THE TAP FREE ENERGY FOR HIGH-DIMENSIONAL LINEAR REGRESSION

BY JIAZE QIU$^1$, SUBHABRATA SEN$^2$

$^1$Department of Statistics, Harvard University, jiazeqiu@g.harvard.edu
$^2$Department of Statistics, Harvard University, subhabratasen@fas.harvard.edu

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. Our representation holds when the variance of the additive noise is sufficiently large, which corresponds to a high-temperature condition in statistical physics. This rigorously establishes the Thouless-Anderson-Palmer (TAP) approximation arising from spin glass theory, and proves a conjecture of [23] in the special case of the spherical prior (at sufficiently high temperature).

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 [33, 11]. 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 [24].

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. Throughout, we set \( \varepsilon = (\varepsilon_1, \cdots, \varepsilon_n) \top \in \mathbb{R}^n \). Given a prior distribution \( \pi_0 \) for \( \beta \), setting \( y = (y_1, \cdots, y_n) \in \mathbb{R}^n \) and \( X^\top = [x_1, \cdots, x_n] \in \mathbb{R}^{p \times n} \), the scientist constructs the posterior

\[
\frac{d\mathbb{P}}{d\pi_0}(\beta | y, X) = \frac{1}{Z_p} e^{-\frac{1}{2\Delta} \|y - X\beta\|^2},
\]

where \( Z_p \) is the normalizing constant of the posterior distribution. Borrowing terminology from statistical physics, we will refer to \( Z_p \) as the partition function of the model.

Using the classical Gibbs variational principle [33], one has

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

MSC2020 subject classifications: Primary 60F99, 62C10; secondary 82B44.

Keywords and phrases: Linear regression, TAP approximation, spin glasses.
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 a valid lower bound:

$$\log Z_p \geq \sup_{Q=\Pi_i Q_i} \left( E_Q \left[ -\frac{1}{2\Delta} \| y - X\beta \|^2 \right] - D_{KL}(Q\|\pi_0) \right).$$

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

$$\log Z_p - \sup_{Q=\Pi_i Q_i} \left( E_Q \left[ -\frac{1}{2\Delta} \| y - X\beta \|^2 \right] - D_{KL}(Q\|\pi_0) \right) = o(p).$$

In recent joint work with Sumit Mukherjee [26], 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 $\pi_0 = \prod_{i=1}^p \pi$ is a product distribution. In this setting, it is easy to see that the supremum in (2) is obtained by $Q^* = \Pi_i Q_i^*$, with $\frac{dQ^*_i}{d\pi}(\beta_i) \propto \exp \left( -\frac{(\beta_i - R_i)^2}{2\Sigma_i^2} \right)$ (here $R_i$ and $\Sigma_i$ are variational parameters which are subsequently optimized). The approximation (2) can thus be equivalently parametrized in terms of $a_i = E_{Q_i^*}[\beta_i]$ and $c_i = \text{Var}_{Q_i^*}[\beta_i]$, so that

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

if the NMF approximation is accurate to leading order.

In the special case where $x_i$’s are iid standard gaussian, [26] 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 [23]. In fact, using the Thouless-Anderson-Palmer (TAP) formulation, one obtains the following prediction for the log-partition function.

**Conjecture 1.1 ([23]).** Assume $\pi_0 = \prod_{i=1}^p \pi$, where $\pi$ is a probability distribution on $[-1, 1]$. Then

$$\log Z_p = \max_{a, c, \Delta} \left[ -\frac{\| y - Xa \|^2}{2\Delta} - \frac{n}{2} \log \left( 1 + \frac{\bar{c}}{\Delta} \right) + \sum_{i=1}^p \log \mathbb{E}_{\pi} \left[ \exp\left( -\frac{(\beta_i - R_i)^2}{2\Sigma_i^2} \right) \right] \right] + o(p),$$

where the parameters $R_i$, $c_i$ and $\Sigma_i$ are specified as follows: define $(dQ_i/d\pi)(\beta_i) \propto \exp(- (\beta_i - R_i)^2/(2\Sigma_i^2))$. $R_i$ is chosen such that $E_{Q_i}[\beta_i] = a_i$. Set $c_i = \text{Var}_{Q_i}[\beta_i]$ and let $\bar{c} = 1/p \sum_{i=1}^p c_i$. Finally, the parameter $\Sigma_i$ depends on $\Delta$ and $\alpha$, 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 $\pi$ 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 [7]. 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 \( \beta \). 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 \( \pi_0 \) 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 \( n/p \to \alpha \in (0, \infty) \). 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

\[
Z_p = \int_{S^{p-1}(\sqrt{p})} e^{-\frac{\alpha}{2} \|y - X \beta\|^2} d\pi_0(\beta).
\]

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

\[
F_p(y, X) = \frac{1}{p} \log Z_p(y, X)
\]

as the free energy and

\[
\mathcal{H}(\beta) = \frac{1}{2\Delta} \|y - X \beta\|^2
\]

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 \( p \to \infty \),

\[
\left| F_p - \sup_{\alpha \in \mathbb{R}^p, \| \alpha \| \leq \sqrt{p}} f_{\text{TAP}}(\alpha) \right| \xrightarrow{p} 0,
\]

where

\[
f_{\text{TAP}}(a) := -\frac{1}{2\Delta p} \|y - X a\|^2 - \frac{\alpha}{2} \log \left( 1 + \frac{1 - \|a\|^2/p}{\Delta \alpha} \right) + \frac{1}{2} \log \left( 1 - \frac{\|a\|^2}{p} \right).
\]
**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 $\alpha$ 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 $\alpha$, with $\Delta > 0$ sufficiently large, but note that all results transfer immediately to the fixed $\Delta$ large $\alpha$ 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 [30], 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 [13], cavity method [29], 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 [22].

The progress on the TAP formula for the free energy is more recent. [14] verified the TAP representation by evoking the Parisi formula. More recently, direct proofs of the TAP representation have been derived in the literature. [10] 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 [15, 16]. 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 [15, 16]. 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 [27, 8].

The behavior of the TAP free energy has been examined in the setting of finite rank spiked matrix models in [19]. [19] 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 [17, 12], 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 [19].

1.3. Proof strategy. Our proof has two separate arguments—first, we lower bound the log-partition function $\log Z_p$ 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 $Z_p$ 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 $[\log(1 - \|a\|^2/p)]/2$—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 $-(\alpha/2)\log[1 + (1 - \|a\|^2/p)/(\Delta \alpha)]$ 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 $(\log Z_p)/p \geq f_{\text{TAP}}(a)$ 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 [10] 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 [6]. 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 $\beta^1, \ldots, \beta^m$ 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 [28] 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 $\log Z_p$ 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 $(\log Z_p)/p \leq f_{\text{TAP}}(a) + o(1)$ 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. [9]). 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 in our lower bound argument to calculate the free energy on the band. This condition ensures that every band is at a high-temperature. In principle, using more sophisticated Random Matrix Theory (RMT) ideas [4], it is possible to calculate the free energy of the system restricted to the band, even if the
system is at a low temperature. We refrain from exploring this avenue in this paper for two reasons—first, the results in [4] suggest that the form of the TAP correction would change if the corresponding band is at a low temperature. In this scenario, our current upper bound would no longer match the lower bound derived. We believe that the lower bound is actually tight, and the upper bound should be improved to match this approximation. However, there is an additional and perhaps more fundamental reason why we do not explore this direction. Physicists expect that for these planted models, it should always be possible to restrict only to bands at high-temperature. This is related to Plefka’s condition in the context of the SK model (see e.g. the discussion in [10]). Our current techniques fall short in this regard. We believe this would be an important issue to resolve in future investigations.

(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 [20] 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 $f_{\text{TAP}}$ directly by some out-of-the-box optimization algorithm such as gradient descent. Of course, $f_{\text{TAP}}$ 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 [18, 12]. 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 $X_p \overset{P}{\to} 0$ as $p \to \infty$.

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

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

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

$$\lim_{p \to \infty} \mathbb{P}((X, \epsilon, \beta_0) \in \Omega) = 1.$$
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 Appendix C.

2. Lower bound on the free energy. To get a lower bound of the free energy \( F_p = \frac{1}{p} \log Z_p \), 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 \( \eta > 0 \), \( e \in (0, 1) \),

\[
\mathbb{P}\left( \frac{1}{p} \log Z_p \leq \sup_{a \in \mathbb{R}^p, \|a\| \leq (1-e)\sqrt{p}} f_{\text{TAP}}(a) - \eta \right) \rightarrow 0.
\]

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 \( e \) and \( a \in \mathbb{R}^p \), define

\[
B(a, e) = \{ \beta \in S^{p-1}(\sqrt{p}) : v_i^\top (\beta - a) \in (-pe, pe), \forall i \in \{1, 2\} \}.
\]

There exists \( \Delta_0 > 0 \) such that for all \( \Delta > \Delta_0 \), for any \( e \in (0, 1) \), there exists \( C > 0 \) such that for any \( e > 0 \), as \( p \rightarrow \infty \),

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

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

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

\[
\|X^\top (y - Xa)\|^2 \leq \sigma_{\max}^2(X) \|X(\beta_0 - a) + \epsilon\|^2
\]

(10)

\[
\leq \sigma_{\max}^2(X) \left[ 4p\sigma_{\max}^2(X) + \|\epsilon\|^2 + 4\sqrt{p}\sigma_{\max}(X)\|\epsilon\| \right] = O(p).
\]

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

\[
Z_p = \int_{S^{p-1}(\sqrt{p})} e^{-\frac{1}{2\alpha}\|y - X\beta\|^2} d\pi_0(\beta) \geq \int_{B(a, e)} e^{-\frac{1}{2\alpha}\|y - X\beta\|^2} d\pi_0(\beta).
\]

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 \( B(a, e) \), together with (10), we have

\[
\|y - X\beta\|^2 = \|y - Xa\|^2 + \|X(\beta - a)\|^2 - 2(X^\top (y - Xa))^\top (\beta - a)
\]

\[
= \|y - Xa\|^2 + \|X(\beta - a)\|^2 + 2\epsilon O(p)
\]
for any $\beta \in B(a, \epsilon)$. Thus we have

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

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

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

Note that for any 2-dimensional linear subspace $L \subset \mathbb{R}^p$, the projection $\beta_{\text{proj}}$ of $\beta$ onto $L$ divided by $\sqrt{p}$ has density

$$f_{\text{proj}}(x) = \frac{1}{\pi} \frac{\Gamma \left( \frac{p}{2} \right)}{\Gamma \left( \frac{p-2}{2} \right)} (1 - \|x\|^2)^{\frac{p-4}{2}} = \frac{p-2}{2\pi} (1 - \|x\|^2)^{\frac{p-4}{2}}, x \in \mathbb{R}^2, \|x\| \leq 1$$

with respect to the Lebesgue measure on $L$ (see for instance of [10, (2.8)]). Note that with a slight abuse of notation, we may treat $f_{\text{proj}}$ as a function of $\|x\|$ as well. Without loss of generality, we only consider the non-degenerate case, namely when $a$ and $X^T (y - Xa)$ are not linearly dependent. Recall that $\text{Vol}(a, \epsilon)$ is defined as the volume of $B(a, \epsilon)$ on the sphere $S^{p-1}(\sqrt{p})$. Thus without of loss of generality, assuming $\sqrt{2} \epsilon < c$ and noting $\beta_{\perp} := \beta - \beta_{\text{proj}}$ is orthogonal to both $v_1$ and $v_2$,

$$\frac{1}{p} \log \text{Vol}(a, \epsilon) = \frac{1}{p} \log \mathbb{P}_{\beta \sim \pi_0} (\beta \in B(a, \epsilon))$$

$$= \frac{1}{p} \log \mathbb{P}_{\beta \sim \pi_0} \left( \|v_i^T (\beta_{\text{proj}} - a)\| < p\epsilon, \forall i \in \{1, 2\} \right)$$

$$\geq \frac{1}{p} \log \left( 4\epsilon^2 \cdot f_{\text{proj}} \left( \frac{\|a\|}{\sqrt{p}} + \sqrt{2} \epsilon \right) \right)$$

$$= \frac{1}{2} \log \left( 1 - \frac{\|a\|^2}{p} \right) + o(1),$$

where (i) follows upon observing that $f_{\text{proj}}(x)$ is decreasing in $\|x\|$ and $\|\beta_{\text{proj}}/\sqrt{p}\| \leq \|a\|/\sqrt{p} + \sqrt{2} \epsilon$ for any $\beta \in B(a, \epsilon)$. 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 $\Omega_p$,

$$\mathbb{P} \left( \frac{1}{p} \log Z_p \geq \sup_{a \in \mathbb{R}^p, \|a\| \leq (1 - c) \sqrt{p}} f_{\text{TAP}}(a) - 2(C + C') \epsilon + o(1) \right)$$

$$\geq \mathbb{P} \left( \left\{ \frac{1}{p} \log Z_p \geq \sup_{a \in \mathbb{R}^p, \|a\| \leq (1 - c) \sqrt{p}} f_{\text{TAP}}(a) - 2(C + C') \epsilon + o(1) \right\} \cap \Omega \right)$$

$$\geq \mathbb{P} \left( \left\{ \frac{1}{p} \log Z_p \geq -\frac{1}{2\Delta} \|y - Xa\|^2 - C' \epsilon + \log \left( \int_{B(a, \epsilon)} e^{-\frac{1}{2\Delta} \|X(\beta - a)\|^2} d\pi_0(\beta) \right), \right. \right.$$

$$\forall a \in \mathbb{R}^p : \|a\| \leq (1 - c) \sqrt{p} \left. \right\} \cap \Omega \cap \Omega_p \right)$$

$$= \mathbb{P}(\Omega \cap \Omega_p) \rightarrow 1,$$
where $C$ is the one in statement of Theorem 3. In other words, since $\epsilon$ can be chosen to be arbitrarily small, we have for any $\eta > 0$,

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

\[ \text{\Box} \]

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 $\beta^T X^T X \beta$ instead of $\|X\beta\|^2$ in this subsection since we would like to emphasize the fact that our proofs depend strongly on the properties of the sample covariance matrix $X^T X$.

**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 $p \to \infty$

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

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 $p \to \infty$

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

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 $p \to \infty$

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

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 $\Omega$ 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 $\beta = \beta - a$. Define

$$H(\hat{\beta}) := \frac{1}{2\Delta} \hat{\beta}^T X^T X \hat{\beta} = \frac{1}{2\Delta} (\beta - a)^T X^T X (\beta - a).$$
Note that $H$ is different from the Hamiltonian $\mathcal{H}$ defined in (6). By (10), $\|X^\top (y - X a)\| = O(\sqrt{p})$. By the definition of $B(a, \epsilon)$ in (9), for any $\beta \in B(a, \epsilon)$, one has
\[
|a^\top \beta| = |a^\top (\beta - a)| < c'p\epsilon
\]
and
\[
|(X^\top (y - X a))^\top \beta)| = |(y - X a)^\top X(\beta - a)| < c'p\epsilon,
\]
for some positive constant $c'$. Therefore
\[
\|\beta\|^2 = \|\beta\|^2 - \|a\|^2 - 2(\beta - a)^\top a = p - \|a\|^2 + O(p\epsilon)
\]
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 $\beta^{\perp} = \sqrt{p}\hat{\beta}^{\perp}/\|\beta^{\perp}\|$ and $\gamma = \|\beta^{\perp}\|/\sqrt{p}$. Note that, again by the definition of $B(a, \epsilon)$
\[
\|\hat{\beta} - \gamma\beta^{\perp}\| = \|\beta - \beta^{\perp}\| < c'\sqrt{p}\epsilon,
\]
and
\[
\gamma = \hat{\gamma} + O(\epsilon),
\]
where $\hat{\gamma} := \sqrt{p} - \|a\|^2/\sqrt{p}$. Therefore,
\[
H(\beta) = H(\gamma\beta^{\perp}) + O(p\epsilon) = H(\hat{\gamma}\beta^{\perp}) + O(p\epsilon),
\]
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
\[
\int_{B(a, \epsilon)} e^{-\frac{1}{2\Delta}(\beta - a)^\top X^\top X(\beta - a)} d\pi_0(\beta) = \mathbb{E}_{\pi_0}[1_{B(a, \epsilon)}] e^{-H(\beta^{\perp}) + O(p\epsilon)}.
\]
Since $\beta^{\perp}$ is uniformly distributed on $S^{p-1}(\sqrt{p})$ intersected with $\text{Span}(a, X^\top (y - X a))^{\perp}$ under $\pi_0$ and $\gamma$ is independent of $1_{\beta \in B(a, \epsilon)}$, one has
\[
\begin{align*}
\mathbb{E}_{\pi_0}[1_{B(a, \epsilon)}] e^{-H(\beta^{\perp}) + O(p\epsilon)} \\
= e^{O(p\epsilon)} e_{\pi_0}[1_{B(a, \epsilon)}] \int_{S^{p-1}(\sqrt{p} - \|a\|^2)} e^{-\frac{1}{2\Delta}\beta^\top X^\top X\beta} d\pi(\{v_1, v_2\}^\perp (\hat{\gamma}^{-1}\beta)) \\
= e^{O(p\epsilon)} \text{Vol}(a, \epsilon) \int_{S^{p-1}(\sqrt{p}) \cap \text{Span}(v_1, v_2)^\perp} e^{-\frac{1}{2\Delta}\beta^\top X^\top X\beta} d\pi(\{v_1, v_2\}^\perp (\beta)) \\
= e^{O(p\epsilon)} \text{Vol}(a, \epsilon) \mathbb{E}_{\text{Span}(v_1, v_2)} \left[ \exp \left( -\frac{1}{2\Delta} \beta^\top X^\top X\beta \right) \right] ,
\end{align*}
\]
where
\[
\hat{\Delta} = \frac{p}{p - \|a\|^2} \Delta.
\]
Note that for any $a$ such that $\|a\| \leq (1 - \epsilon)\sqrt{p}$, we have $\hat{\Delta} > \Delta > \Delta_0$. By Lemma 3 and the fact that $\mathbb{P}(\Omega) \to 1$ as $p \to \infty$, we finally get with high probability
\[
\frac{1}{p} \log \int_{B(a, \epsilon)} e^{-\frac{1}{2\Delta}(\beta - a)^\top X^\top X(\beta - a)} d\pi_0(\beta) \\
= O(\epsilon) + \frac{1}{p} \log \text{Vol}(a, \epsilon) - \frac{\alpha}{2} \log (1 + \frac{(1 - \|a\|^2/p)}{\Delta \alpha}) + o(1),
\]
in which the $o(1)$ terms are uniform with respect to the choice of $\hat{\Delta}$, equivalently, with respect to the choice of $a$. \qed
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 \( Z \) 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 \( Z_p \) defined in (5). Let

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

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

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

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 \( \Delta, \alpha, \delta \)) such that for any \( p > p_0 \)

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

**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

\[
K_p = \frac{1}{\Delta}(1 + \alpha^{-1/2} + t/\sqrt{n}).
\]

By triangle inequality, for any \( X_1, X_2 \in \mathcal{G} \),

\[
|H(X_1, \beta) - H(X_2, \beta)| \leq \frac{1}{2\Delta} \left( \|\beta^\top (X_1 - X_2)^\top X_2\beta\| + \|\beta^\top X_1^\top (X_1 - X_2)\beta\| \right)
\]

\[
\leq \frac{1}{2\Delta} \|\beta\|^2 \left[ \|(X_1 - X_2)^\top X_2\|_2 + \|X_1^\top (X_1 - X_2)\|_2 \right]
\]

\[
\leq \frac{\sqrt{p}}{2\Delta} \left[ (\sigma_{\max}(X_1) + \sigma_{\max}(X_2)) \sqrt{p}X_1 - \sqrt{p}X_2 \right]_2
\]

\[
\leq \frac{\alpha^{-1/2}}{\Delta} (\sqrt{p} + \sqrt{n} + t) \sqrt{p}X_1 - \sqrt{p}X_2 \|_F,
\]

which implies

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

\[
= \frac{\alpha^{-1/2}(1 + \sqrt{\alpha} + t/\sqrt{p})}{\Delta\sqrt{p}} \sqrt{p}X_1 - \sqrt{p}X_2 \|_F.
\]
On the other hand, by Lemma 15, \( \mathbb{P}(G^c) \leq 2e^{-t^2} \). Note that \( \Phi(\Delta, 0) = 0 \) and

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

Therefore, by Lemma 14, there exists a positive constant \( C \) such that for any \( r > 6(C + \sqrt{npK_p}) \sqrt{\mathbb{P}(G^c)} \) where \( K_p := \tilde{K}_p / \sqrt{p} \) we have

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

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

\[
\lim_{p \to \infty} \phi_p(\Delta) = -\frac{\alpha}{2} \log(1 + \frac{1}{\Delta\alpha}).
\]

**Proof.** We compute the LHS explicitly, namely

\[
\phi_p(\Delta) = \frac{1}{p} \log \mathbb{E}_X \mathbb{E}_\beta e^{-\frac{1}{2\Delta} \beta^T X^T X \beta} = \frac{1}{p} \log \mathbb{E}_\beta \mathbb{E}_X e^{-\frac{1}{2\Delta} \text{tr}(X^T X \beta^T \beta)}
\]

\[
= \frac{1}{p} \log \mathbb{E}_\beta \mathbb{E}_X e^{-\frac{1}{2\Delta} \|\beta\|^2(X^T X)} = \frac{1}{p} \log \mathbb{E}_X e^{-\frac{1}{2\Delta} \text{tr}(X^T X)} = \frac{1}{p} \log \mathbb{E}_V \sim \chi^2(n) e^{-\frac{1}{2\Delta} V}.
\]

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

\[
\phi_p(\Delta) = \frac{1}{p} \log((1 + \frac{p}{\Delta n})^{-n/2}) \to -\frac{\alpha}{2} \log(1 + \frac{1}{\Delta\alpha}).
\]

\( \square \)

A point worth noting is that there is no non-trivial restriction on the value of \( \Delta \). Now we proceed to use second moment method to show \( \Phi(\Delta, X) \) also converges to the same quantity.

Let \( \sigma \) and \( \beta \) be independent samples from the prior distribution \( \pi_0 \). Then

\[
\mathbb{E}(Z_p(\Delta, X)^2) = \mathbb{E}_X \mathbb{E}_\sigma e^{-\frac{1}{2\Delta} H(\sigma)} \mathbb{E}_\beta e^{-\frac{1}{2\Delta} H(\beta)} = \mathbb{E}_X e^{-\frac{1}{2\Delta} (H(\sigma) + H(\beta))}
\]

\[
= \mathbb{E}_X \mathbb{E}_\beta \mathbb{E}_\sigma e^{-\frac{1}{2\Delta} H(\sigma + \beta)} = \mathbb{E}_X \mathbb{E}_\beta \mathbb{E}_\sigma \exp \left( -\frac{1}{2\Delta} \text{tr} \left\{ X^T X \left( \sigma \sigma^T + \beta \beta^T \right) \right\} \right)
\]

\[
= \mathbb{E}_X \mathbb{E}_\beta \mathbb{E}_\sigma \exp \left( -\frac{1}{2\Delta} \left\{ (p + \beta^T \sigma)(X^T X)_{11} + (p - \beta^T \sigma)(X^T X)_{22} \right\} \right),
\]
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,

\[
\mathbb{E}(Z_p(\Delta, X)^2) = \mathbb{E}_V \mathbb{E}_{\beta, \sigma} \exp \left( \frac{-1}{2\Delta} \left\{ (p + \beta^\top \sigma)V_1 + (p - \beta^\top \sigma)V_2 \right\} \right)
\]

\[
= \mathbb{E}_{\beta, \sigma} \mathbb{E}_V \exp \left( \frac{-1}{2\Delta} (\lambda_1 V_1 + \lambda_2 V_2) \right).
\]

Again, using MGF of \( \chi^2 \) distribution, we further have

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

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 \( \gamma_0 \).

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

\[
\lim_{p \to \infty} \frac{1}{p} \log \gamma_0 = 0.
\]

**PROOF.** Due to symmetry, \( \beta^\top \sigma \equiv \Sigma_{i=1}^p \beta_i \), equivalently, we can assume \( \sigma = (1, 1, \ldots, 1) \) without loss of generality. By Lemma 16, letting \( Q = \beta^\top \sigma/p \in (-1, 1) \), for \( t \in (0, 1) \),

\[
\mathbb{P}_{\beta, \sigma}(Q > t) < e^{\pi - \mu^2/4}.
\]

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

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

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

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

for any \( t \in (0, 1) \). Note that \( r \to 0 \) for any fixed \( \alpha \) as \( \Delta \to \infty \), and thus \( C_Q > 0 \) for \( \Delta > 0 \) sufficiently large. We have \( g(t) \leq -C_Q t^2/2 + g(0) = -C_Q t^2/2 - \alpha \log(1 + r) \). By (12),
integration by parts gives
\[
E(Z_p(\Delta, X)^2) = - \int_0^1 e^{pf(t)} d\mathbb{P}_{\beta, \sigma}(|Q| > t)
\]
\[
\leq - \left[ e^{pf(t)} \mathbb{P}_{\beta, \sigma}(|Q| > t) \right]_{t=0}^{t=1} - \int_0^1 [e^{pf(t)}] e^{-pt^2/4} dt
\]
\[
\leq e^{pf(0)} + e^\pi \int_0^1 \frac{p\alpha \gamma_0^2}{(1+r)^2 - t^2 r^2} e^{pf(t)} dt
\]
\[
\leq e^{pf(0)} + e^\pi \frac{p\alpha \gamma_0^2}{(1+r)^2 - r^2} \int_0^1 e^{pf(t)} dt
\]
\[
\leq e^{\pi - p\alpha \log(1+r)} \left[ 1 + Cp \int_0^1 e^{-pC_q t^2/2} dt \right]
\]
\[
\leq e^{\pi - p\alpha \log(1+r)} (C\sqrt{n} + 1).
\]
Thus
\[
1/p \log \gamma_0 \leq 1/p \log \left\{ \frac{4e^\pi (1+r)^{-n}(C\sqrt{n} + 1)}{(1+r)^{-n}} \right\} \to 0,
\]
as \(p \to \infty\). \(\square\)

For the readers’ convenience, we re-state [25, Lemma 4.1.1] here.

**Lemma 7.**
\[
\mathbb{P} \left( |\Phi_p(\Delta, X) - \phi_p(\Delta)| < \frac{1}{p} \log \gamma_0 \right) \geq \frac{1}{\gamma_0}.
\]

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 \(p > p_0\)
\[
\mathbb{P} \left( |\Phi_p(\Delta, X) - \mathbb{E}\Phi_p(\Delta, X)| \leq \delta \right) + \mathbb{P} \left( |\Phi_p(\Delta, X) - \phi_p(\Delta)| < q_p \right) \geq 1 - e^{C_p} + e^{q_p} > 1,
\]
where \(q_p := \frac{1}{p} \log \gamma_0 \to 0\) as \(p \to \infty\). It implies
\[
|\mathbb{E}\Phi_p(\Delta, X) - \phi_p(\Delta)| \leq \delta,
\]
which further implies
\[
\limsup_{p \to \infty} \left| \mathbb{E}\Phi_p(\Delta, X) - \phi_p(\Delta) \right| \leq \delta.
\]
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 \(w_1, w_2, \ldots, w_p\) be an orthonormal basis such that \(\text{Span}(u, v) = \text{Span}(w_{p-1}, w_p)\). Let \(\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{p-2}\}\) be the eigenvalues of \(Q\), where \(Q\) is the top left \((p-2)\) by \((p-2)\) minor of \(X^TX\) under this basis. Namely, let \(W = (w_1, w_2, \ldots, w_p) \in \mathbb{R}^{p \times p}, X = WX\) and \(\beta = W\hat{\beta}\), where \(\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)\), with
\( \tilde{\beta}_p = \tilde{\beta}_{p-1} = 0 \) for \( \beta \in \text{Span}(u,v) \). Equivalently, \( \tilde{X} \) and \( \tilde{\beta} \) are nothing but \( X \) and \( \beta \) under this new basis and \( Q = (\tilde{X}^\top \tilde{X})_{1:(p-2),1:(p-2)} \).

\[
\mathbb{E}^\text{Span}(u,v)^+ \left[ \exp(-H(\beta)) \right] = \mathbb{E}^\text{Span}(u,v)^+ \left[ \exp \left( -\frac{1}{2\Delta} \beta^\top X^\top X \beta \right) \right]
\]

\[
= \mathbb{E}_\beta^\text{Span}(u,v)^+ \left[ \exp \left( -\frac{1}{2\Delta} \tilde{\beta}^\top \tilde{X}^\top \tilde{X} \tilde{\beta} \right) \right]
\]

\[
= \mathbb{E}_\beta^\text{Span}(u,v)^+ \left[ \exp \left( \frac{1}{2\Delta} \sum_{i=1}^{p-2} \lambda_i \tilde{\beta}_i^2 \right) \right]
\]

\[
= \mathbb{E}_\beta^{p-2} \left[ \exp \left( \frac{1}{2\Delta} \sum_{i=1}^{p-2} \lambda_i \beta_i^2 \right) \right],
\]

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 \((p-2)\times(p-2)\) minor of \( X^\top X \) when written in the standard basis \( \{e_1,e_2,\ldots,e_p\} \). let \( s_1 \geq s_2 \geq \cdots \geq s_{p-2} \) be the corresponding eigenvalues of \( S \). For \( \beta \in \mathbb{R}^{p-2} \), by Lemma 1 we have

\[
\mathbb{E}_\beta^{p-2} e^{-\frac{1}{2\Delta} \beta^\top \beta} = e^{\rho(-\frac{n}{2}\log(1+\frac{1}{\Delta})+o(1))},
\]

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

\[
\mathbb{E}_\beta^{p-2} e^{-\frac{1}{2\Delta} \beta^\top \beta} = \mathbb{E}_\beta^{p-2} e^{-\frac{1}{\Delta} \sum_{i=1}^{p-2} s_i \beta_i^2}.
\]

By the eigenvalue interlacing inequality [21], we have almost surely

\[
\sup_{1=1,2,\ldots,p} |\lambda_i - s_i| \to 0.
\]

Therefore almost surely

\[
\sup_{\beta \in \mathbb{R}^{p-2}(\sqrt{p})} \frac{1}{p} \sum_{i=1}^{p-2} \lambda_i \beta_i^2 - \frac{1}{p} \sum_{i=1}^{p-2} s_i \beta_i^2 \leq \frac{\|\beta\|^2}{p} \sup_{1=1,2,\ldots,p} |\lambda_i - s_i| \to 0.
\]

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

\[
\mathbb{E}^\text{Span}(u,v)^+ e^{-\frac{1}{2\Delta} \beta^\top X^\top X \beta} = e^{\rho(-\frac{1}{2}\log(1+\frac{1}{\Delta})+o(1))}.
\]

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 \( p \to \infty \),

\[
\mathbb{P} \left( \frac{1}{p} \log \mathcal{Z}_p \geq \sup_{\alpha \in \mathbb{R}^p, \|\alpha\| \leq \sqrt{p}} f_{\text{TAP}}(\alpha) + \eta \right) \to 0.
\]
3.1. Perturbed model and overlap concentration. Assume that apart from \( X \) and \( y \), we observe the following additional side information

\[
y_{\text{Pert}} = \sqrt{\lambda_0} \epsilon \beta_0 + Z,
\]

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

\[
\frac{d\mathbb{P}}{d\pi_0}(\beta | y_{\text{Pert}}, y, X) = \frac{1}{Z_{\text{Pert}}} e^{-\mathcal{H}_p^{\text{Pert}}(\beta)},
\]

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

\[
Z_{\text{Pert}} := \int_{S_p^{-1}(\sqrt{p})} e^{-\mathcal{H}_p^{\text{Pert}}(\beta)} d\pi_0(\beta).
\]

Finally, the free energy of the perturbed model is

\[
P_{\text{Pert}}(\lambda_0) := \frac{1}{p} \log Z_{\text{Pert}}(y_{\text{Pert}}, y, X).
\]

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

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

\[
\frac{1}{p} \left| \mathbb{E} \log Z_{\text{Pert}} - \mathbb{E} \log Z_p \right| \leq O(\epsilon_p).
\]

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

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

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

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

**Lemma 8.** For any \( \Delta, \alpha > 0 \), as \( p \to \infty \), there exists \( C > 0 \) depending on \( \alpha \) and \( \Delta \) such that

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

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

\begin{equation}
S_p(m, c) = \left\{ (\beta^1, \ldots, \beta^m) : \frac{1}{p} (\beta^j)^\top \beta^j - \mathbb{E}(R_{1,2}) \leq c, \forall i \neq j \right\}.
\end{equation}

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

\begin{equation}
\frac{1}{p m_p} \log \left( \int e^{-\sum_i \mathcal{H}(\beta^i)} \mathbb{1}_{\{(\beta^1, \ldots, \beta^m) \in S\}} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) = \frac{1}{p} \log Z_p + o(1).
\end{equation}

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

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

\begin{equation}
\mathbb{E}( (R_{1,2} - \mathbb{E}(R_{1,2}))^2 \leq a_p,
\end{equation}

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

\begin{equation}
\mathbb{P}( (R_{1,2} - \mathbb{E}(R_{1,2}))^2 > b_p) \leq \frac{a_p}{b_p}.
\end{equation}

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

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

Using Chebychev inequality, we have, on $C_p^c$

\begin{equation}
\langle 1_{\{\langle R_{1,2} - \mathbb{E}(R_{1,2}) \rangle > b_p \}} \rangle \leq \frac{b_p}{c_p^2}.
\end{equation}

By union bound, on $C_p^c$

\begin{equation}
\langle 1_{S^c} \rangle \leq \sum_{i \neq j} \langle 1_{\{\langle R_{i,j} - \mathbb{E}(R_{i,j}) \rangle > b_p \}} \rangle \leq \sum_{i \neq j} \langle 1_{\{\langle R_{1,2} - \mathbb{E}(R_{1,2}) \rangle > b_p \}} \rangle \leq \frac{m_p^2 b_p}{c_p^2}.
\end{equation}

We choose $m_p \to \infty$ and $c_p \to 0$ such that the RHS above will converge to 0 as $p \to \infty$. This implies, on $C_p^c$,

\begin{equation}
1 \geq \langle S \rangle = \frac{1}{(Z_p^{\text{Pert}})^{m_p}} \int e^{-\sum_i \mathcal{H}(\beta^i)} \mathbb{1}_{\{(\beta^1, \ldots, \beta^m) \in 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).
\end{equation}

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

\begin{equation}
\mathbb{H}(\beta) - \mathbb{H}^{\text{Pert}}(\beta) = \mathbb{H}^{\text{Gauss}}(\beta)
\end{equation}

\begin{equation}
\leq \left| \lambda_0 e_p \beta_0^\top \beta \right| + \left| \sqrt{\lambda_0} e_p^\top Z \beta \right| + \left| \frac{1}{2} \lambda_0 e_p \| \beta \|^2 \right|
\end{equation}

\begin{equation}
\leq O(\sqrt{c_p}) \lambda_0 p.
\end{equation}

This implies that with high probability with respect to $Z$,

\begin{equation}
\log Z_p^{\text{Pert}} - \log Z_p \leq O(\sqrt{c_p}) \lambda_0 p.
\end{equation}
and

$$\frac{1}{pm_p} \log \left( \int e^{-\sum_i H(\beta^i) \mathbf{1}_{(\beta^1, \ldots, \beta^m) \in S}} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) \leq O(\sqrt{\epsilon_p}).$$

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

$$\frac{1}{pm_p} \log \left( \int e^{-\sum_i H(\beta^i) \mathbf{1}_{(\beta^1, \ldots, \beta^m) \in S}} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) = \frac{1}{p} \log Z_p + O(\sqrt{\epsilon_p}).$$

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

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

$$\frac{1}{m_p} \sum_{i=1}^{m_p} ||y - X\beta^i||^2 = ||y - X\tilde{a}||^2 + \frac{1}{m_p} \sum_{i=1}^{m_p} ||X(\beta^i - \tilde{a})||^2 - \frac{2}{m_p} (X^\top (y - X\tilde{a}))^\top \sum_{i=1}^{m_p} (\beta^i - \tilde{a})$$

$$= ||y - X\tilde{a}||^2 + \frac{1}{m_p} \sum_{i=1}^{m_p} ||X(\beta^i - \tilde{a})||^2,$$

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

$$\frac{1}{p} \log Z_p = \frac{1}{pm_p} \log \left( \int e^{-\frac{m_p}{2\Delta} ||y - X\tilde{a}||^2 - \frac{1}{2\Delta} \sum_{i=1}^{m_p} ||X(\beta^i - \tilde{a})||^2} \mathbf{1}_{(\beta^1, \beta^2, \ldots, \beta^m) \in S} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) + o(1)$$

$$\leq \frac{1}{pm_p} \log \left( \int e^{-\frac{m_p}{2\Delta} ||y - X\tilde{a}||^2 - \frac{1}{2\Delta} \sum_{i=1}^{m_p} ||X(\beta^i - \tilde{a})||^2} \mathbf{1}_{(\beta^1, \beta^2, \ldots, \beta^m) \in S} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) + o(1)$$

(29)

$$:= R_p + o(1).$$

In the display above, we drop one of the summands in $\sum_{i=1}^{m_p} ||X(\beta^i - \tilde{a})||^2$ at the expense of a $o(1)$ correction. This might appear ad-hoc to the reader at this point—in our subsequent calculations, we will realize the $\beta^i$‘s as normalized gaussians, and work with their empirical mean. This step ensures that the gaussians (apart from the first one) and their empirical mean have a non-degenerate distribution in $\mathbb{R}^p$—this will be useful for our calculations. 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 $\mathbb{E}[(R_{1,2})]$ converges to a positive constant, there exist $c' > 0$ and $c > 0$, $c + c' < 1$, such that as $k \to \infty$,

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

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 \( k \geq 1 \),
\[
\mathbb{P}
\left( \frac{1}{p_k} \log Z_{p_k} \geq \sup_{a \in \mathbb{R}^{p_k}, \|a\| \leq \sqrt{p_k}} f_{\text{TAP}}(a) + \eta \right) \geq \epsilon.
\]
We can extract a further subsequence such that \( \mathbb{E}[(R_{1,2})] \) converges along the subsequence. For notational simplicity, we refer to this subsequence as \( \{p_k : k \geq 1\} \) as well.

Consider first the case \( \lim \mathbb{E}[(R_{1,2})] > 0 \). Combining (29) and Lemma 10, we obtain that along this subsequence,
\[
\mathbb{P}
\left( \frac{1}{p_k} \log Z_{p_k} \geq \sup_{a \in \mathbb{R}^{p_k}, \|a\| \leq \sqrt{p_k}} f_{\text{TAP}}(a) + \eta \right) \to 0.
\]
This is a contradiction, and completes the proof in this case.

We now turn to the case where \( \lim \mathbb{E}[(R_{1,2})] = 0 \). In this case,
\[
\sup_{a \in \mathbb{R}^{p_k}, \|a\| \leq \sqrt{p_k}} f_{\text{TAP}}(a) \geq f_{\text{TAP}}(0) = -\frac{1}{2 \Delta p_k} \|y\|^2 - \frac{\alpha}{2} \log \left( 1 + \frac{1}{2 \Delta} \right).
\]
On the other hand, setting \( a = \sum_{i=1}^{m_p} \beta^i / m_p \), we note that for \( (\beta^1, \cdots, \beta^{m_p}) \in \mathcal{S} \), \( \|a\|^2 / p \leq 1 / m_p + c_p \), where we use that \( \lim \mathbb{E}[(R_{1,2})] = 0 \). Further, using (29), we have
\[
\frac{1}{p_k} \log Z_{p_k} \leq \frac{1}{pkm_p} \log \left( \int e^{-\frac{m_p}{2\Delta} \|y - Xa\|^2 - \frac{1}{2\Delta} \sum_{i=1}^{m_p} \|X(\beta^i - a)\|^2} 1_{\{\beta^1, \beta^2, \ldots, \beta^{m_p} \in \mathcal{S}\}} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) + o(1)
\]
\[
\leq \frac{1}{pkm_p} \log \left( \int e^{-\frac{m_p}{2\Delta} \|y\|^2 - \frac{1}{2\Delta} \sum_{i=1}^{m_p} \|X\beta^i\|^2} 1_{\{\beta^1, \beta^2, \ldots, \beta^{m_p} \in \mathcal{S}\}} \prod_{i=1}^{m_p} d\pi_0(\beta^i) \right) + o(1),
\]
where we have used the fact that \( \|a\| = o(\sqrt{p}) \). In turn, this implies
\[
\frac{1}{p_k} \log Z_{p_k} \leq -\frac{\|y\|^2}{2\Delta p_k} + \frac{m_p - 1}{pkm_p} \log \left( \int \exp \left( -\frac{1}{2\Delta} \|X\beta\|^2 \right) d\pi_0(\beta) \right) + o(1)
\]
\[
\leq -\frac{\|y\|^2}{2\Delta p_k} - \frac{\alpha}{2} \log \left( 1 + \frac{1}{2 \Delta} \right) + o(1),
\]
where the last step uses Lemma 1. Combining this bound with (30) we immediately obtain that along this subsequence,
\[
\mathbb{P}
\left( \frac{1}{p_k} \log Z_{p_k} \geq \sup_{a \in \mathbb{R}^{p_k}, \|a\| \leq \sqrt{p_k}} f_{\text{TAP}}(a) + \eta \right) \to 0
\]
which is a contradiction. This completes the proof.

It remains to prove Lemma 10. To this end, we will assume without loss of generality that as \( p \to \infty \)
(31)
\[
\mathbb{E}[(R_{1,2})] \rightarrow C_0 > 0.
\]
It will be useful to express the uniform distribution on \( S^{p-1}(\sqrt{p}) \) in terms of the gaussian measure. Specifically, let \( g^1, g^2, \ldots, g^{m_p} \sim \mathcal{N}(0, I_p) \) be iid Gaussian vectors and set \( \beta^i = g^i / \|g^i\|^2 \). The rotational symmetry of the gaussian measure ensures that the \( \beta^i \) 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 \( S \) for the \( g^i \)-vectors:

\[
\tilde{S}_\epsilon := \left\{ (g^1, g^2, \ldots, g^{m_p}) : (\sqrt{p} \frac{g^1}{\|g^1\|}, \sqrt{p} \frac{g^2}{\|g^2\|}, \ldots, \sqrt{p} \frac{g^{m_p}}{\|g^{m_p}\|}) \in S, \|g^i\| \in (\sqrt{p}(1 - \epsilon), \sqrt{p}(1 + \epsilon)), \forall i \right\}.
\]

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

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

\[
\left| (\beta^i - a)^\top a \right| \leq C p_c, \quad \text{where } a = \sum_{i=1}^{m_p} \beta^i/m_p \text{ and } \tilde{a} := a\sqrt{p}/\|a\| \text{ is a normalized version of } a \text{ with radius } \sqrt{p}.
\]

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

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

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

**Proof.** Note that, for any \( g \in \tilde{S}_\epsilon \),

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

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

\[
\left| (g^i - b)^\top \tilde{b} \right| \leq \left| (g^i - b)^\top a \right| + \left| (g^i - b)^\top (b - a) \right| \leq \left| (g^i - \beta^i)^\top a \right| + \left| (\beta^i - a)^\top a \right| + \left| (a - b)^\top a \right| + \|g^i - b\|^2 + \|\beta^i - a\| + O(\epsilon p) \leq O(\epsilon p) + 2pc_p \leq (2c_p + O(\epsilon))p.
\]

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

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 \( b, q^2, q^3, \ldots, q^{m_p} \) are independent. Moreover, \( h_2, h_3, \ldots, h_{m_p} \) are conditionally independent given \( b \). 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 $p \to \infty$

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

Proof. We start with

$$
\sum_{i=2}^{m_p} \left( \|X(g^i - b)^2 - \|X(g^i - b + w)^2 \right) 
\leq (m_p - 1)\|Xw\|^2 + 2 \sum_{i=2}^{m_p} \|\langle w^T X^T X(g^i - b)\rangle\|
\leq (m_p - 1)\|X^T X\|_2^2 \|w\|^2 + 2 \sum_{i=2}^{m_p} \|X^T X\|_2 \|\|g^i - b\|.
$$

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

$$
\mathbb{P}_{g,w} \left( (m_p - 1)\|X^T X\|_2^2 \|w\|^2 \leq Cp \right) \to 1,
$$

and

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

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 $\mathbb{E}(R_{1,2})$ 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 $S$ and $\hat{S}$ 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

$$
R_p = \frac{1}{pm_p} \log \left( \mathbb{E}_{g^1, g^2, \ldots, g^{mp}} \left[ e^{-G(g)} \mathbf{1}_{\{(g^1, g^2, \ldots, g^{mp}) \in S\}} \right] \right)
= \frac{1}{pm_p} \log \left( \mathbb{E}_g \left[ e^{-G(g)} \mathbf{1}_{\{(g^1, g^2, \ldots, g^{mp}) \in S\}} \right] \right)
= \frac{1}{pm_p} \log \left( \mathbb{E}_g \left[ e^{-G(g)} \mathbf{1}_{\{g \in \hat{S}\}} \right] \right) + O(\epsilon^2),
$$

where

$$
G(g) = G(g^1, g^2, \ldots, g^{mp}) := \frac{m_p}{2\Delta} \left\| y - X \left( \frac{1}{m_p} \sum_{i=1}^{m_p} \frac{g^i \sqrt{p}}{\|g^i\|} \right) \right\|^2 + \frac{1}{2\Delta} \sum_{i=2}^{m_p} \left\| X \left( \frac{g^i \sqrt{p}}{\|g^i\|} - \frac{1}{m_p} \sum_{i=1}^{m_p} \frac{g^i \sqrt{p}}{\|g^i\|} \right) \right\|^2.
$$
Define
\[
\tilde{G}(g) = \tilde{G}(g^1, g^2, \ldots, g^{m_p}) := \frac{m_p}{2\Delta} \| y - Xb \|^2 + \frac{1}{2\Delta} \sum_{i=2}^{m_p} \| X(g^i - b) \|^2.
\]

Note that, for any \( g \in \tilde{S}_\epsilon \), by (33), we have \( \| b - a \| = O(\epsilon \sqrt{p}) \) and \( \| g^i - \beta^i \| = O(\epsilon \sqrt{p}) \) for any \( i \). Since \( P(\Omega) \to 1 \) as \( p \to \infty \), with high probability
\[
\left| G(g) - \tilde{G}(g) \right| \leq \frac{m_p}{2\Delta} \| y - Xb \|^2 - \| y - Xa \|^2 + \frac{1}{2\Delta} \sum_{i=2}^{m_p} \| X(g^i - b) \|^2 - \| X(g^i - b - \beta^i + a) \|
\]
\[
\leq \frac{m_p}{2\Delta} \| y^\top X(b - a) \| + \frac{m_p}{2\Delta} \| Xb \|^2 - \| Xa \|^2 + \frac{1}{2\Delta} \sum_{i=2}^{m_p} \left\{ (g^i - b)^\top X^\top X(g^i - b - \beta^i + a) \right\}
\]
\[
\leq O(\sqrt{p}m_p) \| b - a \| + \frac{1}{2\Delta} \sum_{i=2}^{m_p} \{ \| X^\top X \|_2 [\| g^i - b \| (\| g^i - \beta^i \| + \| b - a \|)]
\]
\[
+ \| \beta^i - a \| (\| g^i - \beta^i \| + \| b - a \|) \}
= O(\sqrt{p}m_p) \| b - a \| + O(\sqrt{p}m_p) \| g^i - \beta^i \|
= O(m_p \epsilon),
\]

where (i) follows from the fact that \( \| y \| = O(1) \) and \( \| X \|_2 = O(1) \) on \( \Omega \). This allows us to replace \( G(g) \) in (34) by \( \tilde{G}(g) \) with an additional cost \( O(\epsilon) \). Thus with high probability,
\[
(35) \quad R_p = \frac{1}{pm_p} \log \left( \mathbb{E}_g \left[ e^{-\tilde{G}(g)} \mathbf{1}_{\{g \in \tilde{S}_\epsilon]\}} \right]\right) + O(\epsilon).
\]

Let \( \delta_p := c_p/C_2 \to 0 \), 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 \( \tilde{S}_\epsilon \) and (33), we know that there exist \( c', c > 0 \) such that \( \{ g = (g^1, g^2, \ldots, g^{m_p}) : \| (g^i - b, \tilde{b}) \| \leq (C_2 c_p + C_3 \epsilon)p, \forall i \} \cap \{ \| g^i \| \in ((1 - \epsilon)\sqrt{p}, (1 + \epsilon)\sqrt{p}), \forall i \} \cap \{ \| b \| / \sqrt{p} : (c', 1 - c) \} \) contains \( \tilde{S}_\epsilon \), thus
\[
\mathbb{E}_g \left[ e^{-\tilde{G}(g)} \mathbf{1}_{\{g \in \tilde{S}_\epsilon\}} \right]\]
\[
\leq \mathbb{E}_b \left[ \mathbf{1}_{\{ \| b \| \in ((c', 1 - c)\}} \mathbb{E}_g \left[ e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} \mathbf{1}_{\{ \| (g^i - b, \tilde{b}) \| \leq (C_2 \delta_p + O(\epsilon))p, \| g^i \| \in ((1 - \epsilon)\sqrt{p}, (1 + \epsilon)\sqrt{p})\}} \right]\right]
\]
\[
= \mathbb{E}_b \left[ \mathbb{E}_g \left[ e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} \mathbf{1}_{\{ \| (g^i - b, \tilde{b}) \| \leq (C_2 \delta_p + O(\epsilon))p, \| g^i \| \in ((1 - \epsilon)\sqrt{p}, (1 + \epsilon)\sqrt{p})\}} \mathbf{1}_{\{ \| b \| \leq 2\sqrt{\frac{p}{m_p}} \}} \right]\right]
\]
\[
\cdot \mathbb{P} \left( \| w \| \leq 2\sqrt{\frac{p}{m_p}} \right)^{-1},
\]
\[
(36)
\]
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

\[
(37) \quad \mathbb{E}_b \left[ e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} 1 \{ \langle g^i - b, \tilde{b} \rangle \leq (C\delta_p + O(\epsilon))p_i \| g^i \| \in ([1 - \epsilon - \frac{2}{\sqrt{m_p}}] \sqrt{\mathbb{P}}, [1 + \epsilon + \frac{2}{\sqrt{m_p}}] \sqrt{\mathbb{P}}) \} \right] \leq \mathbb{E}_b \left[ e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} 1 \{ \langle h^i - b, \tilde{b} \rangle \leq (C\delta_p + O(\epsilon) + \frac{2}{\sqrt{m_p}})p_i \| h^i \| \in ([1 - \epsilon - \frac{2}{\sqrt{m_p}}] \sqrt{\mathbb{P}}, [1 + \epsilon + \frac{2}{\sqrt{m_p}}] \sqrt{\mathbb{P}}) \} \right]
\]

\[
\leq \mathbb{E}_b \left[ e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} 1 \{ \langle h^i - b, \tilde{b} \rangle \leq (C\delta_p + O(\epsilon) + \frac{2}{\sqrt{m_p}})p_i \| h^i \| \in ([1 - \epsilon - \frac{2}{\sqrt{m_p}}] \sqrt{\mathbb{P}}, [1 + \epsilon + \frac{2}{\sqrt{m_p}}] \sqrt{\mathbb{P}}) \} \right]
\]

where (i) holds because dropping the indicator results in a natural upper bound and it also eliminates dependence on \( w \); in addition, the \( O(p \sqrt{m_p}) \) term in the third line can be deduced by an analogous argument as Lemma 13; (ii) 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 \( \delta_p \to 0 \) and \( m_p \to \infty \) as \( p \to \infty \), for \( p \) large enough we can rewrite the RHS above as (note that for the sake of notational simplicity, we drop the \( O(p \sqrt{m_p}) \) term here since it is a vanishing term once the entire expression is scaled by \( \frac{1}{m_p} \log \) later on)

\[
(38) \quad \mathbb{E}_b \left[ e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} 1 \{ \langle g^i - b, \tilde{b} \rangle \leq (C\delta_p + O(\epsilon))p_i \| g^i \| \in ([1 - \epsilon - \frac{2}{\sqrt{m_p}}] \sqrt{\mathbb{P}}, [1 + \epsilon + \frac{2}{\sqrt{m_p}}] \sqrt{\mathbb{P}}) \} \right] \leq \mathbb{E}_b \left[ e^{-\tilde{G}(g)} \prod_{i=2}^{m_p} 1 \{ \langle h^i - b, \tilde{b} \rangle \leq (C\delta_p + O(\epsilon) + \frac{2}{\sqrt{m_p}})p_i \| h^i \| \in ([1 - \epsilon - \frac{2}{\sqrt{m_p}}] \sqrt{\mathbb{P}}, [1 + \epsilon + \frac{2}{\sqrt{m_p}}] \sqrt{\mathbb{P}}) \} \right]
\]
Finally, combining (36, 37, 38, 39), we have

\[
\mathbb{E}_{q^2} \left( e^{-\frac{1}{2\Delta} \|Xq^2\|^2} 1\{ \|q^2\| \leq O(\epsilon) \} \right) 
= \mathbb{E}_{\beta \sim \text{Unif}(S^{n-1}(\sqrt{p})))} \times \mathbb{P}_{x \sim p} \left( e^{-\frac{1}{2\Delta} \|X\|^2} 1\{ \|\beta\| \leq O(\epsilon) \} \right) 
\leq \mathbb{E}_{\beta \sim \text{Unif}(S^{n-1}(\sqrt{p})))} \left( e^{-\frac{1}{2\Delta} \|X\|^2} 1\{ \|\beta\| \leq O(\epsilon) \} \right) 
\leq \mathbb{E}_{\beta \sim \text{Unif}(S^{n-1}(\sqrt{p})))} \left( e^{-\frac{1}{2\Delta} \|X\|^2} 1\{ \|\beta\| \leq O(\epsilon) \} \right) := \kappa_p(b).
\]

We turn to the proof of Lemma 8. We turn to the proof of Lemma 8. Note that the replicas \( \beta_1 \) and \( \beta_2 \) are iid samples from the posterior distribution. Thus

\[
\langle R_{12} \rangle = \frac{1}{p} \sum_{i=1}^{p} \langle \beta_1 \beta_2 \rangle = \frac{1}{p} \sum_{i=1}^{p} \langle \beta_i \rangle \langle \beta_i \rangle = \frac{p}{p} \langle \beta_i \rangle^2 \geq 0
\]

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**

\[
\tag{40} f_p(\delta) = \lim_{\varepsilon \to 0^+} \frac{1}{p} \log \left\{ \frac{1}{2\varepsilon} \int_{|\beta_1 - \sqrt{p}\delta| < \varepsilon} e^{-\frac{1}{2\varepsilon} ||y-X\beta||^2 + H_p^{\text{Gauss}}(\beta)} d\pi_0(\beta) \right\}.
\]

By the Nishimori identity (see e.g. [9, (2.25)])

\[
\tag{41} \mathbb{E} \langle R_{1,2} \rangle = \frac{1}{p} \mathbb{E} \langle \beta^T \beta_0 \rangle,
\]

so it suffices to show there exists \( C > 0 \) such that

\[
\tag{42} (1 - C) > \frac{1}{p} \mathbb{E} \langle \beta^T \beta_0 \rangle = \mathbb{E} \left\langle \frac{\beta_1}{\sqrt{p}} \right\rangle.
\]

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

\[
f_p(\delta) = \frac{1}{p} \log \left\{ \frac{1}{\sqrt{\pi}} \frac{\Gamma \left( \frac{p}{2} \right)}{\Gamma \left( \frac{p-1}{2} \right)} (1 - \delta^2)^{\frac{p-3}{2}} \int_{\beta_1 = \sqrt{p}\delta} e^{-\frac{1}{2\varepsilon} ||y-X\beta||^2 + H_p^{\text{Gauss}}(\beta)} d\pi_{\beta_1 = \sqrt{p}\delta} (\beta) \right\},
\]

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 \( X = [X_1, X_2, \ldots, X_p] \). Define \( X_{(-1)} := [X_2, \ldots, X_p] \), \( h = \sqrt{p}X_1 + \varepsilon - \beta_1 X_1 \), and \( \beta_{(-1)} := (\beta_2, \beta_3, \ldots, \beta_p) \). Then

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

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

\[
||y - X\beta||^2 = ||h||^2 + \beta_{(-1)}^T X_{(-1)} \beta_{(-1)} - 2h^T X_{(-1)} \beta_{(-1)} = ||h||^2 + p(1 - \delta^2)u^T X_{(-1)}^T X_{(-1)} u - 2\sqrt{p(1 - \delta^2)}h^T X_{(-1)} u.
\]

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

\[
\tag{43} f_p(\delta) - \frac{1}{2} \log(1 - \delta^2)
\]

\[
= \frac{1}{p} \log \left( \int_{S^{p-2}(1)} e^{-\frac{1}{p\Delta} \left( ||h||^2 + p(1 - \delta^2)u^T X_{(-1)}^T X_{(-1)} u - 2\sqrt{p(1 - \delta^2)}h^T X_{(-1)} u \right)} d\mu_{p-2}(u) \right) + o(1)
\]

\[
= -\frac{||h||^2}{2p\Delta} + \frac{1}{p} \log \int_{S^{p-2}(1)} e^{-\frac{1}{p\Delta} \left( p(1 - \delta^2)u^T X_{(-1)}^T X_{(-1)} u - 2\sqrt{p(1 - \delta^2)}h^T X_{(-1)} u \right)} d\mu_{p-2}(u) + o(1).
\]

Note that both \( \pi_{\beta_1 = \sqrt{p}\delta} \) and \( \mu_{p-2} \) are uniform measures so that the change of measure does not result in a Jacobian term. In addition, since \( ||X_{(-1)}^T X_{(-1)}||_{op} = O(1) \) and \( ||h|| = O(\sqrt{p}) \) with high probability, the second term is of order \( O(1) \). By Law of Large Numbers, almost surely

\[
\tag{44} \frac{1}{p} ||h||^2 \to \alpha \Delta + (1 - \delta)^2.
\]

As \( \log(1 - \delta^2) \to -\infty \) as \( \delta \to 1 \), for any \( C_1 > 0 \), there exists \( \delta < 1 \) such that with high probability, \( \sup_{\delta \in [0,1]} f_p(\delta') < -C_1 \). On the other hand, for any fixed \( \delta < 1 \), there exists \( C' > 0 \) (depending on \( \alpha, \Delta \) and \( \delta \)) 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,

\[
\langle 1(\beta_1 > \delta_1 \sqrt{p}) \rangle = \frac{\int_{\delta_1}^1 \exp(pf_p(\delta'))d\delta'}{\int_{1-1}^1 \exp(pf_p(\delta'))d\delta'} \leq \frac{\int_{\delta_1}^1 \exp(pf_p(\delta'))d\delta'}{\int_{1-1}^1 \exp(pf_p(\delta'))d\delta'}
\]

\[
\leq \frac{(1 - \delta_1) \exp(-2C_{\delta_0})}{(\delta_0 + 1) \exp(-pC_{\delta_0})} \leq \frac{1 - \delta_1}{1 + \delta_0} \exp(-pC_{\delta_0}).
\]

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

\[
\langle 1(\beta_1 < \delta_1 \sqrt{p}) \rangle > 1 - \frac{1 - \delta_1}{1 + \delta_0} \exp(-pC_{\delta_0}) \tag{45}
\]

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 \( C := (1 - \delta_1)/2. \)

\[\square\]

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

\[
P\left(\frac{1}{p} \log Z_p \geq \sup_{a \in \mathbb{R}^p, \|a\| \leq \sqrt{p}} f_{\text{TAP}}(a) + \eta\right) \to 0,
\]

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

\[
\sup_{a \in \mathbb{R}^p, \|a\| \leq \sqrt{p}} f_{\text{TAP}}(a) - \sup_{a \in \mathbb{R}^p, \|a\| \leq (1-c) \sqrt{p}} f_{\text{TAP}}(a)
\]

\[
\leq \max \left\{ \sup_{a \in \mathbb{R}^p, \|a\| \leq \sqrt{p}} f_{\text{TAP}}(a) - f_{\text{TAP}}(a^*) \right\}, \tag{46}
\]

the second term of which can be further controlled by

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

\[
\leq \sup_{a \in \mathbb{R}^p, (1-c) \sqrt{p} \leq \|a\| \leq \sqrt{p}} \left\{ \frac{1}{2\Delta p} \left\| y - Xa \right\|^2 - \left\| y - Xa^* \right\|^2 \right\}
\]

\[
- \frac{\alpha}{2} \left[ \log \left( 1 + \frac{1 - \|a\|^2/p}{\Delta \alpha} \right) - \log \left( 1 + \frac{1 - \|a^*\|^2/p}{\Delta \alpha} \right) \right] + \frac{1}{2} \left[ \log \left( 1 - \frac{\|a\|^2}{p} \right) - \log \left( 1 - \frac{\|a^*\|^2}{p} \right) \right]
\]
\[
\begin{align*}
\leq & \sup_{a \in \mathbb{R}^r, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} \left\{ \frac{1}{2\Delta p} \left[ 2y^\top X (a - a^*) \right] + \left\| a^\top Xa - (a^*)^\top Xa^* \right\| \right\} \\
+ & \sup_{a \in \mathbb{R}^r, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} \left\{ \frac{\alpha}{2} \left[ \log \left( 1 + \frac{1 - \|a^*\|^2/p}{\Delta \alpha} \right) - \log \left( 1 + \frac{1 - \|a\|^2/p}{\Delta \alpha} \right) \right] \right\} \\
+ & \sup_{a \in \mathbb{R}^r, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} \left\{ \frac{1}{\Delta} \left[ \log \left( 1 - \frac{\|a\|^2}{p} \right) - \log \left( 1 - \frac{\|a^*\|^2}{p} \right) \right] \right\} \\
\leq & \sup_{a \in \mathbb{R}^r, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} \left\{ \frac{1}{2\Delta p} \left[ 2c\sqrt{p}\|y^\top X\| + 2p(2c - c^2)\|X^\top X\|_2 \right] + \tau(c) \right\},
\end{align*}
\]

where
\[
\tau(c) := \frac{\alpha}{2} \log \left( 1 + \frac{2c - c^2}{\Delta \alpha} \right),
\]
as a deterministic function, goes to 0 as \(c \to 0\). Since with high probability, \(\|X^\top X\|_2 = O(1)\) and \(\|y^\top X\| = O(\sqrt{p})\), we have
\[
\sup_{a \in \mathbb{R}^r, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} (f^\text{TAP}(a) - f^\text{TAP}(a^*)) \to 0
\]
as \(c \to 0\), with high probability. Thus, there exists positive constant \(c^*\) such that if \(c \leq c^*\), then with high probability,
\[
\sup_{a \in \mathbb{R}^r, (1-c)\sqrt{p} \leq \|a\| \leq \sqrt{p}} (f^\text{TAP}(a) - f^\text{TAP}(a^*)) \leq \frac{\eta}{3}.
\]
Finally, choosing \(c \leq c^*\), Theorem 2 together with (46) and (47) give us
\[
\mathbb{P} \left( \frac{1}{p} \log \mathcal{Z} \leq \sup_{a \in \mathbb{R}^r, \|a\| \leq \sqrt{p}} f^\text{TAP}(a) - \frac{\eta}{3} \right) \to 0,
\]
which is exactly the lower bound part of Theorem 1. \(\square\)

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

APPENDIX A: CONCENTRATION AND CONVEXITY INEQUALITIES

**Lemma 14 (Concentration of almost Lipschitz functions [7]).** Let \(U_i \sim \mathcal{N}(0,1)\). Let \(f\) be a Lipschitz function over \(G \subset \mathbb{R}^p\) with Lipschitz constant \(K_p\). Assume \(0 \in G\), \(f(0)^2 \leq C^2\), \(\mathbb{E} \left[ f^2 \right] \leq C^2\) for some \(C > 0\). Then for \(r \geq 6 \left( C + \sqrt{pK_p} \right) \sqrt{\mathbb{P}(G^c)}\) we have
\[
\mathbb{P} \left( |f(U_1, \ldots, U_p) - \mathbb{E} [f(U_1, \ldots, U_p)]| \geq r \right) \leq 2e^{-\frac{r^2}{16K_p^2}} + \mathbb{P}(G^c).
\]

**Lemma 15 (The largest and smallest eigenvalues of Wishart matrices [31]).** 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 \(1 - 2 \exp \left( -t^2/2 \right)\) 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 [32]). Let $S^{n-1}(1)$ be the $(n-1)$-sphere of radius 1. Let $f : S^{n-1}(1) \to \mathbb{R}$ be a Lipschitz function with Lipschitz constant $L$, and let $\sigma$ be a uniform random vector on $S^{n-1}(1)$. Then for any $t > 0,$

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

Lemma 17 (A bound for differences between derivatives of convex functions [5]). 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

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

APPENDIX 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 the sample covariance matrix in statistics and probability literature. Please refer to [3] for complete statements and proofs. Let $\theta_1 < \theta_2 < \cdots < \theta_p$ be the eigenvalues of the matrix $S_p = X^\top X \in \mathbb{R}^{p \times p}$. We have that as $p \to \infty$, almost surely,

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

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

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

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

APPENDIX C: PROOF OF OVERLAP CONCENTRATION RESULTS

C.1. Proof of Theorem 5. In order to avoid possible confusion caused by the similarity between $\epsilon$ and $\varepsilon$, we will instead use $z$ to denote the noise vector i.e. $y - X \beta_0$, in this section. In addition, note that in this subsection we will use $\mathbb{E}$ as a shorthand for $\mathbb{E}_{X,z,Y,\beta_0}$.

Observe that since we are working with a well-specified Bayesian model, the so-called Nishimori identity (see e.g. [9]) holds naturally, which says for any measurable function $f$ and positive integer $k$,

$$\mathbb{E}(f(\beta^1, \beta^2, \ldots, \beta^k, \beta_0)) = \mathbb{E}(f(\beta^1, \beta^2, \ldots, \beta^k, \beta^{k+1})).$$

In the display above, $\langle \cdot \rangle$ denotes the expectation with respect to the posterior distribution. This symmetry identity will play a crucial role in our subsequent arguments. Recall that $H^\text{Gauss}_p$ is the Hamiltonian of the side information channel, namely

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

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

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

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 $L := H' / p$. 
LEMMA 18. Let
\[ \mathcal{L} := \frac{H'}{p} = \frac{1}{p} \frac{d \mathcal{H}_{p}^{Gauss}}{d \lambda_{0,p}}. \]

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

Gaussian integration by parts implies
\[
\frac{1}{p \sqrt{\lambda_{0,p}}} \mathbb{E} \langle R_{1,*} Z^{\top} \beta \rangle = \frac{1}{p} \mathbb{E} \langle R_{1,*} \| \beta \|^{2} \rangle - \frac{1}{p} \mathbb{E} \langle R_{1,*} \beta^{\top} \beta \rangle = \frac{1}{p} \mathbb{E} \langle R_{1,*} \| \beta \|^{2} \rangle - \mathbb{E} \langle (R_{1,*})^{2} \rangle,
\]
in which \( \mathbb{E} \langle R_{1,*} \beta^{\top} \beta \rangle = \frac{1}{p} \mathbb{E} [\langle (\beta_{0}^{\top} \beta_{0}^{\top}) \beta \rangle] = \frac{1}{p} \mathbb{E} [\langle (\beta_{0}^{\top} \beta_{1}^{\top} \beta_{1}^{\top}) \beta \rangle] = \frac{1}{p} \mathbb{E} [\langle (\beta_{1}^{\top} \beta_{0}^{\top} \beta_{0}^{\top}) \beta \rangle] = p \mathbb{E} \langle (R_{1,*})^{2} \rangle \) 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
\[
2 \mathbb{E} \langle R_{1,*} (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle = \mathbb{E} \langle R_{1,*} \rangle^{2} - 2 \mathbb{E} \langle R_{1,*}^{2} \rangle + (\mathbb{E} \langle R_{1,*} \rangle)^{2}
- (\mathbb{E} \langle R_{1,*}^{2} \rangle - (\mathbb{E} \langle R_{1,*} \rangle)^{2}) - (\mathbb{E} \langle R_{1,*}^{2} \rangle - \mathbb{E} \langle R_{1,*} \rangle^{2})
\leq - \mathbb{E} \langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle
\leq 0.
\]

Finally, using Cauchy-Schwarz inequality, we have
\[
\mathbb{E} \langle (R_{1,2} - \mathbb{E} \langle R_{1,2} \rangle)^2 \rangle \leq 2 |\mathbb{E} \langle R_{1,*} (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle|
= 2 |\mathbb{E} \langle (R_{1,*} - \mathbb{E} \langle R_{1,*} \rangle) (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle|
\leq 2 \left\{ \mathbb{E} \langle (R_{1,*} - \mathbb{E} \langle R_{1,*} \rangle)^2 \rangle \mathbb{E} \langle (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle)^2 \rangle \right\}^{1/2},
\]

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

\[ \square \]

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), \( \mathcal{H}_{p}^{\text{Pert}} \) is the Hamiltonian of the whole perturbed model, i.e. including both the original model and the side information channel, and \( F_{p}^{\text{Pert}} \) is the corresponding free energy, as defined in (23). Moreover, let
\[ \mathcal{H} := \frac{d \mathcal{H}_{p}^{\text{Pert}}}{d \lambda_{0,p}}. \]
Please note $\mathcal{H}'$ and $H'$ defined in (48) 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

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

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

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

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

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

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

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

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

We start with the first term. Observe that

$$\int_{\epsilon_p/2}^{\epsilon_p} \mathbb{E} \langle (\mathcal{H}' - \langle \mathcal{H}' \rangle)^2 \rangle \ d\lambda_{0,p}$$

$$\leq \int_{\epsilon_p/2}^{\epsilon_p} \left( p \frac{d^2 F_p^{\text{Pert}}}{d\lambda_{0,p}^2} + p \lambda_{0,p} \right) \ d\lambda_{0,p} = p \left( \frac{dE F_p^{\text{Pert}}}{d\lambda_{0,p}} \bigg|_{\lambda_{0,p}=\epsilon_p} \right) + p \log 2.$$  

Since

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

For any $\lambda_{0,p}$, we naturally have a trivial bound of $\frac{dE F_p^{\text{Pert}}}{d\lambda_{0,p}} \bigg|_{\lambda_{0,p}=\epsilon_p} = 1$, i.e.,

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

Recalling that $L = H'/p = H'/p$,

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

Now we proceed to control the second term using convexity arguments.
We introduce the following two convex functions of $\lambda_{0,p}$,

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

and

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

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

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

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

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

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

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

Note that by using [32, Theorem 3.1.1], there exists a positive constant $C_Z$ such that $\mathbb{E} \left[ (\|Z\| - \mathbb{E} \|Z\|)^2 \right] \leq C_Z$ 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,

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

Note that

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

and thus

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

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

$$
\int_{\tau_p / 2}^{\tau_p} d\lambda_{0,p} \left\{ C^+ (\lambda_{0,p})^2 + C^- (\lambda_{0,p})^2 \right\} \leq \left( 1 + \frac{1}{\sqrt{\tau_p / 2 - \delta}} \right) \int_{\tau_p / 2}^{\tau_p} d\lambda_{0,p} \left\{ C^+ (\lambda_{0,p}) + C^- (\lambda_{0,p}) \right\}
\leq \left( 1 + \frac{1}{\sqrt{\tau_p / 2 - \delta}} \right) \left[ (\mathbb{E} \tilde{F}(\tau_p / 2 + \delta) - \mathbb{E} \tilde{F}(\tau_p / 2 - \delta)) + (\mathbb{E} \tilde{F}(\tau_p - \delta) - \mathbb{E} \tilde{F}(\tau_p + \delta)) \right].
$$
By the mean value theorem,

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

Similarly,

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

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 \( \delta_p/\epsilon_p \to 0 \) and \( \delta_p/\epsilon_p < 1/4 \) for any \( p \). Putting them all together, we have

\[ \int_{\epsilon_p/2}^{\epsilon_p} d\lambda_{0,p} \mathbb{E}(\mathcal{L} - \mathbb{E}(\mathcal{L}))^2 \leq \frac{27\epsilon_p}{2\delta_p^2} (v_p + 2\epsilon_p C_Z) + 9 \int_{\epsilon_p/2}^{\epsilon_p} \frac{C_Z}{4p\lambda_{0,p}} d\lambda_{0,p} + \frac{288\delta_p}{\epsilon_p} \]

\[ = \frac{27\epsilon_p}{2\delta_p^2} (v_p + 2\epsilon_p C_Z) + \frac{288\delta_p}{\epsilon_p} + \frac{9C_Z \log 2}{4p}. \]

Finally, we choose \( \delta_p = \left\{ \left[ 36\epsilon_p^2(v_p + 2\epsilon_p C_Z) \right] / (64p) \right\}^{1/3} \) to optimize the bound, which together with the previous bound on \( \mathbb{E}_{\lambda_0} \mathbb{E}(\mathcal{L} - \mathbb{E}(\mathcal{L}))^2 \) would result in

\[ \mathbb{E}_{\lambda_0} \mathbb{E}(\mathcal{L} - \mathbb{E}(\mathcal{L}))^2 \leq \frac{576}{\epsilon_p} \left( \frac{36\epsilon_p^2(v_p + 2\epsilon_p C_Z)}{64p} \right)^{1/3} + \frac{9C_Z \log 2}{4p} + \frac{1 + \log 2}{p} \]

\[ \leq C \left( \frac{v_p}{p\epsilon_p} + \frac{1}{p} \right)^{1/3}, \]

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

\[ \mathbb{E}_{\lambda_0} \mathbb{E}(\mathcal{L} - \mathbb{E}(\mathcal{L}))^2 \leq C \left( \frac{v_p}{p\epsilon_p} + \frac{1}{p} \right)^{1/3}, \]

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

**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 \( \Delta, \alpha \)) such that for any \( r \geq O(e^{-C_1 p^{1/2}}) \), we have for any fixed \( \beta_0 \) and \( \lambda_0 \in [1/2, 1] \)

\[ \mathbb{P}\left( \frac{1}{p} F_{\epsilon_p}^{\text{Pert}} - \mathbb{E}F_{\epsilon_p}^{\text{Pert}}(\beta_0) \geq r | \beta_0 \right) \leq e^{-C_2 r^2 p^{1/2}}. \]
PROOF. Consider
\[ G = \{ X, z, Z : \max_{\mu} |z_i| \leq \sqrt{D_1} \cap \max_i |Z_i| \leq \sqrt{D_1} \cap \text{ for all } \beta, \beta_0 : \|X(\beta - \beta_0)\|^2 \leq D_2 p \}. \]

In order to use the known results regarding concentration of almost Lipschitz functions, we first try to control \( \mathbb{P}(G^c) \). Since if \( U \sim N(0, I) \) then \( \mathbb{P}(|U| \geq \sqrt{A}) \leq 4e^{-A/4} \), we have by union bound
\[ \mathbb{P}(\max_{\mu} |z_i| \leq \sqrt{D_1} \cap \max_i |\bar{z}_i| \leq \sqrt{D_1}) \leq 4p(1 + \alpha)e^{-D_1/4}, \]
and
\[ \mathbb{P}(\text{for all } \beta, \beta_0 : \|X(\beta - \beta_0)\|^2 \leq D_2 p) = \mathbb{P}(\Lambda_1(X^\top X) \leq D_2/4). \]
Thus
\[ \mathbb{P}(G^c) \leq 4p(1 + \alpha)e^{-D_1/4} + \left( 1 - \mathbb{P} \left( \sigma_{\max}(X)^2 \leq \frac{D_2}{4} \right) \right), \]
which will be made small with a suitable choices of \( D_1 \) and \( D_2 \). Now we proceed to show, for any fixed \( \beta_0 \), free energy of the perturbed model \( F_p^{\text{Pert}}(X, z, Z, \beta_0) \) is Lipschitz on \( G \) with respect to \( (X, z, Z) \), as a vector in \( \mathbb{R}^{np + n + p} \). From now on, we treat \( \beta_0 \) as fixed and omit it in the arguments of \( \mathcal{H}^{\text{Pert}} \) and \( F^{\text{Pert}} \). Recall that, as defined in (23),
\[ \mathcal{H}^{\text{Pert}}(X, z, Z) = \frac{1}{2\Delta} \|y - X\beta\|^2 - \lambda_0 e_p^\top \beta - \sqrt{\lambda_0 e_p} Z^\top \beta + \frac{1}{2} \lambda e_p \|\beta\|^2. \]
Then
\[ |\mathcal{H}^{\text{Pert}}(X, z, Z) - \mathcal{H}^{\text{Pert}}(X', z', Z')| \leq \frac{1}{2\Delta} \|\|X(\beta_0 - \beta) + z\|^2 - \|X'(\beta_0 - \beta) + z'\|^2\| + \sqrt{\lambda e_p} \|Z^\top \beta - (Z')^\top \beta\|, \]
in which
\[ \|\|X(\beta_0 - \beta) + z\|^2 - \|X'(\beta_0 - \beta) + z'\|^2\| = \left( \|X(\beta_0 - \beta)\|^2 + 2z^\top X(\beta_0 - \beta) + \|z\|^2 \right) - \left( \|X'(\beta_0 - \beta)\|^2 + 2z'^\top X'(\beta_0 - \beta) + \|z'\|^2 \right) \]
\[ \leq |(\beta - \beta_0)^\top (X^\top X - X'^\top X')(\beta - \beta_0)| + 2|(z^\top X - z'^\top X')(\beta - \beta_0)| + \|z\|^2 - \|z'\|^2 \]
\[ \leq \|\beta - \beta_0\|^2 \left( \|X^\top (X - X')\| + \|(X - X')^\top X'\| \right) + (O(\sqrt{pD_1}) + O(\sqrt{pD_2})) \sqrt{p} \|X - X'|_F + \|z - z'\| \]
\[ \leq (O(\sqrt{pD_1}) + O(\sqrt{pD_2})) \sqrt{p} \|X - X'|_F + \|z - z'\|, \]
and
\[ \|Z^\top \beta - (Z')^\top \beta\| \leq \sqrt{\lambda_0 e_p \|Z - Z'\|}. \]
Since \( \lambda_0 \in [1/2, 1] \) and \( \lim_{p \to \infty} e_p = 0 \), 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
\[ F_p^{\text{Pert}}(X, z, Z) - F_p^{\text{Pert}}(X', z', Z') \]
\[ = \frac{1}{p} \log \left( \frac{\int d\beta \pi_0(\beta) e^{-H_p^{\text{Pert}}(y)}}{\int d\beta \pi_0(\beta) e^{-H_p^{\text{Pert}}(y')}} \right) \]
\[ \leq \frac{1}{p} \log \left( \frac{\int d\beta \pi_0(\beta) e^{-H_p^{\text{Pert}}(y)} + |H_p^{\text{Pert}}(y) - H_p^{\text{Pert}}(y')|}{\int d\beta \pi_0(\beta) e^{-H_p^{\text{Pert}}(y')}} \right) \]
\[ \leq \frac{1}{p} (O(\sqrt{pD_1}) + O(\sqrt{pD_2}))(\sqrt{p}||X - X'||_F + ||z - z'|| + ||Z - Z'||). \]

By symmetry, we can further conclude that
\[ |F_p^{\text{Pert}}(X, z, Z) - F_p^{\text{Pert}}(X', z', Z')| \]
\[ \leq (O(p^{-1/2}\sqrt{D_1}) + O(p^{-1/2}\sqrt{D_2}))(\sqrt{p}||X - X'||_F + ||z - z'|| + ||Z - Z'||) \]
\[ \leq (O(p^{-1/2}\sqrt{D_1}) + O(p^{-1/2}\sqrt{D_2}))(\sqrt{p}||X, z, Z|| - (\sqrt{p}X', z', Z'||), \]
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 $\mathbb{E}_{X,z,Z}[F_p^{\text{Pert}}(X, z, Z)] = \mathbb{E}[F_p^{\text{Pert}}(X, z, Z)|\beta_0]$ is bounded. In fact, the upper bound constant does not depend on $\beta_0$. 

\[ |F_p^{\text{Pert}}| = \frac{1}{p} \log \int_{S^{n-1}(\sqrt{p})} e^{-\frac{1}{\sqrt{p}}||y - X\beta||^2 - \lambda\epsilon_p \beta_0 \beta - \sqrt{\lambda\epsilon_p}Z^\top \beta} d\pi_0(\beta) \]
\[ \leq \frac{1}{p} \log \int_{S^{n-1}(\sqrt{p})} e^{-\frac{1}{\sqrt{p}}||X(\beta_0 - \beta) + z||^2 - \lambda\epsilon_p \beta_0 \beta - \sqrt{\lambda\epsilon_p}Z^\top \beta} d\pi_0(\beta) \]
\[ \leq \frac{1}{p} \log \int_{S^{n-1}(\sqrt{p})} e^{-\frac{1}{p}((||X(\beta_0 - \beta)||^2 + ||z||^2) - p\lambda\epsilon_p - \sqrt{p\lambda\epsilon_p}Z|| - \frac{1}{2}p\lambda\epsilon_p d\pi_0(\beta) \]
\[ \leq \frac{1}{p} \left[ \frac{4p}{\Delta} ||X^\top X||_2 + \frac{1}{\Delta} ||z||^2 + \frac{3}{2} p\lambda\epsilon_p + \sqrt{p\lambda\epsilon_p} ||Z|| \right], \]
which leads to
\[ \mathbb{E}\left(F_p^{\text{Pert}}(\lambda)^2 \right) \leq \mathbb{E}\left\{ \frac{4}{p^2} \left[ \frac{16p^2}{\Delta^2} \max(X) + \frac{1}{4} \frac{9\lambda\epsilon_p}{4} p^2 + p\lambda\epsilon_p ||Z||^2 \right] \right\} = O(1). \]

By applying Lemma 14 and choosing $D_1 = p^\gamma$, $D_2 = O(1)$, $K_p = O(p^{-1/2})$ (accordingly $\mathbb{P}(\mathcal{G}^c) = O(pe^{-\gamma/4})$), for any $r \geq O(e^{-C_1p^{\gamma}}) \geq O(p^{(5+\gamma)/4}e^{-\gamma/8})$, \]
\[ \mathbb{P}(|F_p^{\text{Pert}}(\lambda) - \mathbb{E}[F_p^{\text{Pert}}(\lambda)|\beta_0]| \geq r |\beta_0| \leq O(e^{-C_2p^{1-\gamma}}) + O(pe^{-\gamma/4}). \]

Finally, the choice $\gamma = 1/2$ renders the optimal bound
\[ \mathbb{P}(|F_p^{\text{Pert}}(\lambda) - \mathbb{E}[F_p^{\text{Pert}}(\lambda)|\beta_0]| \geq r |\beta_0| \leq e^{-C_2p^{1/2}}. \]

We will see that, with an integral argument, Lemma 21 further implies the following lemma, which gives a upper bound on the conditional variance of $F_p^{\text{Pert}}(\lambda)$ giving any fixed $\beta_0$. 

**Lemma 22.** For any fixed $\beta_0 \in SP^{-1}(\sqrt{p})$, 
\[
\mathbb{E} \left[ \left( F_{p}^{\text{Pert}}(\lambda) - \mathbb{E} [F_{p}^{\text{Pert}}(\lambda)|\beta_0] \right)^2 \right] = O \left( \frac{1}{p} \right).
\]

**Proof.** By Lemma 21, there exist positive constants $C_1$, $C_2$, and $C_3$, depending only on $\Delta$ and $\alpha$, such that for any $r \geq t_0 := C_1 e^{-C_2 p^{1/2}}$, we have for any $\beta_0 \in SP^{-1}(\sqrt{p})$, 
\[
\mathbb{P} \left( |F_{p}^{\text{Pert}}(\lambda) - \mathbb{E} [F_{p}^{\text{Pert}}(\lambda)|\beta_0]| > r \right) \leq e^{-C_2 p^{1/2} r}.
\]

Let $V = \frac{|F_{p}^{\text{Pert}}(\lambda) - \mathbb{E} [F_{p}^{\text{Pert}}(\lambda)|\beta_0]|}{p}$. 
\[
\mathbb{E} [V^2 | \beta_0] = \mathbb{E} [V^2 | V \leq t_0 | \beta_0] + \mathbb{E} [V^2 | V > t_0 | \beta_0]
\leq t_0^2 \mathbb{P}(V \leq t_0 | \beta_0) + \int_{t_0}^{\infty} s \mathbb{P}(V > s | \beta_0) ds
\leq t_0^2 + \int_{t_0}^{\infty} s e^{-C_2 s p^{1/2}} ds
= t_0^2 + \frac{1}{C_3 p^{1/2}} \left( t_0^2 + \frac{1}{C_3 p^{1/2}} \right) e^{-C_3 p^{1/2} t_0^2}
= O \left( \frac{1}{p} \right),
\]
where the last equality is because of $t_0^2 = C_1 e^{-2C_2 p^{1/2}} = o(p^{-1/2})$. 

**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 $\beta_0$.

**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 
\[
\frac{1}{p} \langle H^{\lambda}(\beta) - H^{\lambda}(\beta_0) \rangle_{\beta_0} \leq \frac{1}{p} \langle \log \mathcal{Z}^{\text{Pert}}(\beta_0) - \log \mathcal{Z}^{\text{Pert}}(\beta_0) \rangle \leq \frac{1}{p} \langle H^{\text{Pert}}(\beta_0) - H^{\text{Pert}}(\beta_0) \rangle_{\beta_0},
\]
where $\langle f(\beta) \rangle_{\beta_0}$ refers to the expectation of $f(\beta)$ under the distribution proportional to $d\pi(\beta) e^{-H^{\text{Pert}}(\beta, \beta_0)}$ and similarly for $\langle f(\beta) \rangle_{\beta_0}$. Note that $H^{\text{Pert}}(\beta, \beta_0)$ is the same as $H^{\text{Pert}}(\beta)$, only to emphasize the different values of $\beta_0$. Moreover, 
\[
\mathbb{E}_{X, z, Z} \left[ \frac{1}{2\Delta} \left( \| X(\beta_0 - \beta) + z \|^2 - \| X(\beta_0 - \beta) + z \|^2 + \frac{1}{2} \| \lambda_0 \epsilon_p(\beta_0 - \beta_0)^\top \beta \|^2 \right) \right]
\leq \mathbb{E}_{X, z, Z} \left[ \frac{1}{2\Delta} \left( \| X(\beta_0 - \beta_0) \|^2 + 2 \langle X(\beta_0 - \beta_0), X(\beta_0 - \beta) + z \rangle \right) + O(\sqrt{p}) \| \beta_0 - \beta_0 \| \right]
= O(\sqrt{p}) \| \beta_0 - \beta_0 \|.
\]
Upon defining
\[
f(\frac{1}{\sqrt{p}} \beta_0) := \frac{1}{p} \mathbb{E}_{X,z} \log Z^{\text{Pert}}(\beta_0) = \mathbb{E}_{X,z} Z F_p^{\text{Pert}},
\]
we see that there exists a positive constant \( C \) such that
\[
\left| f(\frac{1}{\sqrt{p}} \beta_0) - f(\frac{1}{\sqrt{p}} \tilde{\beta}_0) \right| \leq C \left\| \frac{1}{\sqrt{p}} \beta_0 - \frac{1}{\sqrt{p}} \tilde{\beta}_0 \right\|
\]
Finally, Lévy’s lemma, i.e. Lemma 16, implies
\[
P \left( \left| \mathbb{E}_{X,z} Z F_p^{\text{Pert}} - \mathbb{E}_{X,z} Z F_p^{\text{Pert}} \right| > Ct \right) \leq e^{-pt^2/4},
\]
for any \( t > 0 \), which further leads to
\[
\mathbb{E} \left[ \left( \mathbb{E} \left[ F_p^{\text{Pert}} \mid \beta_0 \right] - \mathbb{E} F_p^{\text{Pert}} \right)^2 \right] \leq \frac{4C^2 e^{-\pi}}{p} = O \left( \frac{1}{p} \right).
\]

C.4. Completing the proof of Theorem 5.

PROOF OF THEOREM 5. By Lemma 21 and 23,
\[
\text{Var}(F_p^{\text{Pert}}(\lambda)) = \mathbb{E} \left[ \text{Var}(F_p^{\text{Pert}}(\lambda) \mid \beta_0) \right] + \text{Var} \left( \mathbb{E} F_p^{\text{Pert}}(\lambda) \mid \beta_0 \right)
\]
\[
\leq O \left( \frac{1}{p} \right) + O \left( \frac{1}{p} \right) = O \left( \frac{1}{p} \right).
\]
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
\[
\left| \frac{d \mathbb{E} F_p^{\text{Pert}}}{d \lambda_0} \right| = \left| \frac{d \mathbb{E} F_p^{\text{Pert}}}{d \lambda_{0,p}} \right| = \left| \frac{\epsilon_p \mathbb{E} \langle \mathcal{H}' \rangle}{p} \right| = \left| \frac{\epsilon_p \mathbb{E} \langle R_{1,2} \rangle}{2} \right| \leq \frac{\epsilon_p}{2}.
\]
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.

Acknowledgments. SS thanks Sumit Mukherjee for his encouragement during the completion of this manuscript.

Funding. SS was partially supported by a Harvard Dean’s Competitive Fund Fellowship.
REFERENCES

[1] Adhikari, A., Brennecke, C., von Soosten, P. and Yau, H.-T. (2021). Dynamical Approach to the TAP Equations for the Sherrington–Kirkpatrick Model. *Journal of Statistical Physics* 183 1–27.

[2] Auffinger, A. and Jagannath, A. (2019). Thouless–Anderson–Palmer equations for generic $p$-spin glasses. *The Annals of Probability* 47 2230–2256.

[3] Bai, Z. and Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices* 20. Springer.

[4] Bai, J. and Lee, J. O. (2016). Fluctuations of the free energy of the spherical Sherrington–Kirkpatrick model. *Journal of Statistical Physics* 165 185–224.

[5] Barbier, J., Krzakala, F., Macris, N., Miolane, L. and Zdeborová, L. (2019). Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences* 116 5451–5460.

[6] Barbier, J. and Macris, N. (2019). The adaptive interpolation method: a simple scheme to prove replica formulas in Bayesian inference. *Probability theory and related fields* 174 1133–1185.

[7] Barbier, J., Macris, N., Dia, M. and Krzakala, F. (2020). Mutual information and optimality of approximate message-passing in random linear estimation. *IEEE Transactions on Information Theory* 66 4270–4303.

[8] Barbier, J., Macris, N., Mailard, A. and Krzakala, F. (2018). The mutual information in random linear estimation beyond iid matrices. In 2018 *IEEE International Symposium on Information Theory (ISIT)* 1390–1394. IEEE.

[9] Barbier, J. and Panchenko, D. (2020). Strong replica symmetry in high-dimensional optimal Bayesian inference. *arXiv preprint arXiv:2005.03115*.

[10] Belius, D. and Kistler, N. (2019). The TAP–Plefka Variational Principle for the Spherical SK Model. *Communications in Mathematical Physics* 367 991–1017.

[11] Blei, D. M., Kucukelbir, A. and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association* 112 859–877.

[12] Celsentano, M., Fan, Z. and Mei, S. (2021). Local convexity of the TAP free energy and AMP convergence for $\mathbb{Z}^2$-synchronization. *arXiv preprint arXiv:2106.11428*.

[13] Chatterjee, S. (2010). Spin glasses and Stein’s method. *Probability theory and related fields* 148 567–600.

[14] Chen, W.-K. and Panchenko, D. (2018). On the TAP free energy in the mixed $p$-spin models. *Communications in Mathematical Physics* 362 219–252.

[15] Chen, W.-K., Panchenko, D. and Subag, E. (2018). The generalized TAP free energy. *arXiv preprint arXiv:1812.05066*.

[16] Chen, W.-K., Panchenko, D. and Subag, E. (2021). The generalized TAP free energy II. *Communications in Mathematical Physics* 381 257–291.

[17] Fan, Z., Mei, S. and Montanari, A. (2021). TAP free energy, spin glasses and variational inference. *The Annals of Probability* 49 1–45.

[18] Fan, Z. and Wu, Y. (2021). The replica-symmetric free energy for Ising spin glasses with orthogonally invariant couplings. *arXiv preprint arXiv:2105.02797*.

[19] Ghorbani, B., Javadi, H. and Montanari, A. (2019). An instability in variational inference for topic models. In *International conference on machine learning* 2221–2231. PMLR.

[20] Guionnet, A., Mai, M. et al. (2005). A Fourier view on the R-transform and related asymptotics of spherical integrals. *Journal of functional analysis* 222 435–490.

[21] Horn, R. A. and Johnson, C. R. (2012). *Matrix analysis*. Cambridge university press.

[22] Jagannath, A. (2017). Approximate ultrametricity for random measures and applications to spin glasses. *Communications on Pure and Applied Mathematics* 70 611–664.

[23] Krzakala, F., Manoel, A., Tramel, E. W. and Zdeborová, L. (2014). Variational free energies for compressed sensing. In 2014 *IEEE International Symposium on Information Theory* 1499–1503. IEEE.

[24] Mezard, M. and Montanari, A. (2009). *Information, physics, and computation*. Oxford University Press.

[25] Montanari, A. (2013). Statistical mechanics and algorithms on sparse and random graphs. *Lectures on Probability Theory and Statistics. Saint-Flour*.

[26] Mukherjee, S. and Sen, S. (2021). Variational Inference in high-dimensional linear regression. *arXiv preprint arXiv:2104.12232*.

[27] Parisi, G. and Potters, M. (1995). Mean-field equations for spin models with orthogonal interaction matrices. *Journal of Physics A: Mathematical and General* 28 5267.

[28] Subag, E. (2018). Free energy landscapes in spherical spin glasses. *arXiv preprint arXiv:1804.10576*. 
[29] TALAGRAND, M. (2010). Mean field models for spin glasses: Volume I: Basic examples 54. Springer Science & Business Media.

[30] THOULESS, D. J., ANDERSON, P. W. and PALMER, R. G. (1977). Solution of ‘solvable model of a spin glass’. Philosophical Magazine 35 593–601.

[31] VERSHYNIN, R. (2010). Introduction to the non-asymptotic analysis of random matrices. arXiv preprint arXiv:1011.3027.

[32] VERSHYNIN, R. (2018). High-dimensional probability: An introduction with applications in data science 47. Cambridge university press.

[33] WAINWRIGHT, M. J. and JORDAN, M. I. (2008). Graphical models, exponential families, and variational inference. Now Publishers Inc.