAP}}(a) - \frac{\eta}{3} \right) \rightarrow 0,$$

which is exactly the lower bound part of Theorem 1. \square

References

- [1] Arka Adhikari, Christian Brennecke, Per von Soosten, and Horng-Tzer Yau. Dynamical approach to the tap equations for the sherrington–kirkpatrick model. *Journal of Statistical Physics*, 183(3):1–27, 2021.
- [2] Antonio Auffinger and Aukosh Jagannath. Thouless–anderson–palmer equations for generic p -spin glasses. *The Annals of Probability*, 47(4):2230–2256, 2019.
- [3] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.
- [4] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116(12):5451–5460, 2019.
- [5] Jean Barbier and Nicolas Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference. *Probability theory and related fields*, 174(3):1133–1185, 2019.
- [6] Jean Barbier, Nicolas Macris, Mohamad Dia, and Florent Krzakala. Mutual information and optimality of approximate message-passing in random linear estimation. *IEEE Transactions on Information Theory*, 66(7):4270–4303, 2020.
- [7] Jean Barbier, Nicolas Macris, Antoine Maillard, and Florent Krzakala. The mutual information in random linear estimation beyond iid matrices. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1390–1394. IEEE, 2018.
- [8] Jean Barbier and Dmitry Panchenko. Strong replica symmetry in high-dimensional optimal bayesian inference. *arXiv preprint arXiv:2005.03115*, 2020.
- [9] David Belius and Nicola Kistler. The tap–plefka variational principle for the spherical sk model. *Communications in Mathematical Physics*, 367(3):991–1017, 2019.
- [10] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.
- [11] Michael Celentano, Zhou Fan, and Song Mei. Local convexity of the tap free energy and amp convergence for z2-synchronization. *arXiv preprint arXiv:2106.11428*, 2021.

- [12] Sourav Chatterjee. Spin glasses and stein’s method. *Probability theory and related fields*, 148(3-4):567–600, 2010.
- [13] Wei-Kuo Chen and Dmitry Panchenko. On the tap free energy in the mixed p-spin models. *Communications in Mathematical Physics*, 362(1):219–252, 2018.
- [14] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized tap free energy. *arXiv preprint arXiv:1812.05066*, 2018.
- [15] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized tap free energy ii. *Communications in Mathematical Physics*, 381(1):257–291, 2021.
- [16] Zhou Fan, Song Mei, and Andrea Montanari. Tap free energy, spin glasses and variational inference. *The Annals of Probability*, 49(1):1–45, 2021.
- [17] Zhou Fan and Yihong Wu. The replica-symmetric free energy for ising spin glasses with orthogonally invariant couplings. *arXiv preprint arXiv:2105.02797*, 2021.
- [18] Behrooz Ghorbani, Hamid Javadi, and Andrea Montanari. An instability in variational inference for topic models. In *International conference on machine learning*, pages 2221–2231. PMLR, 2019.
- [19] Alice Guionnet, M Maı, et al. A fourier view on the r-transform and related asymptotics of spherical integrals. *Journal of functional analysis*, 222(2):435–490, 2005.
- [20] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.
- [21] Aukosh Jagannath. Approximate ultrametricity for random measures and applications to spin glasses. *Communications on Pure and Applied Mathematics*, 70(4):611–664, 2017.
- [22] Florent Krzakala, Andre Manoel, Eric W Tramel, and Lenka Zdeborová. Variational free energies for compressed sensing. In *2014 IEEE International Symposium on Information Theory*, pages 1499–1503. IEEE, 2014.
- [23] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.
- [24] Andrea Montanari. Statistical mechanics and algorithms on sparse and random graphs. *Lectures on Probability Theory and Statistics. Saint-Flour*, 2013.
- [25] Sumit Mukherjee and Subhabrata Sen. Variational inference in high-dimensional linear regression. *arXiv preprint arXiv:2104.12232*, 2021.
- [26] Giorgio Parisi and Marc Potters. Mean-field equations for spin models with orthogonal interaction matrices. *Journal of Physics A: Mathematical and General*, 28(18):5267, 1995.
- [27] Eliran Subag. Free energy landscapes in spherical spin glasses. *arXiv preprint arXiv:1804.10576*, 2018.
- [28] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010.
- [29] David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of ’solvable model of a spin glass’. *Philosophical Magazine*, 35(3):593–601, 1977.
- [30] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.
- [31] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

- [32] Martin J Wainwright and Michael Irwin Jordan. *Graphical models, exponential families, and variational inference*. Now Publishers Inc, 2008.

Appendices

In Section A we collect some standard technical results that are used in our proofs. In the same vein, Section B collects some facts about sample covariance matrices with iid gaussian entries. Finally, Section C establishes Theorem 5.

A Concentration and convexity inequalities

Lemma 14 (Concentration of almost Lipschitz functions [6]). *Let $U_i \stackrel{iid}{\sim} N(0, 1)$. Let f be a Lipschitz function over $G \subset \mathbb{R}^p$ with Lipschitz constant K_p . Assume $0 \in G$, $f(0)^2 \leq C^2$, $\mathbb{E}[f^2] \leq C^2$ for some $C > 0$. Then for $r \geq 6(C + \sqrt{pK_p})\sqrt{\mathbb{P}(G^c)}$ we have*

$$\mathbb{P}(|f(U_1, \dots, U_p) - \mathbb{E}[f(U_1, \dots, U_p)]| \geq r) \leq 2e^{-\frac{r^2}{16K_p^2}} + \mathbb{P}(G^c).$$

Lemma 15 (The largest and smallest eigenvalues of Wishart matrices [30]). *Let A be an $n \times p$ matrix whose entries are independent standard normal random variables. Then for every $t \geq 0$, with probability at least $1 - 2\exp(-t^2/2)$ one has*

$$\sqrt{n} - \sqrt{p} - t \leq \sigma_{\min}(A) \leq \sigma_{\max}(A) \leq \sqrt{n} + \sqrt{p} + t.$$

Lemma 16 (Lévy's lemma [31]). *Let $S^{n-1}(1)$ be the $(n-1)$ -sphere of radius 1. Let $f : S^{n-1}(1) \rightarrow \mathbb{R}$ be a Lipschitz function with Lipschitz constant L , and let σ be a uniform random vector on $S^{n-1}(1)$. Then for any $t > 0$,*

$$\mathbb{P}(|f(\sigma) - \mathbb{E}f(\sigma)| \geq Lt) \leq \exp(\pi - nt^2/4).$$

Lemma 17 (A bound for differences between derivatives of convex functions [4]). *Let $G(x)$ and $g(x)$ be convex functions. Let $\delta > 0$ and define $C_\delta^-(x) := g'(x) - g'(x - \delta) \geq 0$ and $C_\delta^+(x) := g'(x + \delta) - g'(x) \geq 0$. Then*

$$|G'(x) - g'(x)| \leq \delta^{-1} \sum_{u \in \{x-\delta, x, x+\delta\}} |G(u) - g(u)| + C_\delta^+(x) + C_\delta^-(x).$$

B Facts about the random matrix $X^\top X$

In this section, we list some well-known facts about the random matrix $X^\top X$, which is usually referred to as *sample covariance matrix* in statistics and probability literature. Please refer to [3] for complete statements and proofs. Let $\theta_1 < \theta_2 < \dots < \theta_p$ be the eigenvalues of the matrix $S_p = X^\top X \in \mathbb{R}^{p \times p}$. We have that as $p \rightarrow \infty$, almost surely,

$$\frac{1}{p} \sum_{i=1}^p \delta_{\theta_i} \rightarrow \mu(x) dx \text{ in distribution,}$$

where $\mu(\cdot)$ is the Marchenko-Pastur law. Moreover, if we let θ_t be the t -quantile of μ , then

$$\theta^i = \theta_{i/p} + o(1) \text{ for } i \in \{1, 2, \dots, p\},$$

where the $o(1)$ terms above converge to 0 almost surely, uniformly in i .

C Proof of overlap concentration results

C.1 Proof of Theorem 5

First of all, please note that in this subsection we will use \mathbb{E} as a shorthand for $\mathbb{E}_{X,z,Z,\beta_0}$. Recall that $\mathcal{H}_p^{\text{Gauss}}$ is the Hamiltonian of the side information channel, namely

$$\mathcal{H}_p^{\text{Gauss}} = -\lambda_0 \epsilon_p \beta_0^\top \beta - \sqrt{\lambda_0 \epsilon_p} Z^\top \beta + \frac{1}{2} \lambda_0 \epsilon_p \|\beta\|^2,$$

which can be seen as only keeping the β -related terms in $\|Y^{\text{Gauss}} - \sqrt{\lambda_0 \epsilon_p} \beta\|^2/2$. Let $\lambda_{0,p} = \lambda_0 \epsilon_p$. Furthermore, let

$$H' := \frac{d\mathcal{H}_p^{\text{Gauss}}}{d\lambda_{0,p}} = -\beta_0^\top \beta - \frac{Z^\top \beta}{2\sqrt{\lambda_{0,p}}} + \frac{\|\beta\|^2}{2}. \quad (47)$$

We begin the proof of overlap concentration, i.e. Theorem 5, by introducing the following lemma, which upper bounds the fluctuations of $R_{1,2}$ by the fluctuations of $\mathcal{L} := H'/p$.

Lemma 18. *Let*

$$\mathcal{L} := \frac{H'}{p} = \frac{1}{p} \frac{d\mathcal{H}_p^{\text{Gauss}}}{d\lambda_{0,p}}.$$

Then

$$\mathbb{E} \langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle \leq 4\mathbb{E} \langle (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle)^2 \rangle.$$

Proof: Let $R_{1,*} = \beta_0^\top \beta/p$. Using the definition of H' gives

$$\begin{aligned} 2\mathbb{E} \langle R_{1,*} (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle &= \mathbb{E} \left[\frac{1}{p} \langle R_{1,*} \|\beta\|^2 \rangle - 2 \langle R_{1,*}^2 \rangle - \frac{1}{p\sqrt{\lambda_{0,p}}} \langle R_{1,*} Z^\top \beta \rangle \right] \\ &\quad - \mathbb{E} \langle R_{1,*} \rangle \mathbb{E} \left[\frac{1}{p} \langle \|\beta\|^2 \rangle - 2 \langle R_{1,*} \rangle - \frac{1}{p\sqrt{\lambda_{0,p}}} Z^\top \langle \beta \rangle \right]. \end{aligned}$$

Gaussian integration by parts implies

$$\frac{1}{p\sqrt{\lambda_{0,p}}} \mathbb{E} \langle R_{1,*} Z^\top \beta \rangle = \frac{1}{p} \mathbb{E} \langle R_{1,*} \|\beta\|^2 \rangle - \frac{1}{p} \mathbb{E} \langle R_{1,*} \beta^\top \langle \beta \rangle \rangle = \frac{1}{p} \mathbb{E} \langle R_{1,*} \|\beta\|^2 \rangle - \mathbb{E} \left[\langle R_{1,*} \rangle^2 \right],$$

in which $\mathbb{E} \langle R_{1,*} \beta^\top \langle \beta \rangle \rangle = \frac{1}{p} \mathbb{E} [\langle \beta_0^\top \beta \beta^\top \rangle \langle \beta \rangle] = \frac{1}{p} \mathbb{E} [\langle \beta_0^\top \beta^1 (\beta^1)^\top \beta^2 \rangle] = \frac{1}{p} \mathbb{E} [\langle (\beta^1)^\top \beta_0 \beta_0^\top \beta^2 \rangle] = p\mathbb{E} \left[\langle R_{1,*} \rangle^2 \right]$ is due to the Nishimori identity. Again, using Gaussian integration by parts

$$\frac{1}{p\sqrt{\lambda_{0,p}}} \mathbb{E} \langle Z^\top \beta \rangle = \frac{1}{p} \mathbb{E} \langle \|\beta\|^2 \rangle - \mathbb{E} \langle R_{1,*} \rangle.$$

Putting them together, we have

$$\begin{aligned} 2\mathbb{E} \langle R_{1,*} (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle &= \mathbb{E} \langle R_{1,*} \rangle^2 - 2\mathbb{E} \langle R_{1,*}^2 \rangle + (\mathbb{E} \langle R_{1,*} \rangle)^2 \\ &= -(\mathbb{E} \langle R_{1,*}^2 \rangle - (\mathbb{E} \langle R_{1,*} \rangle)^2) - (\mathbb{E} \langle R_{1,*}^2 \rangle - \mathbb{E} \langle R_{1,*} \rangle^2) \\ &\leq -\mathbb{E} \langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle \\ &\leq 0. \end{aligned}$$

Finally, using Cauchy-Schwarz inequality, we have

$$\begin{aligned} \mathbb{E} \langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle &\leq 2|\mathbb{E} \langle R_{1,*} (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle| \\ &= 2|\mathbb{E} \langle (R_{1,*} - \mathbb{E} \langle R_{1,*} \rangle) (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle| \\ &\leq 2 \left\{ \mathbb{E} \langle (R_{1,*} - \mathbb{E} \langle R_{1,*} \rangle)^2 \rangle \mathbb{E} \langle (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle)^2 \rangle \right\}^{1/2}, \end{aligned}$$

which implies

$$\mathbb{E} \langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle = \mathbb{E} \langle (R_{1,*} - \mathbb{E} \langle R_{1,*} \rangle)^2 \rangle \leq 4 \mathbb{E} \langle (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle)^2 \rangle.$$

□

Since the above lemma basically says the fluctuations of the overlaps are upper bounded by those of \mathcal{L} , it is sufficient to prove that \mathcal{L} will concentrate. Before moving on, let us recall a few notations. As in (21), $\mathcal{H}^{\text{Pert}}$ is the Hamiltonian of the whole perturbed model, i.e. including both the original model and the side information channel, and F_p^{Pert} is the corresponding free energy, as defined in (23). Moreover, let

$$\mathcal{H}' := \frac{d\mathcal{H}^{\text{Pert}}}{d\lambda_{0,p}}.$$

Please note \mathcal{H}' and H' defined in (47) are essentially the same since $\mathcal{H} = \mathcal{H}^{\text{Pert}} - \mathcal{H}^{\text{Gauss}}$ does not depend on $\lambda_{0,p}$. Armed with these notations, we observe,

$$\frac{dF_p^{\text{Pert}}}{d\lambda_{0,p}} = -\frac{1}{p} \langle \mathcal{H}' \rangle, \quad \text{and} \quad \frac{d^2 F_p^{\text{Pert}}}{d\lambda_{0,p}^2} = \frac{1}{p} \left\{ \langle (\mathcal{H}' - \langle \mathcal{H}' \rangle)^2 \rangle - \frac{\langle \beta \rangle^\top Z}{4\lambda_{0,p}^{3/2}} \right\}.$$

Exchanging expectation and derivatives, we obtain

$$\frac{d\mathbb{E} F_p^{\text{Pert}}}{d\lambda_{0,p}} = -\frac{1}{p} \mathbb{E} \langle \mathcal{H}' \rangle = -\frac{1}{2} \mathbb{E} \langle R_{1,2} \rangle, \quad \text{and} \quad \frac{d^2 \mathbb{E} F_p^{\text{Pert}}}{d\lambda_{0,p}^2} = \frac{1}{p} \left\{ \mathbb{E} \langle (\mathcal{H}' - \langle \mathcal{H}' \rangle)^2 \rangle - \frac{1}{4\lambda_{0,p}} \mathbb{E} \langle \|\beta - \langle \beta \rangle\|^2 \rangle \right\},$$

in which we have utilized both Nishimori identity and the following identity obtained using Gaussian integration by parts

$$\frac{1}{p\sqrt{\lambda_{0,p}}} \mathbb{E} \langle Z^\top \beta \rangle = \mathbb{E} \langle \|\beta\|^2 \rangle - \mathbb{E} \langle R_{1,*} \rangle.$$

Lemma 19 (Fluctuations of \mathcal{L}). *Let $\lambda_0 \sim \text{Unif}[1/2, 1]$. If $\frac{v_p}{p\epsilon_p} \rightarrow 0$ then there exists a constant $C > 0$ such that*

$$\mathbb{E}_{\lambda_0} \mathbb{E} \langle (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle)^2 \rangle \leq \frac{C}{\epsilon_p} \left(\frac{v_p}{p\epsilon_p} + \frac{1}{p} \right)^{1/3}.$$

Proof: To prove this lemma, we consider the following decomposition and split our proof into controlling these two terms respectively,

$$\mathbb{E} \langle (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle)^2 \rangle = \mathbb{E} \langle (\mathcal{L} - \langle \mathcal{L} \rangle)^2 \rangle + \mathbb{E} \langle (\langle \mathcal{L} \rangle - \mathbb{E} \langle \mathcal{L} \rangle)^2 \rangle.$$

We start with the first term. Consider

$$\int_{\epsilon_p/2}^{\epsilon_p} \mathbb{E} \langle (\mathcal{H}' - \langle \mathcal{H}' \rangle)^2 \rangle d\lambda_{0,p} \leq \int_{\epsilon_p/2}^{\epsilon_p} \left(p \frac{d^2 \mathbb{E} F_p^{\text{Pert}}}{d\lambda_{0,p}^2} + \frac{p}{\lambda_{0,p}} \right) d\lambda_{0,p} = p \left(\frac{d\mathbb{E} F_p^{\text{Pert}}}{d\lambda_{0,p}} \Big|_{\lambda_{0,p}=\epsilon_p/2}^{\lambda_{0,p}=\epsilon_p} \right) + p \ln 2.$$

Since

$$\left| \frac{d\mathbb{E} F_p^{\text{Pert}}}{d\lambda_{0,p}} \right| = \frac{1}{p} |\mathbb{E} \langle \mathcal{H}' \rangle| = \left| \frac{1}{2} \mathbb{E} \langle R_{1,2} \rangle \right| \leq \frac{1}{2},$$

For any $\lambda_{0,p}$. We naturally have a trivial bound of $\frac{d\mathbb{E} F_p^{\text{Pert}}}{d\lambda_{0,p}} \Big|_{\lambda_{0,p}=\epsilon_p/2}^{\lambda_{0,p}=\epsilon_p}$, i.e.,

$$\frac{d\mathbb{E} F_p^{\text{Pert}}}{d\lambda_{0,p}} \Big|_{\lambda_{0,p}=\epsilon_p/2}^{\lambda_{0,p}=\epsilon_p} \leq 1.$$

Recalling that $\mathcal{L} = H'/p = \mathcal{H}'/p$,

$$\mathbb{E}_{\lambda_0} \mathbb{E} \langle (\mathcal{L} - \mathbb{E}\langle \mathcal{L} \rangle)^2 \rangle = \frac{1}{p^2} \mathbb{E}_{\lambda_0} \mathbb{E} \langle (\mathcal{H}' - \mathbb{E}\langle \mathcal{H}' \rangle)^2 \rangle \leq \frac{1 + \ln 2}{p}.$$

Now we proceed to control the second term using convexity arguments.

We introduce the following two convex functions of $\lambda_{0,p}$,

$$\tilde{F}(\lambda_{0,p}) := F_p^{\text{Pert}}(\lambda_{0,p}) - \sqrt{\frac{\lambda_{0,p}}{p}} \|Z\|,$$

and

$$\mathbb{E}\tilde{F}(\lambda_{0,p}) := \mathbb{E}F_p^{\text{Pert}}(\lambda_{0,p}) - \sqrt{\frac{\lambda_{0,p}}{p}} \mathbb{E}\|Z\|.$$

Letting $A_p := p^{-1/2} (\|Z\| - \mathbb{E}\|Z\|)$, we have

$$\tilde{F}(\lambda_{0,p}) - \mathbb{E}\tilde{F}(\lambda_{0,p}) = (F_p^{\text{Pert}}(\lambda_{0,p}) - \mathbb{E}F_p^{\text{Pert}}(\lambda_{0,p})) - \sqrt{\lambda_{0,p}} A_p,$$

as well as

$$\tilde{F}'(\lambda_{0,p}) - \mathbb{E}\tilde{F}'(\lambda_{0,p}) = \langle \mathcal{L} \rangle - \mathbb{E}\langle \mathcal{L} \rangle - \frac{A_p}{2\sqrt{\lambda_{0,p}}}.$$

Thus Lemma 17 implies

$$|\langle \mathcal{L} \rangle - \mathbb{E}\langle \mathcal{L} \rangle| \leq \delta^{-1} \sum_{u \in \mathcal{U}} (|F_p^{\text{Pert}}(u) - \mathbb{E}F_p^{\text{Pert}}(u)| + |A_p| \sqrt{u}) + C_\delta^+(\lambda_{0,p}) + C_\delta^-(\lambda_{0,p}) + \frac{|A_p|}{2\sqrt{\lambda_{0,p}}},$$

where $\mathcal{U} := \{\lambda_{0,p} - \delta, \lambda_{0,p}, \lambda_{0,p} + \delta\}$ and

$$C_\delta^-(\lambda_{0,p}) := \mathbb{E}\tilde{F}'(\lambda_{0,p}) - \mathbb{E}\tilde{F}'(\lambda_{0,p} - \delta) \geq 0, \quad C_\delta^+(\lambda_{0,p}) := \mathbb{E}\tilde{F}'(\lambda_{0,p} + \delta) - \mathbb{E}\tilde{F}'(\lambda_{0,p}) \geq 0.$$

Note that by using [31, Theorem 3.1.1], there exists a positive constant C_Z such that $\mathbb{E}[(\|Z\| - \mathbb{E}\|Z\|)^2] \leq C_Z$ independently of p , which ensures $\mathbb{E}(A_p^2) \leq C_Z/p = O(1/p)$. Thus, by the definition of v_p in the statement of Theorem 5,

$$\begin{aligned} \mathbb{E}(\langle \mathcal{L} \rangle - \mathbb{E}\langle \mathcal{L} \rangle)^2 &\leq 9\mathbb{E} \left\{ \delta^{-2} \sum_{u \in \mathcal{U}} \left[|F_p^{\text{Pert}}(u) - \mathbb{E}F_p^{\text{Pert}}(u)|^2 + (|A_p| \sqrt{u})^2 \right] \right. \\ &\quad \left. + C_\delta^+(\lambda_{0,p})^2 + C_\delta^-(\lambda_{0,p})^2 + \left(\frac{|A_p|}{2\sqrt{\lambda_{0,p}}} \right)^2 \right\} \\ &\leq 9 \left\{ \frac{3}{p\delta^2} [v_p + (\lambda_{0,p} + \delta)C_Z] + C_\delta^+(\lambda_{0,p})^2 + C_\delta^-(\lambda_{0,p})^2 + \frac{C_Z}{4p\lambda_{0,p}} \right\}. \end{aligned}$$

Note that

$$|\mathbb{E}\tilde{F}'(\lambda_{0,p})| = \left| \frac{1}{2} \mathbb{E}R_{1,2} + \frac{\mathbb{E}\|Z\|}{2\sqrt{\lambda_{0,p}p}} \right| \leq \frac{1}{2} \left(1 + \frac{1}{\sqrt{\lambda_{0,p}}} \right),$$

and thus

$$|C_\delta^\pm(\lambda_{0,p})| \leq 1 + \frac{1}{\sqrt{\varepsilon_p/2 - \delta}},$$

for $\delta < \epsilon_p/2$, which further gives us

$$\begin{aligned} \int_{\epsilon_p/2}^{\epsilon_p} d\lambda_{0,p} \left\{ C_\delta^+ (\lambda_{0,p})^2 + C_\delta^- (\lambda_{0,p})^2 \right\} &\leq \left(1 + \frac{1}{\sqrt{\epsilon_p/2 - \delta}} \right) \int_{\epsilon_p/2}^{\epsilon_p} d\lambda_{0,p} \left\{ C_\delta^+ (\lambda_{0,p}) + C_\delta^- (\lambda_{0,p}) \right\} \\ &= \left(1 + \frac{1}{\sqrt{\epsilon_p/2 - \delta}} \right) \left[\left(\mathbb{E}\tilde{F}(\epsilon_p/2 + \delta) - \mathbb{E}\tilde{F}(\epsilon_p/2 - \delta) \right) + \left(\mathbb{E}\tilde{F}(\epsilon_p - \delta) - \mathbb{E}\tilde{F}(\epsilon_p + \delta) \right) \right]. \end{aligned}$$

By the mean value theorem,

$$|\mathbb{E}\tilde{F}(\epsilon_p/2 + \delta) - \mathbb{E}\tilde{F}(\epsilon_p/2 - \delta)| \leq 2\delta \max_{x \in [\epsilon_p/2 - \delta, \epsilon_p/2 + \delta]} \mathbb{E}\tilde{F}'(x) \leq \delta \left(1 + \frac{1}{\sqrt{\epsilon_p/2 - \delta}} \right).$$

Similarly,

$$|\mathbb{E}\tilde{F}(\epsilon_p + \delta) - \mathbb{E}\tilde{F}(\epsilon_p - \delta)| \leq 2\delta \max_{x \in [\epsilon_p - \delta, \epsilon_p + \delta]} \mathbb{E}\tilde{F}'(x) \leq \delta \left(1 + \frac{1}{\sqrt{\epsilon_p - \delta}} \right) \leq \delta \left(1 + \frac{1}{\sqrt{\epsilon_p/2 - \delta}} \right).$$

Therefore

$$\int_{\epsilon_p/2}^{\epsilon_p} d\lambda_{0,p} \left\{ C_\delta^+ (\lambda_{0,p})^2 + C_\delta^- (\lambda_{0,p})^2 \right\} \leq 2\delta \left(1 + \frac{1}{\sqrt{\epsilon_p/2 - \delta}} \right)^2 \leq 4\delta \frac{\epsilon_p/2 - \delta + 1}{\epsilon_p/2 - \delta} \leq \frac{8\delta}{\epsilon_p/2 - \delta}$$

where we set $\delta = \delta_p$ so that $\delta_p/\epsilon_p \rightarrow 0$ and $\delta_p/\epsilon_p < 1/4$ for any p . Putting them all together, we have

$$\begin{aligned} \int_{\epsilon_p/2}^{\epsilon_p} d\lambda_{0,p} \mathbb{E} \langle (\mathcal{L} - \mathbb{E}\langle \mathcal{L} \rangle)^2 \rangle &\leq \frac{27\epsilon_p}{2p\delta_p^2} (v_p + 2\epsilon_p C_Z) + 9 \int_{\epsilon_p/2}^{\epsilon_p} \frac{C_Z}{4p\lambda_{0,p}} d\lambda_{0,p} + \frac{288\delta_p}{\epsilon_p} \\ &= \frac{27\epsilon_p}{2p\delta_p^2} (v_p + 2\epsilon_p C_Z) + \frac{288\delta_p}{\epsilon_p} + \frac{9C_Z \ln 2}{4p}. \end{aligned}$$

Finally, we choose $\delta_p = \{ [3\epsilon_p^2(v_p + 2\epsilon_p C_Z)] / (64p) \}^{1/3}$ to optimize the bound, which together with the previous bound on $\mathbb{E}_{\lambda_0} \mathbb{E} \langle (\mathcal{L} - \mathbb{E}\langle \mathcal{L} \rangle)^2 \rangle$ would result in

$$\begin{aligned} \mathbb{E}_{\lambda_0} \mathbb{E} \langle (\mathcal{L} - \mathbb{E}\langle \mathcal{L} \rangle)^2 \rangle &\leq \frac{576}{\epsilon_p} \left(\frac{3\epsilon_p^2(v_p + 2\epsilon_p C_Z)}{64p} \right)^{1/3} + \frac{9C_Z \ln 2}{4p} + \frac{1 + \ln 2}{p} \\ &\leq C \left(\frac{v_p}{p\epsilon_p} + \frac{1}{p} \right)^{1/3}, \end{aligned}$$

for some positive constant C . □

Combining Lemma 18 and Lemma 19 gives us the following lemma, which is essentially the first half of Theorem 5.

Lemma 20 (Overlap concentration: Fluctuations of $R_{1,2}$). *Assuming $\lambda_0 \sim \text{Unif}(\frac{1}{2}, 1)$, there exist a positive constant C such that*

$$\mathbb{E}_{\lambda_0} \mathbb{E} \langle (R_{1,2} - \mathbb{E}\langle R_{1,2} \rangle)^2 \rangle \leq C \left(\frac{v_p}{p\epsilon_p} + \frac{1}{p} \right)^{1/3},$$

where $v_p := p \sup_{\lambda_0 \in [\frac{1}{2}, 1]} \{ \mathbb{E}(F_p^{\text{Pert}}(\lambda_0) - \mathbb{E}F_p^{\text{Pert}}(\lambda_0))^2 \}$.

C.2 Concentration of the free energy with respect to the Gaussian quenched variables

Lemma 21. *There exist two positive constants C_1 and C_2 (depending only on Δ, α) such that for any $r \geq O(e^{-C_1 p^{1/2}})$, we have for any fixed β_0 and $\lambda_0 \in [1/2, 1]$,*

$$\mathbb{P}\left(\frac{1}{p}|F_p^{\text{Pert}} - \mathbb{E}[F_p^{\text{Pert}}|\beta_0]| \geq r|\beta_0\right) \leq e^{-C_2 r^2 p^{1/2}}.$$

Proof: Consider

$$\mathcal{G} = \{X, z, Z : \max_{\mu} |z_i| \leq \sqrt{D_1} \cap \max_i |Z_i| \leq \sqrt{D_1} \cap \text{for all } \beta, \beta_0 : \|X(\beta - \beta_0)\|^2 \leq D_2 p\}.$$

In order to use the known results regarding concentration of almost Lipschitz functions, we first try to control $\mathbb{P}(\mathcal{G}^c)$. Since if $U \sim N(0, 1)$ then $\mathbb{P}(|U| \geq \sqrt{A}) \leq 4e^{-A/4}$, we have by union bound

$$\mathbb{P}(\max_{\mu} |z_i| \leq \sqrt{D_1} \cap \max_i |\tilde{z}_i| \leq \sqrt{D_1}) \leq 4p(1 + \alpha)e^{-D_1/4},$$

and

$$\mathbb{P}(\text{for all } \beta, \beta_0 : \|X(\beta - \beta_0)\|^2 \leq D_2 p) = \mathbb{P}(\Lambda_1(X^\top X) \leq \frac{D_2}{4}).$$

Thus

$$\mathbb{P}(\mathcal{G}^c) \leq 4p(1 + \alpha)e^{-D_1/4} + \left(1 - \mathbb{P}\left(\sigma_{\max}(X)^2 \leq \frac{D_2}{4}\right)\right),$$

which will be made small with a suitable choices of D_1 and D_2 . Now we proceed to show, for any fixed β_0 , free energy of the perturbed model $F_p^{\text{Pert}}(X, z, Z, \beta_0)$ is Lipschitz on \mathcal{G} with respect to (X, z, Z) , as a vector in \mathbb{R}^{np+n+p} . From now on, we treat β_0 as fixed and omit it in the arguments of $\mathcal{H}^{\text{Pert}}$ and F^{Pert} . Recall that, as defined in (23),

$$\mathcal{H}^{\text{Pert}}(X, z, Z) = \frac{1}{2\Delta} \|y - X\beta\|^2 - \lambda_0 \epsilon_p \beta_0^\top \beta - \sqrt{\lambda_0 \epsilon_p} Z^\top \beta + \frac{1}{2} \lambda \epsilon_p \|\beta\|^2.$$

Then

$$|\mathcal{H}^{\text{Pert}}(X, z, Z) - \mathcal{H}^{\text{Pert}}(X', z', Z')| \leq \frac{1}{2\Delta} \left| \|X(\beta_0 - \beta) + z\|^2 - \|X'(\beta_0 - \beta) + z'\|^2 \right| + \sqrt{\lambda \epsilon_p} |Z^\top \beta - (Z')^\top \beta|,$$

in which

$$\begin{aligned} & \left| \|X(\beta_0 - \beta) + z\|^2 - \|X'(\beta_0 - \beta) + z'\|^2 \right| \\ &= \left| (\|X(\beta_0 - \beta)\|^2 + 2z^\top X(\beta_0 - \beta) + \|z\|^2) - (\|X'(\beta_0 - \beta)\|^2 + 2z'^\top X'(\beta_0 - \beta) + \|z'\|^2) \right| \\ &\leq |(\beta - \beta_0)^\top (X^\top X - X'^\top X')(\beta - \beta_0)| + 2|(z^\top X - z'^\top X')(\beta - \beta_0)| + \left| \|z\|^2 - \|z'\|^2 \right| \\ &\leq \|\beta - \beta_0\|_2^2 (\|X^\top (X - X')\| + \|(X - X')^\top X'\|) + (O(\sqrt{pD_1}) + O(\sqrt{pD_2}))(\sqrt{p}\|X - X'\|_F + \|z - z'\|) \\ &\leq (O(\sqrt{pD_1}) + O(\sqrt{pD_2}))(\sqrt{p}\|X - X'\|_F + \|z - z'\|), \end{aligned}$$

and

$$|Z^\top \beta - (Z')^\top \beta| \leq \sqrt{\lambda_0 \epsilon_p p} \|Z - Z'\|.$$

Since $\lambda_0 \in [1/2, 1]$ and $\lim_{p \rightarrow \infty} \epsilon_p = 0$, we get, as long as both D_1 and D_2 are always lower bounded by a positive constant,

$$|\mathcal{H}^{\text{Pert}}(X, z, Z) - \mathcal{H}^{\text{Pert}}(X', z', Z')| \leq (O(\sqrt{pD_1}) + O(\sqrt{pD_2}))(\sqrt{p}\|X - X'\|_F + \|z - z'\| + \|Z - Z'\|).$$

For the difference of free energies, we thus have

$$\begin{aligned}
F_p^{\text{Pert}}(X, z, Z) - F_p^{\text{Pert}}(X', z', Z') &= \frac{1}{p} \ln \left(\frac{\int d\beta \pi_0(\beta) e^{-\mathcal{H}_\lambda^{\text{Pert}}(y)}}{\int d\beta \pi_0(\beta) e^{-\mathcal{H}_\lambda^{\text{Pert}}(y')}} \right) \\
&\leq \frac{1}{p} \ln \left(\frac{\int d\beta \pi_0(\beta) e^{-\mathcal{H}_\lambda^{\text{Pert}}(y') + |\mathcal{H}_\lambda^{\text{Pert}}(y) - \mathcal{H}_\lambda^{\text{Pert}}(y')|}}{\int d\beta \pi_0(\beta) e^{-\mathcal{H}_\lambda^{\text{Pert}}(y')}} \right) \\
&\leq \frac{1}{p} (O(\sqrt{pD_1}) + O(\sqrt{pD_2})) (\sqrt{p} \|X - X'\|_F + \|z - z'\| + \|Z - Z'\|).
\end{aligned}$$

By symmetry, we can further conclude that

$$\begin{aligned}
|F_p^{\text{Pert}}(X, z, Z) - F_p^{\text{Pert}}(X', z', Z')| &\leq (O(p^{-1/2} \sqrt{D_1}) + O(p^{-1/2} \sqrt{D_2})) (\sqrt{p} \|X - X'\|_F + \|z - z'\| + \|Z - Z'\|) \\
&\leq (O(p^{-1/2} \sqrt{D_1}) + O(p^{-1/2} \sqrt{D_2})) \|(\sqrt{p}X, z, Z) - (\sqrt{p}X', z', Z')\|,
\end{aligned}$$

where with a slight abuse of notation both $(\sqrt{p}X, z, Z)$ and $(\sqrt{p}X', z', Z')$ are treated as vectors in \mathbb{R}^{np+n+p} .

In order to apply Lemma 14, the last thing we need to show is that $\mathbb{E}_{X, z, Z} \left[(F_p^{\text{Pert}})^2 \right] = \mathbb{E} \left[(F_p^{\text{Pert}})^2 | \beta_0 \right]$ is bounded. In fact, the upper bound constant does not depend on β_0 .

$$\begin{aligned}
|F_p^{\text{Pert}}| &= \left| \frac{1}{p} \ln \int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{2\Delta} \|y - X\beta\|^2 - \lambda \epsilon_p \beta_0^\top \beta - \sqrt{\lambda \epsilon_p} Z^\top \beta + \frac{1}{2} \lambda \epsilon_p \|\beta\|^2} d\pi_0(\beta) \right| \\
&\leq \left| \frac{1}{p} \ln \int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{2\Delta} \|X(\beta_0 - \beta) + z\|^2 - \lambda \epsilon_p |\beta_0^\top \beta| - \sqrt{\lambda \epsilon_p} |Z^\top \beta| - \frac{1}{2} \lambda \epsilon_p \|\beta\|^2} d\pi_0(\beta) \right| \\
&\leq \left| \frac{1}{p} \ln \int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{\Delta} (\|X(\beta_0 - \beta)\|^2 + \|z\|^2) - p\lambda \epsilon_p - \sqrt{p\lambda \epsilon_p} \|Z\| - \frac{1}{2} p\lambda \epsilon_p} d\pi_0(\beta) \right| \\
&\leq \frac{1}{p} \left[\frac{4p}{\Delta} \|X^\top X\|_2 + \frac{1}{\Delta} \|z\|^2 + \frac{3}{2} p\lambda \epsilon_p + \sqrt{p\lambda \epsilon_p} \|Z\| \right],
\end{aligned}$$

which leads to

$$\mathbb{E} \left[(F_p^{\text{Pert}}(\lambda))^2 \right] \leq \mathbb{E} \left\{ \frac{4}{p^2} \left[\frac{16p^2}{\Delta^2} \sigma_{\max}(X)^4 + \frac{1}{\Delta^2} \|z\|^4 + \frac{9\lambda \epsilon_p}{4} p^2 + p\lambda \epsilon_p \|Z\|^2 \right] \right\} = O(1).$$

By applying lemma 14 and choosing $D_1 = p^\gamma, D_2 = O(1), K_p = O(p^{-(1-\gamma)/2})$ (accordingly $\mathbb{P}(\mathcal{G}^c) = O(pe^{-p^\gamma/4})$), for any $r \geq O(e^{-C_1 p^\gamma}) \geq O(p^{(5+\gamma)/4} e^{-p^\gamma/8})$,

$$\mathbb{P}(|F_p^{\text{Pert}}(\lambda) - \mathbb{E}[F_p^{\text{Pert}}(\lambda)|\beta_0]| \geq r|\beta_0) \leq O(e^{-Cr^2 p^{1-\gamma}}) + O(pe^{-p^\gamma/4}).$$

Finally, the choice $\gamma = 1/2$ renders the optimal bound

$$\mathbb{P}(|F_p^{\text{Pert}}(\lambda) - \mathbb{E}[F_p^{\text{Pert}}(\lambda)|\beta_0]| \geq r|\beta_0) \leq e^{-C_2 r^2 p^{1/2}}.$$

□

We will see that, with an integral argument, lemma 21 further implies the following lemma, which gives an upper bound on the conditional variance of $F_p^{\text{Pert}}(\lambda)$ giving any fixed β_0 .

Lemma 22. *For any fixed $\beta_0 \in S^{p-1}(\sqrt{p})$,*

$$\mathbb{E} \left[(F_p^{\text{Pert}}(\lambda) - \mathbb{E}[F_p^{\text{Pert}}(\lambda)|\beta_0])^2 \middle| \beta_0 \right] = O\left(\frac{1}{p}\right).$$

Proof: By Lemma 21, there exist positive constants C_1, C_2 , and C_3 , depending only on Δ and α , such that for any $r \geq t_0 := C_1 e^{-C_2 p^{1/2}}$, we have for any $\beta_0 \in S^{p-1}(\sqrt{p})$,

$$\mathbb{P} \left(\left| F_p^{\text{Pert}}(\lambda) - \mathbb{E} F_p^{\text{Pert}}(\lambda) \right| > r \middle| \beta_0 \right) \leq e^{-C_3 r^2 p^{1/2}}.$$

Let $V = |F_p^{\text{Pert}}(\lambda) - \mathbb{E} F_p^{\text{Pert}}(\lambda)|/p$.

$$\begin{aligned} \mathbb{E} [V^2 | \beta_0] &= \mathbb{E} [V^2 \mathbf{1}_{V \leq t_0} | \beta_0] + \mathbb{E} [V^2 \mathbf{1}_{V > t_0} | \beta_0] \\ &\leq t_0^2 \mathbb{P}(V \leq t_0 | \beta_0) + \int_{t_0^2}^{\infty} s \mathbb{P}(V^2 > s | \beta_0) ds \\ &\leq t_0^2 + \int_{t_0^2}^{\infty} s e^{-C_3 s p^{1/2}} ds \\ &= t_0^2 + \frac{1}{C_3 p^{1/2}} (t_0^2 + \frac{1}{C_3 p^{1/2}}) e^{-C_3 p^{1/2} t_0^2} \\ &= O\left(\frac{1}{p}\right), \end{aligned}$$

where the last equality is because of $t_0^2 = C_1^2 e^{-2C_2 p^{1/2}} \ll p^{-1/2}$. □

C.3 Concentration of the free energy with respect to the signal

In this subsection, we state and prove Lemma 23, which is an upper bound on $\mathbb{E} [F_p^{\text{Pert}}(\lambda) | \beta_0]$, or equivalently, we prove that free energy of the perturbed model concentrates with respect to the randomness of β_0 .

Lemma 23.

$$\mathbb{E} \left[\left(\mathbb{E} [F_p^{\text{Pert}}(\lambda) | \beta_0] - \mathbb{E} F_p^{\text{Pert}}(\lambda) \right)^2 \right] = O\left(\frac{1}{p}\right).$$

Proof: By Jensen's inequality, we have

$$\frac{1}{p} \langle \mathcal{H}_\lambda^{\text{Pert}}(\tilde{\beta}_0) - \mathcal{H}_\lambda^{\text{Pert}}(\beta_0) \rangle_{\tilde{\beta}_0} \leq \frac{1}{p} (\ln \mathcal{Z}^{\text{Pert}}(\beta_0) - \ln \mathcal{Z}^{\text{Pert}}(\tilde{\beta}_0)) \leq \frac{1}{p} \langle \mathcal{H}_\lambda^{\text{Pert}}(\tilde{\beta}_0) - \mathcal{H}_\lambda^{\text{Pert}}(\beta_0) \rangle_{\beta_0},$$

where $\langle f(\beta) \rangle_{\beta_0}$ refers to the expectation of $f(\beta)$ under the distribution proportional to $d\pi(\beta) e^{-\mathcal{H}^{\text{Pert}}(\beta, \beta_0)}$ and similarly for $\langle f(\beta) \rangle_{\tilde{\beta}_0}$. Note that $\mathcal{H}^{\text{Pert}}(\beta, \beta_0)$ is the same as $\mathcal{H}^{\text{Pert}}(\beta)$, only to emphasize the different values of β_0 . Moreover,

$$\begin{aligned} &\left| \mathbb{E}_{X,z,Z} (\mathcal{H}_\lambda^{\text{Pert}}(\beta, \tilde{\beta}_0) - \mathcal{H}_\lambda^{\text{Pert}}(\beta, \beta_0)) \right| \\ &\leq \mathbb{E}_{X,z,Z} \left\{ \frac{1}{2\Delta} \left| \|X(\beta_0 - \beta) + z\|^2 - \|X(\tilde{\beta}_0 - \beta) + z\|^2 \right| + \frac{1}{2} \left| \lambda_0 \epsilon_p (\beta_0 - \tilde{\beta}_0)^\top \beta \right| \right\} \\ &\leq \mathbb{E}_{X,z,Z} \left\{ \frac{1}{2\Delta} \left(\|X(\beta_0 - \tilde{\beta}_0)\|^2 + 2 \left| \langle X(\beta_0 - \tilde{\beta}_0), X(\tilde{\beta}_0 - \beta) + z \rangle \right| \right) + O(\sqrt{p}) \|\beta_0 - \tilde{\beta}_0\| \right\} \\ &= O(\sqrt{p}) \|\beta_0 - \tilde{\beta}_0\|. \end{aligned}$$

Upon defining

$$f\left(\frac{1}{\sqrt{p}}\beta_0\right) := \frac{1}{p} \mathbb{E}_{X,z,Z} \ln \mathcal{Z}^{\text{Pert}}(\beta_0) = \mathbb{E}_{X,z,Z} F_p^{\text{Pert}},$$

we see that there exists a positive constant C such that

$$\left| f\left(\frac{1}{\sqrt{p}}\beta_0\right) - f\left(\frac{1}{\sqrt{p}}\tilde{\beta}_0\right) \right| \leq C \left\| \frac{1}{\sqrt{p}}\beta_0 - \frac{1}{\sqrt{p}}\tilde{\beta}_0 \right\|.$$

Finally, Lévy's lemma, i.e. Lemma 16, implies

$$\mathbb{P} \left(\left| \mathbb{E}_{X,z,Z} F_p^{\text{Pert}} - \mathbb{E}_{X,z,Z,\beta_0} F_p^{\text{Pert}} \right| > Ct \right) \leq e^{\pi - pt^2/4},$$

for any $t > 0$, which further leads to

$$\mathbb{E} \left[\left(\mathbb{E} [F_p^{\text{Pert}} | \beta_0] - \mathbb{E} F_p^{\text{Pert}} \right)^2 \right] \leq \frac{4C^2 e^\pi}{p} = O \left(\frac{1}{p} \right).$$

□

C.4 Completing the proof of Theorem 5

Proof of Theorem 5: By Lemma 21 and 23,

$$\begin{aligned} \text{Var}(F_p^{\text{Pert}}(\lambda)) &= \mathbb{E} [\text{Var}(F_p^{\text{Pert}}(\lambda) | \beta_0)] + \text{Var}(\mathbb{E}[F_p^{\text{Pert}}(\lambda) | \beta_0]) \\ &\leq O \left(\frac{1}{p} \right) + O \left(\frac{1}{p} \right) = O \left(\frac{1}{p} \right). \end{aligned}$$

Recall that

$$\frac{d\mathbb{E}F_p^{\text{Pert}}}{d\lambda_{0,p}} = \frac{1}{p} \mathbb{E} \langle \mathcal{H}' \rangle = \frac{1}{2} \mathbb{E} \langle R_{1,2} \rangle,$$

which gives

$$\left| \frac{d\mathbb{E}F_p^{\text{Pert}}}{d\lambda_0} \right| = \left| \epsilon_p \frac{d\mathbb{E}F_p^{\text{Pert}}}{d\lambda_{0,p}} \right| = \left| \frac{\epsilon_p}{p} \mathbb{E} \langle \mathcal{H}' \rangle \right| = \left| \frac{\epsilon_p}{2} \mathbb{E} \langle R_{1,2} \rangle \right| \leq \frac{\epsilon_p}{2}.$$

Finally, noting that when $\lambda_0 = 0$, the perturbed model would be the same as our original linear regression model. Thus,

$$\left| \mathbb{E} F_p^{\text{Pert}}(\lambda_0) - \mathbb{E} F_p \right| = \left| \mathbb{E} F_p^{\text{Pert}}(\lambda_0) - \mathbb{E} F_p^{\text{Pert}}(0) \right| \leq \frac{\lambda_0 \epsilon_p}{2} = O(\epsilon_p),$$

since $\lambda_0 \in [1/2, 1]$. Together with Lemma 20, the proof is therefore complete. □